



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

**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

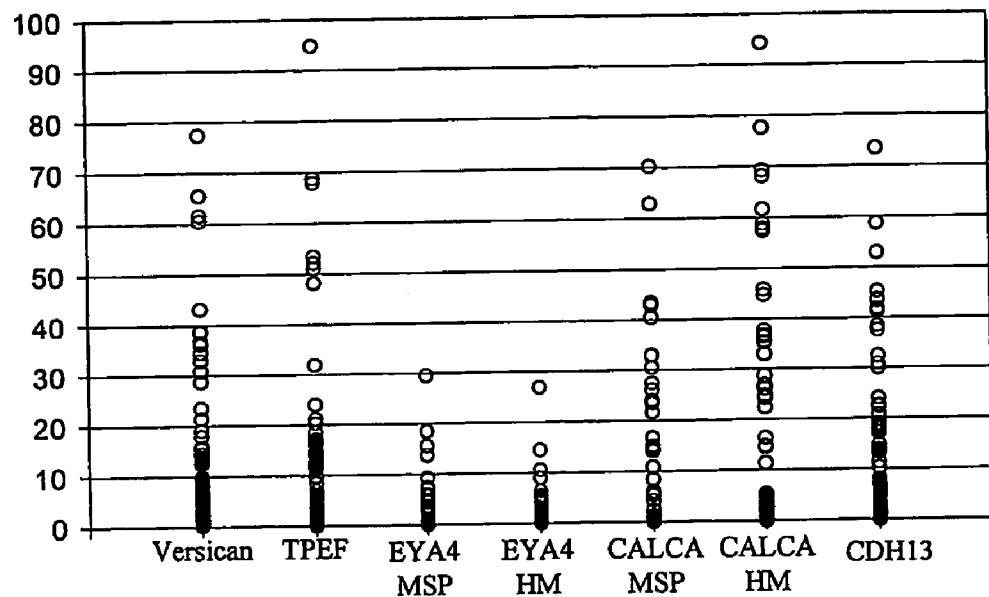


