US 20060234224A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2006/0234224 A1**

Adorjan et al. (43) **Pub. Date: Oct. 19, 2006**

(54) **METHOD AND NUCLEIC ACIDS FOR THE ANALYSIS OF COLON CELL PROLIFERATIVE DISORDERS**

(76) Inventors: **Peter Adorjan**, Berlin (DE); **Matthias Burger**, Berlin (DE); **Sabine Maier**, Berlin (DE); **Ralf Lesche**, Berlin (DE); **Susan Cottrell**, Seattle, WA (US); **Suzanne Mooney**, Seattle, WA (US)

Correspondence Address:
DAVIS WRIGHT TREMAINE, LLP
2600 CENTURY SQUARE
1501 FOURTH AVENUE
SEATTLE, WA 98101-1688 (US)

(21) Appl. No.: **10/506,089**

(22) PCT Filed: **Feb. 27, 2003**

(86) PCT No.: **PCT/EP03/02034**

(30) **Foreign Application Priority Data**

Feb. 27, 2002 (EP) 02004551.4

Publication Classification

(51) **Int. Cl.**
C12Q 1/68 (2006.01)
C07H 21/04 (2006.01)
C07K 14/47 (2006.01)

(52) **U.S. Cl.** **435/6**; 530/350; 536/24.3

(57) **ABSTRACT**

The invention provides methods and nucleic acids for detecting, differentiating or distinguishing between colon cell proliferative disorders by analysis of one or more of the genes Versican, TPEF, H-Cadherin, Calcitonin, and EYA4. The invention further provides novel nucleic acid sequences useful for the cell proliferative disorder specific analysis of said genes as well as methods, assays and kits thereof.

FIGURE 1

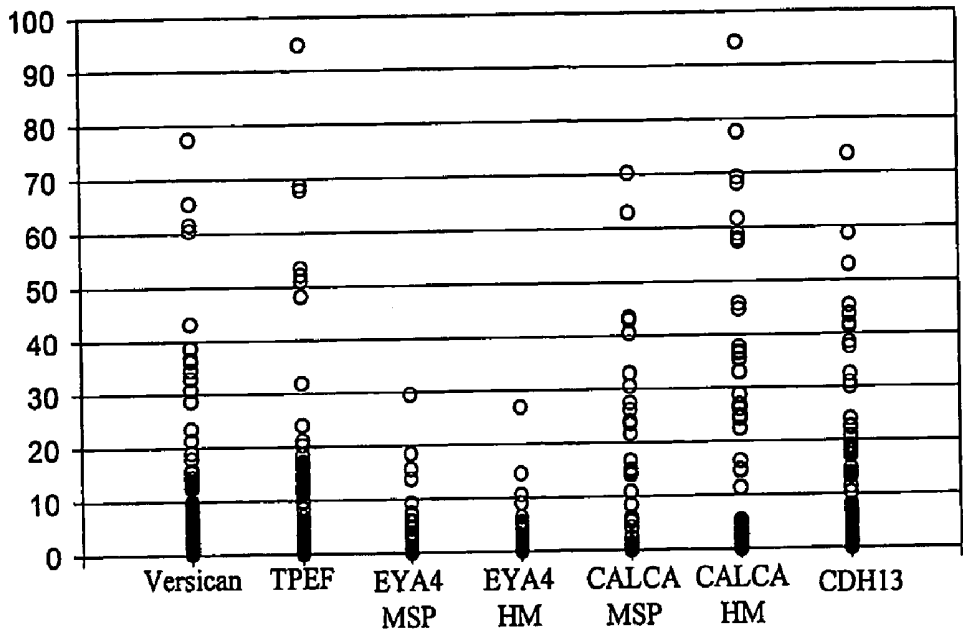

100
90
80
70
60
50
40
30
20
10
0
Versican
TPEF
EYA4 MSP
EYA4 HM
CALCA MSP
CALCA HM
CDH13

FIGURE 2

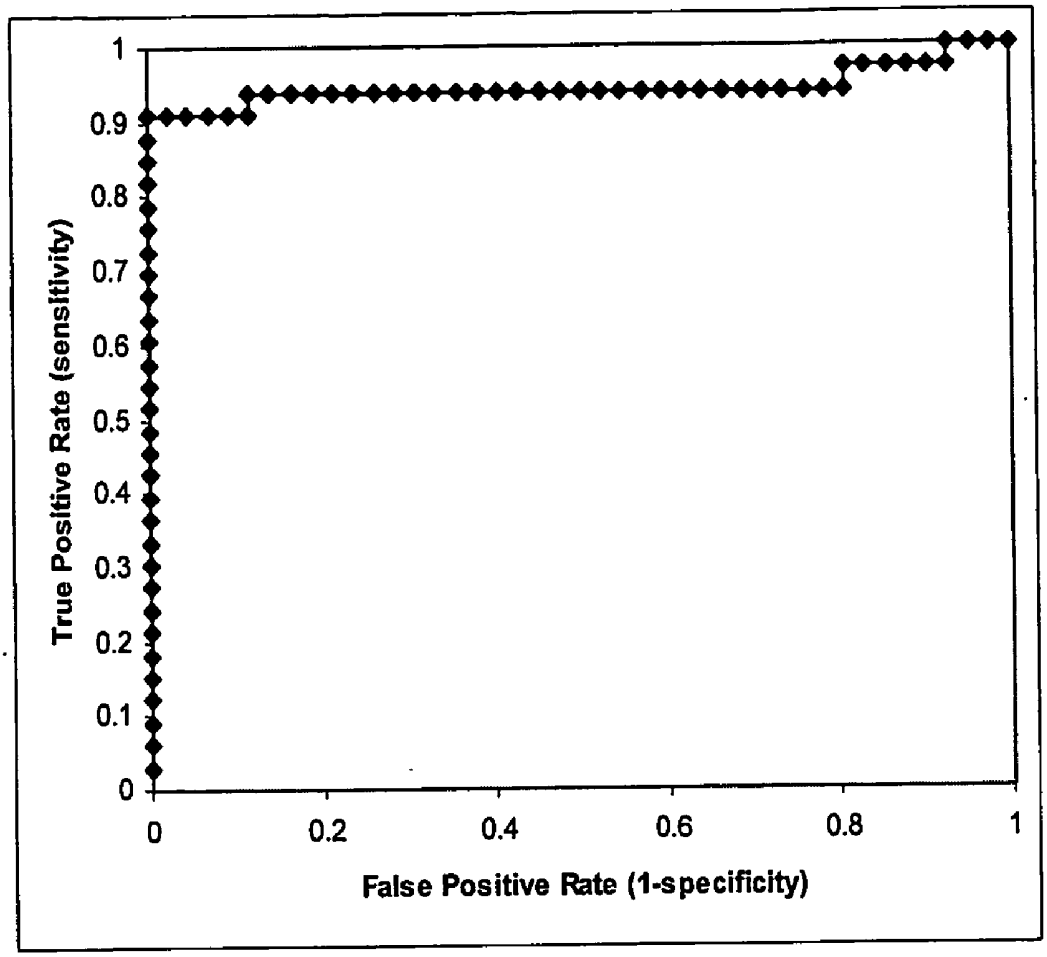

True Positive Rate (sensitivity)
False Positive Rate (1-specificity)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1

FIGURE 3

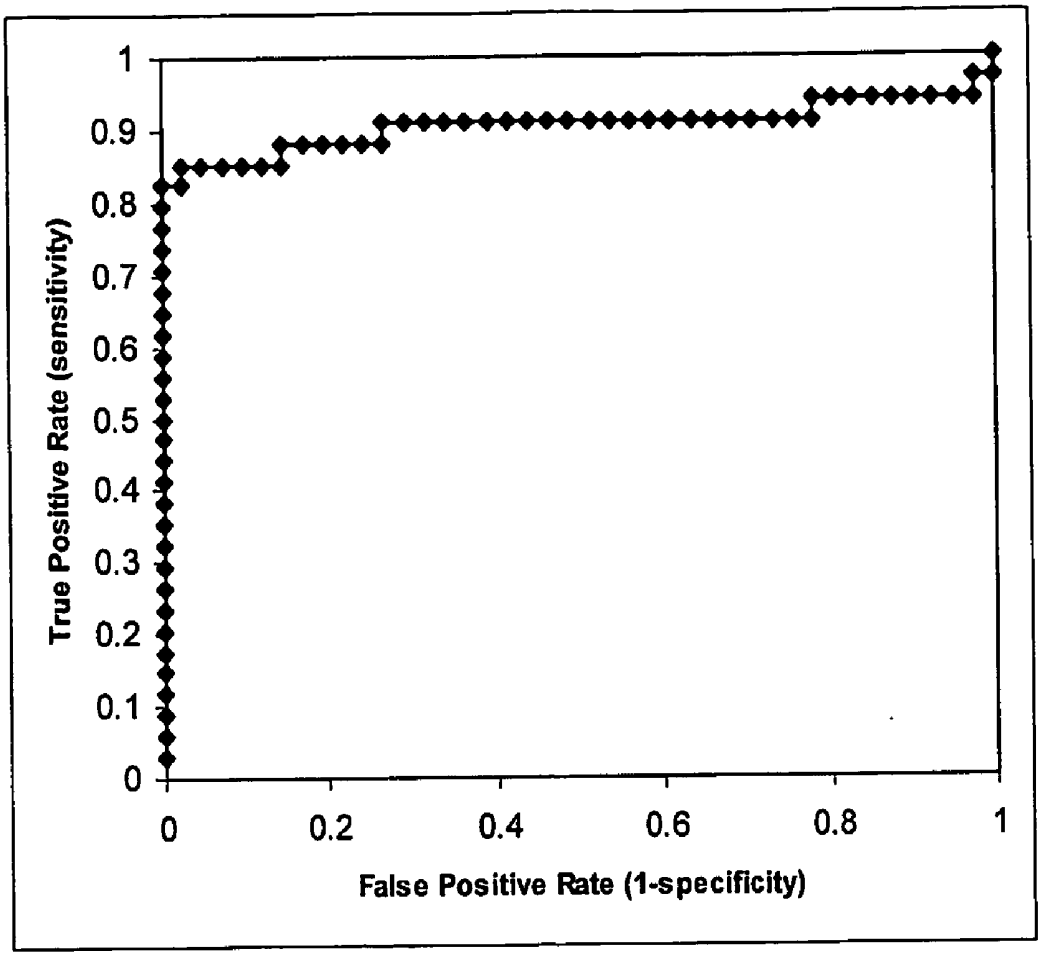

True Positive Rate (sensitivity)
False Positive Rate (1-specificity)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1

FIGURE 4

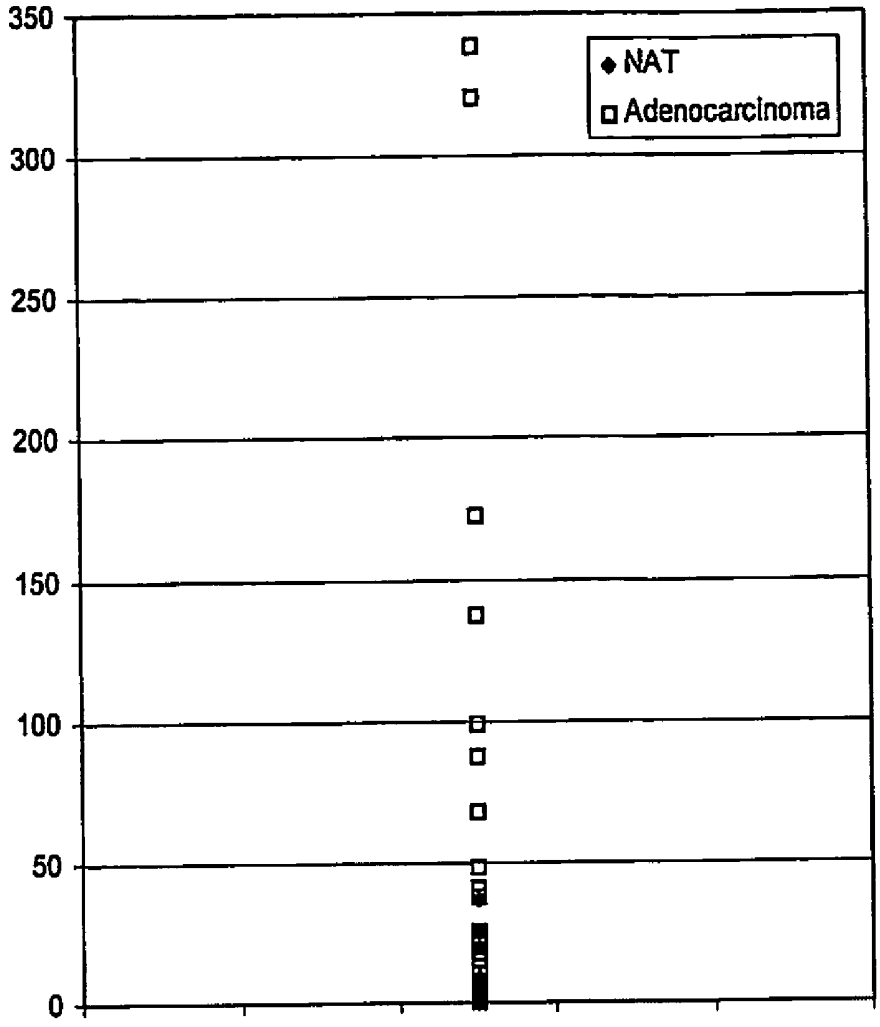


FIGURE 5

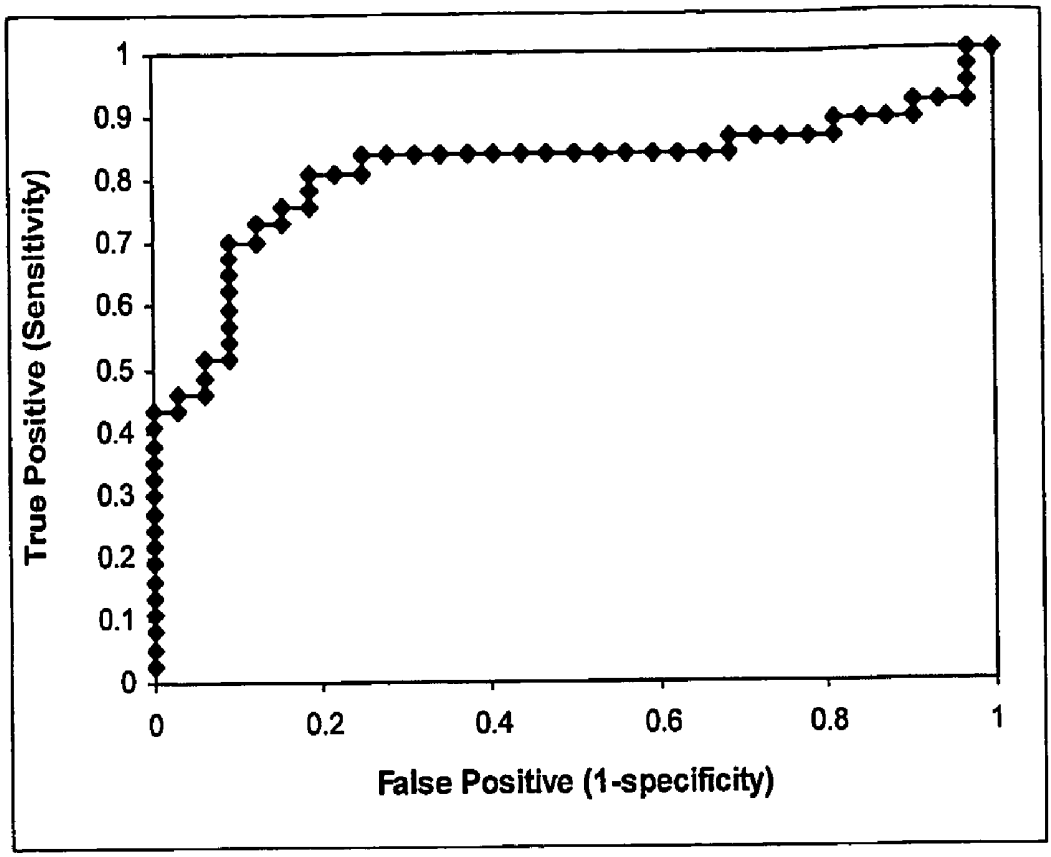

True Positive (Sensitivity)
False Positive (1-specificity)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1

FIGURE 6

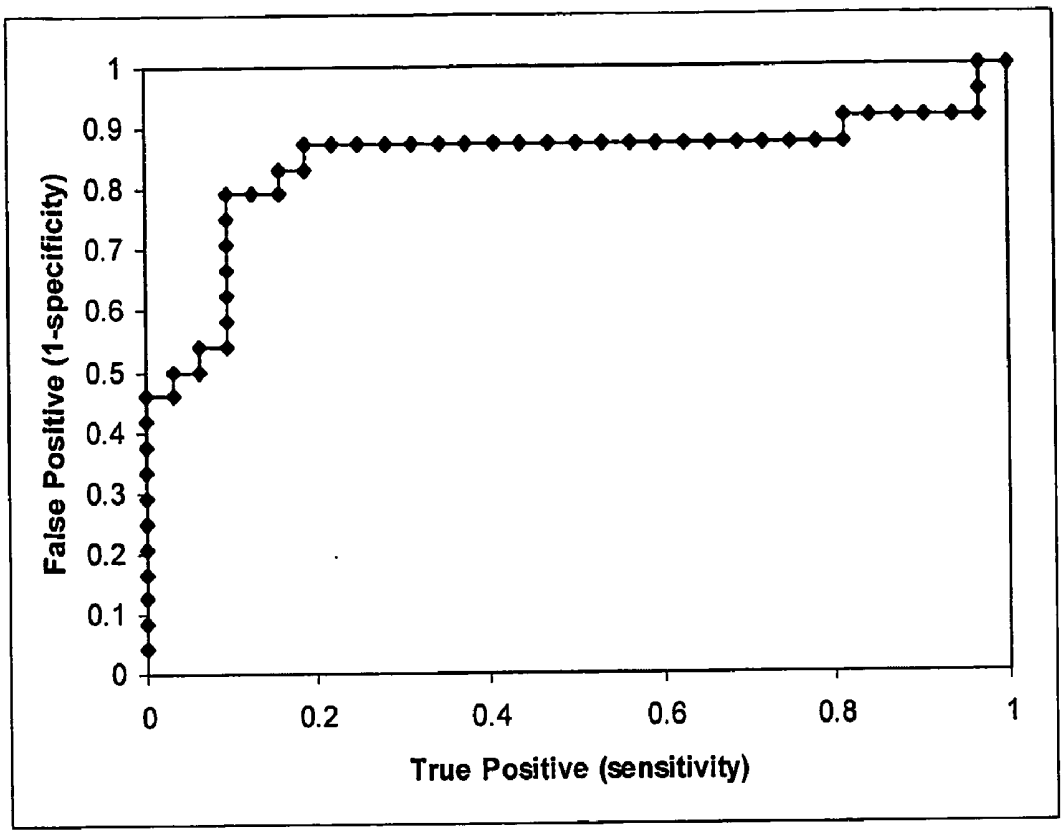

False Positive (1-specificity)
True Positive (sensitivity)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1

FIGURE 7

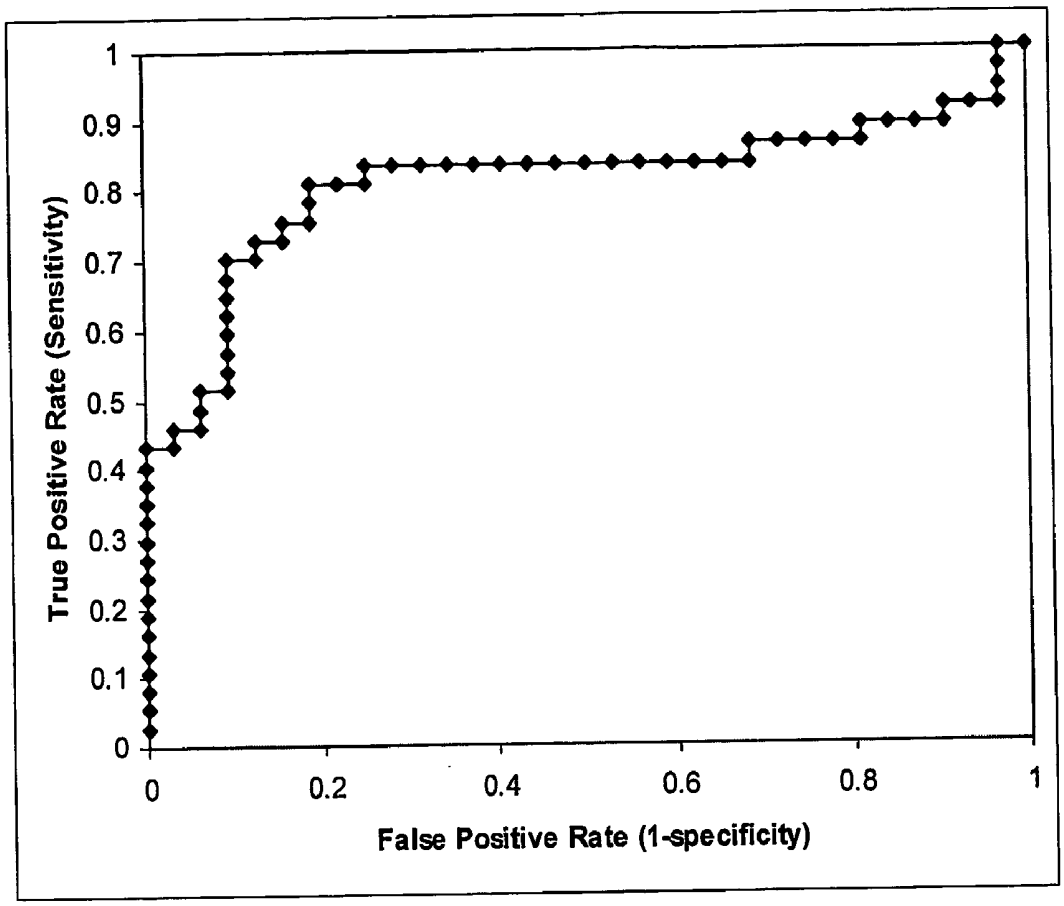

True Positive Rate (Sensitivity)
False Positive Rate (1-specificity)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1

FIGURE 8

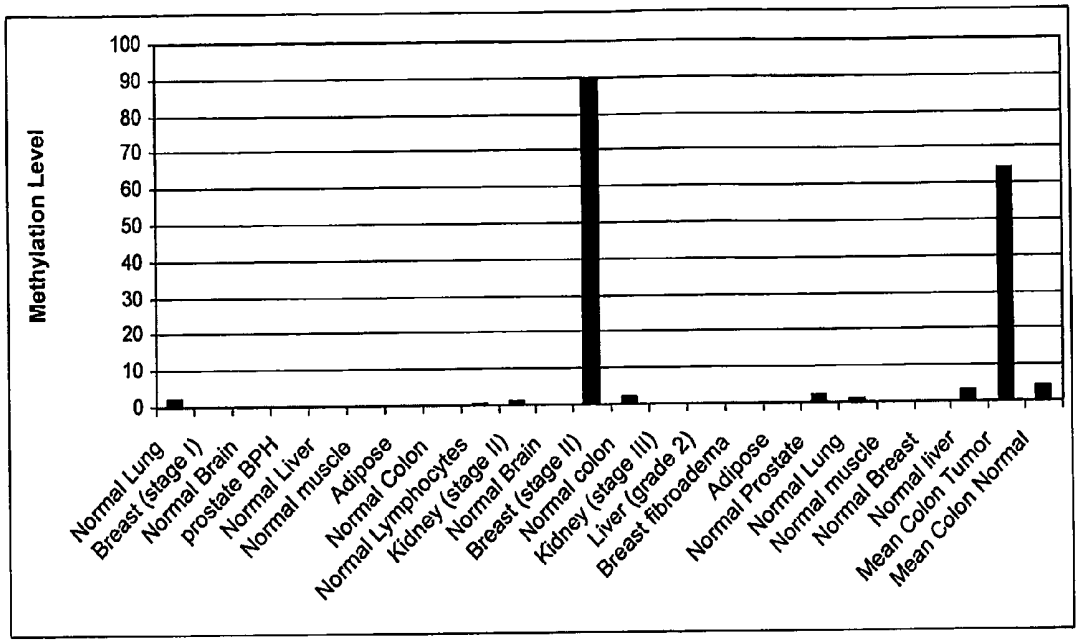

Methylation Level
100
90
80
70
60
50
40
30
20
10
0
Normal Lung
Breast (stage I)
Normal Brain
prostate BPH
Normal Liver
Normal muscle
Adipose
Normal Colon
Normal Lymphocytes
Kidney (stage II)
Normal Brain
Breast (stage II)
Normal colon
Kidney (stage III)
Liver (grade 2)
Breast fibroadema
Adipose
Normal Prostate
Normal Lung
Normal muscle
Normal Breast
Normal liver
Mean Colon Tumor
Mean Colon Normal

FIGURE 9

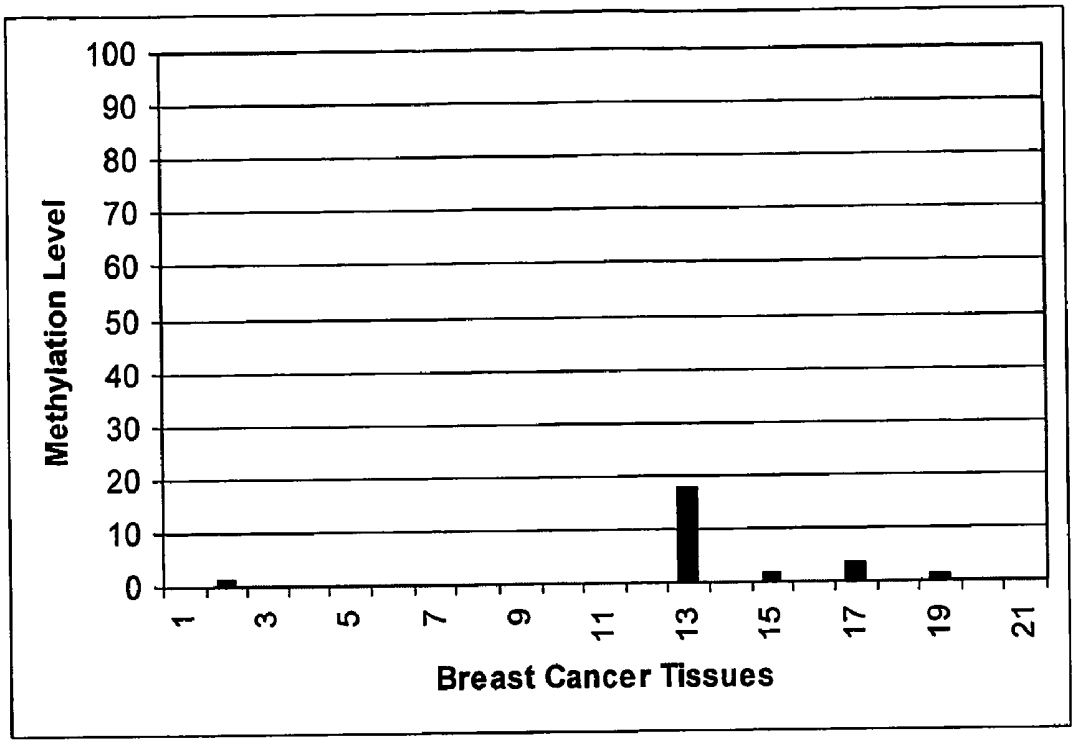

Methylation Level
100
90
80
70
60
50
40
30
20
10
0
1
3
5
7
9
11
13
15
17
19
21
Breast Cancer Tissues

FIGURE 10

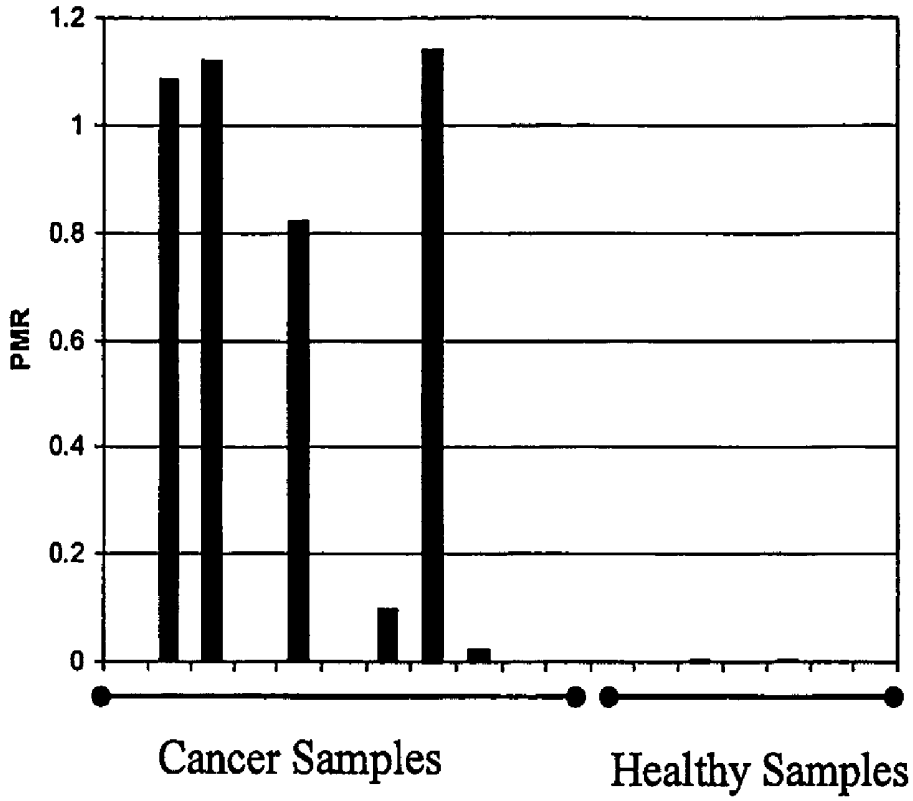


FIGURE 11

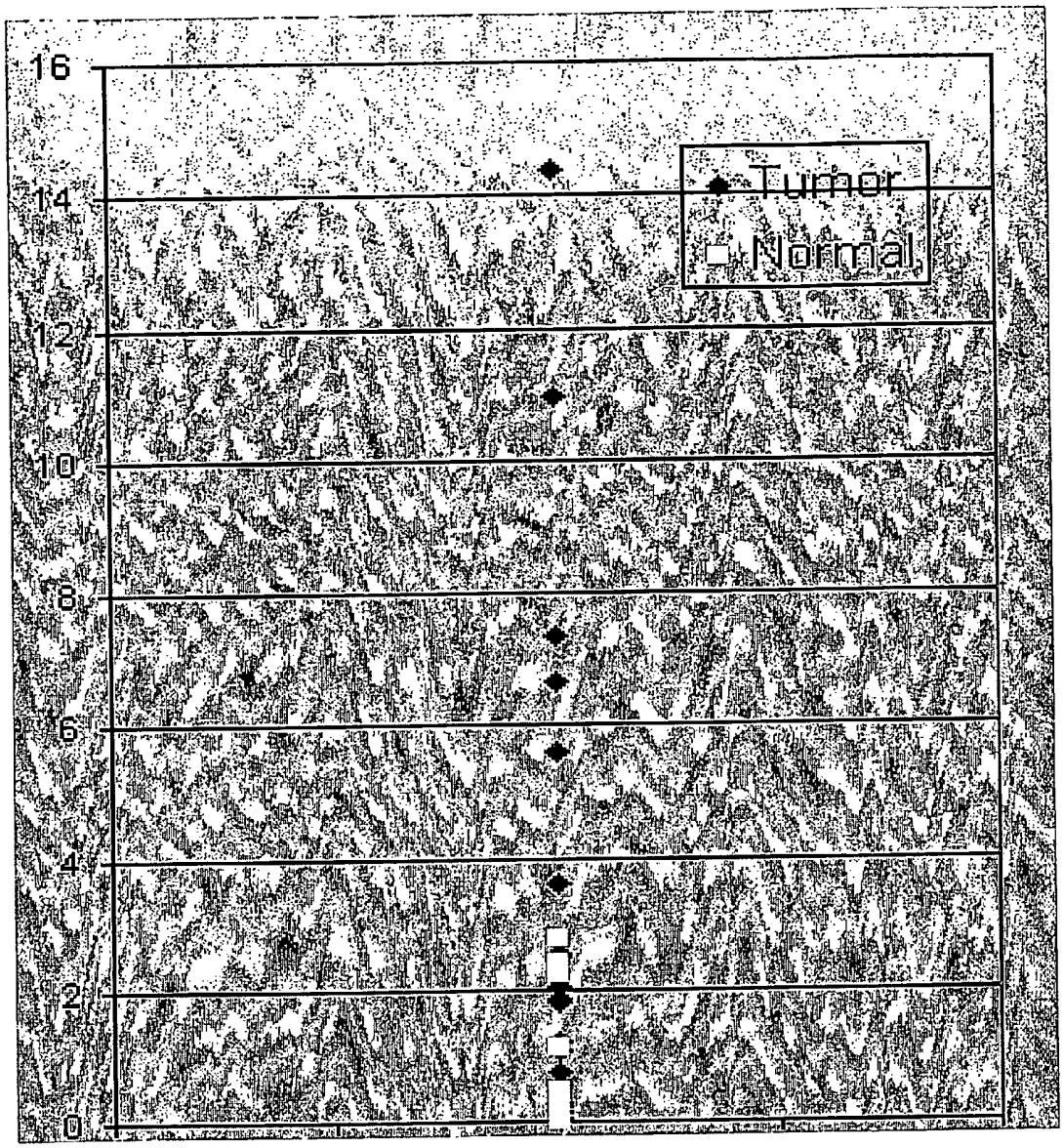


FIGURE 12

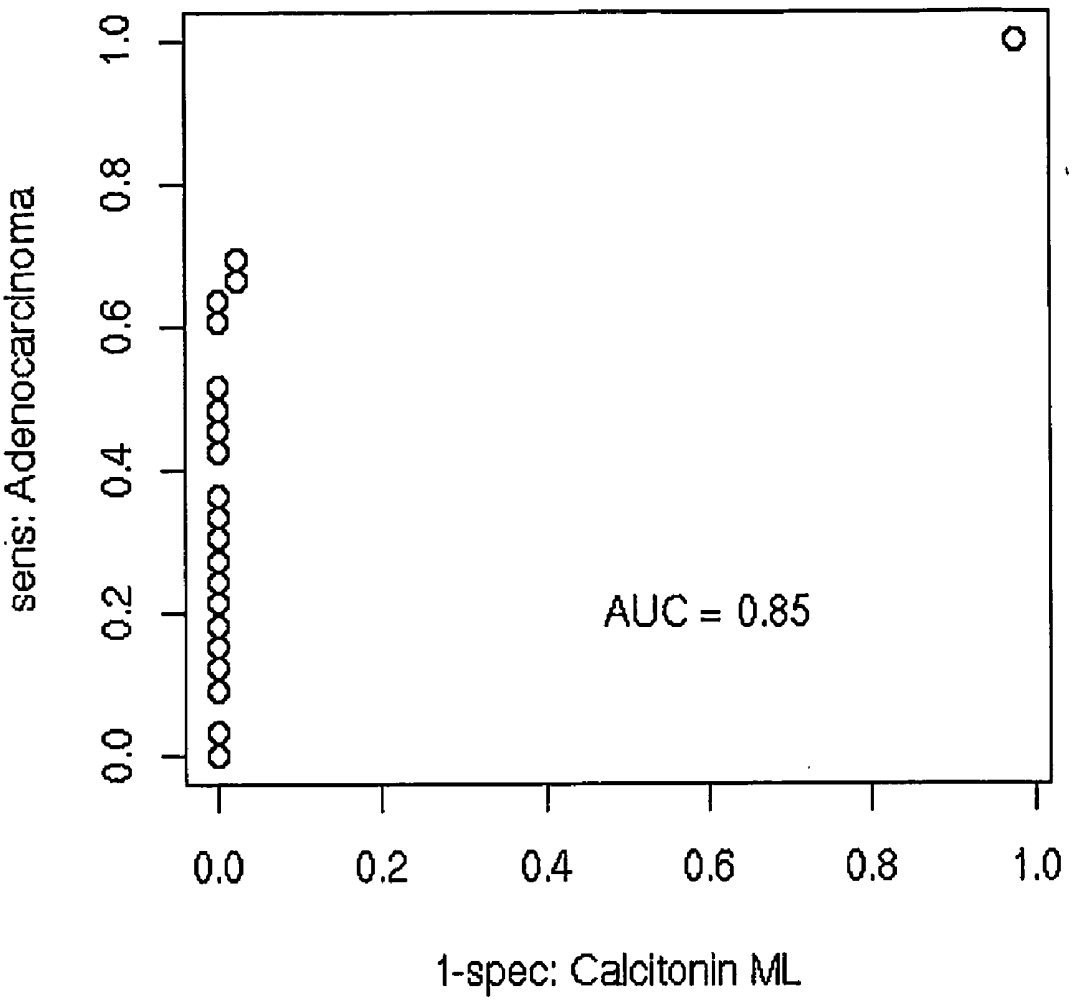

sens: Adenocarcinoma
0.0
0.2
0.4
0.6
0.8
1.0
AUC = 0.85
0.0
0.2
0.4
0.6
0.8
1.0
1-spec: Calcitonin ML

FIGURE 13

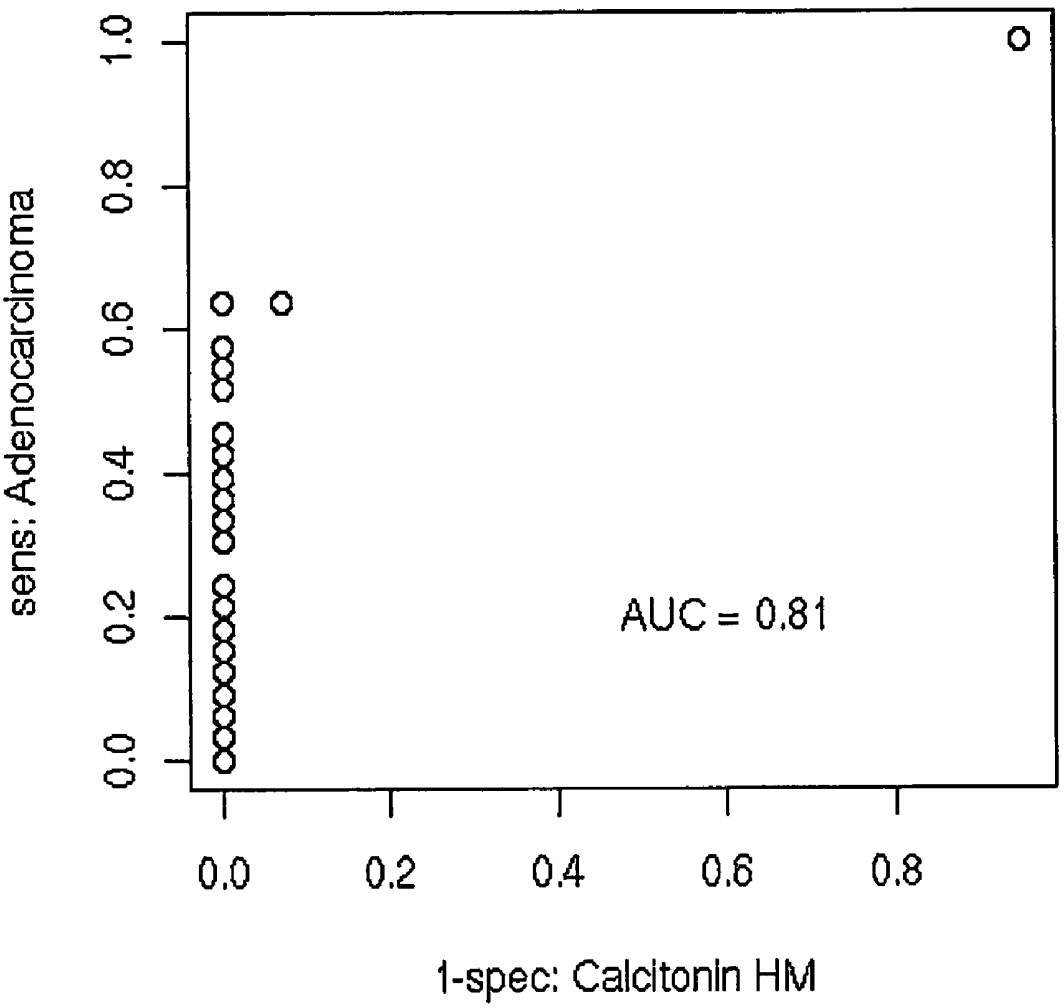

sens: Adenocarcinoma
0.0
0.2
0.4
0.6
0.8
1.0
AUC = 0.81
0.0
0.2
0.4
0.6
0.8
1-spec: Calcitonin HM

Figure 14

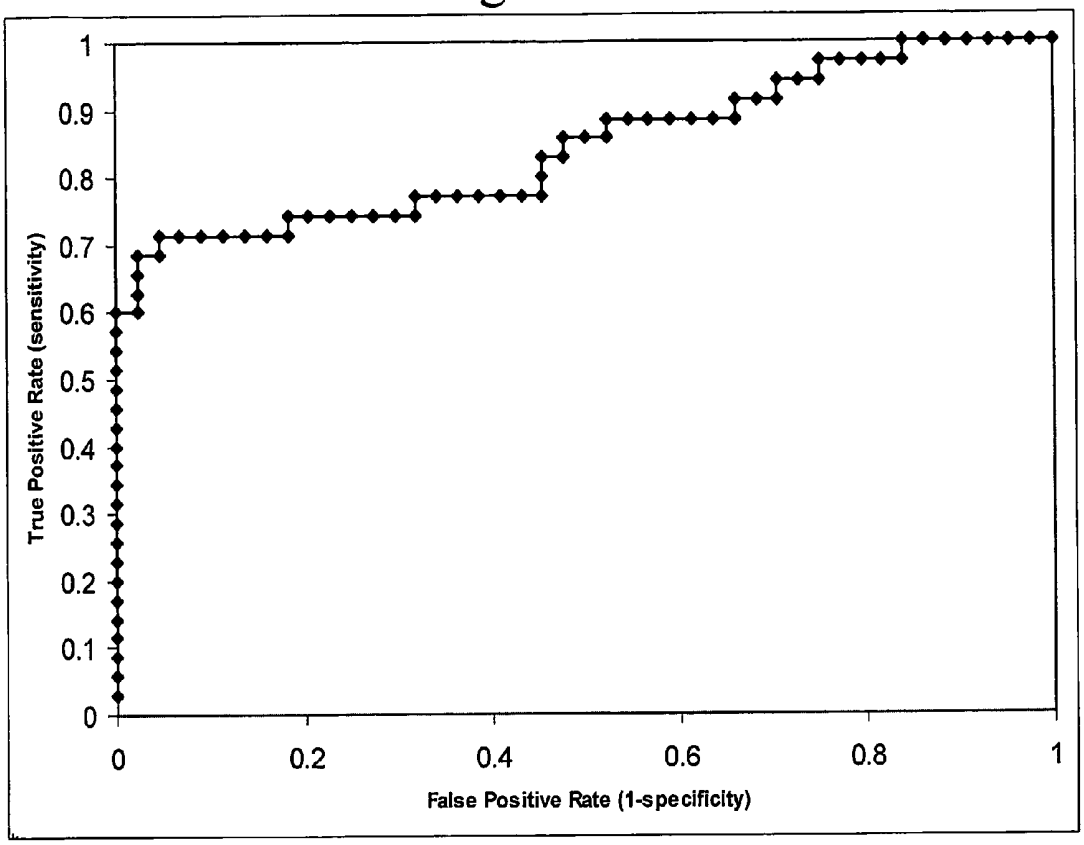

True Positive Rate (sensitivity)
False Positive Rate (1-specificity)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1

Figure 15

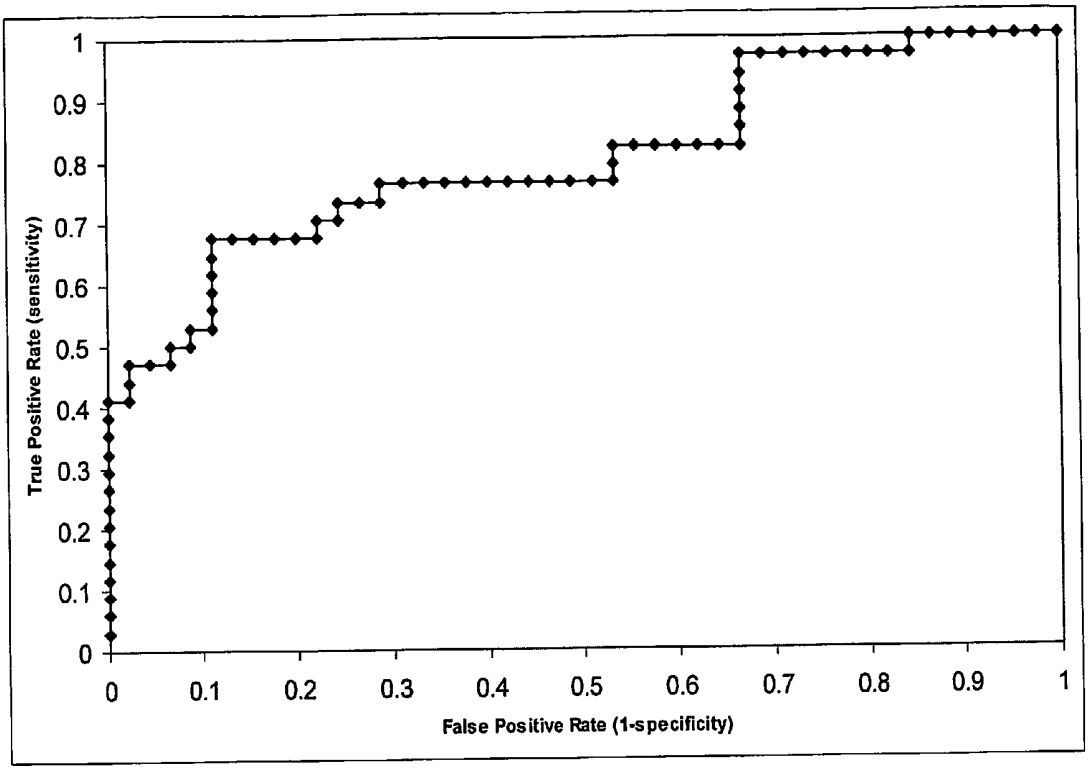

True Positive Rate (sensitivity)
False Positive Rate (1-specificity)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Figure 16

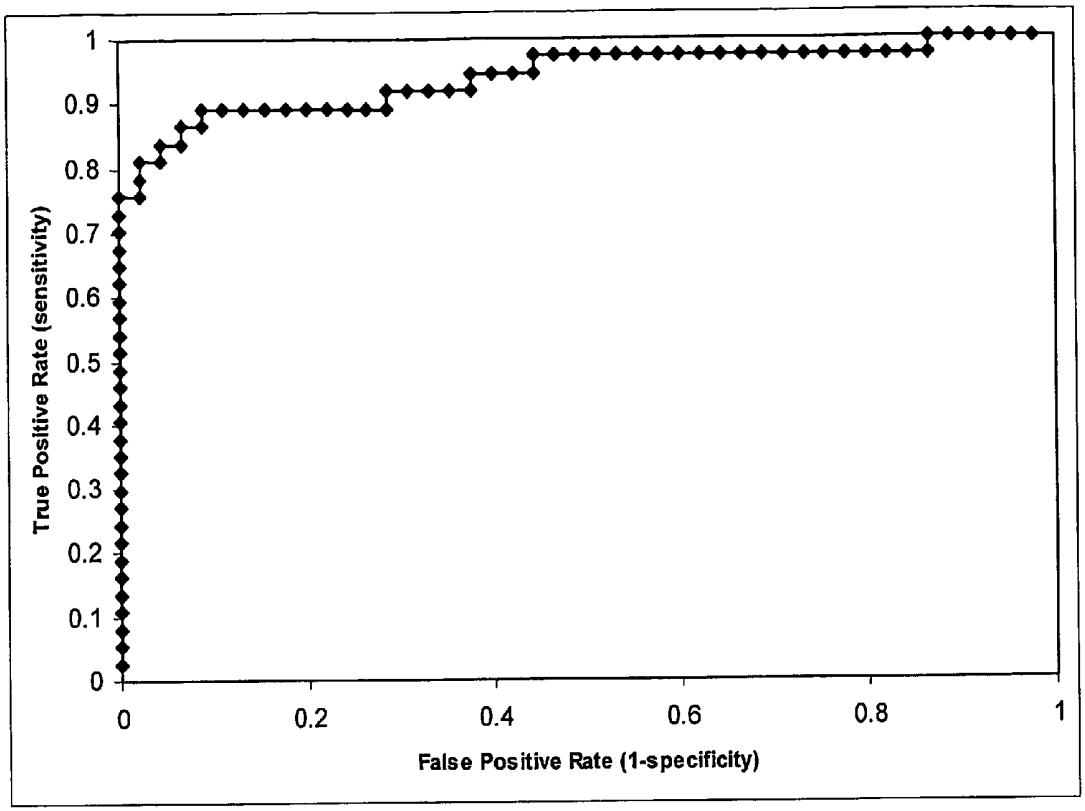

True Positive Rate (sensitivity)
False Positive Rate (1-specificity)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1

FIGURE 17

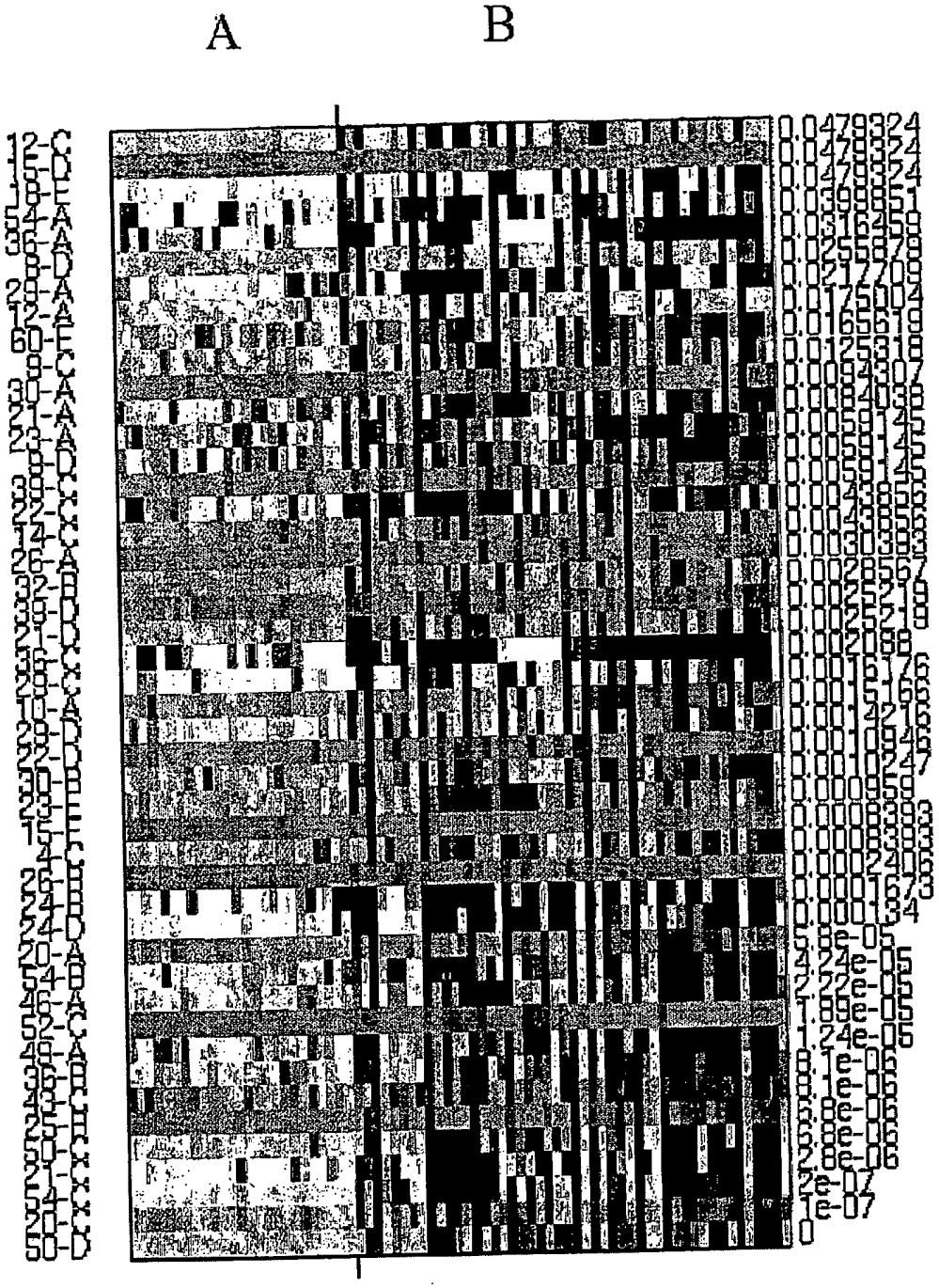


FIGURE 18

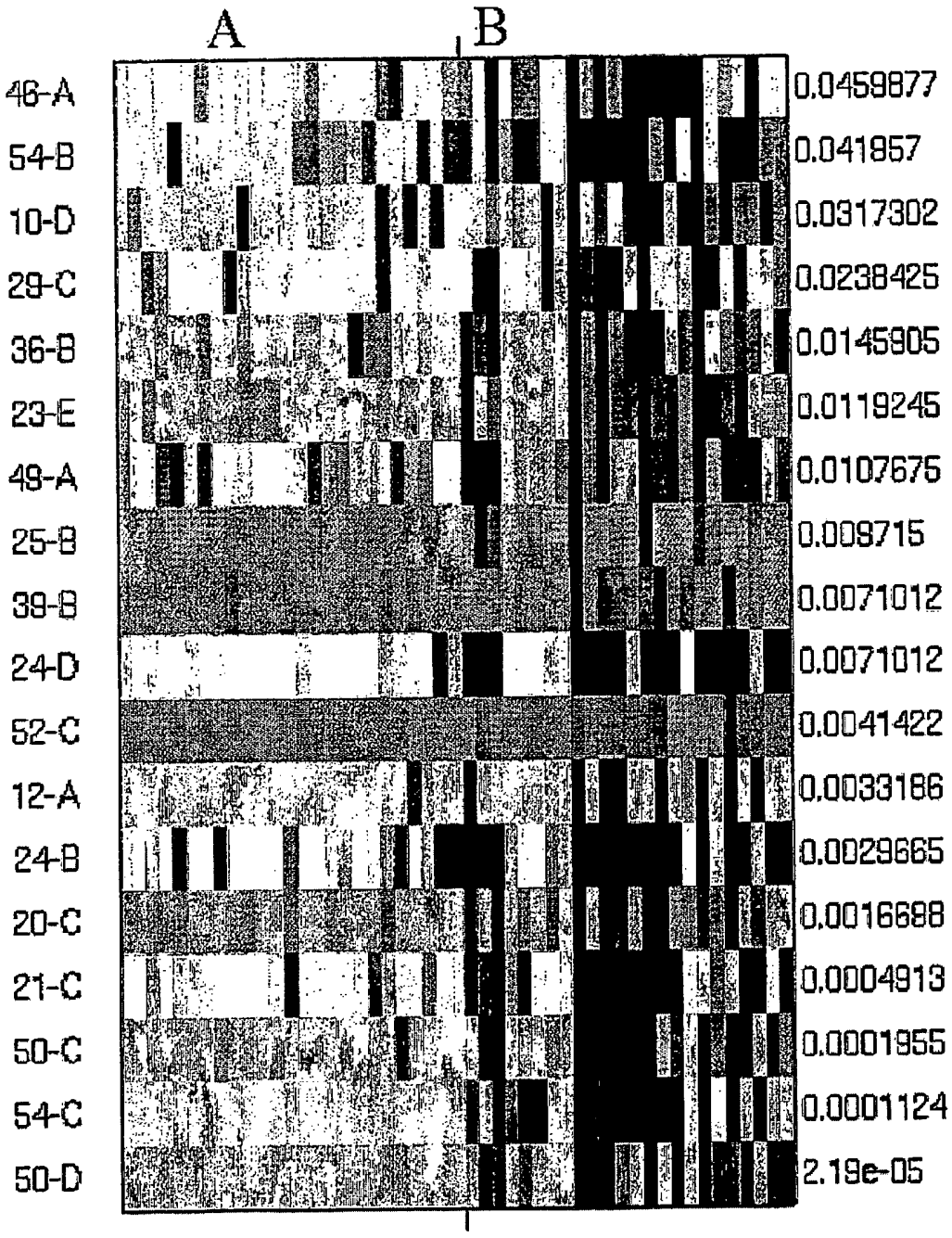

A
B
46-A
54-B
10-D
29-C
36-B
23-E
49-A
25-B
39-B
24-D
52-C
12-A
24-B
20-C
21-C
50-C
54-C
50-D
0.0459877
0.041857
0.0317302
0.0238425
0.0145905
0.0119245
0.0107675
0.009715
0.0071012
0.0071012
0.0041422
0.0033186
0.0029665
0.0016688
0.0004913
0.0001955
0.0001124
2.19e-05

FIGURE 19

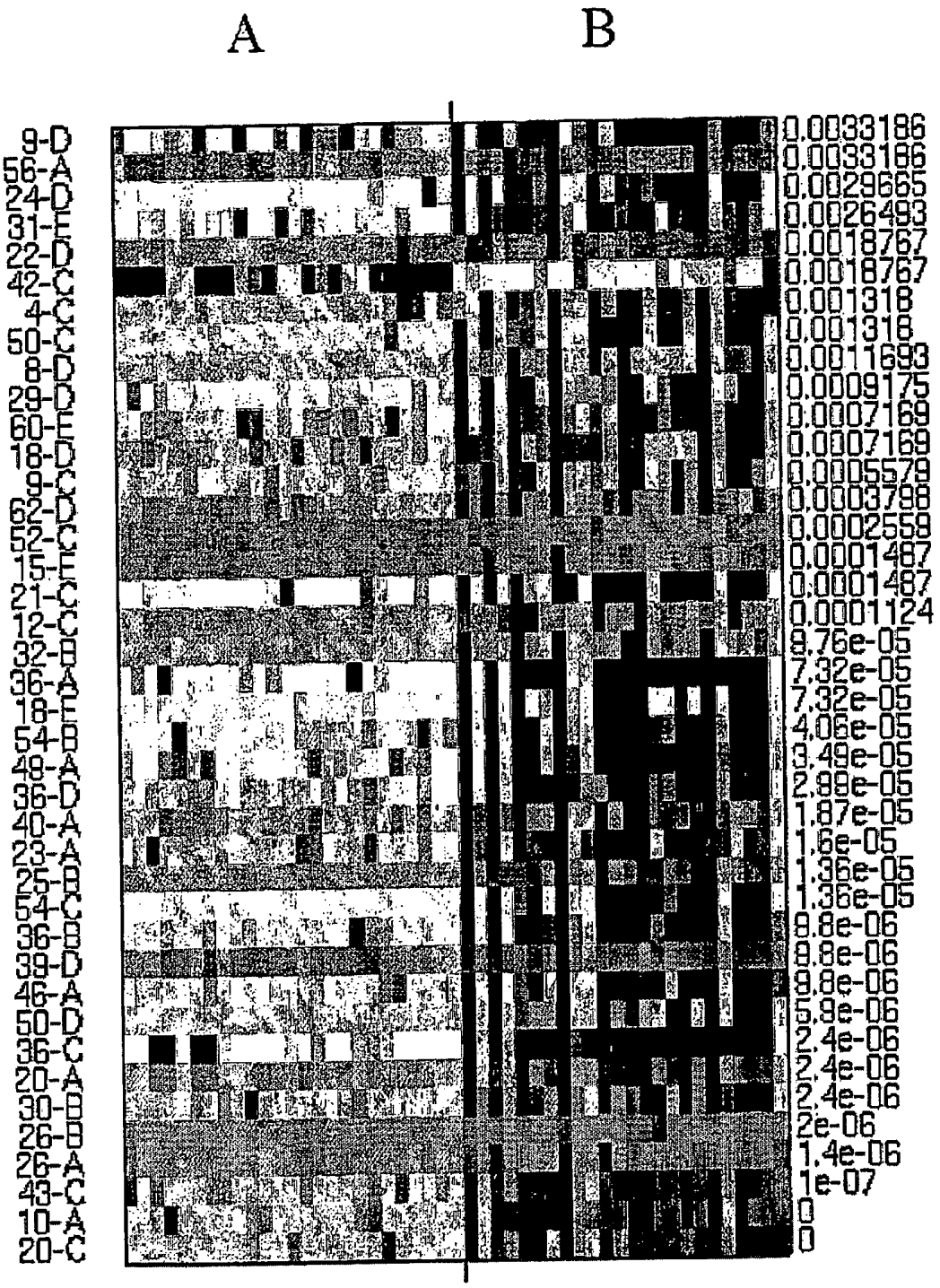

A
B
9-D
56-A
24-D
31-E
22-D
42-C
4-C
50-C
8-D
29-D
60-E
18-D
9-C
62-D
52-C
15-E
21-C
12-C
32-B
36-A
18-E
54-B
48-A
36-D
40-A
23-A
25-B
54-C
36-B
39-D
46-A
50-D
36-C
20-A
30-B
26-B
26-A
43-C
10-A
20-C
0.0033186
0.0033186
0.0029665
0.0026493
0.0018767
0.0018767
0.001318
0.001318
0.0011693
0.0009175
0.0007169
0.0007169
0.0005579
0.0003798
0.0002558
0.0001487
0.0001487
0.0001124
8.76e-05
7.32e-05
7.32e-05
4.06e-05
3.49e-05
2.98e-05
1.87e-05
1.6e-05
1.36e-05
1.36e-05
9.8e-06
9.8e-06
9.8e-06
5.9e-06
2.4e-06
2.4e-06
2.4e-06
2e-06
1.4e-06
1e-07
0
0

FIGURE 20

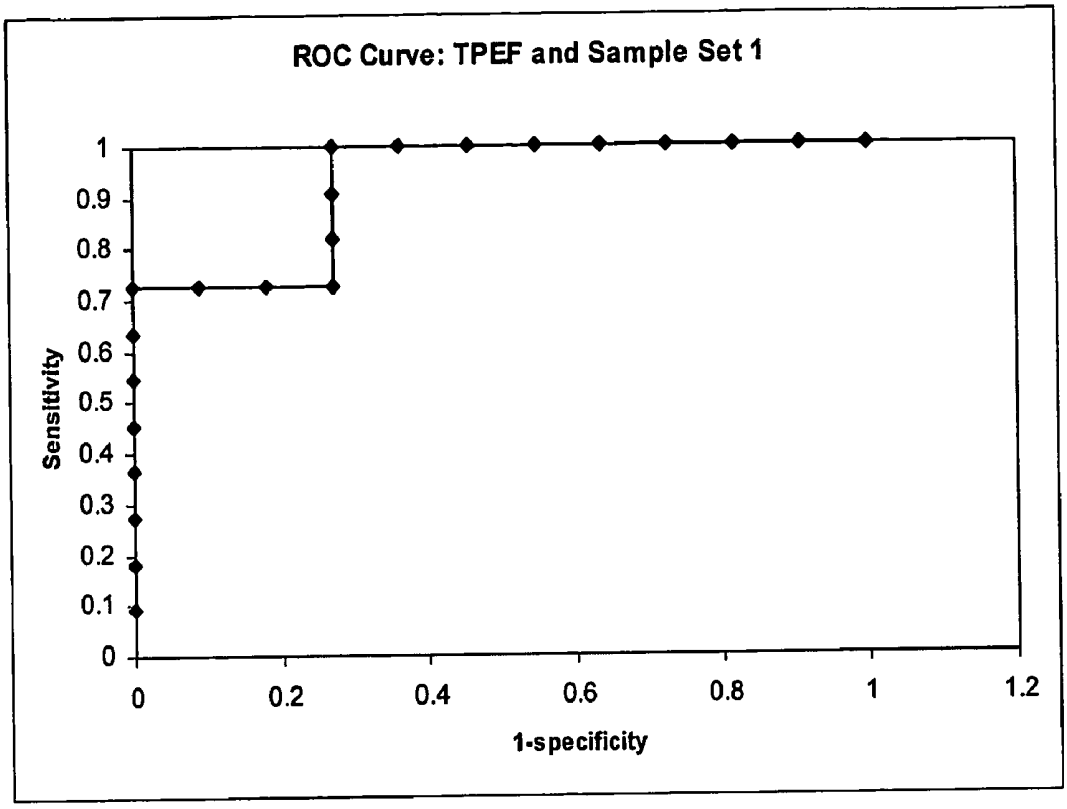

ROC Curve: TPEF and Sample Set 1
Sensitivity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
1.2
1-specificity

FIGURE 21

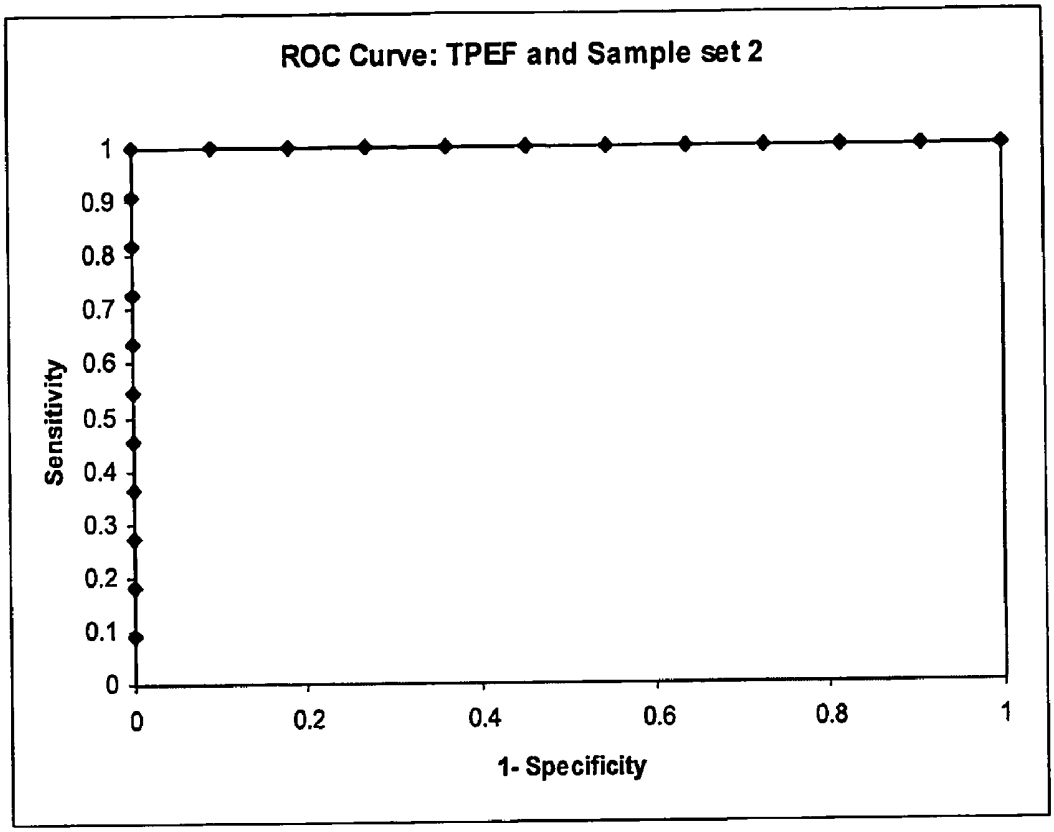

ROC Curve: TPEF and Sample set 2
Sensitivity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
1- Specificity

Figure 22

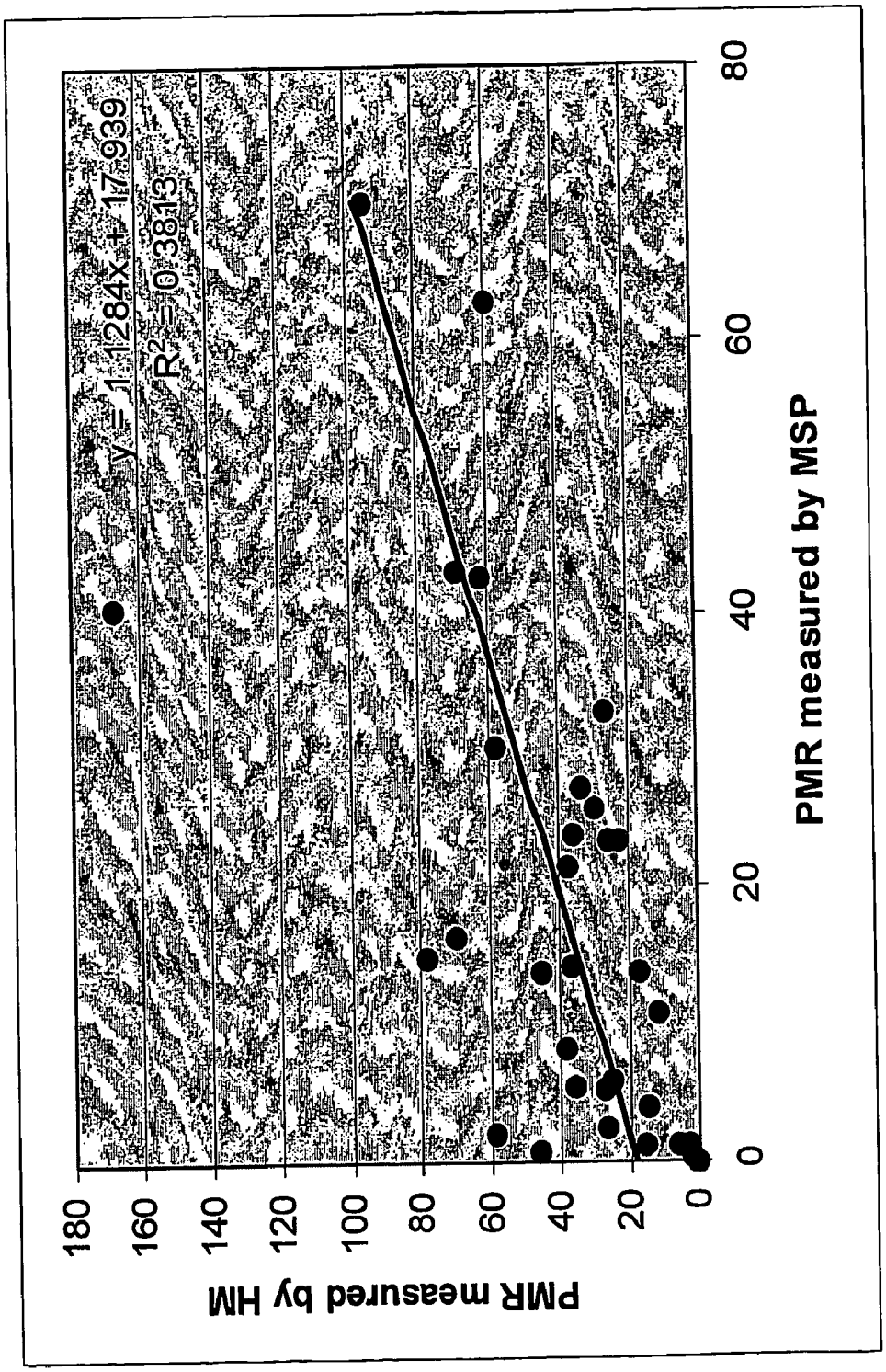
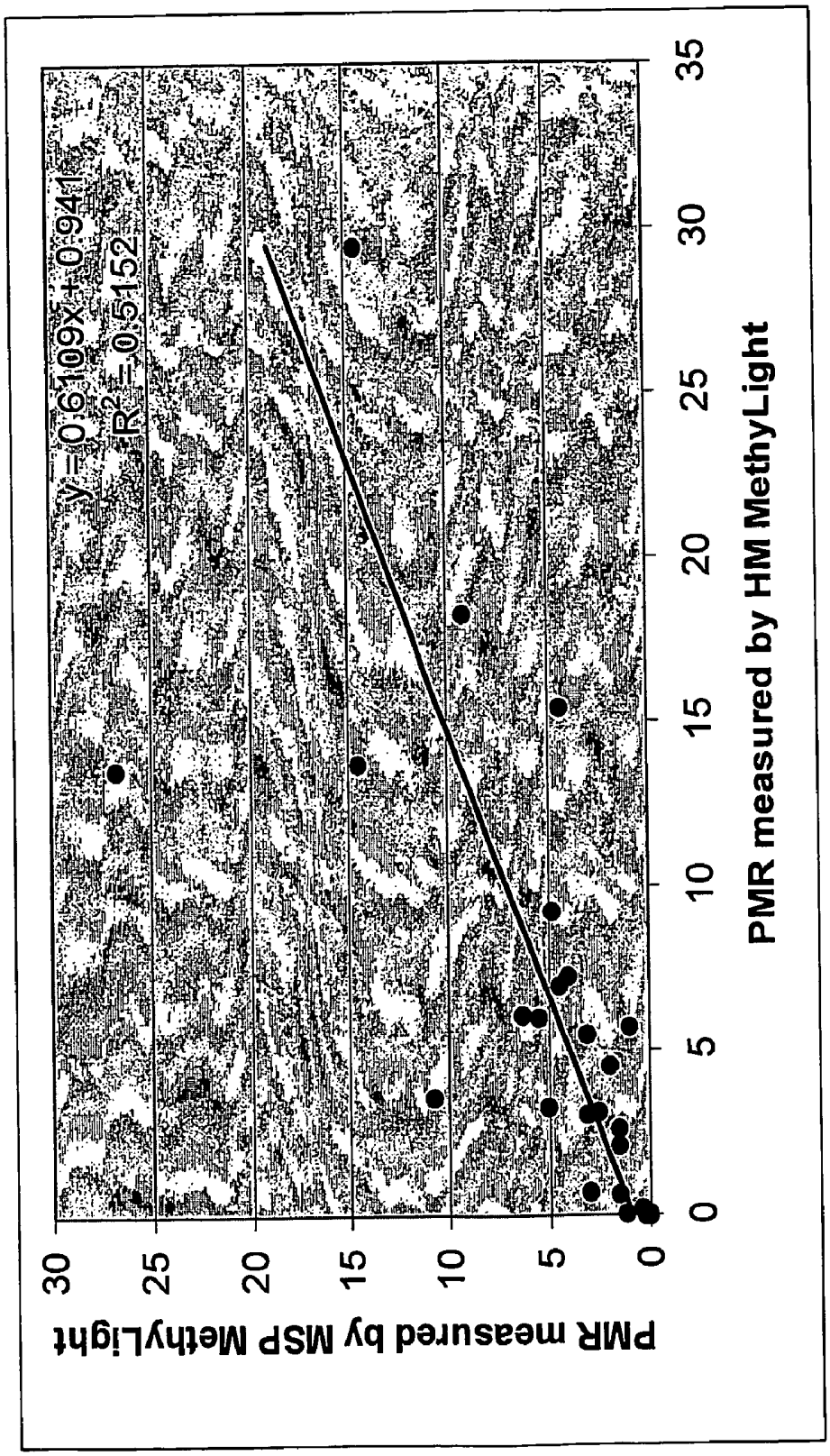

y = 1.1284x + 17.939
R² = 0.3813
PMR measured by HM
PMR measured by MSP
0
20
40
60
80
100
120
140
160
180

Figure 23


y = 0.6109x + 0.941
R² = 0.5152
PMR measured by MSP MethyLight
PMR measured by HM MethyLight
0
5
10
15
20
25
30
35

METHOD AND NUCLEIC ACIDS FOR THE ANALYSIS OF COLON CELL PROLIFERATIVE DISORDERS

FIELD OF THE INVENTION

[0001] Colon cancer is the fourth leading cause of cancer mortality in men and women. The 5-year survival rate is 61% over all stages with early detection being a prerequisite for curative therapy of the disease. Up to 95% of all colorectal cancers are adenocarcinomas of varying differentiation grades.

[0002] Sporadic colon cancer develops in a multistep process starting with the pathological transformation of normal colonic epithelium to an adenoma which consecutively progresses to invasive cancer. The progression rate of colonic adenomas is currently predicted based on their histological appearance, location, degree of spread and extent of bowel involvement. For example, tubular-type benign adenomas rarely progress to malignant tumours, whereas villous benign adenomas, particularly if larger than 2 cm in diameter, have a significant malignant potential.

[0003] During progression from benign proliferative lesions to malignant neoplasms several genetic and epigenetic alterations are known to occur. Somatic mutation of the APC gene seems to be one of the earliest events in 75 to 80% of colorectal adenomas and carcinomas. Activation of K-RAS is thought to be a critical step in the progression towards a malignant phenotype. Consecutively, mutations in other oncogenes as well as alterations leading to inactivation of tumor suppressor genes accumulate.

[0004] Aberrant DNA methylation within CpG islands is among the earliest and most common alterations in human cancers leading to abrogation or overexpression of a broad spectrum of genes. In addition, abnormal methylation has been shown to occur in CpG rich regulatory elements in intronic and coding parts of genes for certain tumours. In contrast to the specific hypermethylation of tumour suppressor genes, an overall hypomethylation of DNA can be observed in tumour cells. This decrease in global methylation can be detected early, far before the development of frank tumour formation. Also, correlation between hypomethylation and increased gene expression was reported for many oncogenes. In colon cancer, aberrant DNA methylation constitutes one of the most prominent alterations and inactivates many tumour suppressor genes such as p14ARF, p16INK4a, THBS1, MINT2, and MINT31 and DNA mismatch repair genes such as hMLH1.

[0005] In the molecular evolution of colorectal cancer, DNA methylation errors have been suggested to play two distinct roles. In normal colonic mucosa cells, methylation errors accumulate as a function of age or as time-dependent events predisposing these cells to neoplastic transformation. For example, hypermethylation of several loci could be shown to be already present in adenomas, particularly in the tubulovillous and villous subtype. At later stages, increased DNA methylation of CpG islands plays an important role in a subset of tumours affected by the so called CpG island methylator phenotype (CIMP). Most CIMP+ tumours, which constitute about 15% of all sporadic colorectal cancers, are characterized by microsatellite instability (MIN) due to hypermethylation of the hMLH1 promoter and other DNA mismatch repair genes. By contrast, CIMP-colon cancers evolve along a more classic genetic instability pathway (CIN), with a high rate of p53 mutations and chromosomal changes.

[0006] However, the molecular subtypes do not only show varying frequencies regarding molecular alterations. According to the presence of either micro satellite instability or chromosomal aberrations, colon cancer can be subclassified into two classes, which also exhibit significant clinical differences. Almost all MIN tumours originate in the proximal colon (ascending and transversum), whereas 70% of CIN tumours are located in the distal colon and rectum. This has been attributed to the varying prevalence of different carcinogens in different sections of the colon. Methylating carcinogens, which constitute the prevailing carcinogen in the proximal colon have been suggested to play a role in the pathogenesis of MIN cancers, whereas CIN tumours are thought to be more frequently caused by adduct-forming carcinogens, which occur more frequently in distal parts of the colon and rectum. Moreover, MIN tumours have a better prognosis than do tumours with a CIN phenotype and respond better to adjuvant chemotherapy.

[0007] The identification of markers for the differentiation of colon carcinoma as well as for early detection are main goals of current research.

[0008] The alpha-calcitonin gene encodes a small family of peptides: calcitonin, katacalcin, and calcitonin gene-related peptide (CGRP). Calcitonin is concerned with skeletal integrity, the secretion of calcitonin is, in part, oestrogen dependent, and it appears likely that a postmenopausal decline in calcitonin secretion is a factor in the development of postmenopausal osteoporosis.

[0009] Investigation of the Calcitonin gene has revealed that hypermethylation of the promoter region of the gene is present in neoplastic cells of several cancer types, particularly acute leukemias. The major part of said research was carried out using methylation sensitive enzyme based methods, this identified the general phenomenon of hypermethylation within the promoter and first exon regions of the gene in multiple types of cancers. However said methods do not allow for a targeted analysis of selected CpG positions. The observations of hypermethylation were enabled only by the assumption of comethylation within the region. Comethylation is a phenomenon whereby methylation of one CpG position is taken as indicative of methylation of all CpG positions within the region. Examples of research carried out using restriction enzyme based methods include the following:

[0010] Hiltunen M O, Koistinaho J, Alhonen L, Myohanen S, Marin S, Kosma V M, Paakkonen M, Janne J. Hypermethylation of the WT1 and calcitonin gene promoter regions at chromosome 11p in human colorectal cancer. Br J Cancer. 1997; 76(9):1124-30.

[0011] Silverman A L, Park J G, Hamilton S R, Gazdar A F, Luk G D, Baylin S B. Abnormal methylation of the calcitonin gene in human colonic neoplasms. Cancer Res. 1989 Jul. 1; 49(13):3468-73

[0012] The gene 'Versican', (NM_004385) also known as CSPG2 encodes a large chondroitin sulfate proteoglycan. This gene is known to exhibit aberrant methylation patterns and conflicting opinions on this matter have been published. For instance Adany R and lozzo R V. ('Altered methylation

of versican proteoglycan gene in human colon carcinoma.' Biochem Biophys Res Commun 1990 Sep. 28; 171(3):1402-13) observed a correlation between hypomethylation and colonic neoplasms. However, more recently Issa et. al. ('Accelerated Age-related CpG Island Methylation in Ulcerative Colitis' Cancer Research 61, 3573-3577, May 1, 2001) described an observed hypermethylation of dysplastic mucosa as compared to non-UC control mucosa (58% versus 31%, P=0.01) or compared with adjacent uninvolved mucosa (58% versus 35%, P=0.06). Therefore it would seem that although aberrant methylation of this gene has been observed in colorectal cell proliferative disorders, the characterisation of this aberrant methylation is as yet not obvious and it would appear that the commonly held assumption of comethylation does not hold in the case of this gene.

[0013] The gene TPEF (also known as TMEFF2) NM_016192 encodes a transmembrane protein containing EGF and follistatin domains. It was initially identified on the basis of its methylation properties by Jones et. al. ('The Gene for a Novel Transmembrane Protein Containing Epidermal Growth Factor and Follistatin Domains Is Frequently Hypermethylated in Human Tumor Cells' Cancer Research 60, 4907-4912, Sep. 1, 2000). It was therein shown that the 5' region of the gene contained a CpG island, a 3' region of which was shown to exhibit significant hypermethylation in tumor cell lines. Although significant said onbservation was carried out by means of arbitrarily primed PCR, a methodology that is not suitable for application in a clinical or diagnostic setting.

[0014] EYA4 is the most recently identified member of the vertebrate Eya (eyes-absent) gene family, a group of four transcriptional activators that interact with other proteins in a conserved regulatory hierarchy to ensure normal embryologic development. The EYA4 gene is mapped to 6q22.3 and encodes a 640 amino acid protein. The structure of EYA4 conforms to the basic pattern established by EYA1-3, and includes a highly conserved 271 amino acid C-terminus called the eya-homologous region (eyaHR; alternatively referred to as the eya domain or eya homology domain 1) and a more divergent proline-serine-threonine (PST)-rich (34-41%) transactivation domain at the N-terminus. EYA proteins interact with members of the SIX and DACH protein families during early embryonic development. Mutations in the EYA4 gene are responsible for postlingual, progressive, autosomal dominant hearing loss at the DFNA10 locus (Wayne S, Robertson N G, DeClau F, Chen N, Verhoeven K, Prasad S, Tranebjarg L, Morton C C, Ryan A F, Van Camp G, Smith R J: Mutations in the transcriptional activator EYA4 cause late-onset deafness at the DFNA10 locus. Hum Mol Genet 2001 Feb. 1; 10(3):195-200 with further references). A link between the Methylation of Cytosine positions in the EYA 4 gene and cancer has not yet been established.

[0015] The cadherins are a family of cell surface glycoproteins responsible for selective cell recognition and adhesion. Several family members, including CDH1 (E-cadherin) and CDH13 (H-cadherin, NM_001257) are located on the long arm of chromosome 16, while another gene cluster resides on the short arm of chromosome 5. The chromosomal locations of several of the cadherins are sites of frequent loss of heterozygosity in many tumor types. Deletions of 16q are frequent in breast, lung, and other carcinomas. Loss of expression of cadherins has been described in many epithelial cancers, and it may play a role in tumour cell invasion and metastasis CDH13 expression is diminished in breast and lung cancers. In ovarian tumours, the combination of deletion and aberrant methylation has been reported to inactivate CDH13. Aberrant methylation of CDH13 has also been reported in lung cancers.

[0016] 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 since 5-methylcytosine has the same base pairing behaviour as cytosine. Moreover, the epigenetic information carried by 5-methylcytosine is completely lost during PCR amplification.

[0017] A relatively new and currently the most frequently used method for analysing DNA for 5-methylcytosine is based upon the specific reaction of bisulfite with cytosine which, upon subsequent alkaline hydrolysis, is converted to uracil which corresponds to thymidine in its base pairing behaviour. 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 hybridisation behaviour, can now be detected as the only remaining cytosine using "normal" molecular biological techniques, for example, by amplification and hybridisation or sequencing. All of these techniques are based on base pairing which can now be fully exploited. In terms of sensitivity, the prior art is defined by a method which encloses the DNA to be analysed in an agarose matrix, thus preventing the diffusion and renaturation of the DNA (bisulfite only reacts with single-stranded DNA), and which replaces all precipitation and purification steps with fast dialysis (Olek A, Oswald J, Walter J. A modified and improved method for bisulphite based cytosine methylation analysis. Nucleic Acids Res. 1996 Dec. 15; 24(24):5064-6). Using this method, it is possible to analyse individual cells, which illustrates the potential of the method. However, currently only individual regions of a length of up to approximately 3000 base pairs are analysed, a global analysis of cells for thousands of possible methylation events is not possible. However, this method cannot reliably analyse very small fragments from small sample quantities either. These are lost through the matrix in spite of the diffusion protection.

[0018] An overview of the further known methods of detecting 5-methylcytosine may be gathered from the following review article: Rein, T., DePamphilis, M. L., Zorbas, H., Nucleic Acids Res. 1998, 26, 2255.

[0019] To date, barring few exceptions (e.g., Zeschnigk M, Lich C, Buiting K, Doerfler W, Horsthemke B. A single-tube PCR test for the diagnosis of Angelman and Prader-Willi syndrome based on allelic methylation differences at the SNRPN locus. Eur J Hum Genet. 1997 March-April; 5(2):94-8) the bisulfite technique is only used in research. Always, however, short, specific fragments of a known gene are amplified subsequent to a bisulfite treatment and either completely sequenced (Olek A, Walter J. The pre-implantation ontogeny of the H19 methylation imprint.

Nat Genet. 1997 November; 17(3):275-6) or individual cytosine positions are detected by a primer extension reaction (Gonzalgo M L, Jones P A. Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). Nucleic Acids Res. 1997 Jun. 15; 25(12):2529-31, WO 95/00669) or by enzymatic digestion (Xiong Z, Laird P W. COBRA: a sensitive and quantitative DNA methylation assay. Nucleic Acids Res. 1997 Jun. 15; 25(12):2532-4). In addition, detection by hybridization has also been described (Olek et al., WO 99/28498).

[0020] Further publications dealing with the use of the bisulfite technique for methylation detection in individual genes are: Grigg G, Clark S. Sequencing 5-methylcytosine residues in genomic DNA. Bioessays. 1994 June; 16(6):431-6, 431; Zeschnigk M, Schmitz B, Dittrich B, Buiting K, Horsthemke B, Doerfler W. Imprinted segments in the human genome: different DNA methylation patterns in the Prader-Willi/Angelman syndrome region as determined by the genomic sequencing method. Hum Mol Genet. 1997 March; 6(3):387-95; Feil R, Charlton J, Bird A P, Walter J, Reik W. Methylation analysis on individual chromosomes: improved protocol for bisulphite genomic sequencing. Nucleic Acids Res. 1994 Feb. 25; 22(4):695-6; Martin V, Ribieras S, Song-Wang X, Rio M C, Dante R. Genomic sequencing indicates a correlation between DNA hypomethylation in the 5' region of the pS2 gene and its expression in human breast cancer cell lines. Gene. 1995 May 19; 157(1-2):261-4; WO 97/46705 and WO 95/15373.

[0021] 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), published in January 1999, and from the literature cited therein.

[0022] Fluorescently labeled 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 hybridized probes may be carried out, for example via a confocal microscope. Cy3 and Cy5 dyes, besides many others, are commercially available.

[0023] Matrix Assisted Laser Desorption Ionisation Mass Spectrometry (MALDI-TOF) is a very efficient development for the analysis of biomolecules (Karas M, Hillenkamp F. Laser desorption ionisation of proteins with molecular masses exceeding 10,000 daltons. Anal Chem. 1988 Oct. 15; 60(20):2299-301). 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 ionised 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.

[0024] MALDI-TOF spectrometry is excellently suited to the analysis of peptides and proteins. The analysis of nucleic acids is somewhat more difficult (Gut I G, Beck S. DNA and Matrix Assisted Laser Desorption Ionisation Mass Spectrometry. Current Innovations and Future Trends. 1995, 1; 147-57). The sensitivity to nucleic acids is approximately 100 times worse than to peptides and decreases disproportionally with increasing fragment size. For nucleic acids having a multiply negatively charged backbone, the ionisation process via the matrix is considerably less efficient. In MALDI-TOF spectrometry, the selection of the matrix plays an eminently important role. For the desorption of peptides, several very efficient matrixes have been found which produce a very fine crystallisation. There are now several responsive matrixes for DNA, however, the difference in sensitivity has not been reduced. The difference in sensitivity can be reduced by chemically modifying the DNA in such a manner that it becomes more similar to a peptide. 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 I G, Beck S. A procedure for selective DNA alkylation and detection by mass spectrometry. Nucleic Acids Res. 1995 Apr. 25; 23(8):1367-73). The coupling of a charge tag to this modified DNA results in an increase in 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 make the detection of unmodified substrates considerably more difficult.

[0025] Genomic DNA is obtained from DNA of cell, tissue or other test samples using standard methods. This standard methodology is found in references such as Sambrook, Fritsch and Maniatis eds., Molecular Cloning: A Laboratory Manual, 1989.

DESCRIPTION

[0026] The invention provide a method for the analysis of biological samples for features associated with the development of colon cell proliferative disorders, characterised in that the nucleic acid of at least one member of the group comprising Versican, TPEF, H-Cadherin, Calcitonin and EYA4 is/are contacted with a reagent or series of reagents capable of distinguishing between methylated and non methylated CpG dinucleotides within the genomic sequence of interest.

[0027] The genes that form the basis of the present invention may also be used to form a "gene panel", i.e. a collection comprising the particular genetic sequences of the present invention and/or their respective informative methylation sites. The formation of gene panels allows for a quick and specific analysis of specific aspects of breast cancer treatment. The gene panel(s) as described and employed in this invention can be used with surprisingly high efficiency for the the improved diagnosis, treatment and monitoring of colon cell proliferative disorders.

[0028] The invention provides significant improvements over the state of the art in that there are currently no markers used to detect colorectal cancer from body fluid samples. Current methods used to detect and diagnose colon cell proliferative disorders 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. Compared to the previous descriptions of these markers in the literature, the described invention provides significant advantages in terms of sensitivity and specificity due to the advantageous combination of using a gene panel and highly sensitive assay techniques.

[0029] The present invention makes available a method for ascertaining genetic and/or epigenetic parameters of

genomic DNA. The method is for use in the improved diagnosis, treatment and monitoring of colon cell proliferative disorders, more specifically by enabling the improved identification of and differentiation between subclasses of said disorder and the genetic predisposition to said disorders. The invention presents improvements over the state of the art in that it enables a highly specific classification of colon carcinomas, thereby allowing for improved and informed treatment of patients.

[0030] In one aspect of the invention, the disclosed matters provides novel nucleic acid sequences useful for the analysis of methylation within said gene, other aspects provide novel uses of the gene and the gene product as well as methods, assays and kits directed to detecting, differentiating and distinguishing colon cell proliferative disorders. The method and nucleic acids according to the invention may be used for the analysis of colon cell proliferative disorders taken form the group comprising adenocarcinomas, polyps, squamous cell cancers, carcinoid tumours, sarcomas and lymphomas.

[0031] In one embodiment the method discloses the use of one or more genes selected from the group comprising Versican, TPEF, H-Cadherin, Calcitonin and EYA4 as markers for the differentiation, detection and distinguishing of colon cell proliferative disorders. Said use of the gene may be enabled by means of analysis of the methylation status of one or more genes selected from the group comprising Versican, TPEF, H-Cadherin, Calcitonin and EYA4 and their promoter or regulatory elements.

[0032] The objective of the invention may be achieved by analysis of the methylation state of the CpG dinucleotides within one or more of the genomic sequence according to SEQ ID NO 1 to SEQ ID NO 5 and sequences complementary thereto. SEQ ID NO 1 to SEQ ID NO 5 disclose the nucleic acid sequences of the genes from the group consisting of Versican, TPEF, H-Cadherin, Calcitonin and EYA4 and their promoter and regulatory elements, wherein said fragment comprises CpG dinucleotides exhibiting a disease specific methylation pattern. Due to the degeneracy of the genetic code, the sequence as identified in SEQ ID NO 1 to SEQ ID NO 5 should be interpreted so as to include all substantially similar and equivalent sequences upstream of the promoter region of a gene which encodes a polypeptide with the biological activity of that encoded by the genes Versican, TPEF, H-Cadherin, Calcitonin and EYA4.

[0033] In a preferred embodiment of the method, the objective of the invention is achieved by analysis of a nucleic acid comprising a sequence of at least 18 bases in length according to one of SEQ ID NO 6 to SEQ ID NO 25 and sequences complementary thereto.

[0034] The sequences of SEQ ID NO 6 to SEQ ID NO 25 provide modified versions of the nucleic acid according to SEQ ID NO 1 to SEQ ID NO 5, wherein the conversion of said sequence results in the synthesis of a nucleic acid having a sequence that is unique and distinct from SEQ ID NO 1 to SEQ ID NO 5 as follows. (see also the following TABLE 1): SEQ ID NO 1 to SEQ ID NO 5, sense DNA strand of Versican, TPEF, H-Cadherin, Calcitonin and EYA4 and their promoter and regulatory elements; SEQ ID NO 6 to SEQ ID NO 15, converted SEQ ID NO 1 to SEQ ID NO 5 and complementary sequences, wherein "C" or "T," but "CpG" remains "CpG" (i.e., corresponds to case where all "C" residues of CpG dinucleotide sequences are methylated and are thus not converted); SEQ ID NO 16 to SEQ ID NO 25, converted SEQ ID NO 1 to SEQ ID NO 5 and complementary sequences, wherein "C" converted to "T" for all "C" residues, including those of "CpG" dinucleotide sequences (i.e., corresponds to case where, for SEQ ID NO 1 to SEQ ID NO 5, all "C" residues of CpG dinucleotide sequences are unmethylated);

TABLE 1

Description of SEQ ID NO 1 to SEQ ID NO 25

SEQ ID NO	Relationship to SEQ ID NO 1 to SEQ ID NO 5	Nature of cytosine base conversion
SEQ ID NO 1 to SEQ ID NO 5	Sense strand (Versican, TPEF, H-Cadherin, Calcitonin and EYA4 including promoter and regulatory elements)	None; untreated sequence
SEQ ID NO 6 to 15	Converted methylated strand	"C"to"T," but "CpG" remains "CpG" (all "C" residues of CpGs are methylated)
SEQ ID NO 16 to 25	Converted sense strand	"C" to "T" for all "C" residues (all "C" residues of CpGs are unmethylated)

[0035] Significantly, heretofore, the nucleic acid sequences and molecules according to SEQ ID NO 6 to SEQ ID NO 25 were not implicated in or connected with the ascertainment of colon cell proliferative disorders.

[0036] The described invention further disclose an oligonucleotide or oligomer for detecting the cytosine methylation state within pretreated DNA, according to SEQ ID NO 6 to SEQ ID NO 25. Said oligonucleotide or oligomer comprising a nucleic acid sequence having a length of at least nine (9) nucleotides which hybridizes, under moderately stringent or stringent conditions (as defined herein above), to a pretreated nucleic acid sequence according to SEQ ID NO 6 to SEQ ID NO 25 and/or sequences complementary thereto.

[0037] Thus, the present invention includes nucleic acid molecules (e.g., oligonucleotides and peptide nucleic acid (PNA) molecules (PNA-oligomers)) that hybridise under moderately stringent and/or stringent hybridisation conditions to all or a portion of the sequences of SEQ ID NO 6 to SEQ ID NO 25, or to the complements thereof. The hybridising portion of the hybridising 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.

[0038] Preferably, the hybridising portion of the inventive hybridising nucleic acids is at least 95%, or at least 98%, or 100% identical to the sequence, or to a portion thereof of SEQ ID NO 6 to SEQ ID NO 25, or to the complements thereof.

[0039] Hybridising 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, hybridisation 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.

[0040] For target sequences that are related and substantially identical to the corresponding sequence of SEQ ID NO 1 to SEQ ID NO 5 (such as allelic variants and SNPs), rather than identical, it is useful to first establish the lowest temperature at which only homologous hybridisation 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 hybridisation 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.

[0041] Examples of inventive oligonucleotides of length X (in nucleotides), as indicated by polynucleotide positions with reference to, e.g., SEQ ID NO 1 to SEQ ID NO 5, include those corresponding to 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:

[0042] n to (n+(X−1));

[0043] where n=1, 2, 3, . . . (Y−(X−1));

[0044] where Y equals the length (nucleotides or base pairs) of SEQ ID NO 1 to SEQ ID NO 5;

[0045] 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

[0046] 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=2,785−19=2,766 for either sense or antisense sets of SEQ ID NO 1 to SEQ ID NO 5, where X=20.

[0047] Preferably, the set is limited to those oligomers that comprise at least one CpG, TpG or CpA dinucleotide.

[0048] The present invention encompasses, for each of SEQ ID NO 6 to SEQ ID NO 25 (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.

[0049] The oligonucleotides or oligomers according to the present invention constitute effective tools useful to ascertain genetic and epigenetic parameters of the genomic sequence corresponding to SEQ ID NO 1 to SEQ ID NO 5. 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 25 (and to the complements thereof). Preferably, said oligomers comprise at least one CpG, TpG or CpA dinucleotide. Included in these preferred sets are the preferred oligomers corresponding to SEQ ID NO 11 to SEQ ID NO 15.

[0050] Particularly preferred oligonucleotides or oligomers according to the present invention are those in which the cytosine of the CpG dinucleotide (or the thymine of the TpG or the adenosine of the CpA dinucleotide) 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.

[0051] 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, U.S. Pat. Nos. 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, hybridisation-triggered cross-linking agent, transport agent, hybridisation-triggered cleavage agent, etc.

[0052] The oligonucleotide may also comprise at least one art-recognised modified sugar and/or base moiety, or may comprise a modified backbone or non-natural internucleoside linkage.

[0053] The oligomers according to the present invention are normally used in so called "sets" which in one embodiment contain at least one oligomer for analysis of each of the CpG dinucleotides of a genomic sequence comprising SEQ ID NO 1 to SEQ ID NO 5 and sequences complementary thereto or to their corresponding CG, TG or CA dinucleotide within the pretreated nucleic acids according to SEQ ID NO 6 to SEQ ID NO 25 and sequences complementary thereto. Preferred is a set which contains at least one oligomer for each of the CpG dinucleotides within one or more genes selected from the group comprising Versican, TPEF, H-Cadherin, Calcitonin and EYA4 and their promoter and regulatory elements in both the pretreated and genomic versions of each gene, SEQ ID NO 1 to SEQ ID NO 25 respectively. 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 contents of the set of oligonucleotides should be altered accordingly. Therefore, the present invention moreover relates to a set of at least 4 oligonucleotides and/or PNA-oligomers used for detecting the cytosine methylation state in pretreated genomic DNA (SEQ ID NO 6 to SEQ ID NO 25 and sequences complementary thereto) and genomic DNA (SEQ ID NO 1 to SEQ ID NO 5 and sequences complementary thereto). These probes enable diagnosis and/or therapy of genetic and epigenetic parameters of cell proliferative disorders. The set of oligomers may also be used for detecting single nucleotide polymorphisms (SNPs) in pretreated genomic DNA (SEQ ID NO 6 to SEQ ID NO 25, and sequences complementary thereto) and genomic DNA (SEQ ID NO 1 to SEQ ID NO 5, and sequences complementary thereto).

[0054] Moreover, the present invention makes available a set of at least two oligonucleotides which can be used as so-called "primer oligonucleotides" for amplifying DNA sequences of one of SEQ ID NO 6 to SEQ ID NO 25 and sequences complementary thereto, or segments thereof.

[0055] In the case of the sets of oligonucleotides according to the present invention, it is preferred that at least one and more preferably all members of the set of oligonucleotides is bound to a solid phase.

[0056] According to the present invention, it is preferred that an arrangement of different oligonucleotides and/or PNA-oligomers (a so-called "array") made available by the present invention is present in a manner that it is likewise bound to a solid phase. This array of different oligonucleotide- and/or PNA-oligomer sequences can be characterized in that it is arranged on the solid phase in the form of a rectangular or hexagonal lattice. The solid phase surface is preferably composed of silicon, glass, polystyrene, aluminium, steel, iron, copper, nickel, silver, or gold. However, 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.

[0057] Therefore, a further subject matter of the present invention is a method for manufacturing an array fixed to a carrier material for analysis in connection with cell proliferative disorders, in which method at least one oligomer according to the present invention is coupled to a solid phase. Methods for manufacturing such arrays are known, for example, from U.S. Pat. No. 5,744,305 by means of solid-phase chemistry and photolabile protecting groups.

[0058] A further subject matter of the present invention relates to a DNA chip for the analysis of cell proliferative disorders. DNA chips are known, for example, in U.S. Pat. No. 5,837,832.

[0059] The present invention further provides a method for conducting an assay in order to ascertain genetic and/or epigenetic parameters of one or more genes selected from the group comprising Versican, TPEF, H-Cadherin, Calcitonin and EYA4 and their promoter and regulatory elements. Most preferably the assays according to the following method are used in order to detect methylation within one or more genes selected from the group comprising Versican, TPEF, H-Cadherin, Calcitonin and EYA4 wherein said methylated nucleic acids are present in a solution further comprising an excess of background DNA, wherein the background DNA is present in between 100 to 1000 times the concentration of the DNA to be detected. Said method comprises contacting a nucleic acid sample obtained from a 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.

[0060] Preferably, said method comprises the following steps: In the first step, a sample of the tissue to be analysed is obtained. The source may be any suitable source, preferably, the source of the sample is selected from the group consisting of histological slides, biopsies, paraffin-embedded tissue, bodily fluids, stool, blood, serum, plasma, urine, sputum and combinations thereof. Preferably, the source is biopsies, bodily fluids, urine, or blood.

[0061] The DNA is then isolated from the sample. Extraction may be by means that are standard to one skilled in the art, including the use of detergent lysates, sonification and vortexing with glass beads. Once the nucleic acids have been extracted, the genomic double stranded DNA is used in the analysis.

[0062] 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 hybridisation behavior. This will be understood as 'pretreatment' herein.

[0063] The above described treatment of genomic DNA is preferably carried out with bisulfite (hydrogen sulfite, disulfite) and subsequent alkaline hydrolysis which results in a conversion of non-methylated cytosine nucleobases to uracil or to another base which is dissimilar to cytosine in terms of base pairing behaviour. Enclosing the DNA to be analysed 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 further preferred that the bisulfite treatment is carried out in the presence of a radical trap or DNA denaturing agent.

[0064] In the third step of the method, fragments of the pretreated DNA are amplified. Wherein the source of the DNA is free DNA from serum, or DNA extracted from paraffin it is particularly preferred that the size of the amplificate fragment is between 100 and 200 base pairs in length, and wherein said DNA source is extracted from cellular sources (e.g. tissues, biopsies, cell lines) it is preferred that the amplificate is between 100 and 350 base pairs in length. It is particularly preferred that said amplificates comprise at least one 20 base pair sequence comprising at least three CpG dinucleotides. Said amplification is carried out using sets of primer oligonucleotides according to the present invention, and a preferably heat-stable polymerase. The amplification of several DNA segments can be carried out simultaneously in one and the same reaction vessel, in one embodiment of the method preferably two or more fragments are amplified simultaneously. Typically, the amplification is carried out using a polymerase chain reaction (PCR). 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 18-base-pair long segment of the base sequences of SEQ ID NO 6 to SEQ ID NO 25 and sequences complementary thereto.

[0065] In an alternate embodiment of the method, the methylation status of preselected CpG positions within the nucleic acid sequences comprising SEQ ID NO 6 to SEQ ID NO 25 may be detected by use of methylation-specific primer oligonucleotides. This technique (MSP) has been described in U.S. Pat. No. 6,265,171 to Herman. The use of methylation status specific primers for the amplification of bisulfite treated DNA allows the differentiation between methylated and unmethylated nucleic acids. MSP 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, TpG or CpA dinucleotide. MSP primers specific for non-methylated DNA contain a "T' at the 3' 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 18 nucleotides which hybridizes to a pretreated nucleic acid sequence according to SEQ ID NO 6 to SEQ ID NO 25 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG, TpG or CpA dinucleotide. In this embodiment of the method according to the invention it is particularly preferred that the MSP primers comprise between 2 and 5 CpG, TpG or CpA dinucleotides. It is further preferred that said dinucleotides are located within the 3' half of the primer e.g. wherein a primer is 18 bases in length the specified dinucleotides are located within the first 9 bases form the 3' end of the molecule. In addition to the CpG, TpG or CpA dinucleotides it is further preferred that said primers should further comprise several bisulfite converted bases (i.e. cytosine converted to thymine, or on the hybridizing strand, guanine converted to adenosine). In a further preferred embodiment said primers are designed so as to comprise no more than 2 cytosine or guanine bases.

[0066] In one embodiment of the method the primers may be selected from the group consisting of SEQ ID NO 34 to SEQ ID NO 49, 96, 97, 101, 102, 106 and 107.

[0067] 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 labelled amplificates have a single positive or negative net charge, allowing for better detectability in the mass spectrometer. The detection may be carried out and visualised by means of, e.g., matrix assisted laser desorption/ionisation mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI).

[0068] Matrix Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-TOF) is a very efficient development for the analysis of biomolecules (Karas & 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 vapour phase in an unfragmented manner. The analyte is ionised 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 & 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 ionisation process via the matrix is considerably less efficient. In MALDI-TOF spectrometry, the selection of the matrix plays an eminently important role. For the desorption of peptides, several very efficient matrixes have been found which produce a very fine crystallisation. 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 & 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.

[0069] In a particularly preferred embodiment of the method the amplification of step three is carried out in the presence of at least one species of blocker oligonucleotides. The use of such blocker oligonucleotides has been described by Yu et al., *BioTechniques* 23:714-720, 1997. The use of blocking oligonucleotides enables the improved specificity of the amplification of a sub-population of nucleic acids. Blocking probes hybridised to a nucleic acid suppress, or hinder the polymerase mediated amplification of said nucleic acid. In one embodiment of the method blocking oligonucleotides are designed so as to hybridise to background DNA. In a further embodiment of the method said oligonucleotides are designed so as to hinder or suppress the amplification of unmethylated nucleic acids as opposed to methylated nucleic acids or vice versa.

[0070] 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 'TpG' at the position in question, as opposed to a 'CpG.' In one embodiment of the method the sequence of said blocking oligonucleotides should be identical or complementary to molecule is complementary or identical to a sequence at least 18 base pairs in length selected from the group consisting of SEQ ID NO 6 to SEQ ID NO 25, preferably comprising one or more CpG, TpG or CpA dinucleotides. In one embodiment of the method the sequence of said oligonucleotides is selected from the group consisting SEQ ID NO 85 to SEQ ID NO 87, 98, 103 and 108 and sequences complementary thereto.

[0071] 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.

[0072] 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.

[0073] 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.

[0074] In one embodiment of the method, the binding site of the blocking oligonucleotide is identical to, or overlaps with that of the primer and thereby hinders the hybridisation of the primer to its binding site. In a further preferred embodiment of the method, two or more such blocking oligonucleotides are used. In a particularly preferred embodiment, the hybridisation of one of the blocking oligonucleotides hinders the hybridisation of a forward primer, and the hybridisation of another of the probe (blocker) oligonucleotides hinders the hybridisation of a reverse primer that binds to the amplificate product of said forward primer.

[0075] In an alternative embodiment of the method, the blocking oligonucleotide hybridises to a location between the reverse and forward primer positions of the treated background DNA, thereby hindering the elongation of the primer oligonucleotides.

[0076] It is particularly preferred that the blocking oligonucleotides are present in at least 5 times the concentration of the primers.

[0077] In the fourth step of the method, the amplificates obtained during the third step of the method are analysed in order to ascertain the methylation status of the CpG dinucleotides prior to the treatment.

[0078] In embodiments where the amplificates were obtained by means of MSP amplification and/or blocking oligonucleotides, the presence or absence of an amplificate is in itself indicative of the methylation state of the CpG positions covered by the primers and or blocking oligonucleotide, according to the base sequences thereof. All possible known molecular biological methods may be used for this detection, including, but not limited to gel electrophoresis, sequencing, liquid chromatography, hybridisations, real time PCR analysis or combinations thereof. This step of the method further acts as a qualitative control of the preceding steps.

[0079] In the fourth step of the method amplificates obtained by means of both standard and methylation specific PCR are further analysed in order to determine the CpG methylation status of the genomic DNA isolated in the first step of the method. This may be carried out 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.

[0080] In one embodiment of the method, the amplificates synthesised in step three are subsequently hybridised to an array or a set of oligonucleotides and/or PNA probes. In this context, the hybridisation takes place in the following manner: the set of probes used during the hybridisation is preferably composed of at least 2 oligonucleotides or PNA-oligomers; in the process, the amplificates serve as probes which hybridise to oligonucleotides previously bonded to a solid phase; the non-hybridised 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 SEQ ID NO 2 to SEQ ID NO 5; and the segment comprises at least one CpG, TpG or CpA dinucleotide.

[0081] 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 the sequence according to SEQ ID NO 1 to SEQ ID NO 5, and the equivalent positions within SEQ ID NO 6 to SEQ ID NO 25. Said oligonucleotides may also be in the form of peptide nucleic acids. The non-hybridised amplificates are then removed. The hybridised amplificates are 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 one embodiment of the method said oligonucleotides may be selected from the group comprising SEQ ID NOS 50-77, 88 and 89.

[0082] In yet a further embodiment of the method, the genomic methylation status of the CpG positions may be ascertained by means of oligonucleotide probes that are hybridised to the bisulfite treated DNA concurrently with the PCR amplification primers (wherein said primers may either be methylation specific or standard).

[0083] 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 U.S. Pat. No. 6,331,393). There are two preferred embodiments of utilising this method. One embodiment, known as the TaqMan™ assay employs a dual-labelled fluorescent oligonucleotide probe. The TaqMan™ PCR reaction employs the use of a nonextendible interrogating oligonucleotide, called a TaqMan™ probe, which is designed to hybridise to a GpC-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. Hybridised probes are displaced and broken down by the polymerase of the amplification reaction thereby leading to an increase in fluorescence. For analysis of methylation within nucleic acids subsequent to bisulfite treatment, it is required that the probe be methylation specific, as described in U.S. Pat. No. 6,331,393, (hereby incorporated by reference in its entirety) also known as the MethylLight™ assay. The second preferred embodiment of this technology is the use of dual-probe technology (Lightcycler™), each carrying donor or recipient fluorescent moieties, hybridisation of two probes in proximity to each other is indicated by an increase or

fluorescent amplification primers. 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 one embodiment of the method the sequence of said probe oligonucleotides may be selected from the group comprising SEQ ID NO 78-84, 90, 99, 100, 104, 105, 109 and 110.

[0084] 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 & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997. In said embodiment it is preferred that the Ms-SNuPE primer is identical or complementary to a sequence at least nine but preferably no more than twenty five nucleotides in length of one or more of the sequences taken from the group of SEQ ID NO 2 to SEQ ID NO 5.

[0085] 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).

[0086] 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 5) without the need for pretreatment.

[0087] In the first step of such additional embodiments, the genomic DNA sample is isolated from tissue or cellular sources. Preferably, such sources include cell lines, histological slides, body fluids, or tissue embedded in paraffin. Extraction may be by means that are standard to one skilled in the art, including but not limited to the use of detergent lysates, sonification and vortexing with glass beads. Once the nucleic acids have been extracted, the genomic double-stranded DNA is used in the analysis.

[0088] In a preferred embodiment, the DNA may be cleaved prior to the treatment, and this may be by any means standard in the state of the art, in particular with methylation-sensitive restriction endonucleases.

[0089] In the second step, the DNA 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.

[0090] In the third 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.

[0091] In the final 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, hybridisation analysis, incorporation of detectable tags within the PCR products, DNA array analysis, MALDI or ESI analysis.

[0092] The present invention enables diagnosis and/or prognosis of events which are disadvantageous to patients or individuals in which important genetic and/or epigenetic parameters within the Versican, TPEF, H-Cadherin, Calcitonin and EYA4 and their promoter or regulatory elements 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.

[0093] Specifically, the present invention provides for diagnostic and/or prognostic cancer assays based on measurement of differential methylation of Versican, TPEF, H-Cadherin, Calcitonin and/or EYA4 CpG dinucleotide sequences. Preferred gene sequences useful to measure such differential methylation are represented herein by SEQ ID NO 1 to SEQ ID NO 25. Typically, such assays involve obtaining a tissue sample from a test tissue, performing an assay to measure the methylation status of at least one of the inventive Versican, TPEF, H-Cadherin, Calcitonin and/or EYA4 specific CpG dinucleotide sequences derived from the tissue sample, relative to a control sample, and making a diagnosis or prognosis based thereon.

[0094] In particular preferred embodiments, inventive oligomers are used to assess Versican, TPEF, H-Cadherin, Calcitonin and/or EYA4 specific CpG dinucleotide methylation status, such as those based on SEQ ID NO 1 to SEQ ID NO 25, including the representative preferred oligomers corresponding to SEQ ID NOS: ALL OLIGOS, or arrays thereof, as well as a kit based thereon are useful for the diagnosis and/or prognosis of cancer and/or other prostate cell proliferative disorders.

[0095] The present invention moreover relates to a diagnostic agent and/or therapeutic agent for the diagnosis and/or therapy colon cell proliferative disorders, the diagnostic agent and/or therapeutic agent being characterised in that at least one primer or probe based on SEQ ID NO 1 to SEQ ID NO 25 is used for manufacturing it, possibly together with suitable additives and ancillary agents.

[0096] Moreover, an additional aspect of the present invention is a kit comprising, for example: a bisulfite-containing reagent as well as at least one oligonucleotide whose sequences in each case correspond, are complementary, or hybridise under stringent or highly stringent conditions to a 18-base long segment of the sequences SEQ ID NO 1 to SEQ ID NO 5. Said kit may further comprise instructions for carrying out and evaluating the described method. In a further preferred 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.

[0097] 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 methylation-altered DNA sequence or CpG island); restriction enzyme and appropriate buffer; gene-hybridisation oligo; control hybridisation oligo; kinase labelling kit for oligo probe; and radioactive 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.

[0098] 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 methylation-altered DNA sequence or CpG island); TaqMan® probes; optimized PCR buffers and deoxynucleotides; and Taq polymerase.

[0099] 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 methylation-altered DNA sequence or CpG island); optimised PCR buffers and deoxynucleotides; gel extraction kit; positive control primers; Ms-SNuPE primers for specific gene; reaction buffer (for the Ms-SNuPE reaction); and radioactive 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.

[0100] 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 methylation-altered DNA sequence or CpG island), optimized PCR buffers and deoxynucleotides, and specific probes.

Definitions:

[0101] 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 in length.

[0102] 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 palindromic CpG methylation sites (each having two CpG CpG dinucleotide sequences) within a DNA sequence include "unmethylated,""fully-methylated" and "hemi-methylated."

[0103] The term "hemi-methylation" or "hemimethylation" refers to the methylation state of a palindromic CpG methylation site, where only a single cytosine in one of the two CpG dinucleotide sequences of the palindromic CpG methylation site is methylated (e.g., 5'-CC^{M}GG-3' (top strand): 3'-GGCC-5' (bottom strand)).

[0104] 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.

[0105] 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.

[0106] The term "microarray" refers broadly to both "DNA microarrays," and 'DNA chip(s),' as recognised 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.

[0107] "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).

[0108] "Epigenetic parameters" are, in particular, cytosine methylations. 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.

[0109] 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.

[0110] 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.

[0111] The term "MS.AP-PCR" (Methylation-Sensitive Arbitrarily-Primed Polymerase Chain Reaction) refers to the art-recognised 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.

[0112] The term "MethyLight" refers to the art-recognised fluorescence-based real-time PCR technique described by Eads et al., *Cancer Res.* 59:2302-2306, 1999.

[0113] The term "HeavyMethyl" assay, in the embodiment thereof implemented herein, refers to a HeavyMethyl™ MethylLight assay, which is a variation of the MethylLight□ assay, wherein the MethylLight assay is combined with methylation specific blocking probes covering CpG positions between the amplification primers.

[0114] The term "Ms-SNuPE" (Methylation-sensitive Single Nucleotide Primer Extension) refers to the art-recognised assay described by Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997.

[0115] 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 U.S. Pat. No. 5,786,146.

[0116] The term "COBRA" (Combined Bisulfite Restriction Analysis) refers to the art-recognized methylation assay described by Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997.

[0117] The term "hybridisation" 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.

[0118] "Stringent hybridisation conditions," as defined herein, involve hybridising at 68° C. in 5×SSC/5× Denhardt's solution/1.0% SDS, and washing in 0.2×SSC/0.1% SDS at room temperature, or involve the art-recognised equivalent thereof (e.g., conditions in which a hybridisation is carried out at 60° C. in 2.5×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 3×SSC at 42°

C., or the art-recognised 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 & Sons, N.Y.) at Unit 2.10.

[0119] "Background DNA" as used herein refers to any nucleic acids which originate from sources other than colon cells.

LEGENDS TO FIGURES

[0120] **FIG. 1** shows the level of methylation determined by different MSP MethyLight assays and HeavyMethyl MethyLight assays. The Y-axis shows the degree of methylation. Tumor samples are represented by white points, and normal colon tissue samples by black points. A significantly higher degree of methylation was observed in tumor samples than in healthy tissue samples.

[0121] **FIG. 2** shows the Receiver Operating Characteristic curve (ROC curve) of the EYA4-MSP-Methyl-Light-Assay for adenocarcinomas according to Example 1. The AUC for the MSP-Methyl-Light-Assay is: 0.94.

[0122] **FIG. 3** shows the Receiver Operating Characteristic curve (ROC curve) of the EYA4-HM-Methyl-Light-Assay for Adenocarcinoma according to Example 2. The AUC for the HM-Methyl-Light-Assay is: 0.91.

[0123] **FIG. 4** shows the level of methylation determined by a EYA4-HeavyMethyl MethyLight™ assay according to example 2, testing an additional set of colon samples (25 adenocarcinoma, 33 normals, and 13 adenomas). The Y-axis shows the degree of methylation within the region of the EYA4 gene investigated. Adenocarcinoma samples are represented by white squares, and normal colon tissue samples by black diamonds. A significantly higher degree of methylation was observed in tumor samples than in healthy tissue samples. The level of significance as measured using a t-test was 0.00424.

[0124] **FIG. 5** shows the Receiver Operating Characteristic curve (ROC curve) of the EYA4-HM-Methyl-Light-Assay for Adenocarcinoma and Adenoma according to Example 2 (additional sets of samples). The area under an ROC curve (AUC) is a measure for the accuracy of a diagnostic test. The AUC for the HM-Methyl-Light-Assay is 0.81.

[0125] **FIG. 6** shows the Receiver Operating Characteristic curve (ROC curve) of the EYA4-HM-Methyl-Light-Assay for Adenocarcinoma only according to Example 2 (additional sets of samples). The area under an ROC curve (AUC) is a measure for the accuracy of a diagnostic test. The AUC for the HM-Methyl-Light-Assay is: 0.844.

[0126] **FIG. 7** shows the Receiver Operating Characteristic curve (ROC curve) of the EYA4-HM-Methyl-Light-Assay for Adenenomas according to Example 2 (additional sets of samples). The area under an ROC curve (AUC) is a measure for the accuracy of a diagnostic test. The AUC for the HM-Methyl-Light-Assay is: 0.748.

[0127] **FIG. 8** shows the level of methylation in different tumor and healthy tissues determined by a EYA4-HeavyMethyl MethyLight™ assay according to example 3. The Y-axis shows the degree of methylation within the region of the EYA4 gene investigated. Besides the colon cancer samples only one of the two breast cancer tissues were methylated.

[0128] **FIG. 9** shows the level of methylation in different breast cancer tissues determined by a EYA4-HeavyMethyl MethyLight™ assay according to example 3. Only one was methylated.

[0129] **FIG. 10** shows the level of methylation in serum samples determined by a EYA4-HeavyMethyl MethyLight™ assay according to example 4. The Y-axis shows the degree of methylation within the region of the EYA4 gene investigated.

[0130] **FIG. 11** shows the Receiver Operating Characteristic curve (ROC curve) of the Calcitonin-MSP-Methyl-Light-Assay according to Example 5. The area under an ROC curve (AUC) is a measure for the accuracy of a diagnostic test. The AUC for the HM-Methyl-Light-Assay is: 0.85.

[0131] **FIG. 12** shows the ROC curve of the Calcitonin-HM-Methyl-Light-Assay according to Example 6. The AUC is: 0.81.

[0132] **FIG. 13** shows the ROC curve of the Versican-MSP-Methyl-Light-Assay according to Example 9. The AUC is: 0.84.

[0133] **FIG. 14** shows the ROC curve of the TBEF-MSP-Methyl-Light-Assay according to Example 10. The AUC is: 0.80.

[0134] **FIG. 15** shows the ROC curve of the Cadherin-MSP-Methyl-Light-Assay according to Example 12. The AUC is: 0.94.

[0135] **FIG. 16** shows the differentiation of healthy tissue from non healthy tissue wherein the non healthy specimens are obtained from either colon adenoma or colon carcinoma tissue (Example 13). The evaluation is carried out using informative CpG positions from 27 genes. Informative CpG positions from the genes Versican, TPEF, EYA4 and H-Cadherin are further described in Table 3.

[0136] **FIG. 17** shows the differentiation of healthy tissue from carcinoma tissue using informative CpG positions from 15 genes (Example 13). Informative CpG positions from the genes Versican, TPEF, EYA4 and H-Cadherin are further described in Table 4.

[0137] **FIG. 18** shows the differentiation of healthy tissue from adenoma tissue using informative CpG positions from 40 genes (Example 13). Informative CpG positions from the genes Versican, TPEF, EYA4 and H-Cadherin are further described in Table 5.

[0138] **FIG. 19** shows the ROC curve of the TPEF-MSP-Methyl-Light-Assay according to Example 11 (first sample set). The AUC is: 0.93.

[0139] **FIG. 20** shows the ROC curve of the TPEF-MSP-Methyl-Light-Assay according to Example 11 (second sample set). The AUC is: 1.

[0140] **FIG. 21** shows the ROC curve of a combined EYA4-Calcitonin-Heavymethyl-Methyl-Light-Assay according to Example 6. The AUC is: 0.97.

[0141] **FIG. 22** shows the regression plot of the percentage methylation within the EYA 4 gene calculated in each sample using the MSP and HeavyMethyl variants of the MethyLight assay.

[0142] **FIG. 23** shows the regression plot of the percentage methylation within the Calcitonin gene calculated in each sample using the MSP and HeavyMethyl variants of the MethyLight assay.

EXAMPLES

[0143] The following examples describe the analysis of the methylation status of the genes EYA 4, Calcitonin, TPEF, H-Cadherin and Versican in healthy and sick colon cell proliferative disorder samples. The initial link between said genes and colon cell proliferative disorders was initially carried by means of hybridisation analysis as described in examples 13 onwards. The genes EYA 4, Calcitonin, TPEF, H-Cadherin and Versican were then selected from the larger set of genes analysed in said examples, and the correlation between methylation status and colon cell proliferative disorder states was validated by analysis of samples using other methylation analysis techniques, namely the MSP-MethyLight and HeavyMethyl MethyLight assays. Please not that the term 'MethyLight' is used to describe real time PCR analysis of bisulfite treated DNA using probes of both the Taqman ingle sprobe) and Lightcycler (dual probe) technologies.

Example 1

[0144] Analysis of methylation within colon cancer using an MSP-MethyLight assay (EYA4) DNA was extracted from 33 colon adenocarcinoma samples and 43 colon normal adjacent tissues using a Qiagen extraction kit. The DNA from each sample was treated using a bisulfite solution (hydrogen sulfite, disulfite) according to the agarose-bead method (Olek et al 1996). The treatment is such that all non methylated cytosines within the sample are converted to thymidine. Conversely, 5-methylated cytosines within the sample remain unmodified.

[0145] The methylation status was determined with a MSP-MethyLight assay designed for the CpG island of interest and a control fragment from the beta actin gene (Eads et al., 2001). The CpG island assay covers CpG sites in both the primers and the Taqman style probe, while the control gene does not. The control gene is used as a measure of total DNA concentration, and the CpG island assay (methylation assay) determines the methylation levels at that site.

[0146] Methods: The EYA4 gene CpG island assay was performed using the following primers and probes: Forward Primer: CGGAGGGTACGGAGATTACG (SEQ ID NO:40); Reverse Primer: CGACGACGCGCGAAA (SEQ ID NO:41); and Probe: CGAAACCCTAAATATCCCGAATAACGCCG (SEQ ID NO:81). The corresponding control assay was performed using the following primers and probes: Primer: TGGTGATGGAGGAGGTTTAGTAAGT (SEQ ID NO:91); Primer: AACCAATAAAACCTACTCCTCCCTTAA (SEQ ID NO:92); and Probe: ACCACCACCCAACACACAATAACAAACACA (SEQ ID NO:93) The reactions were run in triplicate on each DNA sample with the following assay conditions: Reaction solution: (900 nM primers; 300 nM probe; 3.5 mM Magnesium Chloride; 1 unit of taq polymerase; 200 μM dNTPs; 7 μl of DNA, in a final reaction volume of 20 μl); Cycling conditions: (95° C. for 10 minutes; then 50 cycles of: 95° C. for 15 seconds; 60° C. for 1 minute).

[0147] The data was analysed using a PMR calculation previously described in the literature (Eads et al 2001). Results. The mean PMR for normal samples was 0.15, with a standard deviation of 0.18. The mean PMR for tumour samples was 17.98, with a standard deviation of 18.18. The overall difference in methylation levels between tumour and normal samples is significant in a t-test (p=0.00000312). The results are shown in **FIG. 1**. A Receiver Operating Characteristic curve (ROC curve) of the assay was also determined. A ROC is a plot of the true positive rate against the false positive rate for the different possible cutpoints of a diagnostic test. It shows the tradeoff between sensitivity and specificity depending on the selected cutpoint (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 AUC for the MSP-Methyl-Light-Assay is: 0.94 (**FIG. 2**)

Example 2

[0148] Methylation within colon cancer was analysed using a EYA4-HeavyMethyl MethyLight assay. The same DNA samples were also used to analyse methylation of the CpG island with a HeavyMethyl MethyLight (or HM MethyLight) assay, also referred to as the HeavyMethyl assay. The methylation status was determined with a HM MethyLight assay designed for the CpG island of interest and the same control gene assay described above. The CpG island assay covers CpG sites in both the blockers and the Taqman style probe, while the control gene does not.

[0149] Methods. The CpG island assay (methylation assay) was performed using the following primers and probes:

```
Forward Primer:
GGTGATTGTTTATTGTTATGGTTTG      (SEQ ID NO:44)

Reverse Primer:
CCCCTCAACCTAAAAACTACAAC        (SEQ ID NO:45)

Forward Blocker:
GTTATGGTTTGTGATTTTGTGTGGG      (SEQ ID NO:87)

Reverse Blocker:
AAACTACAACCACTCAAATCAACCCA     (SEQ ID NO:86)

Probe:
AAAATTACGACGACGCCACCCGAAA      (SEQ ID NO:84)
```

[0150] The reactions were each run in triplicate on each DNA sample with the following assay conditions:

[0151] Reaction solution: (400 nM primers; 400 nM probe; 10 μM both blockers; 3.5 mM magnesium chloride;

1×ABI Taqman buffer; 1 unit of ABI TaqGold polymerase; 200 μM dNTPs; and 7 μl of DNA, in a final reaction volume of 20 μl);

[0152] Cycling conditions: (95□C for 10 minutes); (95°□C. for 15 seconds, 64° C. for 1 minute (2 cycles)); (95° C. for 15 seconds, 62° C. for 1 minute (2 cycles); (95° C. for 15 seconds, 60° C. for 1 minute (2 cycles)); and (95° C. for 15 seconds, 58° C. for 1 minute, 60° C. for 40 seconds (41 cycles)).

[0153] Results. The mean PMR for normal samples was 1.12 with a standard deviation of 1.45. The mean PMR for tumour samples was 38.23 with a standard deviation of 33.22. The overall difference in methylation levels between tumour and normal samples is significant in a t-test (p=0.000000326). The results are shown in **FIG. 1**.

[0154] A ROC curve of the assay was also determined. The AUC for the MSP-Methyl-Light-Assay is 0.91 (**FIG. 3**)

[0155] The assay was tested on an additional set of colon samples (25 adenocarcinoma, 33 normals, and 13 adenomas). The results showed a significant difference again (**FIG. 4**). The ROC are shown in **FIG. 5-7**.

[0156] The MSP and HeavyMethyl variants of the MethyLight assay were determined to be equivalent for the analysis of methylation in the gene EYA4, **FIG. 22** shows the regression plot of the percentage methylation detected in each sample using the two methods.

Example 3

[0157] The EYA4-HeavyMethyl-MethyLight-assay was also tested against a panel of other tissues (**FIG. 8**). Besides the colon cancer samples only one of the two breast cancer tissues were methylated. However, on a panel of 21 additional breast tumours (different stages), only one was methylated (**FIG. 9**). So the marker is specific for colon tumour samples. All primers, probes, blockers and reaction conditions were identical to those used in the analysis of the colon cancer samples (Example 2).

Example 4

[0158] Twelve of the colon tissues analysed by real-time PCR also had paired serum taken before surgery. We extracted DNA from 1 ml of that serum using a Qiagen UltraSens□ DNA extraction kit, bisulfite treated the DNA sample, and ran the EYA4-HeavyMethyl-MethyLight-assay on those samples. The control gene did not amplify for three of the cancer serum samples and three of the normal serum samples, so we can conclude that the sample preparation did not work in these cases. In the other cases, there was evidence of higher methylation in the cancer samples than the normal samples (**FIG. 10**).

Example 5

[0159] Analysis of methylation within colon cancer using a Calcitonin-MSP-MethyLight Assay The colon cancer samples described in Example 1 were also analysed using a Calcitonin-MSP-MethyLight Assay, with a Taqman® style probe. The sample preparation was carried out as described above (Example 1) The assay was performed using the following primers and probes:

```
Primer:
AGGTTATCGTCGTGCGAGTGT;          (SEQ ID NO:34)

Primer:
TCACTCAAACGTATCCCAAACCTA;       (SEQ ID NO:35)
and

Probe:
CGAATCTCTCGAACGATCGCATCCA.      (SEQ ID NO:78)
```

[0160] The corresponding control assay was performed as described above (Example 1)

[0161] The reactions were run in triplicate on each DNA sample with the following assay conditions:

[0162] Reaction solution: (900 nM primers; 300 nM probe; 3.5 mM Magnesium Chloride; 1 unit of taq polymerase; 200 μM dNTPs; 7 μl of DNA, in a final reaction volume of 20 μl);

[0163] Cycling conditions: (95° C. for 10 minutes; 95□C for 15 seconds; 67° C. for 1 minute (3 cycles)); (95° C. for 15 seconds, 64° C. for 1 minute (3 cycles)); (95° C. for 15 seconds, 62° C. for 1 minute (3 cycles)); and (95° C. for 15 seconds, 60° C. for 1 minute (40 cycles)).

[0164] The data was analysed using a PMR calculation previously described in the literature (Eads et al 2001).

[0165] Results. The mean PMR for normal samples was 0.19, with a standard deviation of 0.79. None of the normal samples was greater than 2 standard deviations about the normal mean, while 18 of 33 tumour samples reached this level of methylation. The overall difference in methylation levels between tumour and normal samples is significant in a t-test (p=0.002). The results are shown in **FIG. 1**. Significantly, the tumour samples are substantially hypermethylated relative to normal control tissue. A ROC curve of the assay was also determined. The AUC for the MSP-Methyl-Light-Assay is 0.80 (**FIG. 11**)

Example 6

[0166] Methylation within colon cancer was analysed using a Calcitonin-HeavyMethyl MethyLight assay. The same DNA samples were also used to analyse methylation of the Calcitonin-CpG island with a HeavyMethyl MethyLight assay using a Taqman® style probe (see Example 2).

[0167] The CpG island assay (methylation assay) was performed using the following primers and probes:

```
Primer:
GGATGTGAGAGTTGTTGAGGTTA;        (SEQ ID NO: 46)

Primer:
ACACACCCAAACCCATTACTATCT;       (SEQ ID NO: 47)

Probe:
ACCTCCGAATCTCTCGAACGATCGC;      (SEQ ID NO: 83)
and

Blocker:
TGTTGAGGTTATGTGTAATTGGGTGTGA.   (SEQ ID NO: 85)
```

[0168] The reactions were each run in triplicate on each DNA sample with the following assay conditions:

[0169] Reaction solution: (300 nM primers; 450 nM probe; 3.5 mM magnesium chloride; 2 units of taq polymerase; 400 μM dNTPs, 5 μM blocker; and 7 μl of DNA, in a final reaction volume of 20 μl);

[0170] Cycling conditions: (95° C. for 10 minutes); (95° C. for 15 seconds, 67° C. for 1 minute (3 cycles)); (95° C. for 15 seconds, 64° C. for 1 minute (3 cycles); (95° C. for 15 seconds, 62° C. for 1 minute (3 cycles)); and (95° C. for 15 seconds, 60° C. for 1 minute (40 cycles)).

[0171] The corresponding control assay was performed as described above (Example 2)

[0172] Results. The mean PMR for normal samples was 0.13 with a standard deviation of 0.58. None of the normal samples was greater than 2 standard deviations about the normal mean, while 19 of 33 tumour samples reached this level of methylation. The overall difference in methylation levels between tumour and normal samples is significant in a t-test (p=0.0004). The results are shown in **FIG. 1**. A ROC curve of the assay was also determined. The AUC for the HM-Methyl-Light-Assay is 0.84 (**FIG. 12**)

[0173] In order to estimate the sensitivity and specificity of a real time assay analysing a gene panel comprising the genes Calcitonin and EYA 4, the ROC of said assay was in sillico determined by combining the ROCs of the 2 genes (as described above) using a logistics model. The AUC of said curve (**FIG. 21**) is 0.97.

[0174] The MSP and HeavyMethyl variants of the MethyLight assay were determined to be equivalent for the analysis of methylation in the gene Calcitonin, **FIG. 23** shows the regression plot of the percentage methylation detected in each sample using the two methods.

Example 7

[0175] Serum analysis with Calcitonin-HM MethyLight assay. Twelve of the colon tissues analysed by real-time PCR also had paired serum taken before surgery. We extracted DNA from 1 ml of that serum using a Qiagen UltraSens® DNA extraction kit, bisulfite treated the DNA sample, and ran the HeavyMethyl-MethyLight-assay on those samples (see Example 3). Calcitonin was not methylated in all tumours, but of the five patients with the highest levels of methylation in the tumours, we were able to detect methylation in the serum of four of them. In contrast, we did not detect methylation in any of the 11 serum samples taken from healthy donors.

Example 8

[0176] Identification of the methylation status of a CpG site within the Calcitonin Gene. A fragment of the upstream region of the calcitonin gene (SEQ ID NO: 1) was amplified by PCR using the primers CCTTAGTCCCTACCTCTGCT (SEQ ID NO:94) and CTCATTTACACACACCCAAAC (SEQ ID NO:95). The resultant amplificate, 378 bp in length, contained an informative CpG at position 165. The amplificate DNA was digested with the methylation sensitive restriction endonuclease Nar I; recognition motif GGCGCC. Hydrolysis by said endonuclease is blocked by methylation of the CpG at position 165 of the amplificate. The digest was used as a control.

[0177] Genomic DNA was isolated from the samples using the DNA wizzard® DNA isolation kit (Promega□). Each sample was digested using Nar I according to manufacturer's recommendations (New England Biolabs).

[0178] About 10 ng of each genomic digest was then amplified using PCR primers CCTTAGTCCCTACCTCTGCT (SEQ ID NO: 94) and CTCATTTACACACACCCAAAC (SEQ ID NO: 95). The PCR reactions were performed using a thermocycler (Eppendorf GmbH) using 10 ng of DNA, 6 pmole of each primer, 200 μM of each dNTP, 1.5 mM $MgCl_2$ and 1 U of Hotstart®Taq (Qiagen AG). The other conditions were as recommended by the Taq polymerase manufacturer.

[0179] Using the above mentioned primers, gene fragments were amplified by PCR performing a first denaturation step for 14 min at 96° C., followed by 30-45 cycles (step 2: 60 sec at 96° C., step 3: 45 sec at 52° C., step 4: 75 sec at 72° C.) and a subsequent final elongation of 10 min at 72° C. The presence of PCR products was analysed by agarose gel electrophoresis.

[0180] PCR products were detectable, with Nar I-hydrolyzed DNA isolated wherein the tissue in question contained upmethylated DNA, when step 2 to step 4 of the cycle program were repeated 34, 37, 39, 42 and 45 fold. In contrast, PCR products were only detectable with Nar I-hydrolysed DNA isolated from downmethylated tissue when steps 2 to step 4 of the cycle program were repeated 42- and 45-fold.

Example 9

[0181] Analysis of methylation within colon cancer using a Versican-MSP-MethyLight Assay. The colon cancer samples described in Example 1 were also analysed using a Versican-MSP-MethyLight Assay with a Taqman® style probe. The sample preparation was carried out as described above (Example 1) The assay was performed using the following primers and probes:

```
Forward Primer:
TGGGATTAAGATTTTCGGTTAGTTTC          (SEQ ID No 36)

Reverse Primer:
CACTACAACGCTACGCGACTAAA             (SEQ ID No 37)

Probe:
TCGACGTTACCCAAACGAATCACATAAAAAAC    (SEQ ID No 79)
```

[0182] The corresponding control assay was performed as described above (Example 1)

[0183] The reactions were run in triplicate on each DNA sample with the following assay conditions:

[0184] Reaction solution: (900 nM primers; 300 nM probe; 3.5 mM magnesium chloride; 1 units of taq polymerase; 200 μM dNTPs, 5 μM blocker; and 7 μl of DNA, in a final reaction volume of 20 μl);

[0185] Cycling conditions: 95° C. for 10 minutes; (95° C. for 15 seconds, 60° C. for 1 minute) 50 cycles

[0186] The data was analysed using a PMR calculation previously described in the literature (Eads et al 2001).

[0187] Results. The results are shown in **FIG. 1**. The mean PMR for normal samples was 3.93, with a standard deviation of 3.57. The mean PMR for tumour samples was 23.06, with a standard deviation of 20.23 The overall difference in methylation levels between tumour and normal samples is significant in a t-test (p=0.000003063). The ROC curve of the assay is shown in **FIG. 13**. The AUC is 0.84.

[0188] This was further confirmed using a Versican-HeavyMethyl MethyLight assay, using dual Lightcycler probes.

[0189] Methods. The CpG island assay (methylation assay) was performed using the following primers and probes:

```
Forward Primer:
TGGATAGGAGTTGGGATTAAGATTTT          (SEQ ID NO:96)

Reverse Primer:
CTTATTACAATTTAAAAAAAAAATTCACTACAA  (SEQ ID NO:97)

Blocker:
AAATTCACTACAACACTACACAACTAAATTCAAC (SEQ ID NO:98)
ATTAC

Probe:
TTTTCGTATTTTTTTTCGGGTTATTACGTTTT-  (SEQ ID NO:99)
Fluor

Probe:
LC640-ATGTGATTCGTTTGGGTAACGTCGA-   (SEQ ID NO:100)
Phos
```

[0190] The reactions were each run in triplicate on each DNA sample with the following assay conditions:

[0191] Reaction conditions:

[0192] 500 nM primers

[0193] 10 uM blocker

[0194] 250 nM probes

LightCycler FastStart Hybridization Probes Mix

[0195] 4 mM Magnesium Chloride

[0196] Cycling profile:

[0197] 95 degree denaturation for 10 minutes

[0198] 50 cycles: 95 degrees 10 seconds, 57 degrees 30 seconds, 72 degrees 20 seconds

Example 10

[0199] Analysis of methylation within colon cancer using a TPEF-MSP-MethyLight Assay. The colon cancer samples described in Example 1 were also analysed using a TPEF-MSP-MethyLight Assay with a Taqman® style probe. The sample preparation was carried out as described above (Example 1) The assay was performed using the following primers and probes:

```
Forward Primer:
TTTTTTTTTCGGACGTCGTTG      (SEQ ID No 38)

Reverse Primer:
CCTCTACATACGCCGCGAAT       (SEQ ID No 39)
```

-continued

```
Probe:
AATTACCGAAAACATCGACCGA     (SEQ ID No 80)
```

[0200] The reactions were run in triplicate on each DNA sample with the following assay conditions:

[0201] Reaction solution: (900 nM primers; 300 nM probe; 3.5 mM magnesium chloride; 1 units of taq polymerase; 200 μM dNTPs, 5 μM blocker; and 7 μl of DNA, in a final reaction volume of 20 μl);

[0202] Cycling conditions: 95° C. for 10 minutes; (95° C. for 15 seconds, 60° C. for 1 minute) 50 cycles

[0203] The corresponding control assay was performed as described above (Example 1)

[0204] The data was analysed using a PMR calculation previously described in the literature (Eads et al 2001).

[0205] Results. The results are shown in **FIG. 1**. The mean PMR for normal samples was 3.04, with a standard deviation of 4.21. The mean PMR for tumour samples was 21.38, with a standard deviation of 24.08 The overall difference in methylation levels between tumour and normal samples is significant in a t-test (p=0.0000101973). The ROC curve of the assay is shown in **FIG. 14**. The AUC is 0.80.

[0206] This was further confirmed using a TPEF-HeavyMethyl MethyLight assay (using dual labeled Lightcycler probes.

[0207] Methods. The CpG island assay (methylation assay) was performed using the following primers and probes:

```
Forward Primer:
GTAGGGTTATTGTTTGGGTTAATAAAT        (SEQ ID NO:101)

Reverse Primer:
TAAAAAAAAAAAAAAAACTCCTCTACATAC     (SEQ ID NO:102)

Blocker:
AACTCCTCTACATACACCACAAATAAATT       (SEQ ID NO:103)

Probe:
CGAAAACATCGACCGAACAACG-Fluor        (SEQ ID NO:104)

Probe:
LC640-GTCCGAAAAAAAAAAAACGAACTCC-   (SEQ ID NO:105)
Phos
```

[0208] The reactions were each run in triplicate on each DNA sample with the following assay conditions:

[0209] Reaction conditions:

Forward primer:	600 nM
Reverse primer:	300 nM
Blocker:	10 uM
Probes:	500 nM
Taq polymerase:	0.1 U/ul
dNTPs:	0.2 mM each
Magnesium Chloride:	4 mM
BSA:	0.25 mg/ml
Roche buffer with no MgCl:	1x

[0210] Cycling conditions:

[0211] 95-degree denaturation for 10 minutes

[0212] 50 cycles: 95-degrees for 10 seconds, 57-degrees for 25 seconds, 72 degrees for 10 seconds

Example 11

[0213] Analysis of methylation within colon cancer using a TPEF-MSP-MethyLight Assay. An additional assay for TPEF was tested on colon samples. The assay was tested on two sets of tissues, each with 12 colon adenocarcinomas and 12 normal adjacent tissue samples.

[0214] The sample preparation was carried out as described above (Example 1) The assay was performed using the following primers and probes:

```
Forward Primer:
GGACGTTTTTTATCGAAGGCG      (SEQ ID No 48)

Reverse Primer:
GCCACCCAACCGCGA            (SEQ ID No 49)

Probe:
ACCCGAAATCACGCGCGAAAAA      (SEQ ID No 90)
```

[0215] The reactions were run in triplicate on each DNA sample with the following assay conditions:

[0216] Reaction solution: (900 nM primers; 300 nM probe; 3.5 mM magnesium chloride; 1 units of taq polymerase; 200 μM dNTPs, 5 μM blocker; and 7 μl of DNA, in a final reaction volume of 20 μl);

[0217] Cycling conditions: 95° C. for 10 minutes; (95° C. for 15 seconds, 60° C. for 1 minute) 50 cycles

[0218] The corresponding control assay was performed as described above (Example 1)

[0219] The data was analysed using a PMR calculation previously described in the literature (Eads et al 2001). In both cases, TPEF was significantly more methylated in the cancer samples The ROC curves of the assays are shown in **FIG. 19-20**. The AUC are 0.93 and 1.

Example 12

[0220] Analysis of methylation within colon cancer using a H Cadherin-MSP-MethyLight Assay The colon cancer samples described in Example 1 were also analysed using a H Cadherin-MSP-MethyLight Assay. The sample preparation was carried out as described above (Example 1) The assay was performed using the following primers and probes:

```
Forward Primer:
GACGGATTTTTTTTAACGTTTTTTC     (SEQ ID No 42)

Reverse Primer:
AAATAAAATACCACCTCCGCGA         (SEQ ID No 43)

Probe:
GCTCCTCGCGAAATACTCACCCCG       (SEQ ID No 82)
```

[0221] The reactions were run in triplicate on each DNA sample with the following assay conditions:

[0222] Reaction solution: (900 nM primers; 300 nM probe; 3.5 mM magnesium chloride; 1 units of taq polymerase; 200 μM dNTPs, 5 μM blocker; and 7 μl of DNA, in a final reaction volume of 20 μl);

[0223] Cycling conditions: 95° C. for 10 minutes; (95° C. for 15 seconds, 60° C. for 1 minute) 50 cycles

[0224] The corresponding control assay was performed as described above (Example 1)

[0225] The data was analysed using a PMR calculation previously described in the literature (Eads et al 2001).

[0226] Results. The results are shown in **FIG. 1**. The mean PMR for normal samples was 2.25, with a standard deviation of 2.42. The mean PMR for tumour samples was 25.67, with a standard deviation of 17.57 The overall difference in methylation levels between tumour and normal samples is significant in a t-test (p=0.00000000118). The ROC curve of the assay is shown in **FIG. 15**. The AUC is 0.94

[0227] This was further confirmed using a H Cadherin-HeavyMethyl MethyLight assay, using dual Lightcycler probes using Lightcycler style dual probe technology.

[0228] Methods. The CpG island assay (methylation assay) was performed using the following primers and probes:

```
Forward Primer:
GTTAGTTAGTTAATTTTTTAAATAGATTAGTAG  (SEQ ID NO:106)

Reverse Primer:
CAAAAAAACAAATAAAATACCACCTCC        (SEQ ID NO:107)

Blocker:
CCTCCACAAAACTCACTCCTCACAAAATAC     (SEQ ID NO:108)

Probe: red640
TTTCGTTTTGTATGGTAGATACGGGGTGA-     (SEQ ID NO:109)
phosphate

Probe:
ATTAATGGTTTTATAAGACGGATTTTTTTTTAAC (SEQ ID NO:110)
GT-fluorescine
```

[0229] The reactions were each run in triplicate on each DNA sample with the following assay conditions:

[0230] Reaction conditions:

Forward primer:	600 nM
Reverse primer:	300 nM
Blocker:	10 uM
Probes:	500 nM
Taq polymerase:	0.1 U/ul
dNTPs:	0.2 mM each
Magnesium Chloride:	4 mM
BSA:	0.25 mg/ml
Roche buffer with no MgCl:	1x

[0231] Cycling conditions:

[0232] 95-degree denaturation for 10 minutes

[0233] 50 cycles: 95-degrees for 10 seconds, 57-degrees for 25 seconds, 72 degrees for 10 seconds

Example 13

[0234] Multiplex-PCR of colon cancer samples. In the first step the genomic DNA was isolated from the cell samples using the Wizzard kit from (Promega). The isolated genomic DNA from the samples are treated using a bisulfite solution (hydrogen sulfite, disulfite). The treatment is such that all non methylated cytosines within the sample are converted to thiamine, conversely 5-methylated cytosines within the sample remain unmodified. The treated nucleic acids were then amplified using multiplex PCRs, amplifying 8 fragments per reaction with Cy5 fluorescently labelled primers. PCR primers used are described in Table 1. PCR conditions were as follows.

[0235] Reaction solution:

[0236] 10 ng bisulfite treated DNA

[0237] 3.5 mM MgCl2

[0238] 400 μM dNTPs

[0239] 2 pmol each primer

[0240] 1 U Hot Star Taq (Qiagen)

[0241] Forty cycles were carried out as follows. Denaturation at 95° C. for 15 min, followed by annealing at 55° C. for 45 sec., primer elongation at 65° C. for 2 min. A final elongation at 65° C. was carried out for 10 min.

[0242] All PCR products from each individual sample were then hybridised to glass slides carrying a pair of immobilised oligonucleotides for each CpG position under analysis. Each of these detection oligonucleotides was designed to hybridise to the bisulphite converted sequence around one CpG site which was either originally unmethylated (TG) or methylated (CG). See Table 2 for further details of all hybridisation oligonucleotides used (both informative and non-informative) Hybridisation conditions were selected to allow the detection of the single nucleotide differences between the TG and CG variants.

[0243] 5 μl volume of each multiplex PCR product was diluted in 10×Ssarc buffer (10×Ssarc: 230 ml 20×SSC, 180 ml sodium lauroyl sarcosinate solution 20%, dilute to 1000 ml with dH2O). The reaction mixture was then hybridised to the detection oligonucleotides as follows. Denaturation at 95° C., cooling down to 10° C., hybridisation at 42° C. overnight followed by washing with 10×Ssarc and dH2O at 42° C.

[0244] Fluorescent signals from each hybridised oligonucleotide were detected using genepix scanner and software. Ratios for the two signals (from the CG oligonucleotide and the TG oligonucleotide used to analyse each CpG position) were calculated based on comparison of intensity of the fluorescent signals.

[0245] The data was then sorted into a ranked matrix (as shown in FIGS. **16** to **18**) according to CpG methylation differences between the two classes of tissues, using an algorithm. The most significant CpG positions are at the bottom of the matrix with significance decreasing towards the top. Black indicates total methylation at a given CpG position, white represents no methylation at the particular position, with degrees of methylation represented in gray, from light (low proportion of methylation) to dark (high proportion of methylation). Each row represents one specific CpG position within a gene and each column shows the methylation profile for the different CpGs for one sample. On the left side a CpG and gene identifier is shown this may be cross referenced with the accompanying tables (Table 1 to 6) in order to ascertain the gene in question and the detection oligomer used. On the right side p values for the individual CpG positions are shown. The p values are the probabilities that the observed distribution occurred by chance in the data set.

[0246] For selected distinctions, we trained a learning algorithm (support vector machine, SVM). The SVM (as discussed by F. Model, P. Adorjan, A. Olek, C. Piepenbrock, Feature selection for DNA methylation based cancer classification. Bioinformatics. 2001 June; 17 Suppl 1:S157-64) constructs an optimal discriminant between two classes of given training samples. In this case each sample is described by the methylation patterns (CG/TG ratios) at the investigated CpG sites. The SVM was trained on a subset of samples of each class, which were presented with the diagnosis attached. Independent test samples, which were not shown to the SVM before were then presented to evaluate, if the diagnosis can be predicted correctly based on the predictor created in the training round. This procedure was repeated several times using different partitions of the samples, a method called crossvalidation. Please note that all rounds are performed without using any knowledge obtained in the previous runs. The number of correct classifications was averaged over all runs, which gives a good estimate of our test accuracy (percent of correct classified samples over all rounds).

TABLES

Table 1: PCR primers and products

Gene	Primers	Amplificate length
Versican (SEQ ID NO 2)	GGATAGGAGTTGGGATTA AGAT AAATCTTTTTCAACACCA AAAT	414
EYA4 (SEQ ID NO 3)	GGAAGAGGTGATTAAATG GAT CCCAAAAATCAAACAACA A	226
H-Cadherin (SEQ ID NO 4)	TTTGTATTAGGTTGGAAGT GGT CCCAAATAAATCAACAAC AACA	286
TPEF (SEQ ID NO 5)	TTGTTTGGGTTAATAAATG GA CTTCTCTCTTCTCCCCTCT C	295

[0247]

Table 2: Hybridisation oligonucleotides

Gene	Oligomer sequence
Versican (SEQ ID NO 2)	AAGATTTTCGGTTAGTTT (SEQ ID NO 88)

-continued

Table 2: Hybridisation oligonucleotides

Gene	Oligomer sequence
Versican (SEQ ID NO 2)	AAGATTTTTGGTTAGTTT (SEQ ID NO 89)
Versican (SEQ ID NO 2)	ATGTGATTCGTTTGGGTA (SEQ ID NO 50)
Versican (SEQ ID NO 2)	ATGTGATTTGTTTGGGTA (SEQ ID NO 51)
Versican (SEQ ID NO 2)	GGGTAACGTCGAATTTAG (SEQ ID NO 52)
Versican (SEQ ID NO 2)	GGGTAATGTTGAATTTAG (SEQ ID NO 53)
Versican (SEQ ID NO 2)	AAAAATTCGCGAGTTTAG (SEQ ID NO 54)
Versican (SEQ ID NO 2)	AAAAATTTGTGAGTTTAG (SEQ ID NO 55)
EYA4 (SEQ ID NO 3)	TATATATACGTGTGGGTA (SEQ ID NO 56)
EYA4 (SEQ ID NO 3)	TATATATATGTGTGGGTA (SEQ ID NO 57)
EYA4 (SEQ ID NO 3)	AGTGTATGCGTAGAAGGT (SEQ ID NO 58)
EYA4 (SEQ ID NO 3)	AGTGTATGTGTAGAAGGT (SEQ ID NO 59)
EYA4 (SEQ ID NO 3)	TTTAGATACGAAATGTTA (SEQ ID NO 60)
EYA4 (SEQ ID NO 3)	TTTAGATATGAAATGTTA (SEQ ID NO 61)
EYA4 (SEQ ID NO 3)	AAGTAAGTCGTTGTTGTT (SEQ ID NO 62)
EYA4 (SEQ ID NO 3)	AAGTAAGTTGTTGTTGTT (SEQ ID NO 63)
H-Cadherin (SEQ ID NO 4)	GAAGTGGTCGTTAGTTTT (SEQ ID NO 64)
H-Cadherin (SEQ ID NO 4)	GAAGTGGTTGTTAGTTTTT (SEQ ID NO 65)
H-Cadherin (SEQ ID NO 4)	TTGTTTAGCGTGATTTGT (SEQ ID NO 66)
H-Cadherin (SEQ ID NO 4)	TTGTTTAGTGTGATTTGT (SEQ ID NO 67)
H-Cadherin (SEQ ID NO 4)	AAGGAATTCGTTTTGTAA (SEQ ID NO 68)
H-Cadherin (SEQ ID NO 4)	AAGGAATTTGTTTTGTAA (SEQ ID NO 69)
H-Cadherin (SEQ ID NO 4)	AATGTTTTCGTGATGTTG (SEQ ID NO 70)
H-Cadherin (SEQ ID NO 4)	AATGTTTTTGTGATGTTG (SEQ ID NO 71)
TPEF (SEQ ID NO 5)	ATTTGTTTCGATTAATTT (SEQ ID NO 72)

-continued

Table 2: Hybridisation oligonucleotides

Gene	Oligomer sequence
TPEF (SEQ ID NO 5)	ATTTGTTTTGATTAATTT (SEQ ID NO 73)
TPEF (SEQ ID NO 5)	ATAGGTTACGGGTTGGAG (SEQ ID NO 74)
TPEF (SEQ ID NO 5)	ATAGGTTATGGGTTGGAG (SEQ ID NO 75)
TPEF (SEQ ID NO 5)	AATTTGCGAACGTTTGGG
TPEF (SEQ ID NO 5)	AATTTGTGAATGTTTGGG

[0248]

Table 3: Oligonucleotides used in differentiation between colon adenomas or carcinoma tissue and healthy colon tissue.

Gene	Oligo:
H-Cadherin (SEQ ID NO 4)	AATGTTTTCGTGATGTTG (SEQ ID NO: 70)
H-Cadherin (SEQ ID NO 4)	AATGTTTTTGTGATGTTG (SEQ ID NO: 71)
TPEF (SEQ ID NO 5)	AATTTGCGAACGTTTGGG (SEQ ID NO 76)
TPEF (SEQ ID NO 5)	AATTTGTGAATGTTTGGG (SEQ ID NO 77)
Versican (SEQ ID NO 2)	GGGTAACGTCGAATTTAG (SEQ ID NO: 52)
Versican (SEQ ID NO 2)	GGGTAATGTTGAATTTAG (SEQ ID NO: 53)
H-Cadherin (SEQ ID NO 4)	AAGGAATTCGTTTTGTAA (SEQ ID NO 68)
H-Cadherin (SEQ ID NO 4)	AAGGAATTTGTTTTGTAA (SEQ ID NO 69)
TPEF (SEQ ID NO 5)	ATAGGTTACGGGTTGGAG (SEQ ID NO 74)
TPEF (SEQ ID NO 5)	ATAGGTTATGGGTTGGAG (SEQ ID NO 75)
EYA4 (SEQ ID NO 3)	AAGTAAGTCGTTGTTGTT (SEQ ID NO 62)
EYA4 (SEQ ID NO 3)	AAGTAAGTTGTTGTTGTT (SEQ ID NO 63)
EYA4 (SEQ ID NO 3)	AGTGTATGCGTAGAAGGT (SEQ ID NO 58)
EYA4 (SEQ ID NO 3)	AGTGTATGTGTAGAAGGT (SEQ ID NO 59)
Versican (SEQ ID NO 2)	AAAAATTCGCGAGTTTAG (SEQ ID NO 54)

-continued

Table 3: Oligonucleotides used in differentiation between colon adenomas or carcinoma tissue and healthy colon tissue.

Gene	Oligo:
Versican (SEQ ID NO 2)	AAAAATTTGTGAGTTTAG (SEQ ID NO 55)
Versican (SEQ ID NO 2)	AAGATTTTCGGTTAGTTT (SEQ ID NO 88)
Versican (SEQ ID NO 2)	AAGATTTTTGGTTAGTTT (SEQ ID NO 89)
TPEF (SEQ ID NO 5)	ATTTGTTTCGATTAATTT (SEQ ID NO 72)
TPEF (SEQ ID NO 5)	ATTTGTTTTGATTAATTT (SEQ ID NO 73)

[0249]

TABLE 4: OLIGONUCLEOTIDES USED IN DIFFERENTIATION BETWEEN COLON CARCINOMA TISSUE AND HEALTHY COLON TISSUE.

Gene	Oligo:
H-Cadherin (SEQ ID NO 4)	AATGTTTTCGTGATGTTG (SEQ ID NO 70)
H-Cadherin (SEQ ID NO 4)	AATGTTTTTGTGATGTTG (SEQ ID NO 71)
TPEF (SEQ ID NO 5)	AATTTGCGAACGTTTGGG (SEQ ID NO 76)
TPEF (SEQ ID NO 5)	AATTTGTGAATGTTTGGG (SEQ ID NO 77)
H-Cadherin (SEQ ID NO 4)	AAGGAATTCGTTTTGTAA (SEQ ID NO 68)
H-Cadherin (SEQ ID NO 4)	AAGGAATTTGTTTTGTAA (SEQ ID NO 69)
Versican (SEQ ID NO 2)	GGGTAACGTCGAATTTAG (SEQ ID NO: 52)
Versican (SEQ ID NO 2)	GGGTAATGTTGAATTTAG (SEQ ID NO: 53)
EYA4 (SEQ ID NO 3)	AGTGTATGCGTAGAAGGT (SEQ ID NO 58)
EYA4 (SEQ ID NO 3)	AGTGTATGTGTAGAAGGT (SEQ ID NO 59)
EYA4 (SEQ ID NO 3)	AAGTAAGTCGTTGTTGTT (SEQ ID NO 62)
EYA4 (SEQ ID NO 3)	AAGTAAGTTGTTGTTGTT (SEQ ID NO 63)
TPEF (SEQ ID NO 5)	ATAGGTTACGGGTTGGAG (SEQ ID NO 74)

-continued

TABLE 4: OLIGONUCLEOTIDES USED IN DIFFERENTIATION BETWEEN COLON CARCINOMA TISSUE AND HEALTHY COLON TISSUE.

Gene	Oligo:
TPEF (SEQ ID NO 5)	ATAGGTTATGGGTTGGAG (SEQ ID NO 75)

[0250]

Table 5: Oligonucleotides used in differentiation between colon adenoma tissue and healthy colon tissue.

H-Cadherin (SEQ ID NO 4)	AATGTTTTCGTGATGTTG (SEQ ID NO: 70)
H-Cadherin (SEQ ID NO 4)	AATGTTTTTGTGATGTTG (SEQ ID NO: 71)
TPEF (SEQ ID NO 5)	AATTTGCGAACGTTTGGG (SEQ ID NO 76)
TPEF (SEQ ID NO 5)	AATTTGTGAATGTTTGGG (SEQ ID NO 77)
TPEF (SEQ ID NO 5)	ATAGGTTACGGGTTGGAG (SEQ ID NO 74)
TPEF (SEQ ID NO 5)	ATAGGTTATGGGTTGGAG (SEQ ID NO 75)
Versican (SEQ ID NO 2)	GGGTAACGTCGAATTTAG (SEQ ID NO: 52)
Versican (SEQ ID NO 2)	GGGTAATGTTGAATTTAG (SEQ ID NO: 53)
H-Cadherin (SEQ ID NO 4)	AAGGAATTCGTTTTGTAA (SEQ ID NO 68)
H-Cadherin (SEQ ID NO 3)	AAGGAATTTGTTTTGTAA (SEQ ID NO 69)
EYA4 (SEQ ID NO 3)	AAGTAAGTCGTTGTTGTT (SEQ ID NO 62)
EYA4 (SEQ ID NO 3)	AAGTAAGTTGTTGTTGTT (SEQ ID NO 63)
EYA4 (SEQ ID NO 3)	AGTGTATGCGTAGAAGGT (SEQ ID NO 58)
EYA4 (SEQ ID NO 3)	AGTGTATGTGTAGAAGGT (SEQ ID NO 59)
Versican (SEQ ID NO 2)	AAAAATTCGCGAGTTTAG (SEQ ID NO 54)
Versican (SEQ ID NO 2)	AAAAATTTGTGAGTTTAG (SEQ ID NO 55)

[0251]

TABLE 6

Genes analysed according to FIGS. 16 to 18

Number in Figures	Gene name
	Healthy vs Non-Healthy
50-D	H-CADHERIN
20-C	CD44
54-C	TPEF (=TMEFF2; =HPP1)
21-C	VERSICAN
50-C	H-CADHERIN
25-B	GSTP1
43-C	TGFBR2
36-B	N33
49-A	CAV1
52-C	PTGS2
46-A	TP73
54-B	TPEF (=TMEFF2; =HPP1)
20-A	CD44
24-D	EYA4
24-B	EYA4
26-B	GTBP/MSH6
4-C	EGR4
15-E	CDH1
23-E	EGFR
30-B	LKB1
22-D	DAPK1
29-D	IGF2
10-A	HLA-F
29-C	IGF2
36-C	N33
21-D	VERSICAN
39-D	PTEN
32-B	MLH1
26-A	GTBP/MSH6
14-C	CALCA
22-C	DAPK1
39-C	PTEN
9-D	WT1
23-A	EGFR
21-A	VERSICAN
30-A	LKB1
9-C	WT1
60-E	ESR1
12-A	APC
29-A	IGF2
8-D	MYOD1
36-A	N33
54-A	TPEF (=TMEFF2; =HPP1)
18-E	CDKN2a
15-D	CDH1
12-C	APC
	Healthy vs Carcinoma
50-D	H-CADHERIN
54-C	TPEF (=TMEFF2; =HPP1)
50-C	H-CADHERIN
21-C	VERSICAN
20-C	CD44
24-B	EYA4
12-A	APC
52-C	PTGS2
24-D	EYA4
39-B	PGR
25-B	GSTP1
49-A	CAV1
23-E	EGFR
36-B	N33
29-C	IGF2
10-D	HLA-F
54-B	TPEF (=TMEFF2; =HPP1)
46-A	TP73

TABLE 6-continued

Genes analysed according to FIGS. 16 to 18

Number in Figures	Gene name
	Healthy vs Adenoma
20-C	CD44
10-A	HLA-F
43-C	TGFBR2
26-A	GTBP/MSH6
26-B	GTBP/MSH6
30-B	LKB1
20-A	CD44
36-C	N33
50-D	H-CADHERIN
46-A	TP73
39-D	PTEN
36-B	N33
54-C	TPEF (=TMEFF2; =HPP1)
25-B	GSTP1
23-A	EGFR
40-A	RARB
36-D	N33
49-A	CAV1
54-B	TPEF (=TMEFF2; =HPP1)
18-E	CDKN2a
36-A	N33
32-B	MLH1
12-C	APC
21-C	VERSICAN
15-E	CDH1
52-C	PTGS2
62-D	RASSF1
9-C	WT1
18-D	CDKN2a
60-E	ESR1
29-D	IGF2
8-D	MYOD1
50-C	H-CADHERIN
4-C	EGR4
42-C	S100A2
22-D	DAPK1
31-E	MGMT
24-D	EYA4
56-A	CEA
9-D	WT1
7-E	GPIb beta
14-C	CALCA
52-D	PTGS2
8-B	MYOD1
24-B	EYA4
21-D	VERSICAN
38-C	PGR
58-A	PCNA
34-D	MSH3
9-B	WT1
35-B	MYC
27-C	HIC-1
52-B	PTGS2
23-E	EGFR
30-A	LKB1
29-C	IGF2
39-C	PTEN
13-D	BCL2
5-B	AR
15-D	CDH1
	Carcinoma vs Adenoma
18-B	CDKN2a
7-E	GPIb beta

[0252]

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 110

<210> SEQ ID NO 1
<211> LENGTH: 965
<212> TYPE: DNA
<213> ORGANISM: Homo Sapiens

<400> SEQUENCE: 1

gatcaattaa gggcatctta gaagttaggc gttcccgctg cctcctttga gcacggaggc 60

caccaacccc taggggaag agatgtagcg cgaggcaggg gtgtcgtgct aagaaatttc 120

gacgcttctg gggactgagg acaaaggtgc ggacacgacc ccggggtacc tggagttccg 180

tgactcgcgc cacggacggc acacctaggg gctaatttct gctctgcctc aaagaacctc 240

aagctagagt ccttgcctcc gcccacagcc ccgggatgcc gctgctgcgc tcaccgcaca 300

ggcagcgccc ggaccggctg cagcagatcg cgcgctgcgc gttccaccgg gagatggtgg 360

agacgctgaa aagcttcttt cttgccactc tggacgctgt gggcggcaag cgccttagtc 420

cctacctctg ctgagctgaa cgctcaggca cagtggaact gaaacccggt tccctacctc 480

tgctgagctg aacgctcagg cacagtggaa ctgaaacccg gttctgcggg atgtgagagc 540

tgttgaggtc acgcgtaatt gggtgtgatg gagggcgcct gttcgtgatg tgtgcaggtt 600

tgatgcaagc aggtcatcgt cgtgcgagtg tgtggatgcg accgcccgag agactcggag 660

gcaggcttgg gacacgtttg agtgaacacc tcaggatact cttctggcca gtatctgttt 720

tttagtgtct gtgattcaga gtgggcacat gttgggagac agtaatgggt ttgggtgtgt 780

gtaaatgagt gtgaccggaa gcgagtgtga gcttgatcta ggcagggacc acacagcact 840

gtcacacctg cctgctcttt agtagaggac tgaagtgcgg gggtgggggt acggggccgg 900

aatagaatgt ctctgggaca tcttggcaaa cagcagccgg aagcaaaggg gcagctgtgc 960

aaacg 965

<210> SEQ ID NO 2
<211> LENGTH: 16579
<212> TYPE: DNA
<213> ORGANISM: Homo Sapiens

<400> SEQUENCE: 2

tgtttttaa aatgatgtct agctgaacca gccaatggag acaaaaatat ttatatataa 60

ttttcctcta agttcactag tcattttata gatcagttat aactttggct actccaatga 120

gttaatttct ttctagagac attataaact ccttaaagat ggaaaataat tgttacttt 180

ttgaataatg cctactatct aaaacaatgc acataaaagc tgctcaataa acttttgttg 240

aattgagaga aaatataaaa aattgaaagc agtctgtttg taataagtgt ttgctaagtc 300

ttacagatct tctctttttt ttggaattat tgcaacaaat atttgacctt tgatttgtag 360

gattaaaatg tgatttctca tattttaata ggagcaagca ttagtcctgc aattcttttt 420

gtcaatagaa taaaaatgga atgtatgctt gctgcttcta catgccggtt gatacccatg 480

ctgaacactt ttcttacaac gcagaggttt ccttgttttc tgtcttgcgt tctacctttg 540

ctcaccacag agacattaat gatgccatgt gtctttttgc accatcagtg cagcagcttt 600

-continued

cattatgctc cttgtctcgg tcccactgta gtctaattaa ataaatggaa acgggtagtc 660

actgaaaata aacagcaact gcaaaagttt tgcctcttaa gaaaccttct gcattgtatt 720

ccgccctgcc tcattgattc taagatctct accgcttcct tgtgttcaga aaccagactc 780

ttcagtttgt cttgcttcag tcagacaaat gctttataaa aatgacctat atttttcaga 840

ctgtggggag agaagaaaaa gactttgaga aagtgactaa ggagggagta agagtgatac 900

aatctttgat tcaatttagt tggactgtgt tctttatgcc ttcatctttt tctttataaa 960

aaagatagcg tttttgatgc tctgctccag gcagcctccc agggtgagtg cagatgtgct 1020

gagccatttg gctgccttga ggaccggctc agagagcctc tcttctcact aatgaagtca 1080

tctcctctcg gcagcccggc ctctcccatc actccatgtt ggaacttggc tttggcttgc 1140

atagttctgt cgtgatgtag atggtctttc agttaagttc tagagaaatc tgctcaacca 1200

tgaacaggac atttacagca caggagaaaa aaaaatttat taaaacaaat attttactgt 1260

aacattttga tttgggggtt ggagagcaaa aactcaatat ctgatgtctg ggaaaaatct 1320

aatgttttat ggctttctaa ggaaatttgt ctgtgtttaa tttatcagag caatttatat 1380

acttcgggtt aaagcacaat gagattgcca ccatcacttt gccttttgca agatttcttt 1440

tttaagaata aaaagaaggc aataaagacc ttatatttct gcctttccta gaaagccaaa 1500

gagaattcaa agtgctggat gaaatgtctg tccctaattt gcagtgtttt atttaaaaaa 1560

aaaattatga cacaatctta gaggataggc ttccttttc cagggtcttt gaaagttaaa 1620

agtacatttt taatctgtag cattactgta tcatccagtg aggataattt gcttagttgt 1680

gcctaggacc agtcagatca aggagttgaa actaagctct ttctgaattt gaatgggact 1740

ggaatcttgt catccctaga gcagcaatag atgatgccat gtctaagaac acttcacctg 1800

atcaataaaa aaggagaagt ctgaaatgta atcctaaaga acctcagaag gaggtttaga 1860

tgaagataat aaaattccct ggagacacat ctgccatcta caccttcct ctccaagtca 1920

ttggcttgct gcccctgagc tcagccaaat tcccaaagac aagatcaaag caaggaagca 1980

aggcacatag cttctggctt atggtttctt ttagacacaa gcattgtgaa tgaagatctt 2040

gctttattta tacctgggat cttggagttt tcacaaagaa tatccccagt acaaggatgt 2100

gatttaaagg gggactcaca gaataatatt ttaattactg tttgattcat acttgagtgt 2160

aatggacaaa taaatacaat tttagttata aaacctgaaa aatatatatt ttcccattca 2220

aaaacttctg atagttttgt caggcattgg agaatcagaa ataaatgctc tctcattttc 2280

tttctcttga tgaaggaaag tcattaaaaa cagttacata gaaataataa tacaactgta 2340

gactctattc taaatatagc tggtatgttt tccttttct tttttatcca ttctccattt 2400

tagacagcta gaggggaaag tctgccctag agatagaaag aagatatcca tggggtataa 2460

tttcatttta gttatctgat agcctgaaga agtatttaat tttctgaata ataatgttac 2520

aagtcaaaga agagaaaaga caggcagttt aaagcttttc aacattgtgc atattctaat 2580

tccttaagtc agccctgaga aagtgccttt gtcctgtaga agtatttggt caagtgttgt 2640

gaagtctatt ggcattaaaa gaaactaaaa attttggagg attagagaat caaatttggc 2700

aacataaatg aaaacatctc agaatgtaaa aagtatctaa cccatcactt tcaatcatgt 2760

tccattatat acagatggca ttatttatag tgaatgttaa ctattaacat ttccctaaaa 2820

atgggcatta atctgctatt agttgcttgt ttctgatcta gttctccttt gttgaaaggt 2880

-continued

```
tttacttaac tgtatcataa agggacaaaa gggtaaaaga taataagctt ttcagactgg   2940

cagatctgtc acttacaaaa tggccttggg caggttactt gaactctttg aaactcaact   3000

agatttactc aaatttaaga atgagaatac aaatccacat cttgaagtgt ttcacagaaa   3060

ggtctatctt aatgtctgga gtatatattt caatgaacat tcattttatt ttatttctct   3120

ccattcctga atcaagcaat cttgaatcta aagttgctat gattagcact gaaaagacca   3180

ctggactatt aattgtgtga ctttgggaca gtaactttct gcaccttagt ttgtttacat   3240

gttatacatg aaggttgaag tctgattctg ctctgtgact atcattctaa acatctgatg   3300

aaatcaaatt tcagtgtttg gaatggtagt acaataaatt tactaagaat aaataattca   3360

ctgcaaaaac acattgattt ccaaatgatg taactgacag ttatattact gcagagggct   3420

gataaataac aaaagaaatg aaagatgcac atggtgagaa ctgaaattat cctgacaagt   3480

cttctacctg tttatcactt aaaatcaatg accatgctga atgcctacaa attacaaaat   3540

ataaaagaaa tcttataaat gcgcatgtac aggagtctaa gttactaaaa gttttaaagc   3600

ataagtttaa accaaactaa tcaaagaagt tgagaggaaa aattggcttt catctttaat   3660

cactactgtt ttgaggtcct atgtttaata taattttcta agtagaggct tcagagagaa   3720

gagttgtgag gatactttca tatttgtgta gaaggaaaag tttgccatcc attctagtat   3780

ccctagtgtt atactgatgt gcaccttgga tttattttgt tcctattgta taaactcata   3840

cttgacttca aagaaaagga aaatccaaag tccctctttt ctaaggggac agaaatcctt   3900

tgtgtcaact gtttgaccct tttctctgta aggtcctatt ggaaatcttt tgtaacacaa   3960

tgcaggggac tcttccatgt gttgatgctg tttacacagt ggggtgggcc tgactgaaga   4020

aaaaaaatcg catatacgca tgaaagatta tggtcttatt tccggaaagc atgaaaggtg   4080

attgatactt ccaagaagtc cctgttactc aggaaaatta tcaaatattc tactcagaga   4140

tacttggaaa gactgaagga aaggaagaac gaagaaagca gaatctagac ttatgtgggg   4200

agagatttgt ggcagaggaa aagtattctc tttgaatccg acaagggatt tgcctggggg   4260

aatttcctgt ccagcctttt attaccaggg tcttttgaag ccgggctccc cattgggcag   4320

ttccctggga gtgcagtggg gaattcttac actttccctc taggtccccg aaggatctcg   4380

ttttctcagt gtctctttca ggttggcagg agccttgagc ctgacacttc cctttgatgg   4440

gacaggcaag ctctgtgggc gcgtaaacac gctgtaacca agttctttgc tgattttaca   4500

gttttgtgtg ctcccgagaa gaagtgatcg tactcaattg tctattgctg gcctgccccc   4560

taagagcctg ggggctcctt tcccctaacc cagaactagc tgcacggggg gcggggaagt   4620

gggggtgggg aaggagtggg agggcagtgg tttccgcgag cagagcgatg ttactgagtg   4680

agtccctgaa tggggagcgc tgctgtcccc aagccgattg gtacttcttg tcaggaagaa   4740

acgccaagag gtgggagtgc ctggggaggg aggcaggcgg tccctaccgc aggcgcgggg   4800

agctgccttt ccgcccctcc gcctgctttc caagcctgga ctcttaggag tggctgaagc   4860

tgcggagcgc ttttggagcc tgtgaatgaa ccctcctcct ctccctcctc cttcttctcg   4920

ctgagtctcc tcctcggctc tgacggtaca gtgatataat gatgatgggt gtcacaaccc   4980

gcatttgaac ttgcaggcga gctgccccga gcctttctgg ggaagaactc caggcgtgcg   5040

gacgcaacag ccgagaacat taggtgttgt ggacaggagc tgggaccaag atcttcggcc   5100

agccccgcat cctcccgcat cttccagcac cgtcccgcac cctccgcatc cttccccggg   5160
```

-continued

```
ccaccacgct tcctatgtga cccgcctggg caacgccgaa cccagtcgcg cagcgctgca   5220

gtgaattttc cccccaaact gcaataagcc gccttccaag gtaatcacgt ttcttttgtt   5280

ccccccttaa aaaacaaaaa caaaaaactt atagaaaaaa acccgcgagc ttagaaaaaa   5340

gaagcaattg gtagaaggct ttaattaagg caaagagctg taaggcgaag ttaagaaaat   5400

gtaggcactt aaaaaatgca ggtaactttc ataaaggctt ttggggagag gcatacagag   5460

ggaccttggt gttgaaaaag attcagacaa aagaaaccca gggtggggtg ggggtaaaat   5520

gactaacgga attgggggaa gggagggaat aaattgtaaa gaaatcatag aaaagtggtg   5580

ggttcttgag ctggagagaa gagagggacc tttggcactt tgatttttt gttgttgttg   5640

ttcttaacac gctcgaggca aaagtttgaa tggggactac caagacttgc cacagacaag   5700

tccccgaagc cgccttggtg caggccacct ggtttcccag cccctggtgt gtggtcagtg   5760

cctggtgttc ctggaaagcc actcccgggc agctcctgac agtgcgaccc ggcgcccaag   5820

cagcctggga ccttgcgcgg acctgacccc ttcagaccgc aggcagtctg ggaggaggtc   5880

cggccggggg aggtgcagga tccccgccgt gtctctttga cgacttgggg actgtcacgg   5940

ttctctcccg gcgcccctgg gttcttttgt cctgcacgcg gtgcgaaggg gccagcaggg   6000

aaggagcaga ggatgggggg tggggttgtt ggagccccgc ggaggtctgg gaggcccctg   6060

ggcgggaaaa gcctgttctg aatcggcagg gatgtgcaac aatttttctc accctgaaga   6120

gtgaaatagg gtctgtcgct cccatctcaa caagcaaacc ggcacccaga gcgcactgca   6180

gacaaaggtg gctcggggac ccgaatcagg ggcctctggg tcagtcctct cgcccagact   6240

acaggagtcc tccgtttcct tacatattcc ccaccctctc ctagttctcc gctctagttg   6300

agcaacttta ctggcaggtc tagcggcagc ggcgcgcgcg tgtgccggga gccccggggg   6360

acgtcctcgg ctggagcgcc ccaccgtcct gcagaggcgt gggcgctgta gggcgacatc   6420

cctggtgcgt gcagacctag ggcatccggg ttgttctggc ccgcggtctg tgtcactggg   6480

cgtggagcgg tctgggtgtt aggcagagga gagcggggta gaaaaatagt gtcatagcct   6540

aaactatttg ctctccaatc caactgcgtc tcggcaagtc ctgcatgtcc ctgggagatg   6600

ctgcgggaag ggggagaaat ctacacggcg cctggagagt tgtcctgccg cgcgcacacc   6660

cgcggcagag tctctctggg tcgcgttcct catctttgtg tctccctctg tctccatctt   6720

ctctgcctcc cttcccctct cccagcctct gttctctctc tctagcttct gcgtcccctc   6780

ccccacagct gacaaatgaa tggctagtgt gaaatccctg cctttcccgt gctccaaggc   6840

agcagggagg gaggagcgag ggaggcggtg cgctccttcg gatctgcccc cagttcagct   6900

cacaagactt gcagaaccta gatgtctagg aattgggagt tttgcggcgg gtgtgggcgc   6960

cccctgatgg agaagtcccg cacaggcgga gaaaaacaag ccccccagac caagcgtgca   7020

tctttcacaa ccttgtgttc aaggatggaa ggccctagtt ttcccttaag tcattcattt   7080

tgcactccac aatctgcggc gtacatctag gagtttggag gaccctgaaa aaaggtcttg   7140

gtctgtgcaa agtgcagaga tgccttctcg cgagtcggct ggaacgcacg cggcgccctg   7200

ggtccagttc gccgcttagt ggacgaacgc acatggccgg ggtggagcca ggtttcggag   7260

gtctagacgc gccaccctgg gcgtctttaa gaaagataat acacataccg tgctctcaaa   7320

acccatgacc acttgaaccc aacggtaggg ccccgcccca ggatattgca aaagaagggc   7380

tcctaatcca aaacttaaac ttcctcaact tcagggcggg cgtcggacgg gaaagtggag   7440
```

-continued

agaaggcggg cagtgggagg aaaaaagaaa gggaaggaag ggaagtggaa ggaaagagga 7500

ccgtggaggg gaggaaaagg agaaaggaac ggagatgggg cagatggcat tcaaacagcc 7560

aggaacggac cggaagggtt cgtttcgggg tttgggtctg aatttcgaga ctcgttttag 7620

gatactcttt tccctttccc agcggcaaca aatgtgactg cagaggcgtg ccgggatgga 7680

agggtgggaa agaactagac aagcggaggt ggtccagccc cgcccaaagg ggctgggtct 7740

ggggcccgcc cccactcgct cacctcgcgc caagggagga cccgcaacgg gcgctccccg 7800

cagcggctgg gttagccagc cgccacgtgg gttggcgctt cggctcccag ccgagacggg 7860

tgtgacttcc tgggctcaga gactgcagta cagccgcaca ggcggccggc gcagccgagg 7920

cccccaccca ccgaggccaa tgggggcggg gcggggaagg cagggcccag cttaaccgag 7980

agagccgagt ggcctcgccc ccgggattcg gggccggctg cttctggagc tgacccgagt 8040

tccatccaga cagcgctgaa cactcttatt aaaaggctat ttaagtttta ttttatttta 8100

ttttattttt gcttgcttct tgggagtggt aggaggggca tcgtgctttt gttttgaggc 8160

aggggagagg ttaaaattta ctccaagcaa actttttgtc agaaaaatga gactttatac 8220

cgtattccat ttgttgactt tttaaataca ttcaggtcac tctaaagccc attcttagtc 8280

aattaacaga ttaatttcgt tccagactgt cataaagagt tcgcctggaa tggcaaatac 8340

tgccattcga caagactaat agaatttcat gggggcaaac atggaagggc tgcgagtgta 8400

actgaggttc ctgtgcctca atgaggaacc ttagcgggga cctggcgggg actggaagct 8460

cagcacctat ttagggagtg actgagctgg aaacctggag aagcttggtg cagacgagac 8520

ggtgcttctt agagctgcat tttaagtcct cccattttgc tgcatattgg aagaactgga 8580

ttttaaagat cttttccatc acctcagatt ctgtagctaa cagccacgtc tttttctttt 8640

tgtctgctga aagctgtgac ctaattgttt taagcacttc ttgacaattt cgccaaaccc 8700

atcctgaaaa tattcacata aagtcctaaa ataggatgtt gtgagtgtaa tataaaaata 8760

ggaaggtacc aatggaagca acaattagtg ttctgagata gaagcggttc ccagtttgta 8820

acttggaaat gttaagagca gtacatcagg gagaaggtct ccaagagtct gacttcttgt 8880

tctgagctct gagtaaagaa agaaatatgt tagaattgta ttgtctgccc gcagtaatat 8940

tgtaaagggt tccaatggat tccaatgtga ttgagcaatt acttcctcca gcaccaagtc 9000

cacagtcctt aatcttcttg tgagaagatt aggtaggaag cgggatgaga ttcaagacaa 9060

tcatttagga tttatgtcag aattcaaagt tttgattctt ttggtagata aaaacagagt 9120

tggttttaaa ttgtatttcc tgctttcaaa aaaacgaggt ttgaaaggaa tccccctaga 9180

tttacatagg tttacagatg taagtttctt ctttttccct gtgtttgcaa actggaggga 9240

gatgcctttg atattccaga atacctgtat ctatacactt gtaaattctg ctaggaaagg 9300

acctaaacag ctaaaccaaa ggtgaccatg attaatgtta aagatactac ccattgttct 9360

ctcttttgtt gtattattag gcttaaatac tttggagtta atgattggtc acttaattga 9420

gtaacagagc ctttgagagc gcaaagatca actaactctg tctcccaaag tctggcaact 9480

aagccctcct ttcctggtga gatttaaaca tttttgttga cgtcattcat ggacaaacat 9540

cggtaaacac tgtatttaaa atgggatatt gccctgtgat ggagcagaac ttctcaagac 9600

acatggttct tttgaaagaa agaattaatg actgtagtaa ggtatttttt caccacccac 9660

gtagagaaac cagttaggac tgacaaacac tcattgttgg gagttacttg ctctggtctg 9720

-continued

agaaaatatg ctgattgtta cataggtctg tattttacca gaaaacacaa agcatgtact 9780

ctccctgatt tggaatcaga aaaggggat tagatatgga tgtcttggcc aagctccatt 9840

tcaattagtg aatttttgcc agtttcctct tgtagcctag acttggtaaa ttacatatat 9900

ataaagtctg tgtgtgtgtg tggggtgcgt gtgtgtgtgt atgtatacag ggtttttaaa 9960

cactgctaag ttcttaccat tgaatgtgtt agaggtggtt gtatttcaga ggtcagtaca 10020

aagaatcttt gtttaatgtg tcagttctgt ttcctgaaac acacctctgg gtagatttac 10080

caactataaa aggatacaaa tgtatatact ttccacatct gtatacatcc tacctcattc 10140

ttctaaggat tttaagaggt gtttatgtct ataaaatgag attacatatt ttacaataaa 10200

catatttgat aaattataat ctttttatta ttaaattatt gatttgatat tacttgaggt 10260

atagtggaat gagcactaga cttggaatca gaaagtttga gttggactca ctagctgaga 10320

aaaccaagca aagatattac ttttggacct ttagtgtcat aatctattaa atgaatatgc 10380

agttataaag atcaaaagag taagtacttt gaaaaactat acaaaaaagg gatgataata 10440

tattatgtct tgggattcat attgatctaa ttctataatt ttaataataa ctgcattacc 10500

caagccatgt agtttctact gaatctataa aataaatcag agaggggcat attgttttaa 10560

aaattgttta gtttcattaa tcagaagaca gattacagaa attgactcca atgaagttgg 10620

acattcagtg gaaaggaaga attattcaat aaatggtttc aggacaattg atgggtccat 10680

gcaaaataag aatcagtctc tacctgattt ctcatcctaa aagaaatata agatgaatca 10740

aatatgaaat gtagggaata acttttttaca aaatgataaa agaaaatcat gggattaaaa 10800

aaaataagtt aggaatagag aaacctgttc taagcaaaac ataaaaccca gtgccattct 10860

aggaaatact gataggtctg actaagtaaa aattacacaa aagcaaataa tatatttgga 10920

aaacatctac gacatttgag agacaaaaag ctaataactc ttgatattcc atggactttt 10980

atgaataact ttgaaaagtc cacatctatc aaaatatgag caaaaattat gtacaggtag 11040

tggcagaaag agacatagag ttttaaaatt atagaaaaaa tcgctcagcc tcatctgtaa 11100

ttaaataaat agtaattaaa acaccaatga tatgctacac catttttca ccaatcagag 11160

gagtaacaaa aaaaaataat ttttttcttt ttttttttt ttgagatgga gtctgttact 11220

caggttggag tgcagtggca cgatctcagc tcactgcaac ctccgccttc caggttcaag 11280

caattctcct gtgtcagcct ccctagtagc tgggattatg ggcacccacc accatgcttg 11340

gctaattttc gtgtttttag tagagtgggg tttcaccatg ttggccatgg cagttgatct 11400

tgaactcctg acctcaagtg attcacccac ctcagcctcc caaagtgctg ggattacagg 11460

catgagccac tgtgcctggc aaaaatgaaa ttaaaaagtt tgatgcaagt ttgtgttggc 11520

caggttgtgg agaaactggg tctctcatat actgttgtat gaaagtaaat aggttcctct 11580

acttcaaaaa ataatttggc aatcactttt ttttgattct ttatcaaaag tttttaaaag 11640

atgtatatac tctttgattc agcaattcca cttctcgaca ttattatatg gatggatatt 11700

ggtaccaata tcacccaaga caaatggaca aggatgttct ttgcagcatt gcttctaaga 11760

gcaaatatct ggaaacagtg taacaggctc atcataggga accatttaaa taaaagttgc 11820

tgtatccacc taatggaaaa ctatgcagct gtcaaaaata acatggaaaa attgtatttg 11880

ctgatttgga acattctcca aaatatattg tattctaagt gaaaaaattg atagataaaa 11940

ctgagtatat attatacatc taattatgta aaatgggtgt atattcattc ttttaacaaa 12000

-continued

tatttactga gtgttctagg cattgtggat acaggagtga acaaaacatc aaagactgct 12060

ctgttggaac ttttactatg gtaggtagag aaaggcaatt cccaatcaca tatgcaaaat 12120

ggtgatatgt gctaggatga aaatagagca tgggaaaggg aatagagtac agtatgtctt 12180

aggtagagta atgagaaagg agttcaccag gcagagaaca gaggagtgct aaggggcgac 12240

aatgtgaaaa agcttagcca agtccaggtt caggaattca catgatgcta ttacttgtat 12300

gaagtaaatt catgaaattt cgcacctaaa tcacgtgatt tctctctgct gtacaatgat 12360

tctttctgtg agaaagaacc agtgagcaat atgatatgtg atgctcatta catccacttt 12420

gtagctgaag tatcctagaa agttggttgt ctagggtacc taattagcag aatggtggac 12480

aaaaccaccg tgtctctgtc cttaagtcac cccttgtcag ctgagcttcc ccaaaatcaa 12540

gggagaatac taaatatcta tttaatagaa tcttggcttt ctctcggagg ctgcccatcc 12600

ctggtgtgtc ccttatcctc cctgcagaca tttctcatgg ataggagttc cattgacctt 12660

gtcttttatg tctcccaggc aattaaacag tctcagtacc ctgtacctgg attgttcaat 12720

gactttcttc ccttctcatt gacttcactg gagtagaagg ctaaatatag gcattcactc 12780

tcatatctct caggtctttc catgcagaca gagttgttat cactagtaga cggtgcttat 12840

aagagagaca tggaaaagtg gaagaatgga aatagggatt tggagtggtg ctaaaaacaa 12900

agaaagaggc ttctgaaagc tttccattta atatgatcaa acatagaaga caaaagatgt 12960

aacttaatga ccaaggatag ggacatagcc ttgagaaaat tacagaagtg gaatgaacgt 13020

cttggactcc aaagaagagg gaatggaaat gcagttctga ggcagctaaa gtcaataaag 13080

ccctctgata tatttactgt agaaataaat gatcatcttt taattcaagg tttgcaaact 13140

taggtgacta caggggcgag gcatggagtg gaaatgggta agagactaag tgggggcagt 13200

aatggccttg agcacacaga cttaagggga gagactgctg cttggttaat gctcatgatc 13260

agccagctga aatatagact tagtgttgct agaccttctg atttttgaa agaagttaga 13320

gattttatat ttagtgtaaa cagtctaatt tttaaattga gtaacttatt caaacatatt 13380

tatgcaactt attcaaatgt atttagagca tgtctgatct cctccttgta ttcttaccta 13440

gtcttagggg taatcttaag gaaagtgatg cgtatacttg ttgtacctat gtagatggag 13500

gaccattctc ctctccccaa caaaagacac ttttataagg attgagtgtg gcacaaagag 13560

ttcgattttt agtatttaaa ttgagtagga aactcaaata cagaagatcc tttcctgggt 13620

cacacctttg gttttttaa atgagatttg actctctaga tattgacaaa catcatatta 13680

ttgggtcatg gtaaatcttc aagtggactg attcaaaagt gttcaaaagt atttcaggtg 13740

ttaaggaatt gttggcacgg agtttcaaag gtgtttctgc ataagaaata gaggcgttga 13800

tcatgtattt gatctacaaa cattcctcaa ggatataccc tgaacctggc aatgaagaaa 13860

tattgtggga attattaagt tgaataaagt actagctctg ccttccagaa gcctactgtt 13920

ctacagagga gattagacat gtttacagat aatttgaagt acatagcaat aaattatagg 13980

atgatttatt ccaagcataa agtgtatatt gctactatag tggtaaccta ggttataatt 14040

ttcaaatgta agagccccaa atactattct ttttactctt agtgatatga aactatgtta 14100

gaacatctgc agaaagtctg aattagtaat tttaataata aatattgagt actgatatat 14160

tgtatatcac tgagatttat tttatataca ttattttatc aattcctatg atcacttcgg 14220

taagtactaa ttttcacaac taagtagtga cagaaaatga ggtttagaga gatatagtat 14280

-continued

tttacccaaa gtcacagaat ataagtagag aatgcagaac ttgcattctg ggctatttga 14340

cagtggagcc cagactttta aaattaggct tataaagatt tcctggaaaa aaagaaaaat 14400

aaaaaataga gaacaaattt tcctgaagaa aagtaggaaa ttattattta taagtataat 14460

taaagagaga tctttaattt ctttttttt ttttttttg agacggagtc tcgctctgtc 14520

gcccaggcgc aatctcagct cactgcaacc tccacctccc gggttcatgc cataaagaaa 14580

gtttctattt gattggtccg acctgattgt ctactagaac ataggctttc cagtagggtt 14640

aacaagtcat aacagttcta gtgatataag ataaatacaa actaataata aggttaattt 14700

agaattatca acatatgttt ttcattgtga attatcagtt ggtctataga tggaaaatat 14760

aagatatata atgtattagc ttaaaaaatg ctcttcagag agtgtctaga attatagaat 14820

ttaaaaattt tgttttgaaa aggtctaatt cttatttcaa aatatataaa gaaatcaaat 14880

gcaaaaaaat taagagaaaa ggtcttaatt aattaaatgc tgtacaactc caagccctat 14940

ttatgttcaa gatcagtttt gtaatttatg actaatttca gcattagctt gttttcaatc 15000

taggtttgct aaataatata gaatttcttt tccaatgaag tttttcaggg tcccagctga 15060

aaaatatata ccagttagga tgctttcaac tgtaaggaac aaaaagtcac gattttattg 15120

ggtttaaaca ataatgacat ttataatctc atataaaatg aagttcaaag gcatggtcca 15180

ctctaggtct aatccttaag gacttacgcc ttgcctttgc tgattctttc tcagcttcat 15240

ttccagaggt atctcactta ctcatggatg caaaatccct acaacagttc cagcagaaag 15300

attccagcat gataatgtcc agggaagaaa actgaccacc tcttcctttg ttctcttctc 15360

aggatgaatc tatctttccc agggacccct cactggagga agagatcccc tcacatctca 15420

ctggccaaaa ctatattata tactgttcca aatccaatca ttggaaaagg gaatggtacc 15480

tctataattg acttaggata tcaggatcat cccttgggct ggggtttgac ccaggattct 15540

ccaaagtatg tgactgaggg taggcactag aatggaacca gagttgtgtt agaaaggaga 15600

aacgcatggc tattatagaa tagttctaaa tgctaccgag gagggcacaa ctgtaaaaac 15660

caaataatct tttgccattc ttttgaagtt ggcactttga tttctagatg gttccccaac 15720

acaggttctt ctcctcctca tcattatagc tgccttgaaa tttgagctgg aagggaacat 15780

tctgagaccc agattgttaa atgtcttttc caaagtcatg caataaatta aatggcaaag 15840

ccagggcagt ttcttgactc agtacagggt attttctttc attctttact cttgagactt 15900

tagaactgtt ggtactgctt taaaattcat ggcaagaact ggtcactttt gtaattaaca 15960

cctccttata atacatttgt tttgtttgct tagccagcta gaaactacat ggagtctgtg 16020

ctttaaaaag cctgccgaag tccttattct ctgttttggt attatgtgca tgaaccacca 16080

attggttcct tctcacctta cacttgatga agatgccttt ctttcaacat ctttctctat 16140

tgctccccat cttctcttgc tctatttatg atcagctgtc tgtttctaaa tagactttgt 16200

ggtcacccat ttctttttgt gccagctcct atccactagt tacttgaact gtggtctcta 16260

ttgttcttca tacagcctac tcacttgtgg ttgtcacaca catccacatt gttatatgtt 16320

cccaaactgt attctggaat gattttggta atggctgcat tggatgagat tcaaactaat 16380

aattaaagca ttgagatagc ttctgctatt acaagtttac tcctgttttt atagtccaag 16440

aggagctact tcttctactc ttattactta atgcttaata ctacccttat tacatacaat 16500

-continued

gcacaaaaag catgtgattt atgacccact ttaaacctga atgcttgtga tttactttgt 16560

gtttttcttc atttctagg 16579

<210> SEQ ID NO 3
<211> LENGTH: 3000
<212> TYPE: DNA
<213> ORGANISM: Homo Sapiens

<400> SEQUENCE: 3

acatttaaat tgcatatttg gcttgtaata tattacatat ttaaaagaat tactattttt 60

ctgaatacca ttaatactaa taatagcatc acagtgacta agagtggact taaataaagg 120

ctggcttgaa actcaggtct gccatttatt agcaagcttc taaaaattct gagcccttag 180

ttttctcaca tgtgaaatgg agaaaataaa tctgcaaaat taattaaatc tgtccattca 240

tatttttctc tccagtgtat attaactggc attcctcgtt aggccagaat gtgctctcaa 300

ccatgctcca aatccgcttt gtgccaaccc cactgccaga acccttttcta ccttgagaac 360

cagaaaagga aacattatgc ctggcaatgc ctacaccctc caaaataaat ctgcaggaaa 420

gaacacccag taagtgatga gagcagcaac gactgccttt atcattttaa atttacaaca 480

ccacctttc tagagcctct taagcattgt agataattcc ccactcatta aaaaataaat 540

tgtaaccata agtattcagg gttgatactg cttttgaatt agacagtgct catatcagtt 600

gcataagacc aacctaaagt agaggatgaa atctttttc tgaacctttt tcagaacgta 660

acttagtgaa tatattaaaa ctaaactttc tttgaatggg agtaatttct acggattaat 720

ctgtaatctc ttagaccaca cctaaggtaa tgtagaggtt gttgtatcat aggctttgtg 780

attagagacc actggatttg tctttggaaa agtcactcat gattctttgg gctttggttt 840

cctcatcttt tataccggct taataatgac caccgtagag ttgtcatgga agttaaatga 900

actcatgtac cataagtgac caatacaatg attgacactc agtgatttat caataaatta 960

aaacatttat tatcaatatg acagagaagg tgccgctaaa atagacaata ggtttttgga 1020

agaggtgatt aaatggatgc aaaatttatg gattgtttat tccgtctacc tttgctgtgt 1080

cccctggttg tggcatacac acgtgtgggt ataaaatcgt aaatcctatg tagtcgcgta 1140

gtgcatgcgc agaaggctta gacacgaaat gtcatttcag caatgtgcct agagaagctc 1200

tgacgccgcc ttggaagtaa gtcgttgctg cctgaccttt gggcgtctgg gacggatgcc 1260

tatacctgca cccagcagca ctggaagggg cccaggccct tcgcagcaca gcctatcccc 1320

agaccgctta gtccttcata acatatatct ccacggaaaa gggtatttcc tcccgtcaga 1380

aaaagcgccc cagtctggtc tgggttggtt tttatttcac gttgttgcaa gtaggcgaag 1440

tcccttctgt ctcctccctt ggggtaagtg gaaaggagtc cggcaggggg cccgcagtgg 1500

cctgcacagg ggaactgggt agcgagagag ttccaggcaa ttccgggggc tgccccacag 1560

aagcaggtgg ggatcgacag tggctctccg gcccagggag gagagcgcgg tcgcgggtcc 1620

ctcccctcag cctggaggct gcagccgctc gagtcggccc gggtgggggc ggggtggggg 1680

cggcgcggag ggcacggaga ttacggcggc gccacccggg acatccaggg ccccgaggcc 1740

ctgggcggtc cccacgcgag atcgcaaacc atgacaatag gcagtcaccc gaggtcaaat 1800

aaaaacggag tgggtccccc gcgcgccgcc gccccccgcg tccctggcgg cctcccccga 1860

ggcccccggc ggcctcacga gcccgcagta gccggtggcg acgtcgcccc cgccccacct 1920

-continued

ccctgcgcaa gtgcgaggct gccggcagcg cggcgcacgc tccggccgtt cccggcttcc 1980

gcgcaaaact tccatcctgt ccacgtgaag ttgtcgctgc cttagagagg gggaaagagc 2040

tgcgggaaaa gccggggagt gacgactgcg gcggctgggc gcgctctctc attttctttt 2100

cttctccttt ccccccctgtc gcagtccgga gttttggctc ctctcctttc ctcctccccc 2160

tcggagccgg cttctccctc cgccccgctt ctcccccgct tgtgtacgct atttgttgtg 2220

gggtggccga aggggatgtc ctgttttcac cagaggcaca gcgcgaaggg gaaacttcga 2280

cactggaagg aacgagaata aatacttaat tacggacgca ctgaaccgcg gctgggacag 2340

acacttcggg aacccgaggc ggaccgggcg acgaggtgag tgacccctttc ttccaacccc 2400

cgccccaggg ctcccgggggg agcctgagtt gagagaaccc ccaaactttc cgggaaagtg 2460

cgcgaggctc cgccggggac gccgagcgct gggtactgag gacgcgcagc tggacggtgc 2520

gtgggcgcct gcgtccccgg ggggcgcttg gaggccgggt gccccacgcc tgagggcccg 2580

ggccgctcgg accgcagcgg tgctctctgc cctagaagac gtccccaagc cccaagggtc 2640

ccttccgagc ctgcctgtcc cttccggggt cggcgcggag cctgcgcgta acggagttca 2700

tccagcagtc cagcgcgcgg cttctacctg caccccgcct ccacctggca gaggcgcgag 2760

catcggggtc tcccccacat ctttcttatg acgtgtatta ctttctgatg acccccctaga 2820

tggtccaggc gcgaggatgc tgacccagag tccttcggag ggtcacaggc gcctgggctt 2880

tcccggtgcc gggtgcgtgt gtactttaaa ggctcgcgtt ctaatctcca ggcactgatc 2940

gggcttttca actgcggcga tcccacttta atagttttta tgtggcgtgg actgaatgtc 3000

<210> SEQ ID NO 4
<211> LENGTH: 1984
<212> TYPE: DNA
<213> ORGANISM: Homo Sapiens
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (9, 17, 18, 33, 41, 49, 50, 52, 54, 62, 80, 84, 127,
157, 159)
<223> OTHER INFORMATION: unknown base
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (175, 207..209, 214, 1181)
<223> OTHER INFORMATION: unknown base

<400> SEQUENCE: 4

agcttcagnc atcactnnca tctttatcaa tgnccttggg nccacaggnn gngnccttct 60

gnccaaacag cagccaatgn aggnatctta aaaggcacat cactttatat ttattcagtg 120

gttatgncat cacttaatcc tcaccacaac cttgtgngnt cagtacaatt tgtgntttca 180

ttttatctgt gaacaaatgg aggcagnnnt ttgngtaatt tacaagtaag tggcagatcc 240

gggggctgtg cctagacaat gtggctccag actccttgcc cctccctgcc ttctgctgca 300

tcttgattga gtctagccat ctatttcttt aacacacact aattgctgtc tccatgttcc 360

aggccctgtg ccagatgcta gaaatacatc cccatccact ggatcttggc atggtccatt 420

gctcagtgtc ctgctatgtt acgtgttaca taatgcattt attccggtat ttagcccatg 480

gaaaataatg ctgaaagaca ttgtatgcat gtttctacaa caaaactatg taacttatgt 540

ttcttttatt gctttggcta tcaggaagct cttatccaaa tcagagcaaa tacattagaa 600

tttgggcttg tcatttcagt ttgctgaact tttccttctg gcccagattt tctattttgg 660

-continued

ttcataaatt ctattgcaca aatgtccttt attgtaaaac accttaaatt ctttctaagg 720

gaaggctgca tggaaatgat acagtaaggt cttccctgca ttttcttaga ttcctattag 780

ggaaggcaga cctagatgtc cctcttaccc tcagtcccaa agccccccat ttataaaatc 840

ccttaagcag tgactactgc tgttctgagt acctggaggt agttcagagc ttctgaaagg 900

taaaatccat atacaaaaag aagttccttc cgatccagat cccagctttg gtgccagatg 960

cacatttgag gagtaggtgg ctagtcagac tctcacctga gcagttaata aaatctattg 1020

ccccttaatg gaatttttc tgcaagctcg aattgatctg tcatctttgt gatttgtgag 1080

atggcaggga agcaccaaac accatcatga cttgggccac agtggtggga aaaaaggaaa 1140

aaagaaaaaa aaaatccact gccaagcctt gccaggcgta naaagggctg gaactgctgg 1200

ggccatttta tctgatttat tggaaataga gtggatctta ttaacatttt aataaagaga 1260

atctttttgc actaggctgg aagtggccgc cagtcccccg tgcaattcca ttctctggaa 1320

aagtggaatc agctggcatt gcccagcgtg atttgtgagg ctgagcccca acagtccaaa 1380

gaagcaaatg ggatgccacc tccgcggggc tcgctcctcg cgaggtgctc accccgtatc 1440

tgccatgcaa aacgagggag cgttaggaag gaatccgtct tgtaaagcca ttggtcctgg 1500

tcatcagcct ctacccaatg ctttcgtgat gctgctgctg atctatttgg gaagttggct 1560

ggctggcgag gcagagcctc tcctcaaagc ctggctccca cggaaaatat gctcagtgca 1620

gccgcgtgca tgaatgaaaa cgccgccggg cgcttctagt cggacaaaat gcagccgaga 1680

actccgctcg ttctgtgcgt tctcctgtcc caggtaggga agaggggctg ccgggcgcgc 1740

tctgcgcccc gtttctgcat tcggatcgcc cggcacgggc agggtgaggg ggctttcggg 1800

gggtcggggc ctccggtcgc ggcggcgaag acagatcggg gctcggtagg gaggtcattc 1860

cgagcccaga gatcctaggc acccccaca cacaggctcc cactctggcg tgcgtgcgtg 1920

tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtacgtt cgttaacggg 1980

agga 1984

<210> SEQ ID NO 5
<211> LENGTH: 7833
<212> TYPE: DNA
<213> ORGANISM: Homo Sapiens

<400> SEQUENCE: 5

gtctttggtg agatatgtgt tttacaagtt ttaatggaga aaaatgtaag tattttacct 60

cctgaaactt ggctatttga gtaatgagaa aatagtcact ttccccagga cagtggttct 120

caatcatggc tatgtgtttc tccaggaaaa ctttaaaaat atatatatac caatgcttct 180

gtgtcacttc tagggattcc aagtctttga atacgaactc tgcatcagta ttctttaatt 240

atccaggtga ttgtgatgtg aaatcatgac tgagccccac tgctctaaga tgaaataaac 300

tttcctcagc actgaaatca caaacttaaa ctaccaaaat taattaaggg catgggaatc 360

aataaggcat agggaagctt ttacattata aaattatttc tttaaatcac agctcattgt 420

ttatatgtta tttgccattg tagaaaaggg tgaaaaaata gcaaatttaa ttactctcag 480

tttgaaaaat tatccagaaa tgaagatgac gactctgaaa cattgtcaat atcatttgac 540

ctataaataa tgttctaata catttactac acactgatag atacttttc atatgaatat 600

tatacattaa aactaaggca ataatgcatt tagaacattc tatctatatc tatgtatctt 660

-continued

aagtaggcta gaaattaaga tatgagttat taagtatgag atgttaaggt gtggggttag 720

aaattatact gtacttcatt atcaataatc aacatatact tcaatatcac atacatttaa 780

ctttaatttg tacatcttta actattttta attatgtgta taaatataag tacacacatc 840

tttatgtatt tatttattca tacctccatt cacttattta tataggggat ccccccaaat 900

ccactaccat taaaccatac atttttattt taatctttag aacaagccca ggaggcaggt 960

attgttatta ctcacatttt acaaatgagg aaattgtcta cagtcacaaa gttactgtgt 1020

cagacatatt agaagcttaa tacatatttg gtgaacatat gcataaaaac agagagacag 1080

acatgtacaa cagctcatct ttacactgag taaaagcttt taacctgtct cagaaacctc 1140

tctgtgaaaa ctgagcaaaa atcgaggtat cctttcattt gtcatatagg tataggtggt 1200

accttacttc tccaacaagg atgaatattg aaatgtggat cccaaggccc aactccagat 1260

tttctgaatc cctgatagtg ggacttggaa tttgtctatt gtttcaaagt ttctcaagga 1320

attcatatga tcaaccaggt tcagaaatca ctggatctta ttgccgaagt ttgagaatta 1380

aagtttgggc cttactgcgg ctccacagaa agggcaaatg aagtatcatg gacagaactg 1440

atacgttccc agttagtttc ccctctcaga agctaacagg cagcaataca gcagaaatta 1500

gtgacttatg tcttgtgctc tgaagtcagg cagaatttca cagagtccca gcagtgtcac 1560

tgacgagatt tgtttcttgg ggcaagttgc ctgatgcttt caaagccata ttccttttat 1620

ataaaatgag ataatattct ttgtctcata ggggtgtttt aaagattaaa taaaaataac 1680

atgttctatc ctacatggca caatgcctga cacctaagaa gcaaaggata catcttacct 1740

ttattgaagc aatcagaaag tatgaaatca tgaaggagat aagagttctg attggcagtg 1800

tatcttattt tcccaggttc atttatttat cttaaactat tcttgttgga gaataactcc 1860

caagcccccc acttaagctg tgagtaatct cacactttat aatgatgttc tttccatgag 1920

aaaaaaaaat gttcttaagt tttctggaga aaatatatct gcactatttc tactgaaaaa 1980

tctaacaact ggactctgct cctctgcatc aattctagag tgtatatgcc acaaataaag 2040

tgttctagct caagaagatt gaaagtaaat atggtatagt attttaaaat aagaattttg 2100

caaatacatg gtatgattgt gtcatattac tagcaatcat atgatacgca atgcaaagta 2160

cagttcatag acttaaattt aattctaata agtaaactga ttttgccttg ctggggaaaa 2220

gttaaagcac taatccaatt gctaatgcag tcttgtctac ttctttggta cctagtgaca 2280

agtctaaata atgtatatat ttttatttac atattcagta atacaattct ctgctcaatg 2340

agtgatgttc ttctgccact tggtggtgct tgccagtttc agaatttgtt tcttggtggc 2400

actataacac taagtacaga gtaagtgcaa caaaattgca gcattcccat tgaaaaggct 2460

ttgcttcaaa ctgtttaata atttaaagga cctctgtgga agcaaccgca tttgttaacc 2520

agttacaacc agtaattaac tcctttggag ttttaactta cttttggcaa aacgtcttag 2580

gaagagcata tattattaga aagtatgcca aaaatttact tagcagaaaa ttcaaaaaca 2640

gttttcctct gctaagaggt tctctaaaat tctacttaca tagccaaact ctgaaatcct 2700

agcaggtcct gtttcattat cataattact gcataaacac ttttaaggac tttgccttta 2760

gtttcaagca tgacttattt tcataagcct gattagttac cacaccagcc ttgctatgga 2820

aaatgacatg ttctcattct ctgctgtaga gttgttaaat cttgatctat atttatgttg 2880

ccttctctgc tgaaagcctg tagcgaaaga aatttctaat tccttgtttt gcaatattag 2940

-continued

ttggcagctc tatctaatgg gtattctgtt tccttaaaga atttagctgc tctgtctaga 3000

agccgatttt ctgatgcctc caacgtctgg tctaattgat ctgttttaat ggagtcttcg 3060

tcggtgagga gcgagatgcc accgactaga atgctgggat ctgctgctta attgccagga 3120

gtgagagaca ctgagattca gaaatctttg gaggtgggag gggagaggga cagtctcgga 3180

cggaggcgga gatgtaagat aaagggatgg atttcacaca ggaaaaaaaa aaagatttcg 3240

ttgaggcact gaggtgctgc acgatcacat ctctcaaagg agaagttaaa aagcaaggaa 3300

gtgggaggag gttggaggtt aaagtactta aaaggattac tcgggtacaa tttgtttttc 3360

tgctggtgtc tgcaaaggat agatagtccc gttttcaaag tatatgaatg cctcttttaa 3420

gtgattggga atggacacta attgcctgtt aaatgttatc aaatgctctc ctaaattcag 3480

gggacacaga aagaggggca caaaaggaga atttaaatag aaaaagggag gatccggagg 3540

cttttgaaag cgggggggaga agaaggagga gggataacag agaggaatag agaaggagag 3600

cggagagaag ataaacaaaa acaaaaacag gaatcactga ataatcacac accaaaaaga 3660

aagctcttcc ctatggggca tccaaaacac tgagactgca atagtgaccc cggtcatgga 3720

agaaagatgt tcctctccac ccttgtcccc gaaagctctt ggtcccgtta ctggcgacta 3780

aaattccatt aggctaaaga gtgtgtctaa ctgcctgaag aatgcagcag acggaaggcg 3840

ggtcccgcta tgccgtttgc ccttcccgct ggagagaatg aaagaaacgc gcagagccag 3900

agactcctgc cgagttagac cttctctcgt cgccccaggt caccggccat ccggcaaaga 3960

cccgagtaag gaacgcaggg tcactgcctg ggccaacaaa tggagcccgc tctccccttc 4020

ccggacgccg ctgcccggcc gatgctcccg gcaacccacc cgcggcgtat gcagaggagc 4080

ctttctcttt ctctcagacc acttgtcccg accaatctga ccttccaaac acatctgacc 4140

gcacctccca ggtggacaca ctaataggct acgggctgga gaggagcggg tgatgaggag 4200

agggattcaa acctgcgaac gcttgggctg ggtcggagct gcgggggccc tgggaggaga 4260

gaggggagaa gagagaagga aggagagcgc ctgccgggat ggctgagctg cctcggcgag 4320

cagccttggg gttgcacgct cttgtgggag atgctgctgt tgcttccagg tcggcaagag 4380

cggttctaac accatcgcct ctcaccctct ttcctgtaaa tccctagaga aacgtccctg 4440

gcctctccgc cgcgacattc ccagcctgca tccccctaca gcctaggcgg cgcgctcccg 4500

cacgctggag cgccggtcgc cagcaggacg ccctctcccg cgccgactcg cccctctctg 4560

ccctgctgct gctgctcctc tgacacctcc gcccccacca tctccagctc ggagagacgc 4620

cacccagccg cggcccgcac tcgcggcccg gggtcacgcg cggaagaggg gcgctagtcc 4680

ggaccccgcc ttcggtaggg ggcgtcctgg agcggagagt gaggcgaatg gtatatgagt 4740

gtgcgggtag cccaccctga agcccgagct tctcatttga gccatgcccc gcctagcccc 4800

actcgggcca gcgcctggcg agcgagccca tctgtggctt ccgcggccgc ctcctccttg 4860

catccttgca cctactcgtc gacccctccc tcccgggacc tgcatcctgc tccaccaatc 4920

agagcccgac tgcctcttcc cacgtgaccc cgggcgggct gaggacctgc tgcttcccaa 4980

acgccagagg gatgcgggcg gcagagctcg agaggcggct gccgggctgc ggggcgcctt 5040

gactctccct ccaccctgcc tcctcgggct ccactcgtct gcccctggac tcccgtctcc 5100

tcctgtcctc cggcttccca gagctccctc cttatggcag cagcttcccg cgtctccggc 5160

gcagcttctc agcggacgac cctctcgctc cggggctgag cccagtccct ggatgttgct 5220

-continued

gaaactctcg agatcatgcg cgggtttggc tgctgcttcc ccgccgggtg ccactgccac 5280

cgccgccgcc tctgctgccg ccgtccgcgg gatgctcagt agcccgctgc ccggcccccg 5340

cgatcctgtg ttcctcggaa gccgtttgct gctgcagagt tgcacgaact agtcatggtg 5400

ctgtgggagt ccccgcggca gtgcagcagc tggacacttt gcgagggctt ttgctggctg 5460

ctgctgctgc ccgtcatgct actcatcgta gcccgcccgg tgaagctcgc tgctttccct 5520

acctccttaa gtgactgcca aacgcccacc ggctggaatt gctctggtaa gtccagaacc 5580

cccgtccccg acccctttaac tccgcagaag aacacgcgta tccagcacag accagcctac 5640

cctagcgcgc ctcctcagcc cctcacctcc tactgcccta gacccctaat accacccacc 5700

tctatccaga gaaacaaggg gaactgttgc aggcccgggg gtgaggggtg gttctgggat 5760

gggcagaaag tgcaggtgta gcaggaaacc tttgcatgct tgcgcttaca ttggagctgc 5820

gaggattttg agaaatatta aacgggatgg ttttctgggt tcactgtttt gaaagagcac 5880

caatcctagg ggaaacactg aaacagaagc tttgtcatca ttaaagaaaa aagtcttact 5940

aggatgagga agaaataact ttatgagaaa gaatgagcga gaaagcaata aatcaaatgg 6000

tgactgcagg ggaatcgctg attcctggca aaggtgccat gaggtcgcac tggtctcccg 6060

ttgaagacca ggtcacacag attctagagg agctgggttt caatagaatt tctctctctc 6120

tctctctctc tctctctctc tctctctctc tctctctatc tatctatctc tctctctctc 6180

tcattccctt ctctcctagg cggcaaaaga cattggtttt gcagtccaga tatgcccctc 6240

tctttgcttc cctaagcttc aaggtagtac aggggagttg agaaaaagaa cactttgcgg 6300

gtctcccagg ccggagtggg catgactgag gctggtcagg ctccatgtag gcgagccgag 6360

ggcggaaccg acttcagtgg gcgctgactc ctccatttct ggacaggctt ctgtggagtg 6420

ggtcaggcac tcttcttgct cgctcgggtt ccttcagatt ctgacggcga acgcttggca 6480

ggcttcgctc tgctgaagct tcctaattaa atagggccag aggatgggag ttgctgcact 6540

cctagctggc atagcattcg gtttgacagc ctgtagtata gggtgtatgt aatttttcat 6600

cttctgtgaa tataattttg ctgtagttaa atctggctct gaataaagtg tctttcaaag 6660

atgtatataa gctgaagtgt atgtaacttt agagaggagg gaatgaccaa ctgtaactca 6720

gggtgaaagc ctgtatagtt cctagttatt actgatgtaa atgccaaaag gaaaattatt 6780

atgcatcatt ctaatttatc ctttacaaag acaagttgag atatgcaacc ctattagatt 6840

tgggtcaata gattgttctc ttttttggca gtttctaaat ttggcatttt aataaaactc 6900

aacatgtttc tataacttct tgattcatgc gtacatgtgt gttgtttttg aaagaataag 6960

tttcactttg ctattgccta atcacttttt agatgcttta ttatggtaat aattatgagc 7020

ctgcaaaaac aatttttgga aatgttgatg gctttgtagt ccaacacaga ctggtttgct 7080

tcattcctag cccttgcatt gttttaggaa ataactaact taaatgtgaa gttgacattt 7140

gcaatcaaga aattacatat ttaccagata ttttaaaggg gactgcataa actaaagaga 7200

ataaactggt tttgcagata ggttgtcaag aacttggcac cccgcttcca cccctgttaa 7260

cttagaggtg atcaatcttc atttgagcca aacagaccat cacagaaaac actgtgcctg 7320

tttatcttta ttattgaggc tttgtttcct ctttgtctgg atacatttca aataaggggt 7380

tgtttcagtc gttgaagcaa aagaacaatt aaagatgggg aaatggtaaa agggtattca 7440

gagatcatca ctagctcttt tccaaaatgt ggagttttgt ggtcataaat attgtccacc 7500

-continued

taatgagcaa aaaataaaaa taaaaaaaaa acaggaagca aatgttaagc tttcattcac 7560

cactgtcagt attaacgcaa gctttaaaaa atagcactat cagaaaagga tactaaagga 7620

gaattgacta gaaaagaatt gtggaaaatg gaaacgaata ttgatcactt aactagattt 7680

tgaggttatc agtagacagt gaccttgcag tacagctata gttgttggat ttaaaattta 7740

ggacaagtat tttaaagctt caaagtagtg ctttttttg ttaaaaatct gtaagatgtt 7800

ttaatgactg gagtgttctc tttgaatttg agg 7833

<210> SEQ ID NO 6
<211> LENGTH: 965
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 6

gattaattaa gggtatttta gaagttaggc gttttcgttg tttttttga gtacggaggt 60

tattaatttt taggggggaag agatgtagcg cgaggtaggg gtgtcgtgtt aagaaatttc 120

gacgtttttg gggattgagg ataaaggtgc ggatacgatt tcggggtatt tggagtttcg 180

tgattcgcgt tacggacggt atatttaggg gttaattttt gttttgtttt aaagaatttt 240

aagttagagt ttttgttttc gtttatagtt tcgggatgtc gttgttgcgt ttatcgtata 300

ggtagcgttc ggatcggttg tagtagatcg cgcgttgcgc gttttatcgg gagatggtgg 360

agacgttgaa aagtttttt tttgttattt tggacgttgt gggcggtaag cgttttagtt 420

tttatttttg ttgagttgaa cgtttaggta tagtggaatt gaaattcggt tttttatttt 480

tgttgagttg aacgtttagg tatagtggaa ttgaaattcg gttttgcggg atgtgagagt 540

tgttgaggtt acgcgtaatt gggtgtgatg gagggcgttt gttcgtgatg tgtgtaggtt 600

tgatgtaagt aggttatcgt cgtgcgagtg tgtggatgcg atcgttcgag agattcggag 660

gtaggtttgg gatacgtttg agtgaatatt ttaggatatt tttttggtta gtatttgttt 720

tttagtgttt gtgatttaga gtgggtatat gttgggagat agtaatgggt ttgggtgtgt 780

gtaaatgagt gtgatcggaa gcgagtgtga gtttgattta ggtagggatt atatagtatt 840

gttatatttg tttgtttttt agtagaggat tgaagtgcgg gggtgggggt acggggtcgg 900

aatagaatgt ttttgggata ttttggtaaa tagtagtcgg aagtaaaggg gtagttgtgt 960

aaacg 965

<210> SEQ ID NO 7
<211> LENGTH: 965
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 7

cgtttgtata gttgtttttt tgttttcggt tgttgtttgt taagatgttt tagagatatt 60

ttatttcggt ttcgtatttt tattttcgta ttttagtttt ttattaaaga gtaggtaggt 120

gtgatagtgt tgtgtggttt ttgtttagat taagtttata ttcgttttcg gttatattta 180

tttatatata tttaaattta ttattgtttt ttaatatgtg tttattttga attatagata 240

ttaaaaaata gatattggtt agaagagtat tttgaggtgt ttatttaaac gtgttttaag 300

-continued

tttgttttcg agtttttcgg gcggtcgtat ttatatattc gtacgacgat gatttgtttg 360

tattaaattt gtatatatta cgaataggcg ttttttatta tatttaatta cgcgtgattt 420

taatagtttt tatatttcgt agaatcgggt tttagtttta ttgtgtttga gcgtttagtt 480

tagtagaggt agggaatcgg gttttagttt tattgtgttt gagcgtttag tttagtagag 540

gtagggatta aggcgtttgt cgtttatagc gtttagagtg gtaagaaaga agtttttag 600

cgtttttatt atttttcggt ggaacgcgta gcgcgcgatt tgttgtagtc ggttcgggcg 660

ttgtttgtgc ggtgagcgta gtagcggtat ttcggggttg tgggcggagg taaggatttt 720

agtttgaggt tttttgaggt agagtagaaa ttagttttta ggtgtgtcgt tcgtggcgcg 780

agttacggaa ttttaggtat ttcggggtcg tgttcgtatt tttgttttta gtttttagaa 840

gcgtcgaaat tttttagtac gatatttttg tttcgcgtta tatttttttt ttttaggggt 900

tggtggtttt cgtgtttaaa ggaggtagcg ggaacgttta atttttaaga tgtttttaat 960

tgatt 965

<210> SEQ ID NO 8
<211> LENGTH: 16579
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 8

tgtttttttaa aatgatgttt agttgaatta gttaatggag ataaaaatat ttatatataa 60

ttttttttta agtttattag ttattttata gattagttat aattttggtt attttaatga 120

gttaattttt ttttagagat attataaatt ttttaaagat ggaaaataat tgttattttt 180

ttgaataatg tttattattt aaaataatgt atataaaagt tgtttaataa atttttgttg 240

aattgagaga aaatataaaa aattgaaagt agtttgtttg taataagtgt ttgttaagtt 300

ttatagattt ttttttttttt ttggaattat tgtaataaat atttgatttt tgatttgtag 360

gattaaaatg tgatttttta tattttaata ggagtaagta ttagttttgt aatttttttt 420

gttaatagaa taaaaatgga atgtatgttt gttgttttta tatgtcggtt gatatttatg 480

ttgaatattt tttttataac gtagaggttt ttttgttttt tgttttgcgt tttatttttg 540

tttattatag agatattaat gatgttatgt gttttttttgt attattagtg tagtagtttt 600

tattatgttt tttgtttcgg ttttattgta gtttaattaa ataaatggaa acgggtagtt 660

attgaaaata aatagtaatt gtaaaagttt tgtttttttaa gaaatttttt gtattgtatt 720

tcgttttgtt ttattgattt taagattttt atcgtttttt tgtgtttaga aattagattt 780

tttagtttgt tttgttttag ttagataaat gttttataaa aatgatttat attttttaga 840

ttgtggggag agaagaaaaa gattttgaga aagtgattaa ggagggagta agagtgatat 900

aatttttgat ttaatttagt tggattgtgt tttttatgtt tttatttttt tttttataaa 960

aaagatagcg tttttgatgt tttgttttag gtagtttttt agggtgagtg tagatgtgtt 1020

gagttatttg gttgttttga ggatcggttt agagagtttt tttttttatt aatgaagtta 1080

tttttttttcg gtagttcggt tttttttatt attttatgtt ggaatttggt tttggtttgt 1140

atagttttgt cgtgatgtag atggtttttt agttaagttt tagagaaatt tgtttaatta 1200

tgaataggat atttatagta taggagaaaa aaaaatttat taaaataaat attttattgt 1260

-continued

aatattttga tttgggggtt ggagagtaaa aatttaatat ttgatgtttg ggaaaaattt 1320

aatgttttat ggtttttaa ggaaatttgt ttgtgtttaa tttattagag taatttatat 1380

atttcgggtt aaagtataat gagattgtta ttattatttt gttttttgta agatttttt 1440

tttaagaata aaaagaaggt aataaagatt ttatattttt gttttttta gaaagttaaa 1500

gagaatttaa agtgttggat gaaatgtttg tttttaattt gtagtgtttt atttaaaaaa 1560

aaaattatga tataatttta gaggataggt ttttttttt tagggttttt gaaagttaaa 1620

agtatatttt taatttgtag tattattgta ttatttagtg aggataattt gtttagttgt 1680

gtttaggatt agttagatta aggagttgaa attaagtttt ttttgaattt gaatgggatt 1740

ggaattttgt tatttttaga gtagtaatag atgatgttat gtttaagaat attttatttg 1800

attaataaaa aaggagaagt ttgaaatgta attttaaaga attttagaag gaggtttaga 1860

tgaagataat aaaatttttt ggagatatat ttgttattta tatttttttt ttttaagtta 1920

ttggtttgtt gtttttgagt ttagttaaat ttttaaagat aagattaaag taaggaagta 1980

aggtatatag tttttggttt atggtttttt ttagatataa gtattgtgaa tgaagatttt 2040

gttttattta tatttgggat tttggagttt ttataaagaa tatttttagt ataaggatgt 2100

gatttaaagg gggatttata gaataatatt ttaattattg tttgatttat atttgagtgt 2160

aatggataaa taaatataat tttagttata aaatttgaaa aatatatatt tttttattta 2220

aaaatttttg atagttttgt taggtattgg agaattagaa ataaatgttt ttttattttt 2280

tttttttga tgaaggaaag ttattaaaaa tagttatata gaaataataa tataattgta 2340

gattttattt taaatatagt tggtatgttt ttttttttt tttttattta ttttttattt 2400

tagatagtta gaggggaaag tttgttttag agatagaaag aagatattta tggggtataa 2460

ttttatttta gttatttgat agtttgaaga agtatttaat tttttgaata ataatgttat 2520

aagttaaaga agagaaaaga taggtagttt aaagtttttt aatattgtgt atattttaat 2580

tttttaagtt agttttgaga aagtgttttt gttttgtaga agtatttggt taagtgttgt 2640

gaagtttatt ggtattaaaa gaaattaaaa attttggagg attagagaat taaatttggt 2700

aatataaatg aaaatatttt agaatgtaaa aagtatttaa tttattattt ttaattatgt 2760

tttattatat atagatggta ttatttatag tgaatgttaa ttattaatat ttttttaaaa 2820

atgggtatta atttgttatt agttgtttgt ttttgattta gttttttttt gttgaaaggt 2880

tttatttaat tgtattataa agggataaaa gggtaaaaga taataagttt tttagattgg 2940

tagatttgtt atttataaaa tggttttggg taggttattt gaatttttttg aaatttaatt 3000

agatttattt aaatttaaga atgagaatat aaatttatat tttgaagtgt tttatagaaa 3060

ggtttatttt aatgtttgga gtatatattt taatgaatat ttattttatt ttatttttt 3120

ttatttttga attaagtaat tttgaattta aagttgttat gattagtatt gaaaagatta 3180

ttggattatt aattgtgtga ttttgggata gtaatttttt gtattttagt ttgtttatat 3240

gttatatatg aaggttgaag tttgattttg ttttgtgatt attattttaa atatttgatg 3300

aaattaaatt ttagtgtttg gaatggtagt ataataaatt tattaagaat aaataattta 3360

ttgtaaaaat atattgattt ttaaatgatg taattgatag ttatattatt gtagagggtt 3420

gataaataat aaaagaaatg aaagatgtat atggtgagaa ttgaaattat tttgataagt 3480

tttttatttg tttattattt aaaattaatg attatgttga atgtttataa attataaaat 3540

-continued

ataaaagaaa ttttataaat gcgtatgtat aggagtttaa gttattaaaa gttttaaagt 3600

ataagtttaa attaaattaa ttaaagaagt tgagaggaaa aattggtttt tatttttaat 3660

tattattgtt ttgaggtttt atgtttaata taatttttta agtagaggtt ttagagagaa 3720

gagttgtgag gatattttta tatttgtgta gaaggaaaag tttgttattt attttagtat 3780

ttttagtgtt atattgatgt gtattttgga tttattttgt ttttattgta taaatttata 3840

tttgatttta aagaaaagga aaatttaaag tttttttttt ttaaggggat agaaattttt 3900

tgtgttaatt gtttgatttt tttttttgta aggttttatt ggaaattttt tgtaatataa 3960

tgtaggggat ttttttatgt gttgatgttg tttatatagt ggggtgggtt tgattgaaga 4020

aaaaaaatcg tatatacgta tgaaagatta tggttttatt ttcggaaagt atgaaaggtg 4080

attgatattt ttaagaagtt tttgttattt aggaaaatta ttaaatattt tatttagaga 4140

tatttggaaa gattgaagga aaggaagaac gaagaaagta gaatttagat ttatgtgggg 4200

agagatttgt ggtagaggaa aagtattttt tttgaattcg ataagggatt tgtttgggggg 4260

aatttttttgt ttagtttttt attattaggg ttttttgaag tcgggttttt tattgggtag 4320

ttttttggga gtgtagtggg gaatttttat attttttttt taggttttcg aaggatttcg 4380

tttttttagt gttttttttta ggttggtagg agttttgagt ttgatatttt tttttgatgg 4440

gataggtaag ttttgtgggc gcgtaaatac gttgtaatta agtttttttgt tgattttata 4500

gttttgtgtg ttttcgagaa gaagtgatcg tatttaattg tttattgttg gtttgttttt 4560

taagagtttg ggggtttttt ttttttaatt tagaattagt tgtacggggg gcggggaagt 4620

gggggtgggg aaggagtggg agggtagtgg ttttcgcgag tagagcgatg ttattgagtg 4680

agtttttgaa tggggagcgt tgttgttttt aagtcgattg gtatttttttg ttaggaagaa 4740

acgttaagag gtgggagtgt ttggggaggg aggtaggcgg tttttatcgt aggcgcgggg 4800

agttgttttt tcgttttttc gtttgttttt taagtttgga tttttaggag tggttgaagt 4860

tgcggagcgt ttttggagtt tgtgaatgaa ttttttttttt ttttttttttt ttttttttcg 4920

ttgagttttt ttttcggttt tgacggtata gtgatataat gatgatgggt gttataattc 4980

gtatttgaat ttgtaggcga gttgtttcga gttttttttgg ggaagaattt taggcgtgcg 5040

gacgtaatag tcgagaatat taggtgttgt ggataggagt tgggattaag attttcggtt 5100

agtttcgtat tttttcgtat tttttagtat cgtttcgtat ttttcgtatt ttttttcggg 5160

ttattacgtt ttttatgtga ttcgtttggg taacgtcgaa tttagtcgcg tagcgttgta 5220

gtgaattttt tttttaaatt gtaataagtc gttttttaag gtaattacgt tttttttgtt 5280

tttttttttaa aaaataaaaa taaaaaattt atagaaaaaa attcgcgagt ttagaaaaaa 5340

gaagtaattg gtagaaggtt ttaattaagg taaagagttg taaggcgaag ttaagaaaat 5400

gtaggtattt aaaaaatgta ggtaattttt ataaaggttt ttggggagag gtatatagag 5460

ggattttggt gttgaaaaag atttagataa aagaaattta gggtggggtg ggggtaaaat 5520

gattaacgga attgggggaa gggagggaat aaattgtaaa gaaattatag aaaagtggtg 5580

ggtttttgag ttggagagaa gagagggatt tttggtattt tgatttttttt gttgttgttg 5640

tttttaatac gttcgaggta aaagtttgaa tggggattat taagatttgt tatagataag 5700

ttttcgaagt cgttttggtg taggttattt ggtttttttag tttttggtgt gtggttagtg 5760

tttggtgttt ttggaaagtt attttcgggt agtttttgat agtgcgattc ggcgtttaag 5820

-continued

tagtttggga ttttgcgcgg atttgatttt tttagatcgt aggtagtttg ggaggaggtt 5880

cggtcggggg aggtgtagga ttttcgtcgt gttttttga cgatttgggg attgttacgg 5940

tttttttcg gcgtttttgg gttttttgt tttgtacgcg gtgcgaaggg gttagtaggg 6000

aaggagtaga ggatgggggg tggggttgtt ggagtttcgc ggaggtttgg gaggtttttg 6060

ggcgggaaaa gtttgttttg aatcggtagg gatgtgtaat aatttttttt attttgaaga 6120

gtgaaatagg gtttgtcgtt tttattttaa taagtaaatc ggtatttaga gcgtattgta 6180

gataaaggtg gttcggggat tcgaattagg ggtttttggg ttagtttttt cgtttagatt 6240

ataggagttt ttcgtttttt tatatatttt ttatttttt ttagttttc gttttagttg 6300

agtaatttta ttggtaggtt tagcggtagc ggcgcgcgcg tgtgtcggga gtttcggggg 6360

acgttttcgg ttggagcgtt ttatcgtttt gtagaggcgt gggcgttgta gggcgatatt 6420

tttggtgcgt gtagatttag ggtattcggg ttgttttggt tcgcggtttg tgttattggg 6480

cgtggagcgg tttgggtgtt aggtagagga gagcggggta gaaaaatagt gttatagttt 6540

aaattatttg ttttttaatt taattgcgtt tcggtaagtt ttgtatgttt ttgggagatg 6600

ttgcgggaag ggggagaaat ttatacggcg tttggagagt tgttttgtcg cgcgtatatt 6660

cgcggtagag ttttttggg tcgcgttttt tatttttgtg ttttttttg tttttatttt 6720

ttttgttttt ttttttttt tttagttttt gttttttttt tttagttttt gcgttttttt 6780

ttttatagtt gataaatgaa tggttagtgt gaaatttttg ttttttcgt gttttaaggt 6840

agtagggagg gaggagcgag ggaggcggtg cgtttttcg gatttgtttt tagtttagtt 6900

tataagattt gtagaattta gatgtttagg aattgggagt tttgcggcgg gtgtgggcgt 6960

tttttgatgg agaagtttcg tataggcgga gaaaaataag ttttttagat taagcgtgta 7020

ttttttataa ttttgtgttt aaggatggaa ggttttagtt ttttttaag ttatttattt 7080

tgtattttat aatttgcggc gtatatttag gagtttggag gattttgaaa aaaggttttg 7140

gtttgtgtaa agtgtagaga tgtttttttcg cgagtcggtt ggaacgtacg cggcgttttg 7200

ggtttagttc gtcgtttagt ggacgaacgt atatggtcgg ggtggagtta ggtttcggag 7260

gtttagacgc gttattttgg gcgtttttaa gaaagataat atatatatcg tgtttttaaa 7320

atttatgatt atttgaattt aacggtaggg tttcgtttta ggatattgta aaagaagggt 7380

ttttaattta aaatttaaat ttttttaatt ttagggcggg cgtcggacgg gaaagtggag 7440

agaaggcggg tagtgggagg aaaaaagaaa gggaaggaag ggaagtggaa ggaaagagga 7500

tcgtggaggg gaggaaaagg agaaaggaac ggagatgggg tagatggtat ttaaatagtt 7560

aggaacggat cggaagggtt cgtttcgggg tttgggtttg aatttcgaga ttcgttttag 7620

gatatttttt ttttttttt agcggtaata aatgtgattg tagaggcgtg tcgggatgga 7680

agggtgggaa agaattagat aagcggaggt ggtttagttt cgtttaaagg ggttgggttt 7740

ggggttcgtt tttattcgtt tatttcgcgt taagggagga ttcgtaacgg gcgtttttcg 7800

tagcggttgg gttagttagt cgttacgtgg gttggcgttt cggtttttag tcgagacggg 7860

tgtgattttt tgggtttaga gattgtagta tagtcgtata ggcggtcggc gtagtcgagg 7920

tttttattta tcgaggttaa tgggggcggg gcggggaagg tagggtttag tttaatcgag 7980

agagtcgagt ggtttcgttt tcgggattcg gggtcggttg tttttggagt tgattcgagt 8040

tttatttaga tagcgttgaa tatttttatt aaaaggttat ttaagtttta ttttatttta 8100

-continued

ttttatttttgtttgttttttgggagtggtaggaggggtatcgtgtttttgttttgaggt 8160

aggggagaggttaaaaattttatttaagtaaatttttttgttagaaaaaatgagattttatat 8220

cgtatttttattgttgatttttaaatatatttaggttattttaaagtttattttagtt 8280

aattaataga ttaatttcgt tttagattgt tataaagagt tcgtttggaa tggtaaatat 8340

tgttattcga taagattaat agaatttttat gggggtaaat atggaagggt tgcgagtgta 8400

attgaggttt ttgtgtttta atgaggaatt ttagcgggga tttggcgggg attggaagtt 8460

tagtatttat ttagggagtg attgagttgg aaatttggag aagtttggtg tagacgagac 8520

ggtgtttttt agagttgtat tttaagtttt tttattttgt tgtatattgg aagaattgga 8580

ttttaaaagat tttttttatt attttagatt ttgtagttaa tagttacgtt ttttttttt 8640

tgtttgttga aagttgtgat ttaattgttt taagtatttt ttgataattt cgttaaattt 8700

attttgaaaa tatttatata aagttttaaa ataggatgtt gtgagtgtaa tataaaaata 8760

ggaaggtatt aatggaagta ataattagtg ttttgagata gaagcggttt ttagtttgta 8820

atttggaaat gttaagagta gtatattagg gagaaggttt ttaagagttt gatttttgt 8880

tttgagtttt gagtaaagaa agaaatatgt tagaattgta ttgtttgttc gtagtaatat 8940

tgtaaagggt tttaatggat tttaatgtga ttgagtaatt attttttta gtattaagtt 9000

tatagttttt aatttttttg tgagaagatt aggtaggaag cgggatgaga tttaagataa 9060

ttatttagga tttatgttag aatttaaagt tttgattttt ttggtagata aaaatagagt 9120

tggttttaaa ttgtattttt tgtttttaaa aaaacgaggt ttgaaaggaa ttttttaga 9180

tttatatagg tttatagatg taagtttttt ttttttttt gtgtttgtaa attggaggga 9240

gatgtttttg atattttaga atatttgtat ttatatattt gtaaattttg ttaggaaagg 9300

atttaaatag ttaaattaaa ggtgattatg attaatgtta aagatattat ttattgtttt 9360

ttttttgtt gtattattag gtttaaatat tttggagtta atgattggtt atttaattga 9420

gtaatagagt ttttgagagc gtaaagatta attaattttg ttttttaaag tttggtaatt 9480

aagtttttt tttttggtga gatttaaata tttttgttga cgttatttat ggataaatat 9540

cggtaaatat tgtatttaaa atgggatatt gttttgtgat ggagtagaat tttttaagat 9600

atatggtttt tttgaaagaa agaattaatg attgtagtaa ggtatttttt tattatttac 9660

gtagagaaat tagttaggat tgataaatat ttattgttgg gagttatttg ttttggtttg 9720

agaaaatatg ttgattgtta tataggtttg tattttatta gaaaatataa agtatgtatt 9780

ttttttgatt tggaattaga aaaggggat tagatatgga tgttttggtt aagttttatt 9840

ttaattagtg aatttttgtt agttttttt tgtagtttag atttggtaaa ttatatatat 9900

ataaagtttg tgtgtgtgtg tggggtgcgt gtgtgtgtgt atgtatatag ggtttttaaa 9960

tattgttaag tttttattat tgaatgtgtt agaggtggtt gtattttaga ggttagtata 10020

aagaattttt gtttaatgtg ttagttttgt ttttgaaat atatttttgg gtagatttat 10080

taattataaa aggatataaa tgtatatatt ttttatattt gtatatattt tattttattt 10140

ttttaaggat tttaagaggt gtttatgttt ataaaatgag attatatatt ttataataaa 10200

tatatttgat aaattataat ttttttatta ttaaattatt gatttgatat tatttgaggt 10260

atagtggaat gagtattaga tttggaatta gaaagtttga gttggattta ttagttgaga 10320

aaattaagta aagatattat ttttggattt ttagtgttat aatttattaa atgaatatgt 10380

-continued

agttataaag attaaaagag taagtatttt gaaaaattat ataaaaaagg gatgataata 10440

tattatgttt tgggatttat attgatttaa ttttataatt ttaataataa ttgtattatt 10500

taagttatgt agtttttatt gaatttataa aataaattag agaggggtat attgttttaa 10560

aaattgttta gttttattaa ttagaagata gattatagaa attgatttta atgaagttgg 10620

atatttagtg gaaaggaaga attatttaat aaatggtttt aggataattg atgggtttat 10680

gtaaaataag aattagtttt tatttgattt tttattttaa aagaaatata agatgaatta 10740

aatatgaaat gtagggaata atttttttata aaatgataaa agaaaattat gggattaaaa 10800

aaaataagtt aggaatagag aaatttgttt taagtaaaat ataaaattta gtgttatttt 10860

aggaaatatt gataggtttg attaagtaaa aattatataa aagtaaataa tatatttgga 10920

aaatatttac gatatttgag agataaaaag ttaataattt ttgatatttt atggattttt 10980

atgaataatt ttgaaaagtt tatatttatt aaaatatgag taaaaattat gtataggtag 11040

tggtagaaag agatatagag ttttaaaatt atagaaaaaa tcgtttagtt ttatttgtaa 11100

ttaaataaat agtaattaaa atattaatga tatgttatat tatttttta ttaattagag 11160

gagtaataaa aaaaaataat ttttttttt ttttttttt ttgagatgga gtttgttatt 11220

taggttggag tgtagtggta cgattttagt ttattgtaat tttcgttttt taggtttaag 11280

taatttttt gtgttagttt ttttagtagt tgggattatg ggtatttatt attatgtttg 11340

gttaattttc gtgtttttag tagagtgggg ttttattatg ttggttatgg tagttgattt 11400

tgaatttttg attttaagtg atttatttat tttagttttt taaagtgttg ggattatagg 11460

tatgagttat tgtgtttggt aaaaatgaaa ttaaaaagtt tgatgtaagt ttgtgttggt 11520

taggttgtgg agaaattggg ttttttatat attgttgtat gaaagtaaat aggtttttt 11580

attttaaaaa ataatttggt aattattttt ttttgatttt ttattaaaag tttttaaaag 11640

atgtatatat tttttgattt agtaatttta tttttcgata ttattatatg gatggatatt 11700

ggtattaata ttatttaaga taaatggata aggatgtttt ttgtagtatt gtttttaaga 11760

gtaaatattt ggaaatagtg taataggttt attataggga attatttaaa taaaagttgt 11820

tgtatttatt taatggaaaa ttatgtagtt gttaaaaata atatggaaaa attgtatttg 11880

ttgatttgga atatttttta aaatatattg tattttaagt gaaaaaattg atagataaaa 11940

ttgagtatat attatatatt taattatgta aaatgggtgt atatttattt ttttaataaa 12000

tatttattga gtgttttagg tattgtggat ataggagtga ataaaatatt aaagattgtt 12060

ttgttggaat ttttattatg gtaggtagag aaaggtaatt tttaattata tatgtaaaat 12120

ggtgatatgt gttaggatga aaatagagta tgggaaaggg aatagagtat agtatgtttt 12180

aggtagagta atgagaaagg agtttattag gtagagaata gaggagtgtt aaggggcgat 12240

aatgtgaaaa agtttagtta agtttaggtt taggaattta tatgatgtta ttatttgtat 12300

gaagtaaatt tatgaaattt cgtatttaaa ttacgtgatt ttttttgtt gtataatgat 12360

ttttttgtg agaaagaatt agtgagtaat atgatatgtg atgtttatta tatttatttt 12420

gtagttgaag tattttagaa agttggttgt ttagggtatt taattagtag aatggtggat 12480

aaaattatcg tgtttttgtt tttaagttat tttttgttag ttgagttttt ttaaaattaa 12540

gggagaatat taaatattta tttaatagaa ttttggtttt ttttcggagg ttgtttattt 12600

ttggtgtgtt ttttattttt tttgtagata ttttttatgg ataggagttt tattgatttt 12660

-continued

gttttttatg ttttttaggt aattaaatag ttttagtatt ttgtatttgg attgtttaat 12720

gatttttttt ttttttttatt gattttattg gagtagaagg ttaaatatag gtatttattt 12780

ttatattttt taggtttttt tatgtagata gagttgttat tattagtaga cggtgtttat 12840

aagagagata tggaaaagtg gaagaatgga aatagggatt tggagtggtg ttaaaaataa 12900

agaaagaggt ttttgaaagt tttttattta atatgattaa atatagaaga taaaagatgt 12960

aatttaatga ttaaggatag ggatatagtt ttgagaaaat tatagaagtg gaatgaacgt 13020

tttggatttt aaagaagagg gaatggaaat gtagttttga ggtagttaaa gttaataaag 13080

ttttttgata tatttattgt agaaataaat gattattttt taatttaagg tttgtaaatt 13140

taggtgatta taggggcgag gtatggagtg gaaatgggta agagattaag tgggggtagt 13200

aatggttttg agtatataga tttaagggga gagattgttg tttggttaat gtttatgatt 13260

agttagttga aatatagatt tagtgttgtt agatttttttg atttttttgaa agaagttaga 13320

gattttatat ttagtgtaaa tagtttaatt tttaaattga gtaatttatt taaatatatt 13380

tatgtaattt atttaaatgt atttagagta tgtttgattt ttttttgta tttttattta 13440

gttttagggg taattttaag gaaagtgatg cgtatatttg ttgtatttat gtagatggag 13500

gattattttt tttttttttaa taaaagatat ttttataagg attgagtgtg gtataaagag 13560

ttcgattttt agtatttaaa ttgagtagga aatttaaata tagaagattt ttttttgggt 13620

tatatttttg gttttttttaa atgagatttg atttttttaga tattgataaa tattatatta 13680

ttgggttatg gtaaattttt aagtggattg atttaaaagt gtttaaaagt attttaggtg 13740

ttaaggaatt gttggtacgg agttttaaag gtgtttttgt ataagaaata gaggcgttga 13800

ttatgtattt gatttataaa tatttttttaa ggatatattt tgaatttggt aatgaagaaa 13860

tattgtggga attattaagt tgaataaagt attagttttg ttttttagaa gtttattgtt 13920

ttatagagga gattagatat gtttatagat aatttgaagt atatagtaat aaattatagg 13980

atgatttatt ttaagtataa agtgtatatt gttattatag tggtaattta ggttataatt 14040

tttaaatgta agagttttaa atattatttt ttttatttttt agtgatatga aattatgtta 14100

gaatatttgt agaaagtttg aattagtaat tttaataata aatattgagt attgatatat 14160

tgtatattat tgagatttat tttatatata ttattttatt aatttttatg attatttcgg 14220

taagtattaa tttttataat taagtagtga tagaaaatga ggtttagaga gatatagtat 14280

tttatttaaa gttatagaat ataagtagag aatgtagaat ttgtattttg ggttatttga 14340

tagtggagtt tagattttta aaattaggtt tataaagatt ttttggaaaa aaagaaaaat 14400

aaaaaataga gaataaattt ttttgaagaa aagtaggaaa ttattattta taagtataat 14460

taaagagaga tttttaattt ttttttttttt ttttttttttg agacggagtt tcgttttgtc 14520

gtttaggcgt aattttagtt tattgtaatt tttatttttc gggtttatgt tataaagaaa 14580

gtttttattt gattggttcg atttgattgt ttattagaat ataggttttt tagtagggtt 14640

aataagttat aatagtttta gtgatataag ataaatataa attaataata aggttaattt 14700

agaattatta atatatgttt tttattgtga attattagtt ggtttataga tggaaaatat 14760

aagatatata atgtattagt ttaaaaaatg ttttttagag agtgtttaga attatagaat 14820

ttaaaaattt tgttttgaaa aggtttaatt tttatttttaa aatatataaa gaaattaaat 14880

gtaaaaaaat taagagaaaa ggttttaatt aattaaatgt tgtataattt taagttttat 14940

-continued

ttatgtttaa gattagtttt gtaatttatg attaatttta gtattagttt gtttttaatt 15000

taggtttgtt aaataatata gaatttttt tttaatgaag ttttttaggg ttttagttga 15060

aaaatatata ttagttagga tgtttttaat tgtaaggaat aaaaagttac gattttattg 15120

ggtttaaata ataatgatat ttataatttt atataaaatg aagtttaaag gtatggttta 15180

ttttaggttt aatttttaag gatttacgtt ttgtttttgt tgatttttt ttagttttat 15240

ttttagaggt attttattta tttatggatg taaaattttt ataatagttt tagtagaaag 15300

attttagtat gataatgttt agggaagaaa attgattatt ttttttttg ttttttttt 15360

aggatgaatt tatttttttt agggattttt tattggagga agagattttt ttatatttta 15420

ttggttaaaa ttatattata tattgtttta aatttaatta ttggaaaagg gaatggtatt 15480

tttataattg atttaggata ttaggattat tttttgggtt ggggtttgat ttaggatttt 15540

ttaaagtatg tgattgaggg taggtattag aatggaatta gagttgtgtt agaaaggaga 15600

aacgtatggt tattatagaa tagttttaaa tgttatcgag gagggtataa ttgtaaaaat 15660

taaataattt tttgttattt ttttgaagtt ggtattttga tttttagatg gttttttaat 15720

ataggttttt ttttttttta ttattatagt tgttttgaaa tttgagttgg aagggaatat 15780

tttgagattt agattgttaa atgttttttt taaagttatg taataaatta aatggtaaag 15840

ttagggtagt tttttgattt agtatagggt attttttttt atttttatt tttgagattt 15900

tagaattgtt ggtattgttt taaaatttat ggtaagaatt ggttattttt gtaattaata 15960

ttttttttata atatatttgt tttgtttgtt tagttagtta gaaattatat ggagtttgtg 16020

ttttaaaaag tttgtcgaag tttttatttt ttgttttggt attatgtgta tgaattatta 16080

attggttttt ttttatttta tatttgatga agatgttttt tttttaatat tttttttat 16140

tgttttttat tttttttgt tttatttatg attagttgtt tgtttttaaa tagattttgt 16200

ggttatttat tttttttgt gttagttttt atttattagt tatttgaatt gtggttttta 16260

ttgttttttta tatagtttat ttatttgtgg ttgttatata tatttatatt gttatatgtt 16320

tttaaattgt attttggaat gattttggta atggttgtat tggatgagat ttaaattaat 16380

aattaaagta ttgagatagt ttttgttatt ataagtttat ttttgttttt atagtttaag 16440

aggagttatt ttttttattt ttattattta atgtttaata ttatttttat tatatataat 16500

gtataaaaag tatgtgattt atgatttatt ttaaatttga atgtttgtga tttattttgt 16560

gttttttttt atttttagg 16579

<210> SEQ ID NO 9
<211> LENGTH: 16579
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 9

tttagaaatg aagaaaaata taaagtaaat tataagtatt taggtttaaa gtgggttata 60

aattatatgt tttttgtgta ttgtatgtaa taagggtagt attaagtatt aagtaataag 120

agtagaagaa gtagtttttt ttggattata aaaataggag taaatttgta atagtagaag 180

ttattttaat gttttaatta ttagtttgaa ttttatttaa tgtagttatt attaaaatta 240

ttttagaata tagtttggga atatataata atgtggatgt gtgtgataat tataagtgag 300

-continued

taggttgtat gaagaataat agagattata gtttaagtaa ttagtggata ggagttggta 360

taaaaagaaa tgggtgatta taaagtttat ttagaaatag atagttgatt ataaatagag 420

taagagaaga tggggagtaa tagagaaaga tgttgaaaga aaggtatttt tattaagtgt 480

aaggtgagaa ggaattaatt ggtggtttat gtatataata ttaaaataga gaataaggat 540

ttcggtaggt tttttaaagt atagatttta tgtagttttt agttggttaa gtaaataaaa 600

taaatgtatt ataaggaggt gttaattata aaagtgatta gtttttgtta tgaattttaa 660

agtagtatta atagttttaa agttttaaga gtaaagaatg aaagaaaata ttttgtattg 720

agttaagaaa ttgttttggt tttgttattt aatttattgt atgattttgg aaaagatatt 780

taataatttg ggttttagaa tgtttttttt tagtttaaat tttaaggtag ttataatgat 840

gaggaggaga agaatttgtg ttggggaatt atttagaaat taaagtgtta attttaaaag 900

aatggtaaaa gattatttgg tttttatagt tgtgtttttt tcggtagtat ttagaattat 960

tttataatag ttatgcgttt tttttttta atataatttt ggttttattt tagtgtttat 1020

ttttagttat atattttgga gaattttggg ttaaatttta gtttaaggga tgattttgat 1080

attttaagtt aattatagag gtattatttt tttttttaat gattggattt ggaatagtat 1140

ataatatagt tttggttagt gagatgtgag gggatttttt tttttagtga agggtttttg 1200

ggaaagatag atttattttg agaagagaat aaaggaagag gtggttagtt tttttttttg 1260

gatattatta tgttggaatt tttttgttgg aattgttgta gggattttgt atttatgagt 1320

aagtgagata tttttggaaa tgaagttgag aaagaattag taaaggtaag gcgtaagttt 1380

ttaaggatta gatttagagt ggattatgtt tttgaatttt attttatatg agattataaa 1440

tgttattatt gtttaaattt aataaaatcg tgatttttg ttttttatag ttgaaagtat 1500

tttaattggt atatattttt tagttgggat tttgaaaaat tttattggaa aagaaatttt 1560

atattattta gtaaatttag attgaaaata agttaatgtt gaaattagtt ataaattata 1620

aaattgattt tgaatataaa tagggtttgg agttgtatag tatttaatta attaagattt 1680

tttttttaa ttttttgta tttgattttt ttatatattt tgaaataaga attagatttt 1740

tttaaaataa aatttttaaa ttttataatt ttagatattt tttgaagagt atttttaag 1800

ttaatatatt atatatttta tatttttat ttatagatta attgataatt tataatgaaa 1860

aatatatgtt gataatttta aattaatttt attattagtt tgtatttatt ttatattatt 1920

agaattgtta tgatttgtta attttattgg aaagtttatg ttttagtaga taattaggtc 1980

ggattaatta aatagaaatt ttttttatgg tatgaattcg ggaggtggag gttgtagtga 2040

gttgagattg cgtttgggcg atagagcgag atttcgtttt aaaaaaaaaa aaaaaaaaga 2100

aattaaagat tttttttaa ttatatttat aaataataat tttttatttt tttttaggaa 2160

aatttgtttt ttatttttta ttttttttt tttttaggaa atttttataa gtttaatttt 2220

aaaagtttgg gttttattgt taaatagttt agaatgtaag ttttgtattt tttatttata 2280

ttttgtgatt ttgggtaaaa tattatattt ttttaaattt tatttttttgt tattatttag 2340

ttgtgaaaat tagtatttat cgaagtgatt ataggaattg ataaaataat gtatataaaa 2400

taaattttag tgatatataa tatattagta tttaatattt attattaaaa ttattaattt 2460

agatttttttg tagatgtttt aatatagttt tatattatta agagtaaaaa gaatagtatt 2520

tggggttttt atatttgaaa attataattt aggttattat tatagtagta atatatattt 2580

-continued

tatgtttgga ataaattatt ttataattta ttgttatgta ttttaaatta tttgtaaata 2640

tgtttaattt tttttgtaga atagtaggtt tttggaaggt agagttagta ttttatttaa 2700

tttaataatt tttataatat ttttttattg ttaggtttag ggtatatttt tgaggaatgt 2760

ttgtagatta aatatatgat taacgttttt atttttttatg tagaaatatt tttgaaattt 2820

cgtgttaata atttttttaat atttgaaata tttttgaata tttttgaatt agtttatttg 2880

aagatttatt atgatttaat aatatgatgt ttgttaatat ttagagagtt aaattttatt 2940

taaaaaaatt aaaggtgtga tttaggaaag gatttttttgt atttgagttt tttatttaat 3000

ttaaatatta aaaatcgaat tttttgtgtt atatttaatt tttataaaag tgtttttttgt 3060

tggggagagg agaatggttt tttatttata taggtataat aagtatacgt attatttttt 3120

ttaagattat ttttaagatt aggtaagaat ataaggagga gattagatat gttttaaata 3180

tatttgaata agttgtataa atatgtttga ataagttatt taatttaaaa attagattgt 3240

ttatattaaa tataaaattt ttaatttttt ttaaaaaatt agaaggttta gtaatattaa 3300

gtttatattt tagttggttg attatgagta ttaattaagt agtagttttt ttttttaagt 3360

ttgtgtgttt aaggttatta ttgtttttat ttagtttttt atttattttt attttatgtt 3420

tcgtttttgt agttatttaa gtttgtaaat tttgaattaa aagatgatta tttatttttа 3480

tagtaaatat attagagggt tttattgatt ttagttgttt tagaattgta tttttatttt 3540

ttttttttтg gagtttaaga cgtttatttt atttttgtaa ttttttttaag gttatgtttt 3600

tatttttggt tattaagtta tatttttttgt tttttatgtt tgattatatt aaatggaaag 3660

tttttagaag ttttttttttt tgtttttagt attattttaa atttttattt ttatttttttt 3720

atttttttat gttttttttа taagtatcgt ttattagtga taataatttt gtttgtatgg 3780

aaagatttga gagatatgag agtgaatgtt tatatttagt tttttatttt agtgaagtta 3840

atgagaaggg aagaaagtta ttgaataatt taggtatagg gtattgagat tgtttaattg 3900

tttgggagat ataaaagata aggttaatgg aatttttatt tatgagaaat gtttgtaggg 3960

aggataaggg atatattagg gatgggtagt tttcgagaga aagttaagat tttattaaat 4020

agatatttag tatttttttt tgattttggg gaagtttagt tgataagggg tgatttaagg 4080

atagagatac ggtggttttg tttattattt tgttaattag gtattttaga taattaattt 4140

tttaggatat tttagttata aagtggatgt aatgagtatt atatattata ttgtttattg 4200

gttttttttt atagaaagaa ttattgtata gtagagagaa attacgtgat ttaggtgcga 4260

aattttatga atttatttta tataagtaat agtattatgt gaatttttga atttggattt 4320

ggttaagttt ttttatattg tcgtttttta gtatttttтt gttttttgtt tggtgaattt 4380

tttttttatt attttattta agatatattg tattttattt tttttttat gttttatttt 4440

tattttagta tatattatta ttttgtatat gtgattggga attgtttttt tttatttatt 4500

atagtaaaag ttttaataga gtagtttttg atgttttgtt tatttttgta tttataatgt 4560

ttagaatatt tagtaaatat ttgttaaaag aatgaatata tatttatttt atataattag 4620

atgtataata tatatttagt tttatttatt aatttttta tttagaatat aatatatttt 4680

ggagaatgtt ttaaattagt aaatataatt tttttatgtt atttttgata gttgtatagt 4740

tttttattag gtggatatag taatttttat ttaaatggtt ttttatgatg agtttgttat 4800

attgttttta gatatttgtt tttagaagta atgttgtaaa gaatattttt gtttatttgt 4860

-continued

tttgggtgat attggtatta atatttattt atataataat gtcgagaagt ggaattgttg 4920

aattaaagag tatatatatt ttttaaaaat ttttgataaa gaattaaaaa aaagtgattg 4980

ttaaattatt ttttgaagta gaggaattta tttatttta tataatagta tatgagagat 5040

ttagtttttt tataatttgg ttaatataaa tttgtattaa atttttaat tttatttttg 5100

ttaggtatag tggtttatgt ttgtaatttt agtattttgg gaggttgagg tgggtgaatt 5160

atttgaggtt aggagtttaa gattaattgt tatggttaat atggtgaaat tttattttat 5220

taaaaatacg aaaattagtt aagtatggtg gtgggtgttt ataattttag ttattaggga 5280

ggttgatata ggagaattgt ttgaatttgg aaggcggagg ttgtagtgag ttgagatcgt 5340

gttattgtat tttaatttga gtaatagatt ttattttaaa aaaaaaaaaa aagaaaaaaa 5400

ttatttttt ttgttatttt tttgattggt gaaaaaatgg tgtagtatat tattggtgtt 5460

ttaattatta tttatttaat tatagatgag gttgagcgat tttttttata attttaaaat 5520

tttatgtttt tttttgttat tatttgtata taatttttgt ttatattttg atagatgtgg 5580

atttttttaaa gttatttata aaagtttatg gaatattaag agttattagt tttttgtttt 5640

ttaaatgtcg tagatgtttt ttaaatatat tatttgtttt tgtgtaattt ttatttagtt 5700

agatttatta gtatttttta gaatggtatt gggttttatg ttttgtttag aataggtttt 5760

tttatttttta atttattttt tttaatttta tgatttttt ttattatttt gtaaaaagtt 5820

atttttttata ttttatattt gatttatttt atatttttt taggatgaga aattaggtag 5880

agattgattt ttattttgta tggatttatt aattgttttg aaattattta ttgaataatt 5940

ttttttttt attgaatgtt taattttatt ggagttaatt tttgtaattt gttttttgat 6000

taatgaaatt aaataatttt taaaataata tgtttttttt tgatttattt tatagattta 6060

gtagaaatta tatggtttgg gtaatgtagt tattattaaa attatagaat tagattaata 6120

tgaattttaa gatataatat attattattt ttttttgta tagttttta aagtatttat 6180

ttttttgatt tttataattg tatatttatt taatagatta tgatattaaa ggtttaaaag 6240

taatattttt gtttggtttt tttagttagt gagtttaatt taaatttttt gattttaagt 6300

ttagtgttta ttttattata ttttaagtaa tattaaatta ataatttaat aataaaaaga 6360

ttataattta ttaaatatgt ttattgtaaa atatgtaatt ttattttata gatataaata 6420

ttttttaaaa tttttagaag aatgaggtag gatgtatata gatgtggaaa gtatatatat 6480

ttgtattttt ttatagttgg taaatttatt tagaggtgtg ttttaggaaa tagaattgat 6540

atattaaata aagatttttt gtattgattt ttgaaatata attatttta atatatttaa 6600

tggtaagaat ttagtagtgt ttaaaaattt tgtatatata tatatatata cgtattttat 6660

atatatatat agattttata tatatgtaat ttattaagtt taggttataa gaggaaattg 6720

gtaaaaattt attaattgaa atggagtttg gttaagatat ttatatttaa ttttttttt 6780

ttgattttaa attagggaga gtatatgttt tgtgtttttt ggtaaaatat agatttatgt 6840

aataattagt atatttttt agattagagt aagtaatttt taataatgag tgtttgttag 6900

ttttaattgg ttttttacg tgggtggtga aaaaatattt tattatagtt attaattttt 6960

ttttttaaaa gaattatgtg ttttgagaag ttttgtttta ttatagggta atattttatt 7020

ttaaatatag tgtttatcga tgtttgttta tgaatgacgt taataaaaat gtttaaattt 7080

tattaggaaa ggagggttta gttgttagat tttgggagat agagttagtt gatttttgcg 7140

-continued

ttttaaagg ttttgttatt taattaagtg attaattatt aattttaaag tatttaagtt 7200

taataatata ataaaagaga gaataatggg tagtattttt aatattaatt atggttattt 7260

ttggtttagt tgtttaggtt tttttttagt agaatttata agtgtataga tataggtatt 7320

ttggaatatt aaaggtattt ttttttagtt tgtaaatata gggaaaaaga agaaatttat 7380

atttgtaaat ttatgtaaat ttaggggggat tttttttaaa tttcgttttt ttgaaagtag 7440

gaaatataat ttaaaattaa ttttgttttt atttattaaa agaattaaaa ttttgaattt 7500

tgatataaat tttaaatgat tgttttgaat tttatttcgt tttttattta attttttttat 7560

aagaagatta aggattgtgg atttggtgtt ggaggaagta attgtttaat tatattggaa 7620

tttattggaa ttttttataa tattattgcg ggtagataat ataattttaa tatatttttt 7680

tttttattta gagtttagaa taagaagtta gatttttgga gattttttttt ttgatgtatt 7740

gtttttaata tttttaagtt ataaattggg aatcgttttt attttagaat attaattgtt 7800

gtttttattg gtatttttttt atttttatat tatatttata atattttatt ttaggatttt 7860

atgtgaatat ttttaggatg ggtttggcga aattgttaag aagtgtttaa aataattagg 7920

ttatagtttt tagtagataa aaagaaaaag acgtggttgt tagttataga atttgaggtg 7980

atggaaaaga ttttttaaaat ttagtttttt taatatgtag taaaatggga ggatttaaaa 8040

tgtagtttta agaagtatcg tttcgtttgt attaagtttt tttaggtttt tagtttagtt 8100

atttttttaaa taggtgttga gtttttagtt ttcgttaggt tttcgttaag gttttttatt 8160

gaggtatagg aattttagtt atattcgtag tttttttatg tttgttttta tgaaatttta 8220

ttagttttgt cgaatggtag tatttgttat tttaggcgaa ttttttatga tagtttggaa 8280

cgaaattaat ttgttaattg attaagaatg ggttttagag tgatttgaat gtatttaaaa 8340

agttaataaa tggaatacgg tataaagttt tatttttttg ataaaaagtt tgtttggagt 8400

aaattttaat ttttttttttg ttttaaaata aaagtacgat gttttttttta ttatttttaa 8460

gaagtaagta aaaataaaat aaaataaaat aaaatttaaa tagtttttta ataagagtgt 8520

ttagcgttgt ttggatggaa ttcgggttag ttttagaagt agtcggtttc gaatttcggg 8580

ggcgaggtta ttcggttttt tcggttaagt tgggttttgt ttttttcgtt tcgtttttat 8640

tggtttcggt gggtgggggt ttcggttgcg tcggtcgttt gtgcggttgt attgtagttt 8700

ttgagtttag gaagttatat tcgtttcggt tgggagtcga agcgttaatt tacgtggcgg 8760

ttggttaatt tagtcgttgc ggggagcgtt cgttgcgggt ttttttttgg cgcgaggtga 8820

gcgagtgggg gcgggtttta gatttagttt ttttgggcgg ggttggatta ttttcgtttg 8880

tttagttttt ttttattttt ttatttcggt acgtttttgt agttatattt gttgtcgttg 8940

ggaaagggaa aagagtattt taaaacgagt ttcgaaattt agatttaaat ttcgaaacga 9000

atttttttcgg ttcgtttttg gttgtttgaa tgttatttgt tttattttcg ttttttttttt 9060

ttttttttttt tttttacggt ttttttttttt tttatttttt ttttttttttt ttttttttttt 9120

tttttattgt tcgttttttt tttatttttt cgttcgacgt tcgttttgaa gttgaggaag 9180

tttaagtttt ggattaggag ttttttttttt gtaatatttt ggggcggggt tttatcgttg 9240

ggtttaagtg gttatgggtt ttgagagtac ggtatgtgta ttatttttttt taaagacgtt 9300

tagggtggcg cgtttagatt ttcgaaattt ggttttattt cggttatgtg cgttcgttta 9360

ttaagcggcg aattggattt agggcgtcgc gtgcgtttta gtcgattcgc gagaaggtat 9420

-continued

ttttgtattt tgtatagatt aagatttttt tttagggttt tttaaatttt tagatgtacg 9480

tcgtagattg tggagtgtaa aatgaatgat ttaagggaaa attagggttt tttatttttg 9540

aatataaggt tgtgaaagat gtacgtttgg tttgggggggt ttgtttttttt tcgtttgtgc 9600

gggatttttt tattaggggg cgtttatatt cgtcgtaaaa tttttaattt ttagatattt 9660

aggttttgta agttttgtga gttgaattgg gggtagattc gaaggagcgt atcgtttttt 9720

tcgttttttt ttttttgttg ttttggagta cgggaaaggt agggatttta tattagttat 9780

ttatttgtta gttgtggggg aggggacgta gaagttagag agagagaata gaggttggga 9840

gaggggaagg gaggtagaga agatggagat agagggagat ataaagatga ggaacgcgat 9900

ttagagagat tttgtcgcgg gtgtgcgcgc ggtaggataa ttttttaggc gtcgtgtaga 9960

ttttttttttt ttttcgtagt atttttttagg gatatgtagg atttgtcgag acgtagttgg 10020

attggagagt aaatagttta ggttatgata ttatttttttt atttcgtttt tttttgttta 10080

atatttagat cgttttacgt ttagtgatat agatcgcggg ttagaataat tcggatgttt 10140

taggtttgta cgtattaggg atgtcgtttt atagcgttta cgtttttgta ggacggtggg 10200

gcgttttagt cgaggacgtt tttcggggtt ttcggtatac gcgcgcgtcg ttgtcgttag 10260

atttgttagt aaagttgttt aattagagcg gagaattagg agagggtggg gaatatgtaa 10320

ggaaacggag gatttttgta gtttgggcga gaggattgat ttagaggttt ttgattcggg 10380

ttttcgagtt atttttgttt gtagtgcgtt ttgggtgtcg gtttgtttgt tgagatggga 10440

gcgatagatt ttattttatt ttttagggtg agaaaaattg ttgtatattt ttgtcgattt 10500

agaataggtt tttttcgttt aggggttttt tagattttcg cggggtttta ataattttat 10560

ttttttatttt ttgtttttttt tttgttggtt ttttcgtatc gcgtgtagga taaaagaatt 10620

taggggcgtc gggagagaat cgtgatagtt tttaagtcgt taaagagata cggcggggat 10680

tttgtatttt tttcggtcgg atttttttttt agattgtttg cggtttgaag gggttaggtt 10740

cgcgtaaggt tttaggttgt ttgggcgtcg ggtcgtattg ttaggagttg ttcgggagtg 10800

gtttttttagg aatattaggt attgattata tattaggggt tgggaaatta ggtggtttgt 10860

attaaggcgg tttcggggat ttgtttgtgg taagttttgg tagtttttat ttaaattttt 10920

gtttcgagcg tgttaagaat aataataata aaaaaattaa agtgttaaag gttttttttt 10980

tttttttagt ttaagaattt attatttttt tatgattttt ttataattta ttttttttttt 11040

ttttttaatt tcgttagtta ttttatttttt attttatttt gggttttttt tgtttgaatt 11100

ttttttaata ttaaggtttt tttgtatgtt tttttttaaa agtttttatg aaagttattt 11160

gtatttttta agtgtttata tttttttaat ttcgttttat agtttttttgt tttaattaaa 11220

gtttttttatt aattgttttt tttttttaag ttcgcgggtt tttttttata agtttttttgt 11280

ttttgttttt taaggggggga ataaaagaaa cgtgattatt ttggaaggcg gtttattgta 11340

gtttgggggg aaaatttatt gtagcgttgc gcgattgggt tcggcgttgt ttaggcgggt 11400

tatataggaa gcgtggtggt tcggggaagg atgcggaggg tgcgggacgg tgttggaaga 11460

tgcgggagga tgcggggttg gtcgaagatt ttggttttag tttttgttta taatatttaa 11520

tgttttcggt tgttgcgttc gtacgtttgg agtttttttt tagaaaggtt cggggtagtt 11580

cgtttgtaag tttaaatgcg ggttgtgata tttattatta ttatattatt gtatcgttag 11640

agtcgaggag gagatttagc gagaagaagg aggagggaga ggaggagggt ttatttatag 11700

-continued

gttttaaaag cgtttcgtag ttttagttat ttttaagagt ttaggtttgg aaagtaggcg 11760

gaggggcgga aaggtagttt ttcgcgtttg cggtagggat cgtttgtttt tttttttagg 11820

tatttttatt ttttggcgtt tttttttgat aagaagtatt aatcggtttg gggatagtag 11880

cgtttttttat ttagggattt atttagtaat atcgttttgt tcgcggaaat tattgttttt 11940

ttatttttttt tttatttttta ttttttcgtt tttcgtgtag ttagttttgg gttaggggaa 12000

aggagttttt aggtttttag ggggtaggtt agtaatagat aattgagtac gattattttt 12060

tttcgggagt atataaaatt gtaaaattag taaagaattt ggttatagcg tgtttacgcg 12120

tttatagagt ttgtttgttt tattaaaggg aagtgttagg tttaaggttt ttgttaattt 12180

gaaagagata ttgagaaaac gagatttttc ggggatttag agggaaagtg taagaatttt 12240

ttattgtatt tttagggaat tgtttaatgg ggagttcggt tttaaaagat tttggtaata 12300

aaaggttgga taggaaattt ttttaggtaa atttttttgtc ggatttaaag agaatatttt 12360

ttttttgtta taaatttttt tttatataag tttagatttt gtttttttcg tttttttttt 12420

tttttagttt ttttaagtat ttttgagtag aatatttgat aatttttttg agtaataggg 12480

atttttttgga agtattaatt atttttttatg tttttcggaa ataagattat aatttttttat 12540

gcgtatatgc gatttttttt ttttagttag gtttatttta ttgtgtaaat agtattaata 12600

tatggaagag ttttttgtat tgtgttataa aagattttta ataggatttt atagagaaaa 12660

gggttaaata gttgatataa aggatttttg ttttttttaga aaagagggat tttggatttt 12720

ttttttttttt gaagttaagt atgagtttat ataataggaa taaaataaat ttaaggtgta 12780

tattagtata atattaggga tattagaatg gatggtaaat ttttttttttt atataaatat 12840

gaaagtattt ttataatttt tttttttgaa gtttttattt agaaaattat attaaatata 12900

ggattttaaa atagtagtga ttaaagatga aagttaattt tttttttttaa ttttttttgat 12960

tagtttggtt taaatttatg ttttaaaatt tttagtaatt tagatttttg tatatgcgta 13020

tttataagat tttttttata ttttgtaatt tgtaggtatt tagtatggtt attgatttta 13080

agtgataaat aggtagaaga tttgttagga taattttagt ttttattatg tgtatttttt 13140

atttttttttg ttatttatta gtttttttgta gtaatataat tgttagttat attatttgga 13200

aattaatgtg tttttgtagt gaattattta tttttagtaa atttattgta ttattatttt 13260

aaatattgaa atttgatttt attagatgtt tagaatgata gttatagagt agaattagat 13320

tttaattttt atgtataata tgtaaataaa ttaaggtgta gaaagttatt gttttaaagt 13380

tatataatta atagtttagt ggtttttttta gtgttaatta tagtaatttt agatttaaga 13440

ttgtttgatt taggaatgga gagaaataaa ataaaatgaa tgtttattga aatatatatt 13500

ttagatatta agatagattt ttttgtgaaa tattttaaga tgtggatttg tatttttatt 13560

tttaaatttg agtaaattta gttgagtttt aaagagttta agtaatttgt ttaaggttat 13620

tttgtaagtg atagatttgt tagtttgaaa agtttattat tttttatttt tttgtttttt 13680

tatgatatag ttaagtaaaa ttttttaata aaggagaatt agattagaaa taagtaatta 13740

atagtagatt aatgtttatt tttagggaaa tgttaatagt taatatttat tataaataat 13800

gttatttgta tataatggaa tatgattgaa agtgatgggt tagatatttt ttatattttg 13860

agatgttttt atttatgttg ttaaatttga ttttttaatt ttttaaaatt tttagttttt 13920

tttaatgtta atagatttta taatatttga ttaaatattt ttataggata aaggtatttt 13980

-continued

tttagggttg atttaaggaa ttagaatatg tataatgttg aaaagtttta aattgtttgt 14040

ttttttttt ttttgatttg taatattatt atttagaaaa ttaaatattt ttttaggtta 14100

ttagataatt aaaatgaaat tatattttat ggatattttt tttttatttt tagggtagat 14160

ttttttttt agttgtttaa aatggagaat ggataaaaaa gaaaaaggaa aatatattag 14220

ttatatttag aatagagttt atagttgtat tattattttt atgtaattgt ttttaatgat 14280

ttttttttat taagagaaag aaaatgagag agtatttatt tttgattttt taatgtttga 14340

taaaattatt agaagttttt gaatgggaaa atatatattt tttaggtttt ataattaaaa 14400

ttgtatttat ttgtttatta tatttaagta tgaattaaat agtaattaaa atattatttt 14460

gtgagttttt ttttaaatta tatttttgta ttggggatat tttttgtgaa aattttaaga 14520

ttttaggtat aaataaagta agatttttat ttataatgtt tgtgtttaaa agaaattata 14580

agttagaagt tatgtgtttt gtttttttgt tttgattttg tttttgggaa tttggttgag 14640

tttaggggta gtaagttaat gatttggaga ggaaaggtgt agatggtaga tgtgttttta 14700

gggaatttta ttatttttat ttaaattttt ttttgaggtt ttttaggatt atattttaga 14760

ttttttttt tttattgatt aggtgaagtg tttttagata tggtattatt tattgttgtt 14820

ttagggatga taagatttta gttttattta aatttagaaa gagtttagtt ttaatttttt 14880

gatttgattg gttttaggta taattaagta aattattttt attggatgat atagtaatgt 14940

tatagattaa aaatgtattt ttaattttta aagattttgg aaaaaggaag tttatttttt 15000

aagattgtgt tataattttt tttttaaata aaatattgta aattagggat agatatttta 15060

tttagtattt tgaatttttt ttggtttttt aggaaaggta gaaatataag gtttttattg 15120

ttttttttt atttttaaaa aagaaatttt gtaaaaggta aagtgatggt ggtaatttta 15180

ttgtgtttta attcgaagta tataaattgt tttgataaat taaatataga taaatttttt 15240

tagaaagtta taaaatatta gatttttttt agatattaga tattgagttt ttgtttttta 15300

atttttaaat taaaatgtta tagtaaaata tttgttttaa taaatttttt ttttttttgt 15360

gttgtaaatg ttttgtttat ggttgagtag atttttttag aatttaattg aaagattatt 15420

tatattacga tagaattatg taagttaaag ttaagtttta atatggagtg atgggagagg 15480

tcgggttgtc gagaggagat gattttatta gtgagaagag aggttttttg agtcggtttt 15540

taaggtagtt aaatggttta gtatatttgt atttattttg ggaggttgtt tggagtagag 15600

tattaaaaac gttatttttt ttataaagaa aaagatgaag gtataaagaa tatagtttaa 15660

ttaaattgaa ttaaagattg tattattttt attttttttt tagttatttt tttaaagttt 15720

ttttttttt tttttatagt ttgaaaaata taggttattt ttataaagta tttgtttgat 15780

tgaagtaaga taaattgaag agtttggttt ttgaatataa ggaagcggta gagattttag 15840

aattaatgag gtagggcgga atataatgta gaaggttttt taagaggtaa aatttttgta 15900

gttgttgttt atttttagtg attattcgtt tttatttatt taattagatt atagtgggat 15960

cgagataagg agtataatga aagttgttgt attgatggtg taaaaagata tatggtatta 16020

ttaatgtttt tgtggtgagt aaaggtagaa cgtaagatag aaaataagga aatttttgcg 16080

ttgtaagaaa agtgtttagt atgggtatta atcggtatgt agaagtagta agtatatatt 16140

ttatttttat tttattgata aaaagaattg taggattaat gtttgttttt attaaaatat 16200

gagaaattat attttaattt tataaattaa aggttaaata tttgttgtaa taattttaaa 16260

-continued

aaaaagagaa gatttgtaag atttagtaaa tatttattat aaatagattg tttttaattt 16320

tttatatttt tttttaattt aataaaagtt tattgagtag tttttatgtg tattgtttta 16380

gatagtaggt attatttaaa aaagtaataa ttatttttta tttttaagga gtttataatg 16440

tttttagaaa gaaattaatt tattggagta gttaaagtta taattgattt ataaaatgat 16500

tagtgaattt agaggaaaat tatatataaa tatttttgtt tttattggtt ggtttagtta 16560

gatattattt taaaaaata 16579

<210> SEQ ID NO 10
<211> LENGTH: 3000
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 10

atatttaaat tgtatatttg gtttgtaata tattatatat ttaaaagaat tattattttt 60

ttgaatatta ttaatattaa taatagtatt atagtgatta agagtggatt taaataaagg 120

ttggtttgaa atttaggttt gttatttatt agtaagtttt taaaaatttt gagtttttag 180

ttttttttata tgtgaaatgg agaaaataaa tttgtaaaat taattaaatt tgtttattta 240

tatttttttt tttagtgtat attaattggt atttttcgtt aggttagaat gtgtttttaa 300

ttatgtttta aattcgtttt gtgttaattt tattgttaga atttttttta ttttgagaat 360

tagaaaagga aatattatgt ttggtaatgt ttatattttt taaaataaat ttgtaggaaa 420

gaatatttag taagtgatga gagtagtaac gattgttttt attattttaa atttataata 480

ttatttttttt tagagttttt taagtattgt agataatttt ttatttatta aaaaataaat 540

tgtaattata agtatttagg gttgatattg tttttgaatt agatagtgtt tatattagtt 600

gtataagatt aatttaaagt agaggatgaa attttttttt tgaatttttt ttagaacgta 660

atttagtgaa tatattaaaa ttaaattttt tttgaatggg agtaattttt acggattaat 720

ttgtaatttt ttagattata tttaaggtaa tgtagaggtt gttgtattat aggttttgtg 780

attagagatt attggatttg tttttggaaa agttatttat gatttttgg gttttggttt 840

ttttatttttt tatatcggtt taataatgat tatcgtagag ttgttatgga agttaaatga 900

atttatgtat tataagtgat taatataatg attgatattt agtgatttat taataaatta 960

aaatatttat tattaatatg atagagaagg tgtcgttaaa atagataata ggtttttgga 1020

agaggtgatt aaatggatgt aaaatttatg gattgtttat ttcgtttatt tttgttgtgt 1080

tttttggttg tggtatatat acgtgtgggt ataaaatcgt aaattttatg tagtcgcgta 1140

gtgtatgcgt agaaggttta gatacgaaat gttattttag taatgtgttt agagaagttt 1200

tgacgtcgtt ttggaagtaa gtcgttgttg tttgattttt gggcgtttgg gacggatgtt 1260

tatatttgta tttagtagta ttggaagggg tttaggtttt tcgtagtata gtttatttttt 1320

agatcgttta gtttttttata atatatattt ttacggaaaa gggtattttt tttcgttaga 1380

aaaagcgttt tagtttggtt tgggttggtt tttattttac gttgttgtaa gtaggcgaag 1440

tttttttttgt ttttttttttt ggggtaagtg gaaaggagtt cggtaggggg ttcgtagtgg 1500

tttgtatagg ggaattgggt agcgagagag ttttaggtaa tttcgggggt tgttttatag 1560

aagtaggtgg ggatcgatag tggtttttcg gtttagggag gagagcgcgg tcgcgggttt 1620

-continued

tttttttag tttggaggtt gtagtcgttc gagtcggttc gggtgggggc ggggtggggg 1680

cggcgcggag ggtacggaga ttacggcggc gttattcggg atatttaggg tttcgaggtt 1740

ttgggcggtt tttacgcgag atcgtaaatt atgataatag gtagttattc gaggttaaat 1800

aaaaacggag tgggtttttc gcgcgtcgtc gtttttcgcg tttttggcgg tttttttcga 1860

ggttttcggc ggttttacga gttcgtagta gtcggtggcg acgtcgtttt cgttttattt 1920

ttttgcgtaa gtgcgaggtt gtcggtagcg cggcgtacgt ttcggtcgtt ttcggttttc 1980

gcgtaaaatt tttattttgt ttacgtgaag ttgtcgttgt tttagagagg gggaaagagt 2040

tgcgggaaaa gtcggggagt gacgattgcg gcggttgggc gcgttttttt attttttttt 2100

ttttttttt tttttttgtc gtagttcgga gttttggttt ttttttttt ttttttttt 2160

tcggagtcgg ttttttttt cgtttcgttt ttttttcgtt tgtgtacgtt atttgttgtg 2220

gggtggtcga aggggatgtt ttgtttttat tagaggtata gcgcgaaggg gaaatttcga 2280

tattggaagg aacgagaata aatatttaat tacggacgta ttgaatcgcg gttgggatag 2340

atatttcggg aattcgaggc ggatcgggcg acgaggtgag tgatttttt ttttaatttt 2400

cgttttaggg ttttcggggg agtttgagtt gagagaattt ttaaattttt cgggaaagtg 2460

cgcgaggttt cgtcggggac gtcgagcgtt gggtattgag gacgcgtagt tggacggtgc 2520

gtgggcgttt gcgttttcgg ggggcgtttg gaggtcgggt gttttacgtt tgagggttcg 2580

ggtcgttcgg atcgtagcgg tgttttttgt tttagaagac gttttaagt tttaagggtt 2640

tttttcgagt ttgtttgttt ttttcggggt cggcgcggag tttgcgcgta acggagttta 2700

tttagtagtt tagcgcgcgg tttttatttg tatttcgttt ttatttggta gaggcgcgag 2760

tatcggggtt ttttttatat tttttttatg acgtgtatta ttttttgatg attttttaga 2820

tggtttaggc gcgaggatgt tgatttagag tttttcggag ggttataggc gtttgggttt 2880

tttcggtgtc gggtgcgtgt gtattttaaa ggttcgcgtt ttaattttta ggtattgatc 2940

gggtttttta attgcggcga ttttatttta atagttttta tgtggcgtgg attgaatgtt 3000

<210> SEQ ID NO 11
<211> LENGTH: 3000
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 11

gatatttagt ttacgttata taaaaattat taaagtggga tcgtcgtagt tgaaaagttc 60

gattagtgtt tggagattag aacgcgagtt tttaaagtat atacgtattc ggtatcggga 120

aagtttaggc gtttgtgatt tttcgaagga ttttgggtta gtattttcgc gtttggatta 180

tttagggggt tattagaaag taatatacgt tataagaaag atgtggggga gatttcgatg 240

ttcgcgtttt tgttaggtgg aggcggggtg taggtagaag tcgcgcgttg gattgttgga 300

tgaatttcgt tacgcgtagg tttcgcgtcg atttcggaag ggataggtag gttcggaagg 360

gatttttggg gtttggggac gtttttttagg gtagagagta tcgttgcggt tcgagcggtt 420

cgggtttttа ggcgtggggt attcggtttt taagcgtttt tcggggacgt aggcgtttac 480

gtatcgttta gttgcgcgtt tttagtattt agcgttcggc gttttcggcg gagtttcgcg 540

tatttttttcg gaaagtttgg gggttttttt aatttaggtt ttttcgggag ttttggggcg 600

-continued

ggggttggaa gaaggggtta tttatttcgt cgttcggttc gtttcgggtt ttcgaagtgt 660

ttgttttagt cgcggtttag tgcgttcgta attaagtatt tattttcgtt ttttttagtg 720

tcgaagtttt tttttcgcgt tgtgtttttg gtgaaaatag gatatttttt tcggttattt 780

tataataaat agcgtatata agcgggggag aagcggggcg gagggagaag tcggtttcga 840

gggggaggag gaaaggagag gagttaaaat ttcggattgc gatagggggg aaaggagaag 900

aaaagaaaat gagagagcgc gtttagtcgt cgtagtcgtt atttttcggt ttttttcgta 960

gttttttttt tttttttaag gtagcgataa ttttacgtgg ataggatgga agttttgcgc 1020

ggaagtcggg aacggtcgga gcgtgcgtcg cgttgtcggt agtttcgtat ttgcgtaggg 1080

aggtggggcg ggggcgacgt cgttatcggt tattgcgggt tcgtgaggtc gtcgggggtt 1140

tcgggggagg tcgttaggga cgcggggggc ggcggcgcgc gggggattta tttcgttttt 1200

atttgatttc gggtgattgt ttattgttat ggtttgcgat ttcgcgtggg gatcgtttag 1260

ggtttcgggg ttttggatgt ttcgggtggc gtcgtcgtaa ttttcgtgtt tttcgcgtcg 1320

tttttatttc gtttttattc gggtcgattc gagcggttgt agtttttagg ttgaggggag 1380

ggattcgcga tcgcgttttt ttttttgggt cggagagtta ttgtcgattt ttatttgttt 1440

ttgtggggta gttttcggaa ttgtttggaa ttttttcgtt atttagtttt tttgtgtagg 1500

ttattgcggg ttttttgtcg gatttttttt tatttatttt aagggaggag atagaaggga 1560

tttcgtttat ttgtaataac gtgaaataaa aattaattta gattagattg gggcgttttt 1620

tttgacggga ggaaatattt tttttcgtgg agatatatgt tatgaaggat taagcggttt 1680

ggggataggt tgtgttgcga agggtttggg tttttttttag tgttgttggg tgtaggtata 1740

ggtattcgtt ttagacgttt aaaggttagg tagtaacgat ttatttttaa ggcggcgtta 1800

gagtttttttt aggtatattg ttgaaatgat atttcgtgtt taagtttttt gcgtatgtat 1860

tacgcgatta tataggattt acgattttat atttatacgt gtgtatgtta taattagggg 1920

atatagtaaa ggtagacgga ataaataatt tataaatttt gtatttattt aattattttt 1980

tttaaaaatt tattgtttat tttagcggta tttttttttgt tatattgata ataaatgttt 2040

taatttattg ataaattatt gagtgttaat tattgtattg gttatttatg gtatatgagt 2100

ttatttaatt tttatgataa ttttacggtg gttattatta agtcggtata aaagatgagg 2160

aaattaaagt ttaaagaatt atgagtgatt tttttaaaga taaatttagt ggtttttaat 2220

tataaagttt atgatataat aatttttata ttattttagg tgtggtttaa gagattatag 2280

attaattcgt agaaattatt tttatttaaa gaaagtttag ttttaatata tttattaagt 2340

tacgttttga aaaaggttta gaaaaaagat tttatttttt attttaggtt ggttttatgt 2400

aattgatatg agtattgttt aatttaaaag tagtattaat tttgaatatt tatggttata 2460

atttatttttt taatgagtgg ggaattattt ataatgttta agaggtttta gaaaaggtgg 2520

tgttgtaaat ttaaaatgat aaaggtagtc gttgttgttt ttattattta ttgggtgttt 2580

ttttttgtag atttattttg gagggtgtag gtattgttag gtataatgtt ttttttttttg 2640

gtttttaagg tagaaagggt tttggtagtg gggttggtat aaagcggatt tggagtatgg 2700

ttgagagtat attttggttt aacgaggaat gttagttaat atatattgga gagaaaaata 2760

tgaatggata gatttaatta attttgtaga tttatttttt ttattttata tgtgagaaaa 2820

ttaagggttt agaattttta gaagtttgtt aataaatggt agatttgagt tttaagttag 2880

-continued

ttttttattta agtttatttt tagttattgt gatgttatta ttagtattaa tggtatttag 2940

aaaaatagta atttttttaa atatgtaata tattataagt taaatatgta atttaaatgt 3000

<210> SEQ ID NO 12
<211> LENGTH: 1984
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (9, 17, 18, 33, 41, 49, 50, 52, 54, 62, 80, 84, 127, 157, 159)
<223> OTHER INFORMATION: unknown base
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (175, 207..209, 214, 1181)
<223> OTHER INFORMATION: unknown base

<400> SEQUENCE: 12

agttttagnt attattnnta tttttattaa tgnttttggg nttataggnn gngnttttt 60

gnttaaatag tagttaatgn aggnatttta aaaggtatat tattttatat ttatttagtg 120

gttatgntat tatttaattt ttattataat tttgtgngnt tagtataatt tgtgntttta 180

ttttatttgt gaataaatgg aggtagnnnt ttgngtaatt tataagtaag tggtagattc 240

gggggttgtg tttagataat gtggttttag atttttttgtt ttttttgtt ttttgttgta 300

ttttgattga gtttagttat ttatttttt aatatatatt aattgttgtt tttatgtttt 360

aggttttgtg ttagatgtta gaaatatatt tttatttatt ggattttggt atggtttatt 420

gtttagtgtt ttgttatgtt acgtgttata taatgtattt atttcggtat ttagtttatg 480

gaaaataatg ttgaaagata ttgtatgtat gtttttataa taaaattatg taatttatgt 540

ttttttttatt gttttggtta ttaggaagtt tttatttaaa ttagagtaaa tatattagaa 600

tttgggtttg ttattttagt ttgttgaatt ttttttttg gtttagattt tttattttgg 660

tttataaatt ttattgtata aatgtttttt attgtaaaat attttaaatt ttttttaagg 720

gaaggttgta tggaaatgat atagtaaggt tttttttgta ttttttaga tttttattag 780

ggaaggtaga tttagatgtt ttttttattt ttagttttaa agtttttttat ttataaaatt 840

ttttaagtag tgattattgt tgttttgagt atttggaggt agtttagagt ttttgaaagg 900

taaaatttat atataaaaag aagtttttttt cgatttagat tttagttttg gtgttagatg 960

tatatttgag gagtaggtgg ttagttagat ttttatttga gtagttaata aaatttattg 1020

ttttttaatg gaatttttttt tgtaagttcg aattgatttg ttatttttgt gatttgtgag 1080

atggtaggga agtattaaat attattatga tttgggttat agtggtggga aaaaaggaaa 1140

aaagaaaaaa aaaatttatt gttaagtttt gttaggcgta naaagggttg gaattgttgg 1200

ggttatttta tttgatttat tggaaataga gtggatttta ttaatatttt aataaagaga 1260

atttttttgt attaggttgg aagtggtcgt tagtttttcg tgtaatttta ttttttggaa 1320

aagtggaatt agttggtatt gtttagcgtg atttgtgagg ttgagtttta atagtttaaa 1380

gaagtaaatg ggatgttatt ttcgcggggt tcgtttttcg cgaggtgttt atttcgtatt 1440

tgttatgtaa aacgagggag cgttaggaag gaattcgttt tgtaaagtta ttggttttgg 1500

ttattagttt ttatttaatg ttttcgtgat gttgttgttg atttatttgg gaagttggtt 1560

ggttggcgag gtagagtttt tttttaaagt ttggttttta cggaaaatat gtttagtgta 1620

-continued

gtcgcgtgta tgaatgaaaa cgtcgtcggg cgtttttagt cggataaaat gtagtcgaga 1680

atttcgttcg ttttgtgcgt ttttttgttt taggtaggga agaggggttg tcgggcgcgt 1740

tttgcgtttc gtttttgtat tcggatcgtt cggtacgggt agggtgaggg ggttttcggg 1800

gggtcggggt tttcggtcgc ggcggcgaag atagatcggg gttcggtagg gaggttattt 1860

cgagtttaga gattttaggt atttttata tataggtttt tattttggcg tgcgtgcgtg 1920

tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtacgtt cgttaacggg 1980

agga 1984

<210> SEQ ID NO 13
<211> LENGTH: 1984
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (804, 1771, 1776..1778, 1810, 1826, 1828, 1858, 1901, 1905)
<223> OTHER INFORMATION: unknown base
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (1923, 1931, 1933, 1935, 1936, 1944, 1952, 1967, 1968, 1976)
<223> OTHER INFORMATION: unknown base

<400> SEQUENCE: 13

tttttcgtt aacgaacgta tatatatata tatatatata tatatatata tatatatata 60

tatatacgta cgtacgttag agtgggagtt tgtgtgtggg gggtgtttag gatttttggg 120

ttcggaatga ttttttttatc gagtttcgat ttgttttcgt cgtcgcgatc ggaggtttcg 180

atttttcgaa agtttttta ttttgttcgt gtcgggcgat tcgaatgtag aaacggggcg 240

tagagcgcgt tcggtagttt ttttttttta tttgggatag gagaacgtat agaacgagcg 300

gagttttcgg ttgtattttg ttcgattaga agcgttcggc ggcgttttta tttatgtacg 360

cggttgtatt gagtatattt ttcgtgggag ttaggttttg aggagaggtt ttgtttcgtt 420

agttagttaa ttttttaaat agattagtag tagtattacg aaagtattgg gtagaggttg 480

atgattagga ttaatggttt tataagacgg attttttttt aacgtttttt cgttttgtat 540

ggtagatacg gggtgagtat ttcgcgagga gcgagtttcg cggaggtggt attttatttg 600

ttttttgga ttgttggggt ttagttttat aaattacgtt gggtaatgtt agttgatttt 660

atttttttag agaatggaat tgtacggggg attggcggtt atttttagtt tagtgtaaaa 720

agatttttt tattaaaatg ttaataagat ttattttatt tttaataaat tagataaaat 780

ggttttagta gttttagttt tttntacgtt tggtaaggtt tggtagtgga ttttttttt 840

ttttttttt ttttttttat tattgtggtt taagttatga tggtgtttgg tgttttttg 900

ttattttata aattataaag atgatagatt aattcgagtt tgtagaaaaa attttattaa 960

ggggtaatag attttattaa ttgtttaggt gagagtttga ttagttattt atttttaaa 1020

tgtgtatttg gtattaaagt tgggatttgg atcggaagga attttttttt gtatatggat 1080

tttatttttt agaagttttg aattattttt aggtatttag aatagtagta gttattgttt 1140

aagggatttt ataaatgggg ggttttggga ttgagggtaa gagggatatt taggtttgtt 1200

ttttttaata ggaatttaag aaaatgtagg gaagatttta ttgtattatt tttatgtagt 1260

-continued

tttttttag aaagaattta aggtgtttta taataaagga tatttgtgta atagaattta 1320

tgaattaaaa tagaaaattt gggttagaag gaaaagttta gtaaattgaa atgataagtt 1380

taaattttaa tgtatttgtt ttgatttgga taagagtttt ttgatagtta aagtaataaa 1440

agaaatataa gttatatagt tttgttgtag aaatatgtat ataatgtttt ttagtattat 1500

tttttatggg ttaaatatcg gaataaatgt attatgtaat acgtaatata gtaggatatt 1560

gagtaatgga ttatgttaag atttagtgga tggggatgta tttttagtat ttggtatagg 1620

gtttggaata tggagatagt aattagtgtg tgttaaagaa atagatggtt agatttaatt 1680

aagatgtagt agaaggtagg gaggggtaag gagtttggag ttatattgtt taggtatagt 1740

tttcggattt gttatttatt tgtaaattat ntaaannntt gtttttattt gtttatagat 1800

aaaatgaaan tataaattgt attgantnta taaggttgtg gtgaggatta agtgatgnta 1860

taattattga ataaatataa agtgatgtgt tttttaagat ntttntattg gttgttgttt 1920

ggntagaagg ntntnntttg tggntttaag gntattgata aagatgnnag tgatgnttga 1980

agtt 1984

<210> SEQ ID NO 14
<211> LENGTH: 7833
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 14

gtttttggtg agatatgtgt tttataagtt ttaatggaga aaaatgtaag tattttattt 60

tttgaaattt ggttatttga gtaatgagaa aatagttatt ttttttagga tagtggtttt 120

taattatggt tatgtgtttt tttaggaaaa ttttaaaaat atatatatat taatgttttt 180

gtgttatttt tagggatttt aagtttttga atacgaattt tgtattagta ttttttaatt 240

atttaggtga ttgtgatgtg aaattatgat tgagttttat tgttttaaga tgaaataaat 300

tttttttagt attgaaatta taaatttaaa ttattaaaat taattaaggg tatgggaatt 360

aataaggtat agggaagttt ttatattata aaattatttt tttaaattat agtttattgt 420

ttatatgtta tttgttattg tagaaaaggg tgaaaaaata gtaaatttaa ttatttttag 480

tttgaaaaat tatttagaaa tgaagatgac gattttgaaa tattgttaat attatttgat 540

ttataaataa tgttttaata tatttattat atattgatag atatttttt atatgaatat 600

tatatattaa aattaaggta ataatgtatt tagaatattt tatttatatt tatgtatttt 660

aagtaggtta gaaattaaga tatgagttat taagtatgag atgttaaggt gtggggttag 720

aaattatatt gtattttatt attaataatt aatatatatt ttaatattat atatatttaa 780

ttttaatttg tatattttta attattttta attatgtgta taaatataag tatatatatt 840

tttatgtatt tatttattta tatttttatt tatttattta tataggggat ttttttaaat 900

ttattattat taaattatat atttttattt taatttttag aataagttta ggaggtaggt 960

attgttatta tttatatttt ataaatgagg aaattgttta tagttataaa gttattgtgt 1020

tagatatatt agaagtttaa tatatatttg gtgaatatat gtataaaaat agagagatag 1080

atatgtataa tagtttattt ttatattgag taaaagtttt taatttgttt tagaaatttt 1140

tttgtgaaaa ttgagtaaaa atcgaggtat ttttttattt gttatatagg tataggtggt 1200

-continued

atttttatttt tttaataagg atgaatattg aaatgtggat tttaaggttt aattttagat 1260

tttttgaatt tttgatagtg ggatttggaa tttgtttatt gttttaaagt tttttaagga 1320

atttatatga ttaattaggt ttagaaatta ttggatttta ttgtcgaagt ttgagaatta 1380

aagtttgggt tttattgcgg ttttatagaa agggtaaatg aagtattatg gatagaattg 1440

atacgttttt agttagtttt ttttttaga agttaatagg tagtaatata gtagaaatta 1500

gtgatttatg ttttgtgttt tgaagttagg tagaatttta tagagtttta gtagtgttat 1560

tgacgagatt tgtttttgg ggtaagttgt ttgatgtttt taaagttata tttttttat 1620

ataaaatgag ataatatttt ttgttttata ggggtgtttt aaagattaaa taaaaataat 1680

atgttttatt ttatatggta taatgtttga tatttaagaa gtaaaggata tattttattt 1740

ttattgaagt aattagaaag tatgaaatta tgaaggagat aagagttttg attggtagtg 1800

tattttattt ttttaggttt atttatttat tttaaattat ttttgttgga gaataatttt 1860

taagtttttt atttaagttg tgagtaattt tatattttat aatgatgttt tttttatgag 1920

aaaaaaaaat gtttttaagt tttttggaga aaatatattt gtattatttt tattgaaaaa 1980

tttaataatt ggattttgtt tttttgtatt aattttagag tgtatatgtt ataaataaag 2040

tgttttagtt taagaagatt gaaagtaaat atggtatagt attttaaaat aagaattttg 2100

taaatatatg gtatgattgt gttatattat tagtaattat atgatacgta atgtaaagta 2160

tagtttatag atttaaattt aattttaata agtaaattga ttttgttttg ttggggaaaa 2220

gttaaagtat taatttaatt gttaatgtag ttttgtttat ttttttggta tttagtgata 2280

agtttaaata atgtatatat ttttatttat atatttagta atataatttt ttgtttaatg 2340

agtgatgttt ttttgttatt tggtggtgtt tgttagtttt agaatttgtt ttttggtggt 2400

attataatat taagtataga gtaagtgtaa taaaattgta gtatttttat tgaaaaggtt 2460

ttgttttaaa ttgtttaata atttaaagga tttttgtgga agtaatcgta tttgttaatt 2520

agttataatt agtaattaat ttttttggag ttttaattta tttttggtaa aacgttttag 2580

gaagagtata tattattaga aagtatgtta aaaatttatt tagtagaaaa tttaaaaata 2640

gttttttttt gttaagaggt tttttaaaat tttatttata tagttaaatt ttgaaatttt 2700

agtaggtttt gttttattat tataattatt gtataaatat ttttaaggat tttgttttta 2760

gttttaagta tgatttattt ttataagttt gattagttat tatattagtt ttgttatgga 2820

aaatgatatg tttttatttt ttgttgtaga gttgttaaat tttgatttat atttatgttg 2880

tttttttttgt tgaaagtttg tagcgaaaga aatttttaat tttttgtttt gtaatattag 2940

ttggtagttt tatttaatgg gtattttgtt tttttaaaga atttagttgt tttgtttaga 3000

agtcgatttt ttgatgtttt taacgtttgg tttaattgat ttgttttaat ggagttttcg 3060

tcggtgagga gcgagatgtt atcgattaga atgttgggat ttgttgttta attgttagga 3120

gtgagagata ttgagattta gaaatttttg gaggtgggag gggagaggga tagtttcgga 3180

cggaggcgga gatgtaagat aaagggatgg attttatata ggaaaaaaaa aaagatttcg 3240

ttgaggtatt gaggtgttgt acgattatat tttttaaagg agaagttaaa aagtaaggaa 3300

gtgggaggag gttggaggtt aaagtattta aaaggattat tcgggtataa tttgtttttt 3360

tgttggtgtt tgtaaaggat agatagtttc gtttttaaag tatatgaatg tttttttaa 3420

gtgattggga atggatatta attgtttgtt aaatgttatt aaatgttttt ttaaatttag 3480

-continued

gggatataga aagaggggta taaaaggaga atttaaatag aaaaagggag gattcggagg 3540

tttttgaaag cgggggggaga agaaggagga gggataatag agaggaatag agaaggagag 3600

cggagagaag ataaataaaa ataaaaatag gaattattga ataattatat attaaaaaga 3660

aagttttttt ttatggggta tttaaaatat tgagattgta atagtgattt cggttatgga 3720

agaaagatgt tttttttat ttttgttttc gaaagttttt ggtttcgtta ttggcgatta 3780

aaattttatt aggttaaaga gtgtgtttaa ttgtttgaag aatgtagtag acggaaggcg 3840

ggtttcgtta tgtcgtttgt ttttttcgtt ggagagaatg aaagaaacgc gtagagttag 3900

agatttttgt cgagttagat tttttttcgt cgttttaggt tatcggttat tcggtaaaga 3960

ttcgagtaag gaacgtaggg ttattgtttg ggttaataaa tggagttcgt ttttttttt 4020

tcggacgtcg ttgttcggtc gatgttttcg gtaatttatt cgcggcgtat gtagaggagt 4080

ttttttttt tttttagatt atttgtttcg attaatttga ttttttaaat atatttgatc 4140

gtatttttta ggtggatata ttaataggtt acgggttgga gaggagcggg tgatgaggag 4200

agggatttaa atttgcgaac gtttgggttg ggtcggagtt gcgggggggtt tgggaggaga 4260

gaggggagaa gagagaagga aggagagcgt ttgtcgggat ggttgagttg tttcggcgag 4320

tagttttggg gttgtacgtt tttgtgggag atgttgttgt tgtttttagg tcggtaagag 4380

cggttttaat attatcgttt tttatttttt tttttgtaaa tttttagaga aacgtttttg 4440

gtttttttcgt cgcgatattt ttagtttgta tttttttata gtttaggcgg cgcgttttcg 4500

tacgttggag cgtcggtcgt tagtaggacg tttttttttcg cgtcgattcg ttttttttttg 4560

ttttgttgtt gttgtttttt tgatattttc gtttttatta tttttagttc ggagagacgt 4620

tatttagtcg cggttcgtat tcgcggttcg gggttacgcg cggaagaggg gcgttagttc 4680

ggatttcgtt ttcggtaggg ggcgttttgg agcggagagt gaggcgaatg gtatatgagt 4740

gtgcgggtag tttattttga agttcgagtt ttttatttga gttatgtttc gtttagtttt 4800

attcgggtta gcgtttggcg agcgagttta tttgtggttt tcgcggtcgt ttttttttg 4860

tatttttgta tttattcgtc gatttttttt tttcgggatt tgtattttgt tttattaatt 4920

agagttcgat tgtttttttt tacgtgattt cgggcgggtt gaggatttgt tgtttttaa 4980

acgttagagg gatgcgggcg gtagagttcg agaggcggtt gtcgggttgc ggggcgtttt 5040

gatttttttt ttattttgtt ttttcgggtt ttattcgttt gtttttggat tttcgttttt 5100

ttttgttttt cggtttttta gagttttttt tttatggtag tagtttttcg cgttttcggc 5160

gtagtttttt agcggacgat tttttcgttt cggggttgag tttagttttt ggatgttgtt 5220

gaaattttcg agattatgcg cgggtttggt tgttgttttt tcgtcgggtg ttattgttat 5280

cgtcgtcgtt tttgttgtcg tcgttcgcgg gatgtttagt agttcgttgt tcggttttcg 5340

cgattttgtg tttttcggaa gtcgtttgtt gttgtagagt tgtacgaatt agttatggtg 5400

ttgtgggagt tttcgcggta gtgtagtagt tggatatttt gcgagggttt ttgttggttg 5460

ttgttgttgt tcgttatgtt atttatcgta gttcgttcgg tgaagttcgt tgtttttttt 5520

atttttttaa gtgattgtta aacgtttatc ggttggaatt gttttggtaa gtttagaatt 5580

ttcgttttcg atttttttaat ttcgtagaag aatacgcgta tttagtatag attagtttat 5640

tttagcgcgt ttttttagtt ttttatttttt tattgtttta gatttttaat attatttatt 5700

tttatttaga gaaataaggg gaattgttgt aggttcgggg gtgaggggtg gttttgggat 5760

-continued

gggtagaaag tgtaggtgta gtaggaaatt tttgtatgtt tgcgtttata ttggagttgc 5820

gaggattttg agaaatatta aacgggatgg ttttttgggt ttattgtttt gaaagagtat 5880

taattttagg ggaaatattg aaatagaagt tttgttatta ttaaagaaaa aagttttatt 5940

aggatgagga agaaataatt ttatgagaaa gaatgagcga gaaagtaata aattaaatgg 6000

tgattgtagg ggaatcgttg atttttggta aaggtgttat gaggtcgtat tggtttttcg 6060

ttgaagatta ggttatatag attttagagg agttgggttt taatagaatt ttttttttt 6120

ttttttttt ttttttttt ttttttttt ttttttatt tatttatttt ttttttttt 6180

ttatttttt ttttttagg cggtaaaaga tattggtttt gtagtttaga tatgtttttt 6240

tttttgtttt tttaagtttt aaggtagtat aggggagttg agaaaaagaa tattttgcgg 6300

gttttttagg tcggagtggg tatgattgag gttggttagg ttttatgtag gcgagtcgag 6360

ggcggaatcg attttagtgg gcgttgattt ttttatttt ggataggttt ttgtggagtg 6420

ggttaggtat tttttttgtt cgttcgggtt tttttagatt ttgacggcga acgtttggta 6480

ggtttcgttt tgttgaagtt ttttaattaa atagggttag aggatgggag ttgttgtatt 6540

tttagttggt atagtattcg gtttgatagt ttgtagtata gggtgtatgt aatttttat 6600

tttttgtgaa tataattttg ttgtagttaa atttggtttt gaataaagtg ttttttaaag 6660

atgtatataa gttgaagtgt atgtaatttt agagaggagg gaatgattaa ttgtaattta 6720

gggtgaaagt ttgtatagtt tttagttatt attgatgtaa atgttaaaag gaaaattatt 6780

atgtattatt ttaatttatt ttttataaag ataagttgag atatgtaatt ttattagatt 6840

tgggttaata gattgttttt ttttttggta gtttttaaat ttggtatttt aataaaattt 6900

aatatgtttt tataattttt tgatttatgc gtatatgtgt gttgtttttg aaagaataag 6960

ttttattttg ttattgttta attatttttt agatgtttta ttatggtaat aattatgagt 7020

ttgtaaaaat aatttttgga aatgttgatg gttttgtagt ttaatataga ttggtttgtt 7080

ttatttttag tttttgtatt gttttaggaa ataattaatt taaatgtgaa gttgatattt 7140

gtaattaaga aattatatat ttattagata ttttaaaggg gattgtataa attaaagaga 7200

ataaattggt tttgtagata ggttgttaag aatttggtat ttcgttttta tttttgttaa 7260

tttagaggtg attaattttt atttgagtta aatagattat tatagaaaat attgtgtttg 7320

tttatttta ttattgaggt tttgtttttt ttttgtttgg atatatttta aataaggggt 7380

tgttttagtc gttgaagtaa aagaataatt aaagatgggg aaatggtaaa agggtattta 7440

gagattatta ttagtttttt tttaaaatgt ggagttttgt ggttataaat attgtttatt 7500

taatgagtaa aaaataaaaa taaaaaaaaa ataggaagta aatgttaagt ttttatttat 7560

tattgttagt attaacgtaa gttttaaaaa atagtattat tagaaaagga tattaaagga 7620

gaattgatta gaaaagaatt gtggaaaatg gaaacgaata ttgattattt aattagattt 7680

tgaggttatt agtagatagt gattttgtag tatagttata gttgttggat ttaaaattta 7740

ggataagtat tttaaagttt taaagtagtg ttttttttttg ttaaaaattt gtaagatgtt 7800

ttaatgattg gagtgttttt tttgaatttg agg 7833

<210> SEQ ID NO 15
<211> LENGTH: 7833
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 15

ttttaaattt aaagagaata ttttagttat taaaatattt tatagatttt taataaaaaa 60

aagtattatt ttgaagtttt aaaatatttg ttttaaattt taaatttaat aattatagtt 120

gtattgtaag gttattgttt attgataatt ttaaaattta gttaagtgat taatattcgt 180

ttttattttt tataattttt ttttagttaa ttttttttta gtatttttt ttgatagtgt 240

tatttttaa agtttgcgtt aatattgata gtggtgaatg aaagtttaat atttgttttt 300

tgttttttt ttatttttat tttttgttta ttaggtggat aatatttatg attataaaat 360

tttatatttt ggaaaagagt tagtgatgat ttttgaatat ttttttatta tttttttatt 420

tttaattgtt tttttgtttt aacgattgaa ataatttttt atttgaaatg tatttagata 480

aagaggaaat aaagttttaa taataaagat aaataggtat agtgtttttt gtgatggttt 540

gtttggttta aatgaagatt gattattttt aagttaatag gggtggaagc ggggtgttaa 600

gtttttgata atttatttgt aaaattagtt tatttttttt agtttatgta gttttttta 660

aaatatttgg taaatatgta atttttgat tgtaaatgtt aattttatat ttaagttagt 720

tatttttaa aataatgtaa gggttaggaa tgaagtaaat tagtttgtgt tggattataa 780

agttattaat atttttaaaa attgtttttg taggtttata attattatta taataaagta 840

tttaaaaagt gattaggtaa tagtaaagtg aaatttattt ttttaaaaat aatatatatg 900

tacgtatgaa ttaagaagtt atagaaatat gttgagtttt attaaaatgt taaatttaga 960

aattgttaaa aaagagaata atttattgat ttaaatttaa tagggttgta tattttaatt 1020

tgtttttgta aaggataaat tagaatgatg tataataatt tttttttgg tatttatatt 1080

agtaataatt aggaattata taggttttta ttttgagtta tagttggtta ttttttttt 1140

tttaaagtta tatatatttt agtttatata tatttttgaa agatatttta tttagagtta 1200

gatttaatta tagtaaaatt atatttatag aagatgaaaa attatatata ttttatatta 1260

taggttgtta aatcgaatgt tatgttagtt aggagtgtag taatttttat tttttggttt 1320

tatttaatta ggaagtttta gtagagcgaa gtttgttaag cgttcgtcgt tagaatttga 1380

aggaattcga gcgagtaaga agagtgtttg atttatttta tagaagtttg tttagaaatg 1440

gaggagttag cgtttattga agtcggtttc gttttcggtt cgtttatatg gagtttgatt 1500

agttttagtt atgtttattt cggtttggga gattcgtaaa gtgtttttt ttttaatttt 1560

tttgtattat tttgaagttt agggaagtaa agagaggggt atatttggat tgtaaaatta 1620

atgtttttg tcgtttagga gagaagggaa tgagagagag agagagatag atagatagag 1680

agagagagag agagagagag agagagagag agagagagag agaaatttta ttgaaattta 1740

gttttttag aatttgtgtg atttggtttt taacgggaga ttagtgcgat tttatggtat 1800

ttttgttagg aattagcgat ttttttgtag ttattatttg atttattgtt ttttcgttta 1860

ttttttttta taaagttatt ttttttttat tttagtaaga ttttttttt taatgatgat 1920

aaagtttttg ttttagtgtt ttttttagga ttggtgtttt tttaaaatag tgaatttaga 1980

aaattatttc gtttaatatt ttttaaaatt ttcgtagttt taatgtaagc gtaagtatgt 2040

aaaggttttt tgttatattt gtatttttg tttattttag aattattttt tattttcggg 2100

tttgtaatag ttttttttgt ttttttggat agaggtgggt ggtattaggg gtttagggta 2160

-continued

gtaggaggtg aggggttgag gaggcgcgtt agggtaggtt ggtttgtgtt ggatacgcgt 2220

gtttttttgc ggagttaaag ggtcggggac gggggttttg gatttattag agtaatttta 2280

gtcggtgggc gtttggtagt tatttaagga ggtagggaaa gtagcgagtt ttatcgggcg 2340

ggttacgatg agtagtatga cgggtagtag tagtagttag taaaagtttt cgtaaagtgt 2400

ttagttgttg tattgtcgcg gggatttta tagtattatg attagttcgt gtaattttgt 2460

agtagtaaac ggttttcgag gaatatagga tcgcgggggt cgggtagcgg gttattgagt 2520

atttcgcgga cggcggtagt agaggcggcg gcggtggtag tggtattcgg cggggaagta 2580

gtagttaaat tcgcgtatga tttcgagagt tttagtaata tttagggatt gggtttagtt 2640

tcggagcgag agggtcgttc gttgagaagt tgcgtcggag acgcgggaag ttgttgttat 2700

aaggagggag ttttgggaag tcggaggata ggaggagacg ggagtttagg ggtagacgag 2760

tggagttcga ggaggtaggg tggagggaga gttaaggcgt ttcgtagttc ggtagtcgtt 2820

tttcgagttt tgtcgttcgt attttttttgg cgtttgggaa gtagtaggtt tttagttcgt 2880

tcggggttac gtgggaagag gtagtcgggt tttgattggt ggagtaggat gtaggtttcg 2940

ggagggaggg gtcgacgagt aggtgtaagg atgtaaggag gaggcggtcg cggaagttat 3000

agatgggttc gttcgttagg cgttggttcg agtggggtta ggcggggtat ggtttaaatg 3060

agaagttcgg gttttagggt gggttattcg tatatttata tattattcgt tttatttttc 3120

gttttaggac gttttttatc gaaggcgggg ttcggattag cgtttttttt tcgcgcgtga 3180

tttcgggtcg cgagtgcggg tcgcggttgg gtggcgtttt ttcgagttgg agatggtggg 3240

ggcggaggtg ttagaggagt agtagtagta gggtagagag gggcgagtcg gcgcgggaga 3300

gggcgttttg ttggcgatcg gcgttttagc gtgcgggagc gcgtcgttta ggttgtaggg 3360

ggatgtaggt tgggaatgtc gcggcggaga ggttagggac gtttttttag ggatttatag 3420

gaaagagggt gagaggcgat ggtgttagaa tcgtttttgt cgatttggaa gtaatagtag 3480

tatttttat aagagcgtgt aattttaagg ttgttcgtcg aggtagttta gttatttcgg 3540

taggcgtttt ttttttttt ttttttttt ttttttttt ttaggttttt cgtagtttcg 3600

atttagttta agcgttcgta ggtttgaatt ttttttttta ttattcgttt ttttttagtt 3660

cgtagtttat tagtgtgttt atttgggagg tgcggttaga tgtgtttgga aggttagatt 3720

ggtcgggata agtggtttga gagaaagaga aaggtttttt tgtatacgtc gcgggtgggt 3780

tgtcgggagt atcggtcggg tagcggcgtt cgggaagggg agagcgggtt ttatttgttg 3840

gtttaggtag tgattttgcg ttttttattc gggtttttgt cggatggtcg gtgatttggg 3900

gcgacgagag aaggtttaat tcggtaggag tttttggttt tgcgcgtttt ttttattttt 3960

tttagcggga agggtaaacg gtatagcggg attcgttttt cgtttgttgt attttttagg 4020

tagttagata tatttttag tttaatggaa ttttagtcgt tagtaacggg attaagagtt 4080

ttcggggata agggtggaga ggaatatttt ttttttatga tcggggttat tattgtagtt 4140

ttagtgtttt ggatgtttta tagggaagag ttttttttgt ggtgtgtgat tatttagtga 4200

tttttgtttt tgtttttgtt tatttttttt tcgttttttt tttttatttt tttttgttat 4260

ttttttttgt ttttttttgt tcgttttaa aagttttcgg atttttttgt tttttattta 4320

aatttttttt ttgtgttttt ttttttgtgt ttttgaatt taggagagta tttgataata 4380

tttaataggt aattagtgtt tatttttaat tatttaaaag aggtatttat atattttgaa 4440

-continued

aacgggatta tttatttttt gtagatatta gtagaaaaat aaattgtatt cgagtaattt 4500

ttttaagtat tttaattttt aatttttttt tatttttttg ttttttaatt tttttttga 4560

gagatgtgat cgtgtagtat tttagtgttt taacgaaatt ttttttttt ttttgtgtga 4620

aatttatttt tttattttat attttcgttt tcgttcgaga ttgttttttt tttttttat 4680

ttttaaagat ttttgaattt tagtgttttt tatttttggt aattaagtag tagattttag 4740

tattttagtc ggtggtattt cgttttttat cgacgaagat tttattaaaa tagattaatt 4800

agattagacg ttggaggtat tagaaaatcg gtttttagat agagtagtta aatttttaa 4860

ggaaatagaa tatttattag atagagttgt taattaatat tgtaaaataa ggaattagaa 4920

atttttttcg ttataggttt ttagtagaga aggtaatata aatatagatt aagatttaat 4980

aattttatag tagagaatga gaatatgtta ttttttatag taaggttggt gtggtaatta 5040

attaggttta tgaaaataag ttatgtttga aattaaaggt aaagttttta aaagtgttta 5100

tgtagtaatt atgataatga aataggattt gttaggattt tagagtttgg ttatgtaagt 5160

agaattttag agaatttttt agtagaggaa aattgttttt gaatttttg ttaagtaaat 5220

ttttggtata ttttttaata atatatgttt ttttaagac gttttgttaa aagtaagtta 5280

aaattttaaa ggagttaatt attggttgta attggttaat aaatgcggtt gtttttatag 5340

aggtttttta aattattaaa tagtttgaag taaagttttt ttaatgggaa tgttgtaatt 5400

ttgttgtatt tattttgtat ttagtgttat agtgttatta agaaataaat tttgaaattg 5460

gtaagtatta ttaagtggta gaagaatatt atttattgag tagagaattg tattattgaa 5520

tatgtaaata aaaatatata tattatttag atttgttatt aggtattaaa gaagtagata 5580

agattgtatt agtaattgga ttagtgtttt aatttttttt tagtaaggta aaattagttt 5640

atttattaga attaaattta agtttatgaa ttgtattttg tattgcgtat tatatgattg 5700

ttagtaatat gatataatta tattatgtat ttgtaaaatt tttattttaa aatattatat 5760

tatatttatt tttaattttt ttgagttaga atattttatt tgtggtatat atattttaga 5820

attgatgtag aggagtagag tttagttgtt agatttttta gtagaaatag tgtagatata 5880

ttttttttag aaaatttaag aatatttttt ttttttatgg aaagaatatt attataaagt 5940

gtgagattat ttatagttta agtagggggt ttgggagtta ttttttaata agaatagttt 6000

aagataaata aatgaatttg ggaaaataag atatattgtt aattagaatt tttatttttt 6060

ttatgatttt atatttttttg attgttttaa taaaggtaag atgtattttt tgtttttttag 6120

gtgttaggta ttgtgttatg taggatagaa tatgttattt ttatttaatt tttaaaatat 6180

ttttatgaga taaagaatat tattttattt tatataaaag gaatatggtt ttgaaagtat 6240

taggtaattt gttttaagaa ataaatttcg ttagtgatat tgttgggatt ttgtgaaatt 6300

ttgtttgatt ttagagtata agatataagt tattaatttt tgttgtattg ttgtttgtta 6360

gtttttgaga ggggaaatta attgggaacg tattagtttt gtttatgata ttttatttgt 6420

ttttttttgtg gagtcgtagt aaggtttaaa ttttaatttt taaatttcgg taataagatt 6480

tagtgatttt tgaatttggt tgattatatg aatttttttga gaaattttga aataatagat 6540

aaattttaag ttttattatt agggatttag aaaatttgga gttgggtttt gggatttata 6600

ttttaatatt tatttttgtt ggagaagtaa ggtattattt atatttatat gataaatgaa 6660

aggatatttc gatttttgtt tagtttttat agagaggttt ttgagatagg ttaaaagttt 6720

-continued

ttatttagtg taaagatgag ttgttgtata tgtttgtttt tttgttttta tgtatatgtt 6780

tattaaatat gtattaagtt tttaatatgt ttgatatagt aattttgtga ttgtagataa 6840

tttttttatt tgtaaaatgt gagtaataat aatatttgtt ttttgggttt gttttaaaga 6900

ttaaaataaa aatgtatggt ttaatggtag tggatttggg gggatttttt atataaataa 6960

gtgaatggag gtatgaataa ataaatatat aaagatgtgt gtatttatat ttatatatat 7020

aattaaaaat agttaaagat gtataaatta aagttaaatg tatgtgatat tgaagtatat 7080

gttgattatt gataatgaag tatagtataa tttttaattt tatattttaa tattttatat 7140

ttaataattt atattttaat ttttagttta tttaagatat atagatatag atagaatgtt 7200

ttaaatgtat tattgtttta gttttaatgt ataatattta tatgaaaaag tatttattag 7260

tgtgtagtaa atgtattaga atattattta taggttaaat gatattgata atgttttaga 7320

gtcgttattt ttatttttgg ataatttttt aaattgagag taattaaatt tgttattttt 7380

ttatttttt ttataatggt aaataatata taaataatga gttgtgattt aaagaaataa 7440

ttttataatg taaaagtttt tttatgtttt attgattttt atgtttttaa ttaattttgg 7500

tagtttaagt ttgtgatttt agtgttgagg aaagtttatt ttattttaga gtagtggggt 7560

ttagttatga ttttatatta taattatttg gataattaaa gaatattgat gtagagttcg 7620

tatttaaaga tttggaattt ttagaagtga tatagaagta ttggtatata tatattttta 7680

aagttttttt ggagaaatat atagttatga ttgagaatta ttgttttggg gaaagtgatt 7740

atttttttat tatttaaata gttaagtttt aggaggtaaa atatttatat ttttttttat 7800

taaaatttgt aaaatatata ttttattaaa gat 7833

<210> SEQ ID NO 16
<211> LENGTH: 965
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 16

gattaattaa gggtatttta gaagttaggt gtttttgttg ttttttttga gtatggaggt 60

tattaatttt taggggggaag agatgtagtg tgaggtaggg gtgttgtgtt aagaaatttt 120

gatgtttttg gggattgagg ataaaggtgt ggatatgatt ttggggtatt tggagttttg 180

tgatttgtgt tatggatggt atatttaggg gttaattttt gttttgtttt aaagaatttt 240

aagttagagt ttttgttttt gtttatagtt ttgggatgtt gttgttgtgt ttattgtata 300

ggtagtgttt ggattggttg tagtagattg tgtgttgtgt gttttattgg gagatggtgg 360

agatgttgaa aagttttttt tttgttattt tggatgttgt gggtggtaag tgttttagtt 420

tttatttttg ttgagttgaa tgtttaggta tagtggaatt gaaatttggt tttttatttt 480

tgttgagttg aatgtttagg tatagtggaa ttgaaatttg gttttgtggg atgtgagagt 540

tgttgaggtt atgtgtaatt gggtgtgatg gagggtgttt gtttgtgatg tgtgtaggtt 600

tgatgtaagt aggttattgt tgtgtgagtg tgtggatgtg attgtttgag agatttggag 660

gtaggtttgg gatatgtttg agtgaatatt ttaggatatt tttttggtta gtatttgttt 720

tttagtgttt gtgatttaga gtgggtatat gttgggagat agtaatgggt ttgggtgtgt 780

gtaaatgagt gtgattggaa gtgagtgtga gtttgattta ggtagggatt atatagtatt 840

-continued

gttatatttg tttgtttttt agtagaggat tgaagtgtgg gggtgggggt atggggttgg 900

aatagaatgt ttttgggata ttttggtaaa tagtagttgg aagtaaaggg gtagttgtgt 960

aaatg 965

<210> SEQ ID NO 17
<211> LENGTH: 965
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 17

tgtttgtata gttgtttttt tgtttttggt tgttgtttgt taagatgttt tagagatatt 60

ttattttggt tttgtatttt tatttttgta ttttagtttt ttattaaaga gtaggtaggt 120

gtgatagtgt tgtgtggttt ttgtttagat taagtttata tttgtttttg gttatattta 180

tttatatata tttaaattta ttattgtttt ttaatatgtg tttattttga attatagata 240

ttaaaaaata gatattggtt agaagagtat tttgaggtgt ttatttaaat gtgttttaag 300

tttgtttttg agtttttgg gtggttgtat ttatatattt gtatgatgat gatttgtttg 360

tattaaattt gtatatatta tgaataggtg ttttttatta tatttaatta tgtgtgattt 420

taatagtttt tatattttgt agaattgggt tttagtttta ttgtgtttga gtgtttagtt 480

tagtagaggt agggaattgg gttttagttt tattgtgttt gagtgtttag tttagtagag 540

gtagggatta aggtgtttgt tgtttatagt gtttagagtg gtaagaaaga agtttttag 600

tgtttttatt atttttggt ggaatgtgta gtgtgtgatt tgttgtagtt ggtttgggtg 660

ttgtttgtgt ggtgagtgta gtagtggtat tttggggttg tgggtggagg taaggatttt 720

agtttgaggt tttttgaggt agagtagaaa ttagttttta ggtgtgttgt ttgtggtgtg 780

agttatggaa ttttaggtat tttggggttg tgtttgtatt tttgttttta gtttttagaa 840

gtgttgaaat tttttagtat gatatttttg ttttgtgtta tatttttttt ttttaggggt 900

tggtggtttt tgtgtttaaa ggaggtagtg ggaatgttta atttttaaga tgtttttaat 960

tgatt 965

<210> SEQ ID NO 18
<211> LENGTH: 16579
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 18

tgtttttttaa aatgatgttt agttgaatta gttaatggag ataaaaatat ttatatataa 60

ttttttttta agtttattag ttattttata gattagttat aattttggtt attttaatga 120

gttaattttt ttttagagat attataaatt ttttaaagat ggaaaataat tgttattttt 180

ttgaataatg tttattattt aaaataatgt atataaaagt tgtttaataa atttttgttg 240

aattgagaga aaatataaaa aattgaaagt agtttgtttg taataagtgt ttgttaagtt 300

ttatagattt ttttttttt ttggaattat tgtaataaat atttgatttt tgatttgtag 360

gattaaaatg tgatttttta tattttaata ggagtaagta ttagttttgt aatttttttt 420

gttaatagaa taaaaatgga atgtatgttt gttgttttta tatgttggtt gatatttatg 480

-continued

ttgaatattt tttttataat gtagaggttt ttttgttttt tgttttgtgt tttatttttg 540

tttattatag agatattaat gatgttatgt gttttttttgt attattagtg tagtagtttt 600

tattatgttt tttgttttgg ttttattgta gtttaattaa ataaatggaa atgggtagtt 660

attgaaaata aatagtaatt gtaaaagttt tgtttttttaa gaaatttttt gtattgtatt 720

ttgttttgtt ttattgattt taagattttt attgtttttt tgtgtttaga aattagattt 780

tttagtttgt tttgttttag ttagataaat gttttataaa aatgatttat atttttttaga 840

ttgtggggag agaagaaaaa gattttgaga aagtgattaa ggagggagta agagtgatat 900

aatttttgat ttaatttagt tggattgtgt tttttatgtt tttatttttt tttttataaa 960

aaagatagtg tttttgatgt tttgttttag gtagtttttt agggtgagtg tagatgtgtt 1020

gagttatttg gttgttttga ggattggttt agagagtttt tttttttatt aatgaagtta 1080

ttttttttttg gtagtttggt tttttttatt attttatgtt ggaatttggt tttggtttgt 1140

atagttttgt tgtgatgtag atggtttttt agttaagttt tagagaaatt tgtttaatta 1200

tgaataggat atttatagta taggagaaaa aaaaatttat taaaataaat attttattgt 1260

aatattttga tttgggggtt ggagagtaaa aatttaatat ttgatgtttg ggaaaaattt 1320

aatgttttat ggtttttaa ggaaatttgt ttgtgtttaa tttattagag taatttatat 1380

attttgggtt aaagtataat gagattgtta ttattatttt gttttttgta agatttttt 1440

tttaagaata aaaagaaggt aataaagatt ttatattttt gttttttta gaaagttaaa 1500

gagaatttaa agtgttggat gaaatgtttg tttttaattt gtagtgtttt atttaaaaaa 1560

aaaattatga tataatttta gaggataggt ttttttttt tagggttttt gaaagttaaa 1620

agtatatttt taatttgtag tattattgta ttatttagtg aggataattt gtttagttgt 1680

gtttaggatt agttagatta aggagttgaa attaagtttt ttttgaattt gaatgggatt 1740

ggaattttgt tatttttaga gtagtaatag atgatgttat gtttaagaat attttatttg 1800

attaataaaa aaggagaagt ttgaaatgta attttaaaga attttagaag gaggtttaga 1860

tgaagataat aaaatttttt ggagatatat ttgttattta tatttttttt ttttaagtta 1920

ttggtttgtt gtttttgagt ttagttaaat ttttaaagat aagattaaag taaggaagta 1980

aggtatatag tttttggttt atggtttttt ttagatataa gtattgtgaa tgaagatttt 2040

gttttattta tatttgggat tttggagttt ttataaagaa tatttttagt ataaggatgt 2100

gatttaaagg gggatttata gaataatatt ttaattattg tttgatttat atttgagtgt 2160

aatggataaa taaatataat tttagttata aaatttgaaa aatatatatt tttttattta 2220

aaaatttttg atagttttgt taggtattgg agaattagaa ataaatgttt ttttatttttt 2280

ttttttttga tgaaggaaag ttattaaaaa tagttatata gaaataataa tataattgta 2340

gattttattt taaatatagt tggtatgttt ttttttttt tttttattta ttttttattt 2400

tagatagtta gaggggaaag tttgttttag agatagaaag aagatattta tggggtataa 2460

ttttatttta gttatttgat agtttgaaga agtatttaat tttttgaata ataatgttat 2520

aagttaaaga agagaaaaga taggtagttt aaagtttttt aatattgtgt atattttaat 2580

tttttaagtt agttttgaga aagtgttttt gttttgtaga agtatttggt taagtgttgt 2640

gaagtttatt ggtattaaaa gaaattaaaa attttggagg attagagaat taaatttggt 2700

aatataaatg aaaatatttt agaatgtaaa aagtatttaa tttattattt ttaattatgt 2760

-continued

tttattatat atagatggta ttatttatag tgaatgttaa ttattaatat ttttttaaaa 2820

atgggtatta atttgttatt agttgtttgt ttttgattta gtttttttt gttgaaaggt 2880

tttatttaat tgtattataa agggataaaa gggtaaaaga taataagttt tttagattgg 2940

tagatttgtt atttataaaa tggttttggg taggttattt gaatttttg aaatttaatt 3000

agatttattt aaatttaaga atgagaatat aaatttatat tttgaagtgt tttatagaaa 3060

ggtttatttt aatgtttgga gtatatattt taatgaatat ttattttatt ttatttttt 3120

ttatttttga attaagtaat tttgaattta aagttgttat gattagtatt gaaaagatta 3180

ttggattatt aattgtgtga ttttgggata gtaatttttt gtattttagt ttgtttatat 3240

gttatatatg aaggttgaag tttgattttg ttttgtgatt attattttaa atatttgatg 3300

aaattaaatt ttagtgtttg gaatggtagt ataataaatt tattaagaat aaataattta 3360

ttgtaaaaat atattgattt ttaaatgatg taattgatag ttatattatt gtagagggtt 3420

gataaataat aaaagaaatg aaagatgtat atggtgagaa ttgaaattat tttgataagt 3480

tttttatttg tttattattt aaaattaatg attatgttga atgtttataa attataaaat 3540

ataaaagaaa ttttataaat gtgtatgtat aggagtttaa gttattaaaa gttttaaagt 3600

ataagtttaa attaaattaa ttaaagaagt tgagaggaaa aattggtttt tatttttaat 3660

tattattgtt ttgaggtttt atgtttaata taatttttta agtagaggtt ttagagagaa 3720

gagttgtgag gatattttta tatttgtgta gaaggaaaag tttgttattt attttagtat 3780

ttttagtgtt atattgatgt gtattttgga tttattttgt ttttattgta taaatttata 3840

tttgatttta aagaaaagga aaatttaaag ttttttttt ttaaggggat agaaattttt 3900

tgtgttaatt gtttgatttt tttttttgta aggttttatt ggaaatttt tgtaatataa 3960

tgtaggggat ttttttatgt gttgatgttg tttatatagt ggggtgggtt tgattgaaga 4020

aaaaaaattg tatatatgta tgaaagatta tggttttatt tttggaaagt atgaaaggtg 4080

attgatattt ttaagaagtt tttgttattt aggaaaatta ttaaatattt tatttagaga 4140

tatttggaaa gattgaagga aaggaagaat gaagaaagta gaatttagat ttatgtgggg 4200

agagatttgt ggtagaggaa aagtattttt tttgaatttg ataagggatt tgtttggggg 4260

aatttttttgt ttagtttttt attattaggg ttttttgaag ttgggttttt tattgggtag 4320

ttttttggga gtgtagtggg gaatttttat attttttttt taggtttttg aaggattttg 4380

tttttttagt gttttttttа ggttggtagg agttttgagt ttgatatttt ttttgatgg 4440

gataggtaag ttttgtgggt gtgtaaatat gttgtaatta agtttttttgt tgattttata 4500

gttttgtgtg tttttgagaa gaagtgattg tatttaattg tttattgttg gtttgttttt 4560

taagagtttg ggggtttttt ttttttaatt tagaattagt tgtatggggg gtggggaagt 4620

gggggtgggg aaggagtggg agggtagtgg tttttgtgag tagagtgatg ttattgagtg 4680

agtttttgaa tggggagtgt tgttgttttt aagttgattg gtatttttttg ttaggaagaa 4740

atgttaagag gtgggagtgt ttggggaggg aggtaggtgg tttttattgt aggtgtgggg 4800

agttgttttt ttgtttttttt gtttgttttt taagtttgga tttttaggag tggttgaagt 4860

tgtggagtgt ttttggagtt tgtgaatgaa ttttttttttt ttttttttttt ttttttttтg 4920

ttgagttttt tttttggttt tgatggtata gtgatataat gatgatgggt gttataattt 4980

gtatttgaat ttgtaggtga gttgttttga gttttttttgg ggaagaattt taggtgtgtg 5040

-continued

gatgtaatag ttgagaatat taggtgttgt ggataggagt tgggattaag atttttggtt 5100

agttttgtat ttttttgtat ttttagtat tgttttgtat tttttgtatt tttttttggg 5160

ttattatgtt ttttatgtga tttgtttggg taatgttgaa tttagttgtg tagtgttgta 5220

gtgaattttt ttttaaatt gtaataagtt gtttttaag gtaattatgt tttttttgtt 5280

tttttttaa aaaataaaaa taaaaaattt atagaaaaaa atttgtgagt ttagaaaaaa 5340

gaagtaattg gtagaaggtt ttaattaagg taaagagttg taaggtgaag ttaagaaaat 5400

gtaggtattt aaaaaatgta ggtaattttt ataaaggttt ttggggagag gtatatagag 5460

ggattttggt gttgaaaaag atttagataa aagaaattta gggtggggtg ggggtaaaat 5520

gattaatgga attgggggaa gggagggaat aaattgtaaa gaaattatag aaaagtggtg 5580

ggtttttgag ttggagagaa gagagggatt tttggtattt tgatttttt gttgttgttg 5640

tttttaatat gtttgaggta aaagtttgaa tggggattat taagatttgt tatagataag 5700

tttttgaagt tgttttggtg taggttattt ggtttttag tttttggtgt gtggttagtg 5760

tttggtgttt ttggaaagtt atttttgggt agtttttgat agtgtgattt ggtgtttaag 5820

tagtttggga ttttgtgtgg atttgatttt tttagattgt aggtagtttg ggaggaggtt 5880

tggttggggg aggtgtagga tttttgttgt gtttttttga tgatttgggg attgttatgg 5940

ttttttttg gtgtttttgg gttttttgt tttgtatgtg gtgtgaaggg gttagtaggg 6000

aaggagtaga ggatgggggg tggggttgtt ggagttttgt ggaggtttgg gaggtttttg 6060

ggtgggaaaa gtttgttttg aattggtagg gatgtgtaat aatttttttt attttgaaga 6120

gtgaaatagg gtttgttgtt tttattttaa taagtaaatt ggtatttaga gtgtattgta 6180

gataaaggtg gtttggggat ttgaattagg ggtttttggg ttagtttttt tgtttagatt 6240

ataggagttt tttgtttttt tatatatttt ttatttttttt ttagtttttt gttttagttg 6300

agtaatttta ttggtaggtt tagtggtagt ggtgtgtgtg tgtgttggga gttttgggggg 6360

atgtttttgg ttggagtgtt ttattgtttt gtagaggtgt gggtgttgta gggtgatatt 6420

tttggtgtgt gtagatttag ggtatttggg ttgttttggt ttgtggtttg tgttattggg 6480

tgtggagtgg tttgggtgtt aggtagagga gagtggggta gaaaaatagt gttatagttt 6540

aaattatttg ttttttaatt taattgtgtt ttggtaagtt ttgtatgttt ttgggagatg 6600

ttgtgggaag ggggagaaat ttatatggtg tttggagagt tgttttgttg tgtgtatatt 6660

tgtggtagag ttttttgggg ttgtgttttt tatttttgtg ttttttttg tttttatttt 6720

ttttgttttt ttttttttt tttagttttt gttttttttt tttagttttt gtgttttttt 6780

ttttatagtt gataaatgaa tggttagtgt gaaatttttg ttttttttgt gttttaaggt 6840

agtagggagg gaggagtgag ggaggtggtg tgttttttg gatttgtttt tagtttagtt 6900

tataagattt gtagaattta gatgtttagg aattgggagt tttgtggtgg gtgtgggtgt 6960

tttttgatgg agaagttttg tataggtgga gaaaaataag ttttttagat taagtgtgta 7020

ttttttataa ttttgtgttt aaggatggaa ggttttagtt ttttttttaag ttatttattt 7080

tgtattttat aatttgtggt gtatatttag gagtttggag gattttgaaa aaaggttttg 7140

gtttgtgtaa agtgtagaga tgtttttttg tgagttggtt ggaatgtatg tggtgttttg 7200

ggtttagttt gttgtttagt ggatgaatgt atatggttgg ggtggagtta ggttttggag 7260

gtttagatgt gttattttgg gtgtttttaa gaaagataat atatatattg tgtttttaaa 7320

-continued

atttatgatt atttgaattt aatggtaggg ttttgtttta ggatattgta aaagaagggt 7380

ttttaattta aaatttaaat ttttttaatt ttagggtggg tgttggatgg gaaagtggag 7440

agaaggtggg tagtgggagg aaaaaagaaa gggaaggaag ggaagtggaa ggaaagagga 7500

ttgtggaggg gaggaaaagg agaaaggaat ggagatgggg tagatggtat ttaaatagtt 7560

aggaatggat tggaagggtt tgttttgggg tttgggtttg aattttgaga tttgttttag 7620

gatatttttt ttttttttt agtggtaata aatgtgattg tagaggtgtg ttgggatgga 7680

agggtgggaa agaattagat aagtggaggt ggtttagttt tgtttaaagg ggttgggttt 7740

ggggtttgtt tttatttgtt tattttgtgt taagggagga tttgtaatgg gtgttttttg 7800

tagtggttgg gttagttagt tgttatgtgg gttggtgttt tggtttttag ttgagatggg 7860

tgtgattttt tgggtttaga gattgtagta tagttgtata ggtggttggt gtagttgagg 7920

tttttattta ttgaggttaa tgggggtggg gtggggaagg tagggtttag tttaattgag 7980

agagttgagt ggttttgttt ttgggatttg gggttggttg tttttggagt tgatttgagt 8040

tttatttaga tagtgttgaa tatttttatt aaaaggttat ttaagtttta ttttatttta 8100

ttttattttt gtttgttttt tgggagtggt aggaggggta ttgtgttttt gttttgaggt 8160

aggggagagg ttaaaattta ttttaagtaa atttttttgtt agaaaaatga gattttatat 8220

tgtattttat ttgttgattt tttaaatata tttaggttat tttaaagttt atttttagtt 8280

aattaataga ttaattttgt tttagattgt tataaagagt ttgtttggaa tggtaaatat 8340

tgttatttga taagattaat agaattttat gggggtaaat atggaagggt tgtgagtgta 8400

attgaggttt ttgtgtttta atgaggaatt ttagtgggga tttggtgggg attggaagtt 8460

tagtatttat ttagggagtg attgagttgg aaatttggag aagtttggtg tagatgagat 8520

ggtgtttttt agagttgtat tttaagtttt tttattttgt tgtatattgg aagaattgga 8580

ttttaaagat ttttttttatt attttagatt ttgtagttaa tagttatgtt ttttttttt 8640

tgtttgttga aagttgtgat ttaattgttt taagtatttt ttgataattt tgttaaattt 8700

attttgaaaa tatttatata aagttttaaa ataggatgtt gtgagtgtaa tataaaaata 8760

ggaaggtatt aatggaagta ataattagtg ttttgagata gaagtggttt ttagtttgta 8820

atttggaaat gttaagagta gtatattagg gagaaggttt ttaagagttt gatttttgt 8880

tttgagtttt gagtaaagaa agaaatatgt tagaattgta ttgtttgttt gtagtaatat 8940

tgtaaagggt tttaatggat tttaatgtga ttgagtaatt attttttta gtattaagtt 9000

tatagttttt aatttttttg tgagaagatt aggtaggaag tgggatgaga tttaagataa 9060

ttatttagga tttatgttag aatttaaagt tttgattttt ttggtagata aaaatagagt 9120

tggttttaaa ttgtattttt tgtttttaaa aaaatgaggt ttgaaaggaa ttttttttaga 9180

tttatatagg tttatagatg taagtttttt ttttttttt gtgtttgtaa attggaggga 9240

gatgtttttg atattttaga atatttgtat ttatatattt gtaaattttg ttaggaaagg 9300

atttaaatag ttaaattaaa ggtgattatg attaatgtta aagatattat ttattgtttt 9360

tttttttgtt gtattattag gtttaaatat tttggagtta atgattggtt atttaattga 9420

gtaatagagt ttttgagagt gtaaagatta attaattttg ttttttaaag tttggtaatt 9480

aagtttttt tttttggtga gatttaaata tttttgttga tgttatttat ggataaatat 9540

tggtaaatat tgtatttaaa atgggatatt gttttgtgat ggagtagaat tttttaagat 9600

-continued

atatggtttt tttgaaagaa agaattaatg attgtagtaa ggtatttttt tattatttat 9660

gtagagaaat tagttaggat tgataaatat ttattgttgg gagttatttg ttttggtttg 9720

agaaaatatg ttgattgtta tataggtttg tattttatta gaaaatataa agtatgtatt 9780

ttttttgatt tggaattaga aaagggggat tagatatgga tgttttggtt aagttttatt 9840

ttaattagtg aatttttgtt agttttttt tgtagtttag atttggtaaa ttatatatat 9900

ataaagtttg tgtgtgtgtg tggggtgtgt gtgtgtgtgt atgtatatag ggtttttaaa 9960

tattgttaag tttttattat tgaatgtgtt agaggtggtt gtattttaga ggttagtata 10020

aagaattttt gtttaatgtg ttagttttgt tttttgaaat atatttttgg gtagatttat 10080

taattataaa aggatataaa tgtatatatt ttttatattt gtatatattt tattttattt 10140

ttttaaggat tttaagaggt gtttatgttt ataaaatgag attatatatt ttataataaa 10200

tatatttgat aaattataat ttttttatta ttaaattatt gatttgatat tatttgaggt 10260

atagtggaat gagtattaga tttggaatta gaaagtttga gttggattta ttagttgaga 10320

aaattaagta aagatattat ttttggattt ttagtgttat aatttattaa atgaatatgt 10380

agttataaag attaaaagag taagtatttt gaaaaattat ataaaaaagg gatgataata 10440

tattatgttt tgggatttat attgatttaa ttttataatt ttaataataa ttgtattatt 10500

taagttatgt agtttttatt gaatttataa aataaattag agaggggtat attgttttaa 10560

aaattgttta gttttattaa ttagaagata gattatagaa attgatttta atgaagttgg 10620

atatttagtg gaaaggaaga attatttaat aaatggtttt aggataattg atgggtttat 10680

gtaaaataag aattagtttt tatttgattt tttattttaa aagaaatata agatgaatta 10740

aatatgaaat gtagggaata atttttttata aaatgataaa agaaaattat gggattaaaa 10800

aaaataagtt aggaatagag aaatttgttt taagtaaaat ataaaattta gtgttatttt 10860

aggaaatatt gataggtttg attaagtaaa aattatataa aagtaaataa tatatttgga 10920

aaatatttat gatatttgag agataaaaag ttaataattt ttgatatttt atggattttt 10980

atgaataatt ttgaaaagtt tatatttatt aaaatatgag taaaaattat gtataggtag 11040

tggtagaaag agatatagag ttttaaaatt atagaaaaaa ttgtttagtt ttatttgtaa 11100

ttaaataaat agtaattaaa atattaatga tatgttatat tatttttttа ttaattagag 11160

gagtaataaa aaaaaataat ttttttttttt ttttttttttt ttgagatgga gtttgttatt 11220

taggttggag tgtagtggta tgattttagt ttattgtaat ttttgttttt taggtttaag 11280

taatttttttt gtgttagttt ttttagtagt tgggattatg ggtatttatt attatgtttg 11340

gttaattttt gtgtttttag tagagtgggg ttttattatg ttggttatgg tagttgattt 11400

tgaatttttg attttaagtg atttatttat tttagttttt taaagtgttg ggattatagg 11460

tatgagttat tgtgtttggt aaaaatgaaa ttaaaaagtt tgatgtaagt ttgtgttggt 11520

taggttgtgg agaaattggg ttttttatat attgttgtat gaaagtaaat aggttttttt 11580

attttaaaaa ataatttggt aattatttttt ttttgatttt ttattaaaag tttttaaaag 11640

atgtatatat tttttgattt agtaatttta ttttttgata ttattatatg gatggatatt 11700

ggtattaata ttatttaaga taaatggata aggatgtttt ttgtagtatt gtttttaaga 11760

gtaaatattt ggaaatagtg taataggttt attataggga attatttaaa taaaagttgt 11820

tgtatttatt taatggaaaa ttatgtagtt gttaaaaata atatggaaaa attgtatttg 11880

-continued

ttgatttgga atatttttta aaatatattg tattttaagt gaaaaaattg atagataaaa 11940

ttgagtatat attatatatt taattatgta aaatgggtgt atatttattt ttttaataaa 12000

tatttattga gtgttttagg tattgtggat ataggagtga ataaaatatt aaagattgtt 12060

ttgttggaat ttttattatg gtaggtagag aaaggtaatt tttaattata tatgtaaaat 12120

ggtgatatgt gttaggatga aaatagagta tgggaaaggg aatagagtat agtatgtttt 12180

aggtagagta atgagaaagg agtttattag gtagagaata gaggagtgtt aaggggtgat 12240

aatgtgaaaa agtttagtta agtttaggtt taggaattta tatgatgtta ttatttgtat 12300

gaagtaaatt tatgaaattt tgtatttaaa ttatgtgatt tttttttgtt gtataatgat 12360

ttttttgtg agaaagaatt agtgagtaat atgatatgtg atgtttatta tatttatttt 12420

gtagttgaag tattttagaa agttggttgt ttagggtatt taattagtag aatggtggat 12480

aaaattattg tgtttttgtt tttaagttat tttttgttag ttgagttttt ttaaaattaa 12540

gggagaatat taaatattta tttaatagaa ttttggtttt tttttggagg ttgtttattt 12600

ttggtgtgtt ttttattttt tttgtagata ttttttatgg ataggagttt tattgatttt 12660

gttttttatg tttttttaggt aattaaatag ttttagtatt ttgtatttgg attgtttaat 12720

gatttttttt tttttttatt gattttattg gagtagaagg ttaaatatag gtatttattt 12780

ttatattttt taggtttttt tatgtagata gagttgttat tattagtaga tggtgtttat 12840

aagagagata tggaaaagtg gaagaatgga aatagggatt tggagtggtg ttaaaaataa 12900

agaaagaggt ttttgaaagt tttttattta atatgattaa atatagaaga taaaagatgt 12960

aatttaatga ttaaggatag ggatatagtt ttgagaaaat tatagaagtg gaatgaatgt 13020

tttggatttt aaagaagagg gaatggaaat gtagttttga ggtagttaaa gttaataaag 13080

ttttttgata tatttattgt agaaataaat gattattttt taatttaagg tttgtaaatt 13140

taggtgatta taggggtgag gtatggagtg gaaatgggta agagattaag tgggggtagt 13200

aatggttttg agtatataga tttaagggga gagattgttg tttggttaat gtttatgatt 13260

agttagttga aatatagatt tagtgttgtt agatttttttg atttttttgaa agaagttaga 13320

gattttatat ttagtgtaaa tagtttaatt tttaaattga gtaatttatt taaatatatt 13380

tatgtaattt atttaaatgt atttagagta tgtttgattt tttttttgta tttttattta 13440

gttttagggg taattttaag gaaagtgatg tgtatatttg ttgtatttat gtagatggag 13500

gattattttt tttttttaa taaaagatat ttttataagg attgagtgtg gtataaagag 13560

tttgattttt agtatttaaa ttgagtagga aatttaaata tagaagattt ttttttgggt 13620

tatatttttg gttttttttaa atgagatttg atttttttaga tattgataaa tattatatta 13680

ttgggttatg gtaaattttt aagtggattg atttaaaagt gtttaaaagt attttaggtg 13740

ttaaggaatt gttggtatgg agttttaaag gtgtttttgt ataagaaata gaggtgttga 13800

ttatgtattt gatttataaa tatttttttaa ggatatattt tgaatttggt aatgaagaaa 13860

tattgtggga attattaagt tgaataaagt attagttttg ttttttagaa gtttattgtt 13920

ttatagagga gattagatat gtttatagat aatttgaagt atatagtaat aaattatagg 13980

atgatttatt ttaagtataa agtgtatatt gttattatag tggtaattta ggttataatt 14040

tttaaatgta agagttttaa atattatttt ttttatttttt agtgatatga aattatgtta 14100

gaatatttgt agaaagtttg aattagtaat tttaataata aatattgagt attgatatat 14160

-continued

tgtatattat tgagatttat tttatatata ttattttatt aatttttatg attattttgg 14220

taagtattaa tttttataat taagtagtga tagaaaatga ggtttagaga gatatagtat 14280

tttatttaaa gttatagaat ataagtagag aatgtagaat ttgtattttg ggttatttga 14340

tagtggagtt tagattttta aaattaggtt tataaagatt ttttggaaaa aaagaaaaat 14400

aaaaaataga gaataaattt ttttgaagaa aagtaggaaa ttattattta taagtataat 14460

taaagagaga tttttaattt ttttttttt ttttttttg agatggagtt ttgttttgtt 14520

gtttaggtgt aattttagtt tattgtaatt tttatttttt gggtttatgt tataaagaaa 14580

gtttttattt gattggtttg atttgattgt ttattagaat ataggttttt tagtagggtt 14640

aataagttat aatagtttta gtgatataag ataaatataa attaataata aggttaattt 14700

agaattatta atatatgttt tttattgtga attattagtt ggtttataga tggaaaatat 14760

aagatatata atgtattagt ttaaaaaatg ttttttagag agtgtttaga attatagaat 14820

ttaaaaattt tgttttgaaa aggtttaatt tttattttaa aatatataaa gaaattaaat 14880

gtaaaaaaat taagagaaaa ggttttaatt aattaaatgt tgtataattt taagttttat 14940

ttatgtttaa gattagtttt gtaatttatg attaatttta gtattagttt gtttttaatt 15000

taggtttgtt aaataatata gaattttttt tttaatgaag ttttttaggg ttttagttga 15060

aaaatatata ttagttagga tgtttttaat tgtaaggaat aaaaagttat gattttattg 15120

ggtttaaata ataatgatat ttataatttt atataaaatg aagtttaaag gtatggttta 15180

ttttaggttt aatttttaag gatttatgtt ttgtttttgt tgattttttt ttagttttat 15240

ttttagaggt attttattta tttatggatg taaaattttt ataatagttt tagtagaaag 15300

attttagtat gataatgttt agggaagaaa attgattatt ttttttttg ttttttttt 15360

aggatgaatt tatttttttt agggattttt tattggagga agagattttt ttatatttta 15420

ttggttaaaa ttatattata tattgtttta aatttaatta ttggaaaagg gaatggtatt 15480

tttataattg atttaggata ttaggattat tttttgggtt ggggtttgat ttaggatttt 15540

ttaaagtatg tgattgaggg taggtattag aatggaatta gagttgtgtt agaaaggaga 15600

aatgtatggt tattatagaa tagttttaaa tgttattgag gagggtataa ttgtaaaaat 15660

taaataattt tttgttattt ttttgaagtt ggtattttga tttttagatg gttttttaat 15720

ataggttttt ttttttttta ttattatagt tgttttgaaa tttgagttgg aagggaatat 15780

tttgagattt agattgttaa atgttttttt taaagttatg taataaatta aatggtaaag 15840

ttagggtagt tttttgattt agtatagggt attttttttt atttttttatt tttgagattt 15900

tagaattgtt ggtattgttt taaaatttat ggtaagaatt ggttatttttt gtaattaata 15960

tttttttata atatatttgt tttgtttgtt tagttagtta gaaattatat ggagtttgtg 16020

ttttaaaaag tttgttgaag tttttatttt ttgttttggt attatgtgta tgaattatta 16080

attggttttt ttttatttta tatttgatga agatgttttt tttttaatat tttttttat 16140

tgtttttttat ttttttttgt tttatttatg attagttgtt tgtttttaaa tagattttgt 16200

ggttatttat ttttttttgt gttagttttt atttattagt tatttgaatt gtggttttta 16260

ttgtttttta tatagtttat ttatttgtgg ttgttatata tatttatatt gttatatgtt 16320

tttaaattgt attttggaat gattttggta atggttgtat tggatgagat ttaaattaat 16380

aattaaagta ttgagatagt ttttgttatt ataagtttat ttttgttttt atagtttaag 16440

-continued

aggagttatt ttttttattt ttattattta atgtttaata ttatttttat tatatataat 16500

gtataaaaag tatgtgattt atgatttatt ttaaatttga atgtttgtga tttattttgt 16560

gtttttttttt atttttagg 16579

<210> SEQ ID NO 19
<211> LENGTH: 16579
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 19

tttagaaatg aagaaaaata taaagtaaat tataagtatt taggtttaaa gtgggttata 60

aattatatgt tttttgtgta ttgtatgtaa taagggtagt attaagtatt aagtaataag 120

agtagaagaa gtagtttttt ttggattata aaaataggag taaatttgta atagtagaag 180

ttattttaat gttttaatta ttagtttgaa ttttatttaa tgtagttatt attaaaatta 240

ttttagaata tagtttggga atatataata atgtggatgt gtgtgataat tataagtgag 300

taggttgtat gaagaataat agagattata gtttaagtaa ttagtggata ggagttggta 360

taaaaagaaa tgggtgatta taaagtttat ttagaaatag atagttgatt ataaatagag 420

taagagaaga tggggagtaa tagagaaaga tgttgaaaga aaggtatttt tattaagtgt 480

aaggtgagaa ggaattaatt ggtggtttat gtatataata ttaaaataga gaataaggat 540

tttggtaggt tttttaaagt atagatttta tgtagttttt agttggttaa gtaaataaaa 600

taaatgtatt ataaggaggt gttaattata aaagtgatta gtttttgtta tgaattttaa 660

agtagtatta atagttttaa agttttaaga gtaaagaatg aaagaaaata ttttgtattg 720

agttaagaaa ttgttttggt tttgttattt aatttattgt atgattttgg aaaagatatt 780

taataatttg ggttttagaa tgttttttttt tagtttaaat tttaaggtag ttataatgat 840

gaggaggaga agaatttgtg ttggggaatt atttagaaat taaagtgtta attttaaaag 900

aatggtaaaa gattatttgg tttttatagt tgtgtttttt ttggtagtat ttagaattat 960

tttataatag ttatgtgttt ttttttttta atataatttt ggttttattt tagtgtttat 1020

ttttagttat atattttgga gaattttggg ttaaatttta gtttaaggga tgattttgat 1080

attttaagtt aattatagag gtattatttt ttttttttaat gattggattt ggaatagtat 1140

ataatatagt tttggttagt gagatgtgag gggatttttt tttttagtga agggtttttg 1200

ggaaagatag atttatttttg agaagagaat aaaggaagag gtggttagtt tttttttttg 1260

gatattatta tgttggaatt tttttgttgg aattgttgta gggattttgt atttatgagt 1320

aagtgagata tttttggaaa tgaagttgag aaagaattag taaaggtaag gtgtaagttt 1380

ttaaggatta gatttagagt ggattatgtt tttgaatttt attttatatg agattataaa 1440

tgttattatt gtttaaattt aataaaattg tgatttttttg ttttttatag ttgaaagtat 1500

tttaattggt atatattttt tagttgggat tttgaaaaat tttattggaa aagaaatttt 1560

atattattta gtaaatttag attgaaaata agttaatgtt gaaattagtt ataaattata 1620

aaattgattt tgaatataaa tagggtttgg agttgtatag tatttaatta attaagattt 1680

tttttttaa ttttttttgta tttgattttt ttatatattt tgaaataaga attagatttt 1740

tttaaaataa aatttttaaa ttttataatt ttagatattt tttgaagagt attttttaag 1800

-continued

ttaatatatt atatatttta tatttttat ttatagatta attgataatt tataatgaaa 1860

aatatatgtt gataatttta aattaatttt attattagtt tgtatttatt ttatattatt 1920

agaattgtta tgatttgtta attttattgg aaagtttatg ttttagtaga taattaggtt 1980

ggattaatta aatagaaatt ttttttatgg tatgaatttg ggaggtggag gttgtagtga 2040

gttgagattg tgtttgggtg atagagtgag attttgtttt aaaaaaaaaa aaaaaaaaga 2100

aattaaagat tttttttaa ttatatttat aaataataat tttttatttt tttttaggaa 2160

aatttgtttt ttatttttta ttttttttt tttttaggaa atttttataa gtttaatttt 2220

aaaagtttgg gttttattgt taaatagttt agaatgtaag ttttgtattt tttatttata 2280

ttttgtgatt ttgggtaaaa tattatattt ttttaaattt tatttttgt tattatttag 2340

ttgtgaaaat tagtatttat tgaagtgatt ataggaattg ataaaataat gtatataaaa 2400

taaattttag tgatatataa tatattagta tttaatattt attattaaaa ttattaattt 2460

agatttttg tagatgtttt aatatagttt tatattatta agagtaaaaa gaatagtatt 2520

tggggttttt atatttgaaa attataattt aggttattat tatagtagta atatatattt 2580

tatgtttgga ataaattatt ttataattta ttgttatgta ttttaaatta tttgtaaata 2640

tgtttaattt tttttgtaga atagtaggtt tttggaaggt agagttagta ttttatttaa 2700

tttaataatt tttataatat ttttttattg ttaggtttag ggtatatttt tgaggaatgt 2760

ttgtagatta aatatatgat taatgttttt attttttatg tagaaatatt tttgaaattt 2820

tgtgttaata attttttaat atttgaaata tttttgaata tttttgaatt agtttatttg 2880

aagatttatt atgatttaat aatatgatgt ttgttaatat ttagagagtt aaattttatt 2940

taaaaaaatt aaaggtgtga tttaggaaag gatttttgt atttgagttt tttatttaat 3000

ttaaatatta aaaattgaat tttttgtgtt atatttaatt tttataaaag tgttttttgt 3060

tggggagagg agaatggttt tttatttata taggtataat aagtatatgt attatttttt 3120

ttaagattat ttttaagatt aggtaagaat ataaggagga gattagatat gttttaaata 3180

tatttgaata agttgtataa atatgtttga ataagttatt taatttaaaa attagattgt 3240

ttatattaaa tataaaattt ttaatttttt ttaaaaaatt agaaggttta gtaatattaa 3300

gtttatattt tagttggttg attatgagta ttaattaagt agtagttttt ttttttaagt 3360

ttgtgtgttt aaggttatta ttgtttttat ttagtttttt atttattttt attttatgtt 3420

ttgtttttgt agttatttaa gtttgtaaat tttgaattaa aagatgatta tttatttta 3480

tagtaaatat attagagggt tttattgatt ttagttgttt tagaattgta tttttatttt 3540

ttttttttg gagtttaaga tgtttatttt atttttgtaa ttttttaag gttatgtttt 3600

tatttttggt tattaagtta tatttttgt tttttatgtt tgattatatt aaatggaaag 3660

tttttagaag ttttttttt tgtttttagt attattttaa atttttattt ttatttttt 3720

atttttttat gttttttta taagtattgt ttattagtga taataatttt gtttgtatgg 3780

aaagatttga gagatatgag agtgaatgtt tatatttagt tttttatttt agtgaagtta 3840

atgagaaggg aagaaagtta ttgaataatt taggtatagg gtattgagat tgtttaattg 3900

tttgggagat ataaaagata aggttaatgg aatttttatt tatgagaaat gtttgtaggg 3960

aggataaggg atatattagg gatgggtagt ttttgagaga aagttaagat tttattaaat 4020

agatatttag tatttttttt tgattttggg gaagtttagt tgataagggg tgatttaagg 4080

-continued

atagagatat ggtggttttg tttattattt tgttaattag gtattttaga taattaattt 4140

tttaggatat tttagttata aagtggatgt aatgagtatt atatattata ttgtttattg 4200

gtttttttttt atagaaagaa ttattgtata gtagagagaa attatgtgat ttaggtgtga 4260

aattttatga atttatttta tataagtaat agtattatgt gaatttttga atttggattt 4320

ggttaagttt ttttatattg ttgtttttta gtatttttt gtttttgtt tggtgaattt 4380

tttttttatt attttattta agatatattg tattttattt tttttttat gttttatttt 4440

tattttagta tatattatta ttttgtatat gtgattggga attgtttttt tttatttatt 4500

atagtaaaag ttttaataga gtagtttttg atgttttgtt tatttttgta tttataatgt 4560

ttagaatatt tagtaaatat ttgttaaaag aatgaatata tatttatttt atataattag 4620

atgtataata tatatttagt tttatttatt aatttttta tttagaatat aatatatttt 4680

ggagaatgtt ttaaattagt aaatataatt tttttatgtt atttttgata gttgtatagt 4740

tttttattag gtggatatag taatttttat ttaaatggtt ttttatgatg agtttgttat 4800

attgttttta gatatttgtt tttagaagta atgttgtaaa gaatattttt gtttatttgt 4860

tttgggtgat attggtatta atatttattt atataataat gttgagaagt ggaattgttg 4920

aattaaagag tatatatatt ttttaaaaat ttttgataaa gaattaaaaa aaagtgattg 4980

ttaaattatt ttttgaagta gaggaattta tttatttttа tataatagta tatgagagat 5040

ttagtttttt tataatttgg ttaatataaa tttgtattaa atttttttaat tttatttttg 5100

ttaggtatag tggtttatgt ttgtaatttt agtattttgg gaggttgagg tgggtgaatt 5160

atttgaggtt aggagtttaa gattaattgt tatggttaat atggtgaaat tttattttat 5220

taaaaatatg aaaattagtt aagtatggtg gtgggtgttt ataattttag ttattaggga 5280

ggttgatata ggagaattgt ttgaatttgg aaggtggagg ttgtagtgag ttgagattgt 5340

gttattgtat tttaatttga gtaatagatt ttattttaaa aaaaaaaaaa aagaaaaaaa 5400

ttatttttttt ttgttatttt tttgattggt gaaaaaatgg tgtagtatat tattggtgtt 5460

ttaattatta tttatttaat tatagatgag gttgagtgat ttttttttata attttaaaat 5520

tttatgtttt tttttgttat tatttgtata taatttttgt ttatattttg atagatgtgg 5580

atttttttaaa gttatttata aaagtttatg gaatattaag agttattagt tttttgtttt 5640

ttaaatgttg tagatgtttt ttaaatatat tatttgtttt tgtgtaattt ttatttagtt 5700

agatttatta gtatttttta gaatggtatt gggttttatg ttttgtttag aataggtttt 5760

tttatttttа atttatttttt tttaatttta tgatttttttt ttattatttt gtaaaaagtt 5820

atttttttata ttttatattt gatttatttt atatttttttt taggatgaga aattaggtag 5880

agattgattt ttattttgta tggatttatt aattgttttg aaattattta ttgaataatt 5940

ttttttttttt attgaatgtt taattttatt ggagttaatt tttgtaattt gttttttgat 6000

taatgaaatt aaataatttt taaaataata tgtttttttt tgatttattt tatagattta 6060

gtagaaatta tatggtttgg gtaatgtagt tattattaaa attatagaat tagattaata 6120

tgaattttaa gatataatat attattattt ttttttttgta tagtttttta aagtatttat 6180

ttttttgatt tttataattg tatatttatt taatagatta tgatattaaa ggtttaaaag 6240

taatattttt gtttggtttt tttagttagt gagtttaatt taaatttttt gattttaagt 6300

ttagtgttta ttttattata ttttaagtaa tattaaatta ataatttaat aataaaaaga 6360

-continued

ttataattta ttaaatatgt ttattgtaaa atatgtaatt ttattttata gatataaata 6420

ttttttaaaa tttttagaag aatgaggtag gatgtatata gatgtggaaa gtatatatat 6480

ttgtattttt ttatagttgg taaatttatt tagaggtgtg ttttaggaaa tagaattgat 6540

atattaaata aagatttttt gtattgattt ttgaaatata attattttta atatatttaa 6600

tggtaagaat ttagtagtgt ttaaaaattt tgtatatata tatatatata tgtattttat 6660

atatatatat agattttata tatatgtaat ttattaagtt taggttataa gaggaaattg 6720

gtaaaaattt attaattgaa atggagtttg gttaagatat ttatatttaa ttttttttt 6780

ttgattttaa attagggaga gtatatgttt tgtgtttttt ggtaaaatat agatttatgt 6840

aataattagt atatttttt agattagagt aagtaatttt taataatgag tgtttgttag 6900

ttttaattgg ttttttatg tgggtggtga aaaaatattt tattatagtt attaattttt 6960

ttttttaaaa gaattatgtg ttttgagaag ttttgtttta ttatagggta atattttatt 7020

ttaaatatag tgtttattga tgtttgttta tgaatgatgt taataaaaat gtttaaattt 7080

tattaggaaa ggagggttta gttgttagat tttgggagat agagttagtt gatttttgtg 7140

tttttaaagg ttttgttatt taattaagtg attaattatt aattttaaag tatttaagtt 7200

taataatata ataaaagaga gaataatggg tagtattttt aatattaatt atggttattt 7260

ttggtttagt tgtttaggtt ttttttagt agaatttata agtgtataga tataggtatt 7320

ttggaatatt aaaggtattt ttttttagtt tgtaaatata gggaaaaaga agaaatttat 7380

atttgtaaat ttatgtaaat ttaggggggat tttttttaaa ttttgttttt ttgaaagtag 7440

gaaatataat ttaaaattaa ttttgttttt atttattaaa agaattaaaa ttttgaattt 7500

tgatataaat tttaaatgat tgttttgaat tttattttgt tttttattta attttttat 7560

aagaagatta aggattgtgg atttggtgtt ggaggaagta attgtttaat tatattggaa 7620

tttattggaa ttttttataa tattattgtg ggtagataat ataattttaa tatatttttt 7680

tttttattta gagtttagaa taagaagtta gatttttgga gatttttttt ttgatgtatt 7740

gtttttaata tttttaagtt ataaattggg aattgttttt attttagaat attaattgtt 7800

gtttttattg gtatttttt atttttatat tatatttata atattttatt ttaggatttt 7860

atgtgaatat ttttaggatg ggtttggtga aattgttaag aagtgtttaa aataattagg 7920

ttatagtttt tagtagataa aaagaaaaag atgtggttgt tagttataga atttgaggtg 7980

atggaaaaga tttttaaaat ttagtttttt taatatgtag taaaatggga ggatttaaaa 8040

tgtagtttta agaagtattg ttttgtttgt attaagtttt tttaggtttt tagtttagtt 8100

atttttttaaa taggtgttga gtttttagtt tttgttaggt ttttgttaag gttttttatt 8160

gaggtatagg aattttagtt atatttgtag ttttttttatg tttgttttta tgaaatttta 8220

ttagttttgt tgaatggtag tatttgttat tttaggtgaa ttttttatga tagtttggaa 8280

tgaaattaat ttgttaattg attaagaatg ggttttagag tgatttgaat gtatttaaaa 8340

agttaataaa tggaatatgg tataaagttt tatttttttg ataaaaagtt tgtttggagt 8400

aaattttaat ttttttttttg ttttaaaata aaagtatgat gttttttttta ttatttttaa 8460

gaagtaagta aaaataaaat aaaataaaat aaaatttaaa tagtttttta ataagagtgt 8520

ttagtgttgt ttggatggaa tttgggttag ttttagaagt agttggtttt gaattttggg 8580

ggtgaggtta tttggttttt ttggttaagt tgggttttgt ttttttgtt ttgtttttat 8640

-continued

tggttttggt gggtgggggt tttggttgtg ttggttgttt gtgtggttgt attgtagttt 8700

ttgagtttag gaagttatat ttgttttggt tgggagttga agtgttaatt tatgtggtgg 8760

ttggttaatt tagttgttgt ggggagtgtt tgttgtgggt ttttttttgg tgtgaggtga 8820

gtgagtgggg gtgggtttta gatttagttt ttttgggtgg ggttggatta tttttgtttg 8880

tttagttttt ttttattttt ttattttggt atgtttttgt agttatattt gttgttgttg 8940

ggaaagggaa aagagtattt taaaatgagt tttgaaattt agatttaaat tttgaaatga 9000

attttttttgg tttgtttttg gttgtttgaa tgttatttgt tttatttttg tttttttttt 9060

ttttttttttt tttttatggt ttttttttttt tttatttttt ttttttttttt ttttttttttt 9120

ttttattgtt ttgtttttttt tttatttttt tgtttgatgt ttgttttgaa gttgaggaag 9180

tttaagtttt ggattaggag ttttttttttt gtaatatttt ggggtggggt tttattgttg 9240

ggtttaagtg gttatgggtt ttgagagtat ggtatgtgta ttatttttttt taaagatgtt 9300

tagggtggtg tgtttagatt tttgaaattt ggttttattt tggttatgtg tgtttgttta 9360

ttaagtggtg aattggattt agggtgttgt gtgtgtttta gttgatttgt gagaaggtat 9420

ttttgtattt tgtatagatt aagatttttt tttagggttt tttaaatttt tagatgtatg 9480

ttgtagattg tggagtgtaa aatgaatgat ttaagggaaa attagggttt tttatttttg 9540

aatataaggt tgtgaaagat gtatgtttgg tttgggggggt ttgtttttttt ttgtttgtgt 9600

gggatttttt tattaggggg tgtttatatt tgttgtaaaa tttttaattt ttagatattt 9660

aggttttgta agttttgtga gttgaattgg gggtagattt gaaggagtgt attgtttttt 9720

ttgttttttt ttttttgttg ttttggagta tgggaaaggt agggatttta tattagttat 9780

ttatttgtta gttgtggggg aggggatgta gaagttagag agagagaata gaggttggga 9840

gaggggaagg gaggtagaga agatggagat agagggagat ataaagatga ggaatgtgat 9900

ttagagagat tttgttgtgg gtgtgtgtgt ggtaggataa ttttttaggt gttgtgtaga 9960

ttttttttttt tttttgtagt atttttttagg gatatgtagg atttgttgag atgtagttgg 10020

attggagagt aaatagttta ggttatgata ttatttttttt attttgtttt tttttgttta 10080

atatttagat tgttttatgt ttagtgatat agattgtggg ttagaataat ttggatgttt 10140

taggtttgta tgtattaggg atgttgtttt atagtgttta tgtttttgta ggatggtggg 10200

gtgttttagt tgaggatgtt ttttggggtt tttggtatat gtgtgtgttg ttgttgttag 10260

atttgttagt aaagttgttt aattagagtg gagaattagg agagggtggg gaatatgtaa 10320

ggaaatggag gatttttgta gtttgggtga gaggattgat ttagaggttt ttgatttggg 10380

tttttgagtt atttttgttt gtagtgtgtt ttgggtgttg gtttgtttgt tgagatggga 10440

gtgatagatt ttattttatt ttttagggtg agaaaaattg ttgtatattt ttgttgattt 10500

agaataggtt ttttttgttt aggggtttttt tagatttttg tggggtttta ataattttat 10560

ttttttatttt ttgtttttttt tttgttggtt tttttgtatt gtgtgtagga taaaagaatt 10620

taggggtgtt gggagagaat tgtgatagtt tttaagttgt taaagagata tggtggggat 10680

tttgtatttt ttttggttgg attttttttt agattgtttg tggtttgaag gggttaggtt 10740

tgtgtaaggt tttaggttgt ttgggtgttg ggttgtattg ttaggagttg tttgggagtg 10800

gtttttttagg aatattaggt attgattata tattaggggt tgggaaatta ggtggtttgt 10860

attaaggtgg ttttggggat ttgtttgtgg taagttttgg tagtttttat ttaaattttt 10920

-continued

gttttgagtg tgttaagaat aataataata aaaaaattaa agtgttaaag gttttttttt 10980

ttttttttagt ttaagaattt attatttttt tatgattttt ttataattta ttttttttt 11040

ttttttaatt ttgttagtta ttttatttttt attttatttt gggttttttt tgtttgaatt 11100

ttttttaata ttaaggtttt tttgtatgtt tttttttaaa agtttttatg aaagttattt 11160

gtatttttta agtgtttata tttttttaat tttgttttat agtttttgt tttaattaaa 11220

gtttttttatt aattgttttt tttttttaag tttgtgggtt tttttttata agtttttgt 11280

ttttgttttt taaggggga ataaaagaaa tgtgattatt ttggaaggtg gtttattgta 11340

gtttgggggg aaaatttatt gtagtgttgt gtgattgggt ttggtgttgt ttaggtgggt 11400

tatataggaa gtgtggtggt ttggggaagg atgtggaggg tgtgggatgg tgttggaaga 11460

tgtgggagga tgtggggttg gttgaagatt ttggttttag tttttgttta taatatttaa 11520

tgtttttggt tgttgtgttt gtatgtttgg agtttttttt tagaaaggtt tggggtagtt 11580

tgtttgtaag tttaaatgtg ggttgtgata tttattatta ttatattatt gtattgttag 11640

agttgaggag gagatttagt gagaagaagg aggagggaga ggaggagggt ttatttatag 11700

gttttaaaag tgttttgtag ttttagttat ttttaagagt ttaggtttgg aaagtaggtg 11760

gaggggtgga aaggtagttt tttgtgtttg tggtagggat tgtttgtttt tttttttagg 11820

tatttttatt ttttggtgtt tttttttgat aagaagtatt aattggtttg gggatagtag 11880

tgtttttttat ttagggattt atttagtaat attgttttgt ttgtggaaat tattgttttt 11940

ttatttttttt tttatttttа ttttttttgtt ttttgtgtag ttagttttgg gttaggggaa 12000

aggagttttt aggtttttag ggggtaggtt agtaatagat aattgagtat gattattttt 12060

ttttgggagt atataaaatt gtaaaattag taaagaattt ggttatagtg tgtttatgtg 12120

tttatagagt ttgtttgttt tattaaaggg aagtgttagg tttaaggttt ttgttaattt 12180

gaaagagata ttgagaaaat gagatttttt ggggatttag agggaaagtg taagaatttt 12240

ttattgtatt tttagggaat tgtttaatgg ggagtttggt tttaaaagat tttggtaata 12300

aaaggttgga taggaaattt ttttaggtaa attttttgtt ggatttaaag agaatatttt 12360

ttttttgtta taaatttttt tttatataag tttagatttt gttttttttg ttttttttt 12420

tttttagttt ttttaagtat ttttgagtag aatatttgat aatttttttg agtaataggg 12480

atttttggа agtattaatt atttttttatg ttttttggaa ataagattat aatttttat 12540

gtgtatatgt gatttttttt ttttagttag gtttatttta ttgtgtaaat agtattaata 12600

tatggaagag ttttttgtat tgtgttataa aagattttta ataggatttt atagagaaaa 12660

gggttaaata gttgatataa aggatttttg ttttttttaga aaagagggat tttggatttt 12720

ttttttttttt gaagttaagt atgagtttat ataataggaa taaaataaat ttaaggtgta 12780

tattagtata atattaggga tattagaatg gatggtaaat ttttttttt atataaatat 12840

gaaagtattt ttataatttt tttttttgaa gtttttattt agaaaattat attaaatata 12900

ggattttaaa atagtagtga ttaaagatga aagttaattt ttttttttaa ttttttgat 12960

tagtttggtt taaatttatg ttttaaaatt tttagtaatt tagatttttg tatatgtgta 13020

tttataagat tttttttata ttttgtaatt tgtaggtatt tagtatggtt attgatttta 13080

agtgataaat aggtagaaga tttgttagga taattttagt ttttattatg tgtatttttt 13140

atttttttttg ttatttatta gtttttttgta gtaatataat tgttagttat attatttgga 13200

-continued

aattaatgtg tttttgtagt gaattattta tttttagtaa atttattgta ttattatttt 13260

aaatattgaa atttgatttt attagatgtt tagaatgata gttatagagt agaattagat 13320

tttaattttt atgtataata tgtaaataaa ttaaggtgta gaaagttatt gttttaaagt 13380

tatataatta atagtttagt ggtttttttа gtgttaatta tagtaatttt agatttaaga 13440

ttgtttgatt taggaatgga gagaaataaa ataaaatgaa tgtttattga aatatatatt 13500

ttagatatta agatagattt ttttgtgaaa tattttaaga tgtggatttg tatttttatt 13560

tttaaatttg agtaaattta gttgagtttt aaagagttta agtaatttgt ttaaggttat 13620

tttgtaagtg atagatttgt tagtttgaaa agtttattat tttttatttt tttgtttttt 13680

tatgatatag ttaagtaaaa ttttttaata aaggagaatt agattagaaa taagtaatta 13740

atagtagatt aatgtttatt tttagggaaa tgttaatagt taatatttat tataaataat 13800

gttatttgta tataatggaa tatgattgaa agtgatgggt tagatatttt ttatattttg 13860

agatgttttt atttatgttg ttaaatttga ttttttaatt ttttaaaatt tttagttttt 13920

tttaatgtta atagatttta taatatttga ttaaatattt ttataggata aaggtatttt 13980

tttagggttg atttaaggaa ttagaatatg tataatgttg aaaagtttta aattgtttgt 14040

ttttttttтt ttttgatttg taatattatt atttagaaaa ttaaatattt ttttaggtta 14100

ttagataatt aaaatgaaat tatattttat ggatattttt tttttatttt tagggtagat 14160

ttttttttтt agttgtttaa aatggagaat ggataaaaaa gaaaaaggaa aatatattag 14220

ttatatttag aatagagttt atagttgtat tattattttt atgtaattgt ttttaatgat 14280

ttttttttat taagagaaag aaaatgagag agtatttatt tttgattttt taatgtttga 14340

taaaattatt agaagttttt gaatgggaaa atatatattt tttaggtttt ataattaaaa 14400

ttgtatttat ttgtttatta tatttaagta tgaattaaat agtaattaaa atattatttt 14460

gtgagttttt ttttaaatta tatttttgta ttggggatat tttttgtgaa aattttaaga 14520

ttttaggtat aaataaagta agatttttat ttataatgtt tgtgtttaaa agaaattata 14580

agttagaagt tatgtgtttt gtttttttgt tttgattttg tttttgggaa tttggttgag 14640

tttaggggta gtaagttaat gatttggaga ggaaaggtgt agatggtaga tgtgttttta 14700

gggaatttta ttatttttat ttaaattttt ttttgaggtt ttttaggatt atattttaga 14760

ttttttttтt tttattgatt aggtgaagtg tttttagata tggtattatt tattgttgtt 14820

ttagggatga taagatttta gttttattta aatttagaaa gagtttagtt ttaatttttt 14880

gatttgattg gttttaggta taattaagta aattattttt attggatgat atagtaatgt 14940

tatagattaa aaatgtattt ttaattttta aagattttgg aaaaaggaag tttatttttt 15000

aagattgtgt tataattttt tttttaaata aaatattgta aattagggat agatatttta 15060

tttagtattt tgaatttttt ttggtttttt aggaaaggta gaaatataag gtttttattg 15120

ttttttttтt atttttaaaa aagaaatttt gtaaaaggta aagtgatggt ggtaatttta 15180

ttgtgtttta atttgaagta tataaattgt tttgataaat taaatataga taaatttttt 15240

tagaaagtta taaaatatta gatttttttt agatattaga tattgagttt ttgtttttta 15300

atttttaaat taaaatgtta tagtaaaata tttgttttaa taaatttttt ttttttttgt 15360

gttgtaaatg ttttgtttat ggttgagtag atttttttag aatttaattg aaagattatt 15420

tatattatga tagaattatg taagttaaag ttaagtttta atatggagtg atgggagagg 15480

-continued

ttgggttgtt gagaggagat gattttatta gtgagaagag aggtttttg agttggtttt 15540

taaggtagtt aaatggttta gtatatttgt atttattttg ggaggttgtt tggagtagag 15600

tattaaaaat gttatttttt ttataaagaa aaagatgaag gtataaagaa tatagtttaa 15660

ttaaattgaa ttaaagattg tattattttt attttttttt tagttatttt tttaaagttt 15720

ttttttttt tttttatagt ttgaaaaata taggttattt ttataaagta tttgtttgat 15780

tgaagtaaga taaattgaag agtttggttt ttgaatataa ggaagtggta gagattttag 15840

aattaatgag gtagggtgga atataatgta gaaggttttt taagaggtaa aatttttgta 15900

gttgttgttt atttttagtg attatttgtt tttatttatt taattagatt atagtgggat 15960

tgagataagg agtataatga aagttgttgt attgatggtg taaaaagata tatggtatta 16020

ttaatgtttt tgtggtgagt aaaggtagaa tgtaagatag aaaataagga aatttttgtg 16080

ttgtaagaaa agtgtttagt atgggtatta attggtatgt agaagtagta agtatatatt 16140

ttatttttat tttattgata aaaagaattg taggattaat gtttgttttt attaaaatat 16200

gagaaattat attttaattt tataaattaa aggttaaata tttgttgtaa taattttaaa 16260

aaaaagagaa gatttgtaag atttagtaaa tatttattat aaatagattg tttttaattt 16320

tttatatttt tttttaattt aataaaagtt tattgagtag tttttatgtg tattgtttta 16380

gatagtaggt attatttaaa aaagtaataa ttatttttta tttttaagga gtttataatg 16440

tttttagaaa gaaattaatt tattggagta gttaaagtta taattgattt ataaaatgat 16500

tagtgaattt agaggaaaat tatatataaa tatttttgtt tttattggtt ggtttagtta 16560

gatattattt taaaaaata 16579

<210> SEQ ID NO 20
<211> LENGTH: 3000
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 20

atatttaaat tgtatatttg gtttgtaata tattatatat ttaaaagaat tattattttt 60

ttgaatatta ttaatattaa taatagtatt atagtgatta agagtggatt taaataaagg 120

ttggtttgaa atttaggttt gttatttatt agtaagtttt taaaaatttt gagtttttag 180

ttttttttata tgtgaaatgg agaaaataaa tttgtaaaat taattaaatt tgtttattta 240

tatttttttt tttagtgtat attaattggt atttttttgtt aggttagaat gtgtttttaa 300

ttatgtttta aatttgtttt gtgttaattt tattgttaga atttttttta ttttgagaat 360

tagaaaagga aatattatgt ttggtaatgt ttatattttt taaaataaat ttgtaggaaa 420

gaatatttag taagtgatga gagtagtaat gattgttttt attattttaa atttataata 480

ttatttttttt tagagttttt taagtattgt agataatttt ttatttatta aaaaataaat 540

tgtaattata agtatttagg gttgatattg tttttgaatt agatagtgtt tatattagtt 600

gtataagatt aatttaaagt agaggatgaa attttttttt tgaatttttt ttagaatgta 660

atttagtgaa tatattaaaa ttaaattttt tttgaatggg agtaattttt atggattaat 720

ttgtaatttt ttagattata tttaaggtaa tgtagaggtt gttgtattat aggttttgtg 780

attagagatt attggatttg tttttggaaa agttatttat gatttttttgg gttttggttt 840

-continued

ttttattttt tatattggtt taataatgat tattgtagag ttgttatgga agttaaatga 900

atttatgtat tataagtgat taatataatg attgatattt agtgatttat taataaatta 960

aaatatttat tattaatatg atagagaagg tgttgttaaa atagataata ggtttttgga 1020

agaggtgatt aaatggatgt aaaatttatg gattgtttat tttgtttatt tttgttgtgt 1080

tttttggttg tggtatatat atgtgtgggt ataaaattgt aaattttatg tagttgtgta 1140

gtgtatgtgt agaaggttta gatatgaaat gttattttag taatgtgttt agagaagttt 1200

tgatgttgtt ttggaagtaa gttgttgttg tttgattttt gggtgtttgg gatggatgtt 1260

tatatttgta tttagtagta ttggaagggg tttaggtttt ttgtagtata gtttattttt 1320

agattgttta gtttttttata atatatattt ttatggaaaa gggtattttt ttttgttaga 1380

aaaagtgttt tagtttggtt tgggttggtt tttattttat gttgttgtaa gtaggtgaag 1440

tttttttttgt ttttttttttt ggggtaagtg gaaaggagtt tggtaggggg tttgtagtgg 1500

tttgtatagg ggaattgggt agtgagagag ttttaggtaa ttttgggggt tgttttatag 1560

aagtaggtgg ggattgatag tggttttttg gtttagggag gagagtgtgg ttgtgggttt 1620

tttttttttag tttggaggtt gtagttgttt gagttggttt gggtgggggt ggggtggggg 1680

tggtgtggag ggtatggaga ttatggtggt gttatttggg atatttaggg ttttgaggtt 1740

ttgggtggtt tttatgtgag attgtaaatt atgataatag gtagttattt gaggttaaat 1800

aaaaatggag tgggtttttt gtgtgttgtt gttttttgtg tttttggtgg ttttttttga 1860

ggtttttggt ggttttatga gtttgtagta gttggtggtg atgttgtttt tgttttattt 1920

ttttgtgtaa gtgtgaggtt gttggtagtg tggtgtatgt tttggttgtt tttggttttt 1980

gtgtaaaatt tttattttgt ttatgtgaag ttgttgttgt tttagagagg gggaaagagt 2040

tgtgggaaaa gttggggagt gatgattgtg gtggttgggt gtgtttttttt attttttttt 2100

ttttttttttt tttttttgtt gtagtttgga gttttggttt ttttttttttt ttttttttttt 2160

ttggagttgg ttttttttttt tgttttgttt tttttttgtt tgtgtatgtt atttgttgtg 2220

gggtggttga aggggatgtt ttgtttttat tagaggtata gtgtgaaggg gaaattttga 2280

tattggaagg aatgagaata aatatttaat tatggatgta ttgaattgtg gttgggatag 2340

atattttggg aatttgaggt ggattgggtg atgaggtgag tgatttttttt ttttaatttt 2400

tgttttaggg tttttgggggg agtttgagtt gagagaattt ttaaattttt tgggaaagtg 2460

tgtgaggttt tgttggggat gttgagtgtt gggtattgag gatgtgtagt tggatggtgt 2520

gtgggtgttt gtgtttttgg ggggtgtttg gaggttgggt gttttatgtt tgagggtttg 2580

ggttgtttgg attgtagtgg tgttttttgt tttagaagat gtttttaagt tttaagggtt 2640

ttttttgagt ttgtttgttt tttttggggt tggtgtggag tttgtgtgta atggagttta 2700

tttagtagtt tagtgtgtgg tttttatttg tattttgttt ttatttggta gaggtgtgag 2760

tattggggtt ttttttatat tttttttatg atgtgtatta ttttttgatg attttttaga 2820

tggtttaggt gtgaggatgt tgatttagag ttttttggag ggttataggt gtttgggttt 2880

ttttggtgtt gggtgtgtgt gtattttaaa ggtttgtgtt ttaattttta ggtattgatt 2940

gggtttttta attgtggtga ttttatttta atagttttta tgtggtgtgg attgaatgtt 3000

<210> SEQ ID NO 21
<211> LENGTH: 3000

-continued

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 21

gatatttagt ttatgttata taaaaattat taaagtggga ttgttgtagt tgaaaagttt 60

gattagtgtt tggagattag aatgtgagtt tttaaagtat atatgtattt ggtattggga 120

aagtttaggt gtttgtgatt ttttgaagga ttttgggtta gtatttttgt gtttggatta 180

tttaggggggt tattagaaag taatatatgt tataagaaag atgtgggggga gattttgatg 240

tttgtgtttt tgttaggtgg aggtggggtg taggtagaag ttgtgtgttg gattgttgga 300

tgaattttgt tatgtgtagg ttttgtgttg attttggaag ggataggtag gtttggaagg 360

gatttttggg gtttggggat gtttttttagg gtagagagta ttgttgtggt ttgagtggtt 420

tgggttttta ggtgtggggt atttggtttt taagtgtttt ttggggatgt aggtgtttat 480

gtattgttta gttgtgtgtt tttagtattt agtgtttggt gtttttggtg gagttttgtg 540

tatttttttg gaaagtttgg gggttttttt aatttaggtt tttttgggag ttttggggtg 600

ggggttggaa gaaggggtta tttattttgt tgtttggttt gttttgggtt tttgaagtgt 660

ttgttttagt tgtggtttag tgtgtttgta attaagtatt tatttttgtt ttttttagtg 720

ttgaagtttt ttttttgtgt tgtgtttttg gtgaaaatag gatatttttt ttggttattt 780

tataataaat agtgtatata agtgggggag aagtggggtg gagggagaag ttggttttga 840

gggggaggag gaaaggagag gagttaaaat tttggattgt gatagggggg aaaggagaag 900

aaaagaaaat gagagagtgt gtttagttgt tgtagttgtt atttttggt tttttttgta 960

gttttttttt ttttttttaag gtagtgataa ttttatgtgg ataggatgga agttttgtgt 1020

ggaagttggg aatggttgga gtgtgtgttg tgttgttggt agttttgtat ttgtgtaggg 1080

aggtggggtg ggggtgatgt tgttattggt tattgtgggt ttgtgaggtt gttgggggtt 1140

ttgggggagg ttgttaggga tgtgggggggt ggtggtgtgt gggggattta ttttgttttt 1200

atttgatttt gggtgattgt ttattgttat ggtttgtgat tttgtgtggg gattgtttag 1260

ggttttgggg ttttggatgt tttgggtggt gttgttgtaa tttttgtgtt ttttgtgttg 1320

tttttatttt gtttttattt gggttgattt gagtggttgt agtttttagg ttgaggggag 1380

ggatttgtga ttgtgttttt ttttttgggt tggagagtta ttgttgattt ttatttgttt 1440

ttgtggggta gtttttggaa ttgtttggaa tttttttgtt atttagtttt tttgtgtagg 1500

ttattgtggg ttttttgttg gatttttttt tatttatttt aagggaggag atagaaggga 1560

ttttgtttat ttgtaataat gtgaaataaa aattaattta gattagattg gggtgttttt 1620

tttgatggga ggaaatattt ttttttgtgg agatatatgt tatgaaggat taagtggttt 1680

ggggataggt tgtgttgtga agggtttggg ttttttttag tgttgttggg tgtaggtata 1740

ggtatttgtt ttagatgttt aaaggttagg tagtaatgat ttatttttaa ggtggtgtta 1800

gagtttttttt aggtatattg ttgaaatgat attttgtgtt taagtttttt gtgtatgtat 1860

tatgtgatta tataggattt atgattttat atttatatgt gtgtatgtta taattagggg 1920

atatagtaaa ggtagatgga ataaataatt tataaatttt gtatttattt aattattttt 1980

tttaaaaatt tattgtttat tttagtggta tttttttttgt tatattgata ataaatgttt 2040

taatttattg ataaattatt gagtgttaat tattgtattg gttatttatg gtatatgagt 2100

-continued

ttatttaatt tttatgataa ttttatggtg gttattatta agttggtata aaagatgagg 2160

aaattaaagt ttaaagaatt atgagtgatt tttttaaaga taaatttagt ggtttttaat 2220

tataaagttt atgatataat aatttttata ttattttagg tgtggtttaa gagattatag 2280

attaatttgt agaaattatt tttatttaaa gaaagtttag ttttaatata tttattaagt 2340

tatgttttga aaaaggttta gaaaaaagat tttatttttt attttaggtt ggttttatgt 2400

aattgatatg agtattgttt aatttaaaag tagtattaat tttgaatatt tatggttata 2460

atttattttt taatgagtgg ggaattattt ataatgttta agaggtttta gaaaaggtgg 2520

tgttgtaaat ttaaaatgat aaaggtagtt gttgttgttt ttattattta ttgggtgttt 2580

ttttttgtag atttattttg gagggtgtag gtattgttag gtataatgtt tttttttttg 2640

gtttttaagg tagaaagggt tttggtagtg gggttggtat aaagtggatt tggagtatgg 2700

ttgagagtat attttggttt aatgaggaat gttagttaat atatattgga gagaaaaata 2760

tgaatggata gatttaatta attttgtaga tttatttttt ttattttata tgtgagaaaa 2820

ttaagggttt agaattttta gaagtttgtt aataaatggt agatttgagt tttaagttag 2880

tttttattta agtttatttt tagttattgt gatgttatta ttagtattaa tggtatttag 2940

aaaaatagta atttttttaa atatgtaata tattataagt taaatatgta atttaaatgt 3000

<210> SEQ ID NO 22
<211> LENGTH: 1984
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (9, 17, 18, 33, 41, 49, 50, 52, 54, 62, 80, 84, 127,
157, 159)
<223> OTHER INFORMATION: unknown base
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (175, 207..209, 214, 1181)
<223> OTHER INFORMATION: unknown base

<400> SEQUENCE: 22

agttttagnt attattnnta tttttattaa tgnttttggg nttataggnn gngnttttttt 60

gnttaaatag tagttaatgn aggnatttta aaaggtatat tattttatat ttatttagtg 120

gttatgntat tatttaattt ttattataat tttgtgngnt tagtataatt tgtgntttta 180

ttttatttgt gaataaatgg aggtagnnnt ttgngtaatt tataagtaag tggtagattt 240

gggggttgtg tttagataat gtggttttag atttttttgtt ttttttttgtt ttttgttgta 300

ttttgattga gtttagttat ttatttttttt aatatatatt aattgttgtt tttatgtttt 360

aggttttgtg ttagatgtta gaaatatatt tttatttatt ggattttggt atggtttatt 420

gtttagtgtt ttgttatgtt atgtgttata taatgtattt attttggtat ttagtttatg 480

gaaaataatg ttgaaagata ttgtatgtat gtttttataa taaaattatg taatttatgt 540

ttttttttatt gttttggtta ttaggaagtt tttatttaaa ttagagtaaa tatattagaa 600

tttgggtttg ttatttttagt ttgttgaatt tttttttttg gtttagattt tttattttgg 660

tttataaatt ttattgtata aatgtttttt attgtaaaat attttaaatt ttttttaagg 720

gaaggttgta tggaaatgat atagtaaggt tttttttgta ttttttttaga tttttattag 780

-continued

ggaaggtaga tttagatgtt ttttttattt ttagttttaa agtttttttat ttataaaatt 840

ttttaagtag tgattattgt tgttttgagt atttggaggt agtttagagt ttttgaaagg 900

taaaatttat atataaaaag aagtttttttt tgatttagat tttagttttg gtgttagatg 960

tatatttgag gagtaggtgg ttagttagat ttttatttga gtagttaata aaatttattg 1020

ttttttaatg gaatttttttt tgtaagtttg aattgatttg ttatttttgt gatttgtgag 1080

atggtaggga agtattaaat attattatga tttgggttat agtggtggga aaaaaggaaa 1140

aaagaaaaaa aaaatttatt gttaagtttt gttaggtgta naaagggttg gaattgttgg 1200

ggttatttta tttgatttat tggaaataga gtggatttta ttaatatttt aataaagaga 1260

atttttttgt attaggttgg aagtggttgt tagttttttg tgtaatttta ttttttggaa 1320

aagtggaatt agttggtatt gtttagtgtg atttgtgagg ttgagtttta atagtttaaa 1380

gaagtaaatg ggatgttatt tttgtggggt ttgttttttg tgaggtgttt attttgtatt 1440

tgttatgtaa aatgagggag tgttaggaag gaatttgttt tgtaaagtta ttggttttgg 1500

ttattagttt ttatttaatg tttttgtgat gttgttgttg atttatttgg gaagttggtt 1560

ggttggtgag gtagagtttt tttttaaagt ttggttttta tggaaaatat gtttagtgta 1620

gttgtgtgta tgaatgaaaa tgttgttggg tgtttttagt tggataaaat gtagttgaga 1680

attttgtttg ttttgtgtgt ttttttgttt taggtaggga agaggggttg ttgggtgtgt 1740

tttgtgtttt gtttttgtat ttggattgtt tggtatgggt agggtgaggg ggtttttggg 1800

gggttggggt ttttggttgt ggtggtgaag atagattggg gtttggtagg gaggttattt 1860

tgagtttaga gattttaggt atttttttata tataggtttt tattttggtg tgtgtgtgtg 1920

tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtatgtt tgttaatggg 1980

agga 1984

<210> SEQ ID NO 23
<211> LENGTH: 1984
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (804, 1771, 1776..1778, 1810, 1826, 1828, 1858, 1901, 1905)
<223> OTHER INFORMATION: unknown base
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (1923, 1931, 1933, 1935, 1936, 1944, 1952, 1967, 1968, 1976)
<223> OTHER INFORMATION: unknown base

<400> SEQUENCE: 23

tttttttgtt aatgaatgta tatatatata tatatatata tatatatata tatatatata 60

tatatatgta tgtatgttag agtgggagtt tgtgtgtggg gggtgtttag gatttttggg 120

tttggaatga ttttttttatt gagttttgat ttgtttttgt tgttgtgatt ggaggttttg 180

atttttgaa agtttttta ttttgtttgt gttgggtgat ttgaatgtag aaatggggtg 240

tagagtgtgt ttggtagttt ttttttttta tttgggatag gagaatgtat agaatgagtg 300

gagtttttgg ttgtattttg tttgattaga agtgtttggt ggtgttttta tttatgtatg 360

tggttgtatt gagtatattt tttgtgggag ttaggttttg aggagaggtt ttgttttgtt 420

-continued

agttagttaa ttttttaaat agattagtag tagtattatg aaagtattgg gtagaggttg 480

atgattagga ttaatggttt tataagatgg attttttttt aatgtttttt tgttttgtat 540

ggtagatatg gggtgagtat tttgtgagga gtgagttttg tggaggtggt attttatttg 600

ttttttgga ttgttggggt ttagttttat aaattatgtt gggtaatgtt agttgatttt 660

attttttag agaatggaat tgtatggggg attggtggtt atttttagtt tagtgtaaaa 720

agatttttt tattaaaatg ttaataagat ttattttatt tttaataaat tagataaaat 780

ggttttagta gttttagttt tttntatgtt tggtaaggtt tggtagtgga ttttttttt 840

ttttttttt ttttttttat tattgtggtt taagttatga tggtgtttgg tgttttttg 900

ttattttata aattataaag atgatagatt aatttgagtt tgtagaaaaa attttattaa 960

ggggtaatag attttattaa ttgtttaggt gagagtttga ttagttattt attttttaaa 1020

tgtgtatttg gtattaaagt tgggatttgg attggaagga attttttttt gtatatggat 1080

tttatttttt agaagttttg aattattttt aggtatttag aatagtagta gttattgttt 1140

aagggatttt ataaatgggg ggttttggga ttgagggtaa gagggatatt taggtttgtt 1200

ttttttaata ggaatttaag aaaatgtagg gaagatttta ttgtattatt tttatgtagt 1260

ttttttttag aaagaattta aggtgtttta taataaagga tatttgtgta atagaattta 1320

tgaattaaaa tagaaaattt gggttagaag gaaaagttta gtaaattgaa atgataagtt 1380

taaattttaa tgtatttgtt ttgatttgga taagagtttt ttgatagtta aagtaataaa 1440

agaaatataa gttatatagt tttgttgtag aaatatgtat ataatgtttt ttagtattat 1500

tttttatggg ttaaatattg gaataaatgt attatgtaat atgtaatata gtaggatatt 1560

gagtaatgga ttatgttaag atttagtgga tggggatgta tttttagtat ttggtatagg 1620

gtttggaata tggagatagt aattagtgtg tgttaaagaa atagatggtt agatttaatt 1680

aagatgtagt agaaggtagg gaggggtaag gagtttggag ttatattgtt taggtatagt 1740

ttttggattt gttatttatt tgtaaattat ntaaannntt gtttttattt gtttatagat 1800

aaaatgaaan tataaattgt attgantnta taaggttgtg gtgaggatta agtgatgnta 1860

taattattga ataaatataa agtgatgtgt tttttaagat ntttntattg gttgttgttt 1920

ggntagaagg ntntnntttg tggntttaag gntattgata aagatgnnag tgatgnttga 1980

agtt 1984

<210> SEQ ID NO 24
<211> LENGTH: 7833
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 24

gtttttggtg agatatgtgt tttataagtt ttaatggaga aaaatgtaag tattttattt 60

tttgaaattt ggttatttga gtaatgagaa aatagttatt ttttttagga tagtggtttt 120

taattatggt tatgtgtttt tttaggaaaa ttttaaaaat atatatatat taatgttttt 180

gtgttatttt tagggatttt aagtttttga atatgaattt tgtattagta ttttttaatt 240

atttaggtga ttgtgatgtg aaattatgat tgagttttat tgttttaaga tgaaataaat 300

ttttttttagt attgaaatta taaatttaaa ttattaaaat taattaaggg tatgggaatt 360

-continued

aataaggtat agggaagttt ttatattata aaattatttt tttaaattat agtttattgt 420

ttatatgtta tttgttattg tagaaaaggg tgaaaaaata gtaaatttaa ttatttttag 480

tttgaaaaat tatttagaaa tgaagatgat gattttgaaa tattgttaat attatttgat 540

ttataaataa tgttttaata tatttattat atattgatag atatttttt atatgaatat 600

tatatattaa aattaaggta ataatgtatt tagaatattt tatttatatt tatgtatttt 660

aagtaggtta gaaattaaga tatgagttat taagtatgag atgttaaggt gtggggttag 720

aaattatatt gtattttatt attaataatt aatatatatt ttaatattat atatatttaa 780

ttttaatttg tatattttta attattttta attatgtgta taaatataag tatatatatt 840

tttatgtatt tatttattta tatttttatt tatttattta tataggggat ttttttaaat 900

ttattattat taaattatat atttttattt taatttttag aataagttta ggaggtaggt 960

attgttatta tttatatttt ataaatgagg aaattgttta tagttataaa gttattgtgt 1020

tagatatatt agaagtttaa tatatatttg gtgaatatat gtataaaaat agagagatag 1080

atatgtataa tagtttattt ttatattgag taaaagtttt taatttgttt tagaaatttt 1140

tttgtgaaaa ttgagtaaaa attgaggtat ttttttattt gttatatagg tataggtggt 1200

attttatttt tttaataagg atgaatattg aaatgtggat tttaaggttt aattttagat 1260

tttttgaatt tttgatagtg ggatttggaa tttgtttatt gttttaaagt tttttaagga 1320

atttatatga ttaattaggt ttagaaatta ttggatttta ttgttgaagt ttgagaatta 1380

aagtttgggt tttattgtgg ttttatagaa agggtaaatg aagtattatg gatagaattg 1440

atatgttttt agttagtttt tttttttaga agttaatagg tagtaatata gtagaaatta 1500

gtgatttatg ttttgtgttt tgaagttagg tagaatttta tagagtttta gtagtgttat 1560

tgatgagatt tgtttttttgg ggtaagttgt ttgatgtttt taaagttata tttttttat 1620

ataaaatgag ataatatttt ttgttttata ggggtgtttt aaagattaaa taaaaataat 1680

atgttttatt ttatatggta taatgtttga tatttaagaa gtaaaggata tattttattt 1740

ttattgaagt aattagaaag tatgaaatta tgaaggagat aagagttttg attggtagtg 1800

tattttattt ttttaggttt atttatttat tttaaattat ttttgttgga gaataatttt 1860

taagtttttt atttaagttg tgagtaattt tatattttat aatgatgttt tttttatgag 1920

aaaaaaaaat gtttttaagt tttttggaga aaatatattt gtattatttt tattgaaaaa 1980

tttaataatt ggattttgtt tttttgtatt aattttagag tgtatatgtt ataaataaag 2040

tgttttagtt taagaagatt gaaagtaaat atggtatagt attttaaaat aagaattttg 2100

taaatatatg gtatgattgt gttatattat tagtaattat atgatatgta atgtaaagta 2160

tagtttatag atttaaattt aattttaata agtaaattga ttttgttttg ttggggaaaa 2220

gttaaagtat taatttaatt gttaatgtag ttttgtttat ttttttggta tttagtgata 2280

agtttaaata atgtatatat ttttatttat atatttagta atataatttt ttgtttaatg 2340

agtgatgttt ttttgttatt tggtggtgtt tgttagtttt agaatttgtt ttttggtggt 2400

attataatat taagtataga gtaagtgtaa taaaattgta gtatttttat tgaaaaggtt 2460

ttgttttaaa ttgtttaata atttaaagga tttttgtgga agtaattgta tttgttaatt 2520

agttataatt agtaattaat ttttttggag ttttaattta tttttggtaa aatgttttag 2580

gaagagtata tattattaga aagtatgtta aaaatttatt tagtagaaaa tttaaaaata 2640

-continued

gtttttttttt gttaagaggt tttttaaaat tttatttata tagttaaatt ttgaaatttt 2700

agtaggtttt gttttattat tataattatt gtataaatat ttttaaggat tttgttttta 2760

gttttaagta tgatttattt ttataagttt gattagttat tatattagtt ttgttatgga 2820

aaatgatatg tttttatttt ttgttgtaga gttgttaaat tttgatttat atttatgttg 2880

tttttttttgt tgaaagtttg tagtgaaaga aatttttaat tttttgtttt gtaatattag 2940

ttggtagttt tatttaatgg gtattttgtt tttttaaaga atttagttgt tttgtttaga 3000

agttgatttt ttgatgtttt taatgtttgg tttaattgat ttgttttaat ggagtttttg 3060

ttggtgagga gtgagatgtt attgattaga atgttgggat ttgttgttta attgttagga 3120

gtgagagata ttgagattta gaaatttttg gaggtgggag gggagaggga tagttttgga 3180

tggaggtgga gatgtaagat aaagggatgg attttatata ggaaaaaaaa aaagattttg 3240

ttgaggtatt gaggtgttgt atgattatat tttttaaagg agaagttaaa aagtaaggaa 3300

gtgggaggag gttggaggtt aaagtattta aaaggattat ttgggtataa tttgtttttt 3360

tgttggtgtt tgtaaaggat agatagtttt gtttttaaag tatatgaatg tttttttaa 3420

gtgattggga atggatatta attgtttgtt aaatgttatt aaatgttttt ttaaatttag 3480

gggatataga aagaggggta taaaaggaga atttaaatag aaaaagggag gatttggagg 3540

tttttgaaag tgggggggaga agaaggagga gggataatag agaggaatag agaaggagag 3600

tggagagaag ataaataaaa ataaaaatag gaattattga ataattatat attaaaaaga 3660

aagtttttttt ttatggggta tttaaaatat tgagattgta atagtgattt tggttatgga 3720

agaaagatgt ttttttttat ttttgttttt gaaagttttt ggttttgtta ttggtgatta 3780

aaattttatt aggttaaaga gtgtgtttaa ttgtttgaag aatgtagtag atggaaggtg 3840

ggttttgtta tgttgtttgt ttttttgtt ggagagaatg aaagaaatgt gtagagttag 3900

agatttttgt tgagttagat tttttttgt tgttttaggt tattggttat ttggtaaaga 3960

tttgagtaag gaatgtaggg ttattgtttg ggttaataaa tggagtttgt tttttttttt 4020

ttggatgttg ttgtttggtt gatgtttttg gtaatttatt tgtggtgtat gtagaggagt 4080

tttttttttt tttttagatt atttgttttg attaatttga ttttttaaat atatttgatt 4140

gtatttttta ggtggatata ttaataggtt atgggttgga gaggagtggg tgatgaggag 4200

agggatttaa atttgtgaat gtttgggttg ggttggagtt gtgggggggtt tgggaggaga 4260

gaggggagaa gagagaagga aggagagtgt ttgttgggat ggttgagttg ttttggtgag 4320

tagttttggg gttgtatgtt tttgtgggag atgttgttgt tgtttttagg ttggtaagag 4380

tggttttaat attattgttt tttatttttt tttttgtaaa tttttagaga aatgtttttg 4440

gtttttttgt tgtgatattt ttagtttgta tttttttata gtttaggtgg tgtgtttttg 4500

tatgttggag tgttggttgt tagtaggatg ttttttttttg tgttgatttg ttttttttttg 4560

ttttgttgtt gttgtttttt tgatattttt gtttttatta tttttagttt ggagagatgt 4620

tatttagttg tggtttgtat ttgtggtttg gggttatgtg tggaagaggg gtgttagttt 4680

ggattttgtt tttggtaggg ggtgttttgg agtggagagt gaggtgaatg gtatatgagt 4740

gtgtgggtag tttattttga agtttgagtt ttttatttga gttatgtttt gtttagtttt 4800

atttgggtta gtgtttggtg agtgagttta tttgtggttt ttgtggttgt tttttttttg 4860

tatttttgta tttatttgtt gatttttttt ttttgggatt tgtattttgt tttattaatt 4920

-continued

agagtttgat tgtttttttt tatgtgattt tgggtgggtt gaggatttgt tgtttttaa 4980

atgttagagg gatgtgggtg gtagagtttg agaggtggtt gttgggttgt ggggtgtttt 5040

gatttttttt ttattttgtt tttttgggtt ttatttgttt gtttttggat ttttgttttt 5100

ttttgttttt tggtttttta gagttttttt tttatggtag tagttttttg tgtttttggt 5160

gtagtttttt agtggatgat ttttttgttt tggggttgag tttagttttt ggatgttgtt 5220

gaaatttttg agattatgtg tgggtttggt tgttgttttt ttgttgggtg ttattgttat 5280

tgttgttgtt tttgttgttg ttgtttgtgg gatgtttagt agtttgttgt ttggtttttg 5340

tgattttgtg ttttttggaa gttgtttgtt gttgtagagt tgtatgaatt agttatggtg 5400

ttgtgggagt ttttgtggta gtgtagtagt tggatatttt gtgagggttt ttgttggttg 5460

ttgttgttgt ttgttatgtt atttattgta gtttgtttgg tgaagtttgt tgttttttt 5520

attttttaa gtgattgtta aatgtttatt ggttggaatt gttttggtaa gtttagaatt 5580

tttgtttttg attttttaat tttgtagaag aatatgtgta tttagtatag attagtttat 5640

tttagtgtgt ttttttagtt ttttattttt tattgtttta gatttttaat attatttatt 5700

tttatttaga gaaataaggg gaattgttgt aggtttgggg gtgaggggtg gttttgggat 5760

gggtagaaag tgtaggtgta gtaggaaatt tttgtatgtt tgtgtttata ttggagttgt 5820

gaggattttg agaaatatta aatgggatgg ttttttgggt ttattgtttt gaaagagtat 5880

taattttagg ggaaatattg aaatagaagt tttgttatta ttaaagaaaa aagttttatt 5940

aggatgagga agaaataatt ttatgagaaa gaatgagtga gaaagtaata aattaaatgg 6000

tgattgtagg ggaattgttg atttttggta aaggtgttat gaggttgtat tggttttttg 6060

ttgaagatta ggttatatag attttagagg agttgggttt taatagaatt ttttttttt 6120

ttttttttt ttttttttt ttttttttt ttttttatt tatttatttt ttttttttt 6180

ttatttttt ttttttagg tggtaaaaga tattggtttt gtagtttaga tatgtttttt 6240

ttttgtttt tttaagtttt aaggtagtat aggggagttg agaaaaagaa tattttgtgg 6300

gtttttagg ttggagtggg tatgattgag gttggttagg ttttatgtag gtgagttgag 6360

ggtggaattg attttagtgg gtgttgattt ttttatttt ggataggttt ttgtggagtg 6420

ggttaggtat ttttttgtt tgtttgggtt tttttagatt ttgatggtga atgtttggta 6480

ggttttgttt tgttgaagtt ttttaattaa atagggttag aggatgggag ttgttgtatt 6540

tttagttggt atagtatttg gtttgatagt ttgtagtata gggtgtatgt aatttttat 6600

tttttgtgaa tataattttg ttgtagttaa atttggtttt gaataaagtg ttttttaaag 6660

atgtatataa gttgaagtgt atgtaatttt agagaggagg gaatgattaa ttgtaattta 6720

gggtgaaagt ttgtatagtt tttagttatt attgatgtaa atgttaaaag gaaaattatt 6780

atgtattatt ttaatttatt ttttataaag ataagttgag atatgtaatt ttattagatt 6840

tgggttaata gattgttttt ttttttggta gtttttaaat ttggtatttt aataaaattt 6900

aatatgtttt tataattttt tgatttatgt gtatatgtgt gttgtttttg aaagaataag 6960

ttttattttg ttattgttta attatttttt agatgtttta ttatggtaat aattatgagt 7020

ttgtaaaaat aatttttgga aatgttgatg gttttgtagt ttaatataga ttggtttgtt 7080

ttatttttag tttttgtatt gttttaggaa ataattaatt taaatgtgaa gttgatattt 7140

gtaattaaga aattatatat ttattagata ttttaaaggg gattgtataa attaaagaga 7200

-continued

ataaattggt tttgtagata ggttgttaag aatttggtat tttgttttta tttttgttaa 7260

tttagaggtg attaattttt atttgagtta aatagattat tatagaaaat attgtgtttg 7320

tttattttta ttattgaggt tttgtttttt ttttgtttgg atatatttta aataaggggt 7380

tgttttagtt gttgaagtaa aagaataatt aaagatgggg aaatggtaaa agggtattta 7440

gagattatta ttagtttttt tttaaaatgt ggagttttgt ggttataaat attgtttatt 7500

taatgagtaa aaaataaaaa taaaaaaaaa ataggaagta aatgttaagt ttttatttat 7560

tattgttagt attaatgtaa gttttaaaaa atagtattat tagaaaagga tattaaagga 7620

gaattgatta gaaaagaatt gtggaaaatg gaaatgaata ttgattattt aattagattt 7680

tgaggttatt agtagatagt gattttgtag tatagttata gttgttggat ttaaaattta 7740

ggataagtat tttaaagttt taaagtagtg ttttttttttg ttaaaaattt gtaagatgtt 7800

ttaatgattg gagtgttttt tttgaatttg agg 7833

<210> SEQ ID NO 25
<211> LENGTH: 7833
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 25

ttttaaattt aaagagaata ttttagttat taaaatattt tatagatttt taataaaaaa 60

aagtattatt ttgaagtttt aaaatatttg ttttaaattt taaatttaat aattatagtt 120

gtattgtaag gttattgttt attgataatt ttaaaattta gttaagtgat taatatttgt 180

ttttattttt tataattttt ttttagttaa ttttttttta gtatttttt ttgatagtgt 240

tatttttaa agtttgtgtt aatattgata gtggtgaatg aaagtttaat atttgttttt 300

tgttttttt ttatttttat tttttgttta ttaggtggat aatatttatg attataaaat 360

tttatatttt ggaaaagagt tagtgatgat ttttgaatat ttttttatta tttttttatt 420

tttaattgtt tttttgtttt aatgattgaa ataatttttt atttgaaatg tatttagata 480

aagaggaaat aaagttttaa taataaagat aaataggtat agtgtttttt gtgatggttt 540

gtttggttta aatgaagatt gattattttt aagttaatag gggtggaagt ggggtgttaa 600

gtttttgata atttatttgt aaaattagtt tatttttttt agtttatgta gtttttttta 660

aaatatttgg taaatatgta atttttgat tgtaaatgtt aattttatat ttaagttagt 720

tatttttaa aataatgtaa gggttaggaa tgaagtaaat tagtttgtgt tggattataa 780

agttattaat atttttaaaa attgtttttg taggtttata attattatta taataaagta 840

tttaaaaagt gattaggtaa tagtaaagtg aaatttattt ttttaaaaat aatatatatg 900

tatgtatgaa ttaagaagtt atagaaatat gttgagtttt attaaaatgt taaatttaga 960

aattgttaaa aaagagaata atttattgat ttaaatttaa tagggttgta tattttaatt 1020

tgtttttgta aaggataaat tagaatgatg tataataatt tttttttgg tatttatatt 1080

agtaataatt aggaattata taggttttta ttttgagtta tagttggtta ttttttttt 1140

tttaaagtta tatatatttt agtttatata tatttttgaa agatatttta tttagagtta 1200

gatttaatta tagtaaaatt atatttatag aagatgaaaa attatatata ttttatatta 1260

taggttgtta aattgaatgt tatgttagtt aggagtgtag taatttttat tttttggttt 1320

-continued

tatttaatta ggaagtttta gtagagtgaa gtttgttaag tgtttgttgt tagaatttga 1380

aggaatttga gtgagtaaga agagtgtttg atttatttta tagaagtttg tttagaaatg 1440

gaggagttag tgtttattga agttggtttt gtttttggtt tgtttatatg gagtttgatt 1500

agttttagtt atgtttattt tggtttggga gatttgtaaa gtgtttttt ttttaatttt 1560

tttgtattat tttgaagttt agggaagtaa agagaggggt atatttggat tgtaaaatta 1620

atgttttttg ttgtttagga gagaagggaa tgagagagag agagagatag atagatagag 1680

agagagagag agagagagag agagagagag agagagagag agaaatttta ttgaaattta 1740

gtttttttag aatttgtgtg atttggtttt taatgggaga ttagtgtgat tttatggtat 1800

ttttgttagg aattagtgat ttttttgtag ttattatttg atttattgtt tttttgttta 1860

ttttttttta taaagttatt tttttttat tttagtaaga ttttttttt taatgatgat 1920

aaagtttttg ttttagtgtt ttttttagga ttggtgtttt tttaaaatag tgaatttaga 1980

aaattatttt gtttaatatt ttttaaaatt tttgtagttt taatgtaagt gtaagtatgt 2040

aaaggttttt tgttatattt gtatttttg tttattttag aattattttt tatttttggg 2100

tttgtaatag ttttttttgt ttttttggat agaggtgggt ggtattaggg gtttagggta 2160

gtaggaggtg aggggttgag gaggtgtgtt agggtaggtt ggtttgtgtt ggatatgtgt 2220

gtttttttgt ggagttaaag ggttggggat gggggttttg gatttattag agtaatttta 2280

gttggtgggt gtttggtagt tatttaagga ggtagggaaa gtagtgagtt ttattgggtg 2340

ggttatgatg agtagtatga tgggtagtag tagtagttag taaaagtttt tgtaaagtgt 2400

ttagttgttg tattgttgtg gggatttta tagtattatg attagtttgt gtaattttgt 2460

agtagtaaat ggtttttgag gaatatagga ttgtgggggt tgggtagtgg gttattgagt 2520

attttgtgga tggtggtagt agaggtggtg gtggtggtag tggtatttgg tggggaagta 2580

gtagttaaat ttgtgtatga ttttgagagt tttagtaata tttagggatt gggtttagtt 2640

ttggagtgag agggttgttt gttgagaagt tgtgttggag atgtgggaag ttgttgttat 2700

aaggagggag ttttgggaag ttggaggata ggaggagatg ggagtttagg ggtagatgag 2760

tggagtttga ggaggtaggg tggagggaga gttaaggtgt tttgtagttt ggtagttgtt 2820

ttttgagttt tgttgtttgt attttttgg tgtttgggaa gtagtaggtt tttagtttgt 2880

ttggggttat gtgggaagag gtagttgggt tttgattggt ggagtaggat gtaggttttg 2940

ggagggaggg gttgatgagt aggtgtaagg atgtaaggag gaggtggttg tggaagttat 3000

agatgggttt gtttgttagg tgttggtttg agtggggtta ggtggggtat ggtttaaatg 3060

agaagtttgg gttttagggt gggttatttg tatatttata tattatttgt tttatttttt 3120

gttttaggat gttttttatt gaaggtgggg tttggattag tgtttttttt ttgtgtgtga 3180

ttttgggttg tgagtgtggg ttgtggttgg gtggtgtttt tttgagttgg agatggtggg 3240

ggtggaggtg ttagaggagt agtagtagta gggtagagag gggtgagttg gtgtgggaga 3300

gggtgttttg ttggtgattg gtgttttagt gtgtgggagt gtgttgttta ggttgtaggg 3360

ggatgtaggt tgggaatgtt gtggtggaga ggttagggat gttttttttag ggatttatag 3420

gaaagagggt gagaggtgat ggtgttagaa ttgtttttgt tgatttggaa gtaatagtag 3480

tatttttttat aagagtgtgt aattttaagg ttgtttgttg aggtagttta gttattttgg 3540

taggtgtttt ttttttttttt ttttttttttt ttttttttttt ttaggttttt tgtagttttg 3600

-continued

atttagttta agtgtttgta ggtttgaatt ttttttttta ttatttgttt ttttttagtt 3660

tgtagtttat tagtgtgttt atttgggagg tgtggttaga tgtgtttgga aggttagatt 3720

ggttgggata agtggtttga gagaaagaga aaggtttttt tgtatatgtt gtgggtgggt 3780

tgttgggagt attggttggg tagtggtgtt tgggaagggg agagtgggtt ttatttgttg 3840

gtttaggtag tgattttgtg ttttttattt gggtttttgt tggatggttg gtgatttggg 3900

gtgatgagag aaggtttaat ttggtaggag tttttggttt tgtgtgtttt ttttattttt 3960

tttagtggga agggtaaatg gtatagtggg atttgttttt tgtttgttgt atttttttagg 4020

tagttagata tatttttag tttaatggaa ttttagttgt tagtaatggg attaagagtt 4080

tttggggata agggtggaga ggaatatttt ttttttatga ttggggttat tattgtagtt 4140

ttagtgtttt ggatgtttta tagggaagag ttttttttt ggtgtgtgat tatttagtga 4200

tttttgtttt tgtttttgtt tatttttttt ttgttttttt tttttatttt tttttgttat 4260

ttttttttt ttttttttt ttgttttaa aagtttttgg atttttttt tttttattta 4320

aatttttttt ttgtgttttt ttttttgtgt tttttgaatt taggagagta tttgataata 4380

tttaataggt aattagtgtt tatttttaat tatttaaaag aggtatttat atattttgaa 4440

aatgggatta tttatttttt gtagatatta gtagaaaaat aaattgtatt tgagtaattt 4500

ttttaagtat tttaattttt aattttttt tatttttttg ttttttaatt tttttttga 4560

gagatgtgat tgtgtagtat tttagtgttt taatgaaatt ttttttttt ttttgtgtga 4620

aatttatttt tttatttta attttgttt ttgtttgaga ttgttttttt tttttttat 4680

ttttaaagat ttttgaattt tagtgttttt tatttttggt aattaagtag tagattttag 4740

tattttagtt ggtggtattt tgttttttat tgatgaagat tttattaaaa tagattaatt 4800

agattagatg ttggaggtat tagaaaattg gtttttagat agagtagtta aatttttaa 4860

ggaaatagaa tatttattag atagagttgt taattaatat tgtaaaataa ggaattagaa 4920

attttttttg ttataggttt ttagtagaga aggtaatata aatatagatt aagatttaat 4980

aattttatag tagagaatga gaatatgtta ttttttatag taaggttggt gtggtaatta 5040

attaggttta tgaaaataag ttatgtttga aattaaaggt aaagttttta aaagtgttta 5100

tgtagtaatt atgataatga aataggattt gttaggattt tagagtttgg ttatgtaagt 5160

agaattttag agaatttttt agtagaggaa aattgttttt gaatttttg ttaagtaaat 5220

ttttggtata ttttttaata atatatgttt tttttaagat gttttgttaa aagtaagtta 5280

aaattttaaa ggagttaatt attggttgta attggttaat aaatgtggtt gtttttatag 5340

aggtttttta aattattaaa tagtttgaag taaagttttt ttaatgggaa tgttgtaatt 5400

ttgttgtatt tattttgtat ttagtgttat agtgttatta agaaataaat tttgaaattg 5460

gtaagtatta ttaagtggta gaagaatatt atttattgag tagagaattg tattattgaa 5520

tatgtaaata aaaatatata tattatttag atttgttatt aggtattaaa gaagtagata 5580

agattgtatt agtaattgga ttagtgtttt aatttttttt tagtaaggta aaattagttt 5640

atttattaga attaaattta agtttatgaa ttgtattttg tattgtgtat tatatgattg 5700

ttagtaatat gatataatta tattatgtat ttgtaaaatt tttattttaa aatattatat 5760

tatatttatt tttaattttt ttgagttaga atattttatt tgtggtatat atattttaga 5820

attgatgtag aggagtagag tttagttgtt agatttttta gtagaaatag tgtagatata 5880

-continued

tttttttag aaaatttaag aatatttttt ttttttatgg aaagaatatt attataaagt 5940

gtgagattat ttatagttta agtagggggt ttgggagtta ttttttaata agaatagttt 6000

aagataaata aatgaatttg ggaaaataag atatattgtt aattagaatt tttatttttt 6060

ttatgatttt atatttttg attgttttaa taaaggtaag atgtattttt tgttttttag 6120

gtgttaggta ttgtgttatg taggatagaa tatgttattt ttatttaatt tttaaaatat 6180

ttttatgaga taaagaatat tattttattt tatataaaag gaatatggtt ttgaaagtat 6240

taggtaattt gttttaagaa ataaattttg ttagtgatat tgttgggatt ttgtgaaatt 6300

ttgtttgatt ttagagtata agatataagt tattaatttt tgttgtattg ttgtttgtta 6360

gtttttgaga ggggaaatta attgggaatg tattagtttt gtttatgata ttttatttgt 6420

ttttttttgtg gagttgtagt aaggtttaaa ttttaatttt taaattttgg taataagatt 6480

tagtgatttt tgaatttggt tgattatatg aatttttga gaaattttga aataatagat 6540

aaattttaag ttttattatt agggatttag aaaatttgga gttgggtttt gggatttata 6600

ttttaatatt tatttttgtt ggagaagtaa ggtattattt atatttatat gataaatgaa 6660

aggatatttt gatttttgtt tagtttttat agagaggttt ttgagatagg ttaaaagttt 6720

ttatttagtg taaagatgag ttgttgtata tgtttgtttt tttgttttta tgtatatgtt 6780

tattaaatat gtattaagtt tttaatatgt ttgatatagt aattttgtga ttgtagataa 6840

ttttttttatt tgtaaaatgt gagtaataat aatatttgtt ttttgggttt gttttaaaga 6900

ttaaaataaa aatgtatggt ttaatggtag tggatttggg gggatttttt atataaataa 6960

gtgaatggag gtatgaataa ataaatatat aaagatgtgt gtatttatat ttatatatat 7020

aattaaaaat agttaaagat gtataaatta aagttaaatg tatgtgatat tgaagtatat 7080

gttgattatt gataatgaag tatagtataa tttttaattt tatattttaa tattttatat 7140

ttaataattt atattttaat ttttagttta tttaagatat atagatatag atagaatgtt 7200

ttaaatgtat tattgtttta gttttaatgt ataatattta tatgaaaaag tatttattag 7260

tgtgtagtaa atgtattaga atattattta taggttaaat gatattgata atgttttaga 7320

gttgttattt ttatttttgg ataatttttt aaattgagag taattaaatt tgttattttt 7380

ttatttttttt ttataatggt aaataatata taaataatga gttgtgattt aaagaaataa 7440

ttttataatg taaaagtttt tttatgtttt attgattttt atgtttttaa ttaattttgg 7500

tagtttaagt ttgtgatttt agtgttgagg aaagtttatt ttattttaga gtagtggggt 7560

ttagttatga ttttatatta taattatttg gataattaaa gaatattgat gtagagtttg 7620

tatttaaaga tttggaattt ttagaagtga tatagaagta ttggtatata tatattttta 7680

aagtttttttt ggagaaatat atagttatga ttgagaatta ttgttttggg gaaagtgatt 7740

attttttttat tatttaaata gttaagtttt aggaggtaaa atatttatat ttttttttat 7800

taaaatttgt aaaatatata ttttattaaa gat 7833

<210> SEQ ID NO 26
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002044:2092P22

<400> SEQUENCE: 26

-continued

ggataggagt tgggattaag at 22

<210> SEQ ID NO 27
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002044:2462O22

<400> SEQUENCE: 27

aaatcttttt caacaccaaa at 22

<210> SEQ ID NO 28
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002064:1038P21

<400> SEQUENCE: 28

ggaagaggtg attaaatgga t 21

<210> SEQ ID NO 29
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002064:1224O19

<400> SEQUENCE: 29

cccaaaaatc aaacaacaa 19

<210> SEQ ID NO 30
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002383:1287P22

<400> SEQUENCE: 30

tttgtattag gttggaagtg gt 22

<210> SEQ ID NO 31
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002383:1529O22

<400> SEQUENCE: 31

cccaaataaa tcaacaacaa ca 22

<210> SEQ ID NO 32
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002393:1127P21

<400> SEQUENCE: 32

ttgtttgggt taataaatgg a 21

<210> SEQ ID NO 33

-continued

<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002393:1381020

<400> SEQUENCE: 33

cttctctctt ctcccctctc 20

<210> SEQ ID NO 34
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: MSP PRIMER CALCITONIN

<400> SEQUENCE: 34

aggttatcgt cgtgcgagtg t 21

<210> SEQ ID NO 35
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: MSP PRIMER CALCITONIN

<400> SEQUENCE: 35

tcactcaaac gtatcccaaa ccta 24

<210> SEQ ID NO 36
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: MSP PRIMER VERSICAN

<400> SEQUENCE: 36

tgggattaag attttcggtt agtttc 26

<210> SEQ ID NO 37
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: MSP PRIMER VERSICAN

<400> SEQUENCE: 37

cactacaacg ctacgcgact aaa 23

<210> SEQ ID NO 38
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: MSP PRIMER TPEF

<400> SEQUENCE: 38

ttttttttttc ggacgtcgtt g 21

<210> SEQ ID NO 39
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: MSP PRIMER TPEF

<400> SEQUENCE: 39

cctctacata cgccgcgaat 20

<210> SEQ ID NO 40
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: MSP PRIMER EYA4

<400> SEQUENCE: 40

cggagggtac ggagattacg 20

<210> SEQ ID NO 41
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: MSP PRIMER EYA4

<400> SEQUENCE: 41

cgacgacgcg cgaaa 15

<210> SEQ ID NO 42
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: MSP PRIMER HCADHERIN

<400> SEQUENCE: 42

gacggatttt tttttaacgt tttttc 26

<210> SEQ ID NO 43
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: MSP PRIMER HCADHERIN

<400> SEQUENCE: 43

aaataaaata ccacctccgc ga 22

<210> SEQ ID NO 44
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: HM PRIMER EYA4

<400> SEQUENCE: 44

ggtgattgtt tattgttatg gtttg 25

<210> SEQ ID NO 45
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: HM PRIMER EYA4

<400> SEQUENCE: 45

-continued

ccccctcaacc taaaaactac aac 23

<210> SEQ ID NO 46
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: HM PRIMER CALCITONIN

<400> SEQUENCE: 46

ggatgtgaga gttgttgagg tta 23

<210> SEQ ID NO 47
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: HM PRIMER CALCITONIN

<400> SEQUENCE: 47

acacacccaa acccattact atct 24

<210> SEQ ID NO 48
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: TPEF FORWARD PRIMER

<400> SEQUENCE: 48

ggacgttttt tatcgaaggc g 21

<210> SEQ ID NO 49
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: TPEF BACKWARD PRIMER

<400> SEQUENCE: 49

gccacccaac cgcga 15

<210> SEQ ID NO 50
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002044:2183A188

<400> SEQUENCE: 50

atgtgattcg tttgggta 18

<210> SEQ ID NO 51
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002044:2183B188

<400> SEQUENCE: 51

atgtgatttg tttgggta 18

<210> SEQ ID NO 52

-continued

<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002044:2194A186

<400> SEQUENCE: 52

gggtaacgtc gaatttag 18

<210> SEQ ID NO 53
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002044:2194B186

<400> SEQUENCE: 53

gggtaatgtt gaatttag 18

<210> SEQ ID NO 54
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002044:2324A187

<400> SEQUENCE: 54

aaaaattcgc gagtttag 18

<210> SEQ ID NO 55
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002044:2324B187

<400> SEQUENCE: 55

aaaaatttgt gagtttag 18

<210> SEQ ID NO 56
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002064:1101A188

<400> SEQUENCE: 56

tatatatacg tgtgggta 18

<210> SEQ ID NO 57
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002064:1101B188

<400> SEQUENCE: 57

tatatatatg tgtgggta 18

<210> SEQ ID NO 58
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: 0002064:1147A188

<400> SEQUENCE: 58

agtgtatgcg tagaaggt 18

<210> SEQ ID NO 59
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002064:1147B188

<400> SEQUENCE: 59

agtgtatgtg tagaaggt 18

<210> SEQ ID NO 60
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002064:1156A188

<400> SEQUENCE: 60

tttagatacg aaatgtta 18

<210> SEQ ID NO 61
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002064:1156B188

<400> SEQUENCE: 61

tttagatatg aaatgtta 18

<210> SEQ ID NO 62
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002064:1222A188

<400> SEQUENCE: 62

aagtaagtcg ttgttgtt 18

<210> SEQ ID NO 63
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002064:1222B188

<400> SEQUENCE: 63

aagtaagttg ttgttgtt 18

<210> SEQ ID NO 64
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002383:1279A188

<400> SEQUENCE: 64

-continued

gaagtggtcg ttagtttt 18

<210> SEQ ID NO 65
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002383:1279B188

<400> SEQUENCE: 65

gaagtggttg ttagttttt 19

<210> SEQ ID NO 66
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002383:1346A188

<400> SEQUENCE: 66

ttgtttagcg tgatttgt 18

<210> SEQ ID NO 67
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002383:1346B188

<400> SEQUENCE: 67

ttgtttagtg tgatttgt 18

<210> SEQ ID NO 68
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002383:1475A188

<400> SEQUENCE: 68

aaggaattcg ttttgtaa 18

<210> SEQ ID NO 69
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002383:1475B188

<400> SEQUENCE: 69

aaggaatttg ttttgtaa 18

<210> SEQ ID NO 70
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002383:1524A188

<400> SEQUENCE: 70

aatgttttcg tgatgttg 18

<210> SEQ ID NO 71

-continued

<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002383:1524B188

<400> SEQUENCE: 71

aatgtttttg tgatgttg 18

<210> SEQ ID NO 72
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002393:1223A188

<400> SEQUENCE: 72

atttgtttcg attaattt 18

<210> SEQ ID NO 73
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002393:1223B188

<400> SEQUENCE: 73

atttgttttg attaattt 18

<210> SEQ ID NO 74
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002393:1294A188

<400> SEQUENCE: 74

ataggttacg ggttggag 18

<210> SEQ ID NO 75
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002393:1294B188

<400> SEQUENCE: 75

ataggttatg ggttggag 18

<210> SEQ ID NO 76
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002393:1332A186

<400> SEQUENCE: 76

aatttgcgaa cgtttggg 18

<210> SEQ ID NO 77
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: 0002393:1332B186

<400> SEQUENCE: 77

aatttgtgaa tgtttggg 18

<210> SEQ ID NO 78
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Calcitonin ML probe

<400> SEQUENCE: 78

cgaatctctc gaacgatcgc atcca 25

<210> SEQ ID NO 79
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Versican ML probe

<400> SEQUENCE: 79

tcgacgttac ccaaacgaat cacataaaaa ac 32

<210> SEQ ID NO 80
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: TPEF ML probe

<400> SEQUENCE: 80

aattaccgaa aacatcgacc ga 22

<210> SEQ ID NO 81
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: EYA4 ML probe

<400> SEQUENCE: 81

cgaaacccta aatatcccga ataacgccg 29

<210> SEQ ID NO 82
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: HCadherin ML probe

<400> SEQUENCE: 82

gctcctcgcg aaatactcac cccg 24

<210> SEQ ID NO 83
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Calcitonin HM probe

<400> SEQUENCE: 83

-continued

acctccgaat ctctcgaacg atcgc 25

<210> SEQ ID NO 84
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: EYA4 HM probe

<400> SEQUENCE: 84

aaaattacga cgacgccacc cgaaa 25

<210> SEQ ID NO 85
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Calcitonin HM blocker

<400> SEQUENCE: 85

tgttgaggtt atgtgtaatt gggtgtga 28

<210> SEQ ID NO 86
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: EYA4 HM blocker

<400> SEQUENCE: 86

aaactacaac cactcaaatc aaccca 26

<210> SEQ ID NO 87
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: EYA4 HM blocker

<400> SEQUENCE: 87

gttatggttt gtgattttgt gtggg 25

<210> SEQ ID NO 88
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002044:2096A188

<400> SEQUENCE: 88

aagattttcg gttagttt 18

<210> SEQ ID NO 89
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002044:2096B188

<400> SEQUENCE: 89

aagatttttg gttagttt 18

<210> SEQ ID NO 90

-continued

<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: TPEF ML PROBE

<400> SEQUENCE: 90

acccgaaatc acgcgcgaaa aa 22

<210> SEQ ID NO 91
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: control primer

<400> SEQUENCE: 91

tggtgatgga ggaggtttag taagt 25

<210> SEQ ID NO 92
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: control primer

<400> SEQUENCE: 92

aaccaataaa acctactcct cccttaa 27

<210> SEQ ID NO 93
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: control probe

<400> SEQUENCE: 93

accaccaccc aacacacaat aacaaacaca 30

<210> SEQ ID NO 94
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo Sapiens

<400> SEQUENCE: 94

ccttagtccc tacctctgct 20

<210> SEQ ID NO 95
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo Sapiens

<400> SEQUENCE: 95

ctcatttaca cacacccaaa c 21

<210> SEQ ID NO 96
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer

<400> SEQUENCE: 96

-continued

tggataggag ttgggattaa gatttt 26

<210> SEQ ID NO 97
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer

<400> SEQUENCE: 97

cttattacaa tttaaaaaaa aaattcacta caa 33

<210> SEQ ID NO 98
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: blocker

<400> SEQUENCE: 98

aaattcacta caacactaca caactaaatt caacattac 39

<210> SEQ ID NO 99
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Probe

<400> SEQUENCE: 99

ttttcgtatt ttttttcggg ttattacgtt tt 32

<210> SEQ ID NO 100
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Probe

<400> SEQUENCE: 100

atgtgattcg tttgggtaac gtcga 25

<210> SEQ ID NO 101
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer

<400> SEQUENCE: 101

gtagggttat tgtttgggtt aataaat 27

<210> SEQ ID NO 102
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer

<400> SEQUENCE: 102

taaaaaaaaa aaaaaaactc ctctacatac 30

<210> SEQ ID NO 103

-continued

<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: blocker

<400> SEQUENCE: 103

aactcctcta catacaccac aaataaatt 29

<210> SEQ ID NO 104
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Probe

<400> SEQUENCE: 104

cgaaaacatc gaccgaacaa cg 22

<210> SEQ ID NO 105
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Probe

<400> SEQUENCE: 105

gtccgaaaaa aaaaaaacga actcc 25

<210> SEQ ID NO 106
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Forward Primer

<400> SEQUENCE: 106

gttagttagt taatttttta aatagattag tag 33

<210> SEQ ID NO 107
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Reverse Primer

<400> SEQUENCE: 107

caaaaaaaca aataaaatac cacctcc 27

<210> SEQ ID NO 108
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Blocker

<400> SEQUENCE: 108

cctccacaaa actcactcct cacaaaatac 30

<210> SEQ ID NO 109
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: LC Probe

<400> SEQUENCE: 109

tttcgttttg tatggtagat acggggtga 29

<210> SEQ ID NO 110
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LC Probe

<400> SEQUENCE: 110

attaatggtt ttataagacg gatttttttt taacgt 36

1. A nucleic acid comprising a sequence at least 18 bases in length of a segment of the chemically pretreated genomic DNA according to one of the sequences taken from the group comprising SEQ ID NO 6 to SEQ ID NO 25 and sequences complementary thereto.

2. An oligomer, in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising at least one base sequence having a length of at least 9 nucleotides which is complementary to, or hybridises under moderately stringent or stringent conditions to a pretreated genomic DNA according to one of the SEQ ID NO 6 to SEQ ID NO 25 or sequences complementary thereto.

3. The oligomer as recited in claim 2, wherein the base sequence includes at least one CpG, TpG or CpA dinucleotide.

4. The oligomer as recited in claim 3, characterised in that the cytosine of the CpG, or the thymine of the TpG or the adenosine of the CpA dinucleotide is located approximately in the middle third of the oligomer.

5. A set of oligomers, comprising at least two oligomers according to any of claims 2 to 4.

6. A set of at least two oligonucleotides as recited in claims 2 to 5, which can be used as primer oligonucleotides for the amplification of DNA sequences of one of SEQ ID NO 6 to SEQ ID NO 25 and sequences complementary thereto.

7. A set of oligonucleotides as recited in claims 2 to 5, characterised in that at least one oligonucleotide is bound to a solid phase.

8. Use of a set of oligomer probes comprising at least four of the oligomers according to any of claims **5** through **7** for detecting the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) within one of the sequences according to SEQ ID NO 1 to SEQ ID NO 5, or sequences complementary thereto.

9. A method for manufacturing an arrangement of different oligomers (array) fixed to a carrier material for analysing colon cell proliferative disorders associated with the methylation state of the CpG dinucleotides of one of SEQ ID NO 1 to SEQ ID NO 5, and sequences complementary thereto wherein at least one oligomer according to any of the claims **2** through **5** and **7** is coupled to a solid phase.

10. An arrangement of different oligomers (array) obtainable according to claim 7.

11. An array of different oligonucleotide- and/or PNA-oligomer sequences as recited in claim 10, characterised in that these are arranged on a solid phase in the form of a rectangular or hexagonal lattice.

12. The array as recited in any of the claims **10** or **11**, characterised in that the solid phase surface is composed of silicon, glass, polystyrene, aluminium, steel, iron, copper, nickel, silver, or gold.

13. A method for detecting, differentiating or distinguishing between colon cell proliferative disorders associated with at least one gene and/or their regulatory regions from the group comprising Versican, TPEF, H-Cadherin, Calcitonin and EYA4 in a subject, said method comprising contacting a target nucleic acid 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.

14. A method according to claim 13, comprising:

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

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

c) contacting the treated genomic DNA, or the treated fragment thereof, with 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 hybridises under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO 6 to SEQ ID NO 25, or complements thereof, wherein the treated DNA or a fragment thereof is either amplified to produce one or more amplificates, or is not amplified; and

d) determining, based on the presence or absence of, or on a property of said amplificate, the methylation state of at least one CpG dinucleotide sequence of SEQ ID NO 1 to SEQ ID NO 5, or an average, or a value reflecting an average methylation state of a plurality of CpG dinucleotide sequences of SEQ ID NO 1 to SEQ ID NO

5, whereby at least one of detecting or distinguishing between colon cell proliferative disorders is, at least in part, enabled.

15. A method according to claim 13 comprising the following steps of

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

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

c) amplifying one or more fragments of the treated DNA such that only or preferentially DNA originating from colon or colon cell proliferative disorder cells are amplified;

d) detecting the amplificates or characteristics thereof and thereby deducing on the presence or absence of a colon cell proliferative disorder.

16. A method according to claims 13 to 15 wherein said colon cell proliferative disorders are taken from the group comprising adenocarcinomas, polyps, squamous cell cancers, carcinoid tumours, sarcomas and lymphomas.

17. The method of claims 14 to 16, wherein in step a) the biological sample obtained from the subject is selected form the group consisting of histological slides, biopsies, paraffin-embedded tissue, bodily fluid, stool, blood, serum, plasma, urine, sputum and combinations thereof.

18. The method of one of claims 14 to 17, wherein step b) treating the genomic DNA, or the fragment thereof, comprises use of a bisulfite solution.

19. The method of one of claims 14 to 16, wherein treating in b) is subsequent to embedding the DNA in agarose.

20. The method of one of claims 14 to 19, wherein step b) treating the genomic DNA, comprises treating in the presence of at least one of a DNA denaturing agent or a radical scavenger.

21. The method of one of claims 14 to 20, wherein contacting or amplifying in step c) comprises use of at least one method selected from the group consisting of: 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 a amplificate nucleic acid molecule carrying a detectable labels; and combinations thereof.

22. The method of claim 21, wherein the detectable amplificate label is selected from the label group consisting of: fluorescent labels; radionuclides or radiolabels; amplificate mass labels detectable in a mass spectrometer; detachable amplificate fragment mass labels detectable in a mass spectrometer; amplificate, and detachable amplificate fragment mass labels having a single-positive or single-negative net charge detectable in a mass spectrometer; and combinations thereof.

23. The method of claim 22, comprising in step d) use of mass spectrometry for detecting the amplificate, or detachable amplificate fragment mass labels.

24. The method according to one of claims 14 to 23 wherein in step c) of the method 2 or more different fragments are amplified.

25. The method according to one of claims 14 to 24 wherein one or more of said primers comprise sequences taken from the group according to SEQ ID NO 34 to SEQ ID NO 49, 96, 97, 101, 102, 106 and 107.

26. The method according to one of claims 14 to 24 wherein one or more of said primers comprise one or more CpG, TpG or CpA dinucleotides.

27. The method of claim 26 wherein said primers comprise between 2 to 5 CpG, TpG or CpA dinucleotides.

28. The method according to one of claims **26** or **27** wherein said one or more CpG, TpG or CpA dinucleotides are located within the 3' half of the primer.

29. The method according to one of claims 26 to 28 wherein said primers comprise one or more bases which hybridise to positions that were converted in the treatment of step b) in claim 14, or of step c) in claim 15.

30. The method of claim 29 wherein at least one of said bases are located within the 3' half of the primer.

31. The method according to one of claims 14 to 30 wherein said amplificates obtained in step d) comprise at least one 20 base pair sequence that comprises 3 or more CpG, TpG or CpA dinucleotides.

32. The method according to one of claims 14 to 31, further comprising in step c) the use of at least one nucleic acid molecule or peptide nucleic acid molecule at least 18 base pairs in length comprising one or more CpG, TpG or CpA dinucleotides and wherein the sequence of said molecule is complementary or identical to a sequence selected from the group consisting of SEQ ID NO 6 to SEQ ID NO 25, and complements thereof, and wherein said nucleic acid molecule or peptide nucleic acid molecule suppresses amplification of the nucleic acid to which it is hybridised.

33. The method according to claim 32 wherein the sequence of said nucleic acid(s) or peptide nucleic acid(s) is selected from the group consisting SEQ ID NO 85 to SEQ ID NO 87, 98, 103 and 108 and sequences complementary thereto.

34. The method of claim 32, wherein amplification of DNA that was unmethylated prior to treatment of step b) is suppressed.

35. The method of one of claims 32 to 34, wherein said nucleic acid molecule or peptide nucleic acid molecule is in each case modified at the 5'-end thereof to preclude degradation by an enzyme having a 5'-3' exonuclease activity.

36. The method of one of claims 32 to 35, wherein said nucleic acid molecule or peptide nucleic acid molecule in each case lack a 3' hydroxyl group.

37. The method of one of claims 32 to 36, wherein the amplification enzyme is a polymerase lacking 5'-3' exonuclease activity.

38. The method of one of claims 32 to 37, wherein the binding site of the oligonucleotide or PNA oligomer is identical to, or overlaps with that of the primer and thereby hinders hybridisation of the primer to its binding site.

39. The method of one of claims 32 to 38, wherein the binding sites of at least two of the oligonucleotides or PNA oligomers are identical to, or overlap with those of at least two of the primers, and thereby hinder hybridisation of the primers to their binding site.

40. The method of claim 39, wherein hybridisation of at least one of the oligonucleotides or peptide nucleic acid oligomers hinders hybridisation of a forward primer, and the hybridisation of at least one of the oligonucleotides or

peptide nucleic acid oligomers hinders the hybridisation of a reverse primer that binds to the elongation product of said forward primer.

41. The method of one of claims 32 to 38, wherein said oligonucleotide or peptide nucleic acid oligomer hybridises between the binding sites of the forward and reverse primers.

42. The method of one of claims **14** or **15**, wherein determining in step d), comprises hybridisation 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 comprising one or more CpG, TpG or CpA dinucleotides and wherein the sequence of said molecule that is complementary or identical to a sequence selected from the group consisting of SEQ ID NO 6 to SEQ ID NO 25.

43. The method of claim 42, wherein at least one such hybridising nucleic acid molecule or peptide nucleic acid molecule is bound to a solid phase.

44. The method of claim 43, wherein a plurality of such hybridising nucleic acid molecules or peptide nucleic acid molecules are bound to a solid phase in the form of a nucleic acid or peptide nucleic acid array selected from the array group consisting of linear, hexagonal, rectangular, and combinations thereof.

45. The method of one of claims **14** or **15**, wherein determining in step d), comprises sequencing of the amplificate.

46. The method of one of claims **14** or **15**, wherein determining in step d), comprises: hybridising at least one nucleic acid molecule comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridises under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO 6 to SEQ ID NO 25, and complements thereof; and extending at least one such hybridised nucleic acid molecule by at least one nucleotide base.

47. The method according to claim 42 wherein the sequence of said nucleic acid(s) or peptide nucleic acid(s) is selected from the group consisting SEQ ID NO 50 to SEQ ID NO 84, SEQ ID NO 88 to SEQ ID NO 90, 99, 100, 104, 105, 109, 110 and sequences complementary thereto.

48. The method according to claim 42 wherein said oligonucleotides or PNA oligomers are fluorescently labelled, and wherein detection thereof is by either an increase or a decrease in fluorescence or fluorescence polarisation.

49. The method according to claim 42 wherein the hybridisation of the oligonucleotides or PNA oligomers is detectable by fluorescence resonance energy transfer, and wherein the detection is by either an increase or a decrease in fluorescence.

50. The method of one of claims **14** or **15**, wherein the background DNA concentration is at between 100 to 1000 fold excess of the concentration of the DNA to be investigated.

51. A method according to claim 13, comprising:

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

b) extracting the genomic DNA;

c) contacting the genomic DNA, or a fragment thereof, comprising SEQ ID NO 1 to SEQ ID NO 5 or a sequence that hybridises under stringent conditions to SEQ ID NO 1 to SEQ ID NO 5, 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

d) 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 SEQ ID NO 1 to SEQ ID NO 5, or an average, or a value reflecting an average methylation state of a plurality of CpG dinucleotide sequences of SEQ ID NO 1 to SEQ ID NO 5, whereby at least one of detecting the prostate cell proliferative disorder, or distinguishing between a transitional and a peripheral zone of origin of the prostate cell proliferative disorder is, at least in part, afforded.

52. The method of claim 51, further comprising, prior to determining in step d), amplifying of the digested or undigested genomic DNA.

53. The method of claim 52, wherein amplifying comprises use of at least one method selected from the group consisting of: use of a heat resistant DNA polymerase as an amplification enzyme; generation of a amplificate nucleic acid carrying a detectable label; and combinations thereof.

54. The method of claim 53, wherein the detectable amplificate label is selected from the label group consisting of: fluorescent labels; radionuclides or radiolabels; amplificate mass labels detectable in a mass spectrometer; detachable amplificate fragment mass labels detectable in a mass spectrometer; amplificate, and detachable amplificate fragment mass labels having a single-positive or single-negative net charge detectable in a mass spectrometer; and combinations thereof.

55. The method of claim 54, comprising use of mass spectrometry for detecting amplificate, or detachable amplificate fragment mass labels.

56. The method of claim 55, wherein the mass spectrometry is selected from the group consisting of matrix assisted laser desorption/ionisation mass spectrometry (MALDI), electron spray mass spectrometry (ESI), and combinations thereof.

57. The method of claim 51, wherein the biological sample obtained from the subject is selected from the group consisting of histological slides, biopsies, paraffin-embedded tissue, bodily fluid, stool, blood, serum, plasma, urine, sputum and combinations thereof.

58. A kit useful for detecting, differentiating or distinguishing between colon cell proliferative disorders, comprising:

i) a bisulfite reagent; and

ii) 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 hybridises under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO 1 to SEQ ID NO 25, and complements thereof.

59. The kit of claim 58, further comprising standard reagents for performing a methylation assay selected from the group consisting of MS-SNuPE, MSP, MethylLight, HeavyMethyl, COBRA, nucleic acid sequencing, and combinations thereof.

60. The use of a nucleic acid according to claim 1, of an oligonucleotide or PNA-oligomer according to one of the claims **2** through **7**, of a kit according to claims **58** or **59**, of an array according to one of the claims **11** through **12**, of a set of oligonucleotides according to one of claims **5** through **7**, or a method according to claims 13 to 57, for the classification, differentiation and/or diagnosis of colon cell proliferative disorders or the predisposition to colon cell proliferative disorders.

61. The use of a nucleic acid according to claim 1, of an oligonucleotide or PNA-oligomer according to one of the claims **2** through **7**, of a kit according to claims **58** or **59**, of an array according to one of the claims **11** through **12**, of a set of oligonucleotides according to one of claims **5** through **7**, or a method according to claims 13 to 57, for the therapy of colon cell proliferative disorders.

* * * * *

专利名称(译)	用于分析结肠细胞增殖性疾病的方法和核酸		
公开(公告)号	US20060234224A1	公开(公告)日	2006-10-19
申请号	US10/506089	申请日	2003-02-27
当前申请(专利权)人(译)	AG EPIGENOMICS		
[标]发明人	ADORJAN PETER BURGER MATTHIAS MAIER SABINE LESCHE RALF COTTRELL SUSAN MOONEY SUZANNE		
发明人	ADORJAN, PETER BURGER, MATTHIAS MAIER, SABINE LESCHE, RALF COTTRELL, SUSAN MOONEY, SUZANNE		
IPC分类号	C12Q1/68 C07H21/04 C07K14/47 G01N27/62 A61K48/00 A61P35/00 A61P43/00 C12M1/00 C12N15/09 C12Q1/02 G01N21/78 G01N33/53 G01N33/543 G01N33/574 G01N33/58 G01N37/00		
CPC分类号	C12Q1/6886 Y10T436/143333 C12Q2600/154		
优先权	2002004551 2002-02-27 EP		
外部链接	Espacenet USPTO		

摘要(译)

本发明提供了通过分析Versican，TPEF，H-钙粘着蛋白，降钙素和EYA4中的一种或多种基因来检测，区分或区分结肠细胞增殖性疾病的方法和核酸。本发明还提供了可用于所述基因的细胞增殖性疾病特异性分析的新型核酸序列及其方法，测定和试剂盒。

FIGURE 1