FIGURE 2

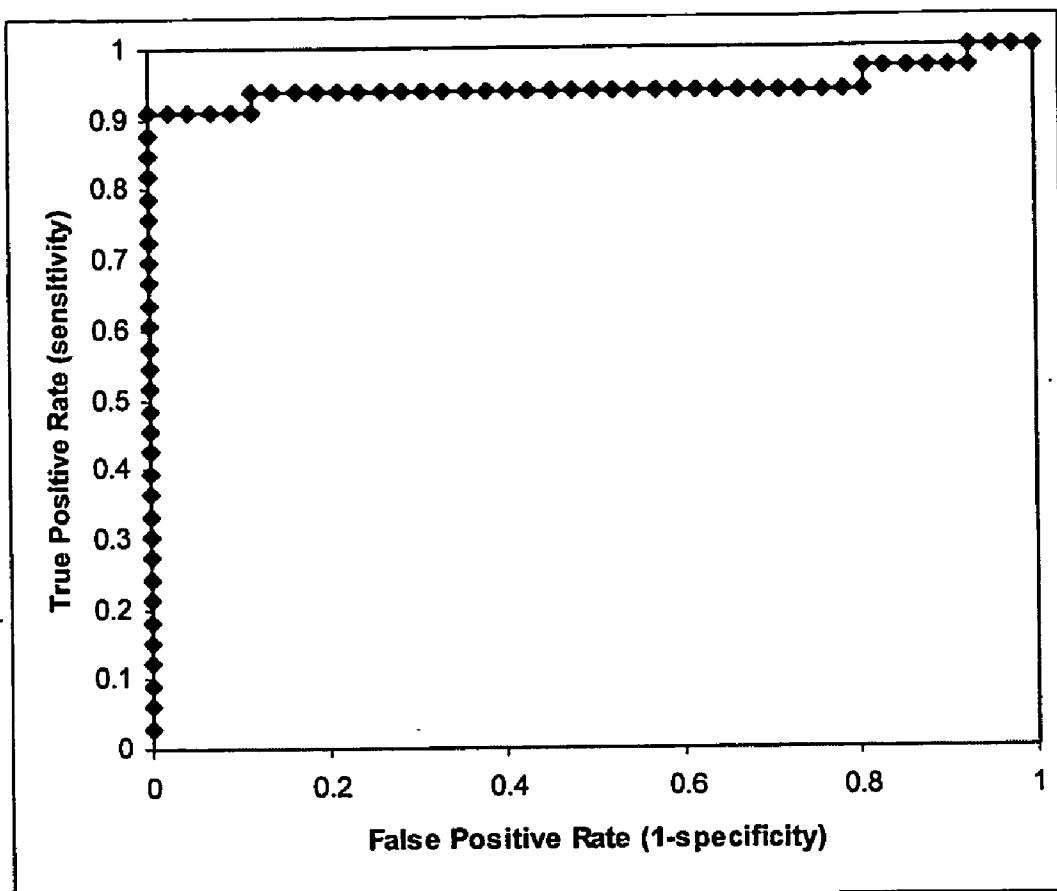
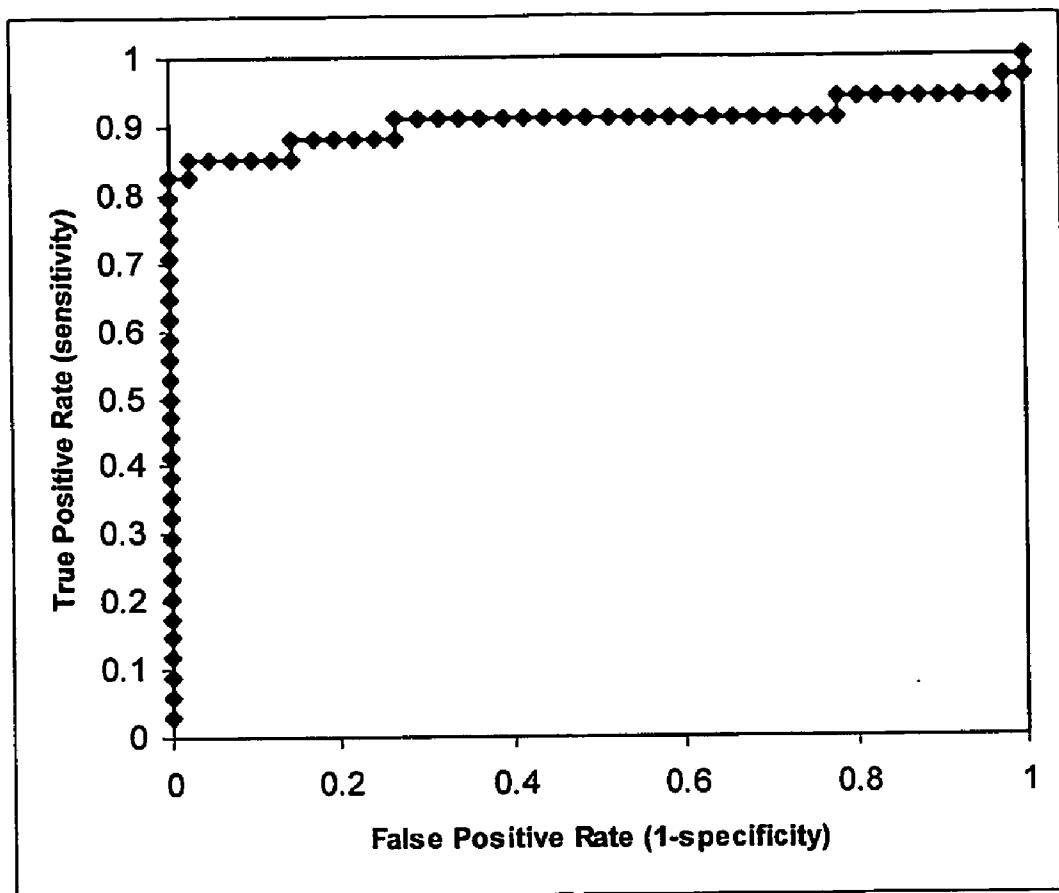
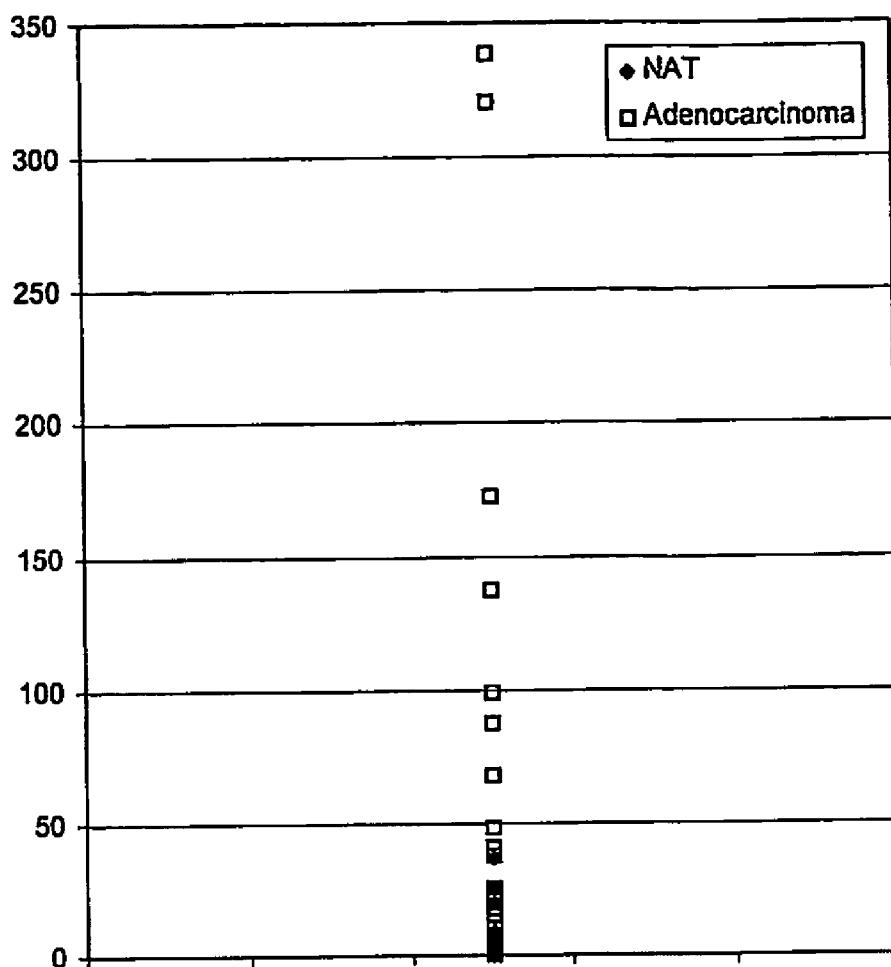


FIGURE 3



**FIGURE 4**

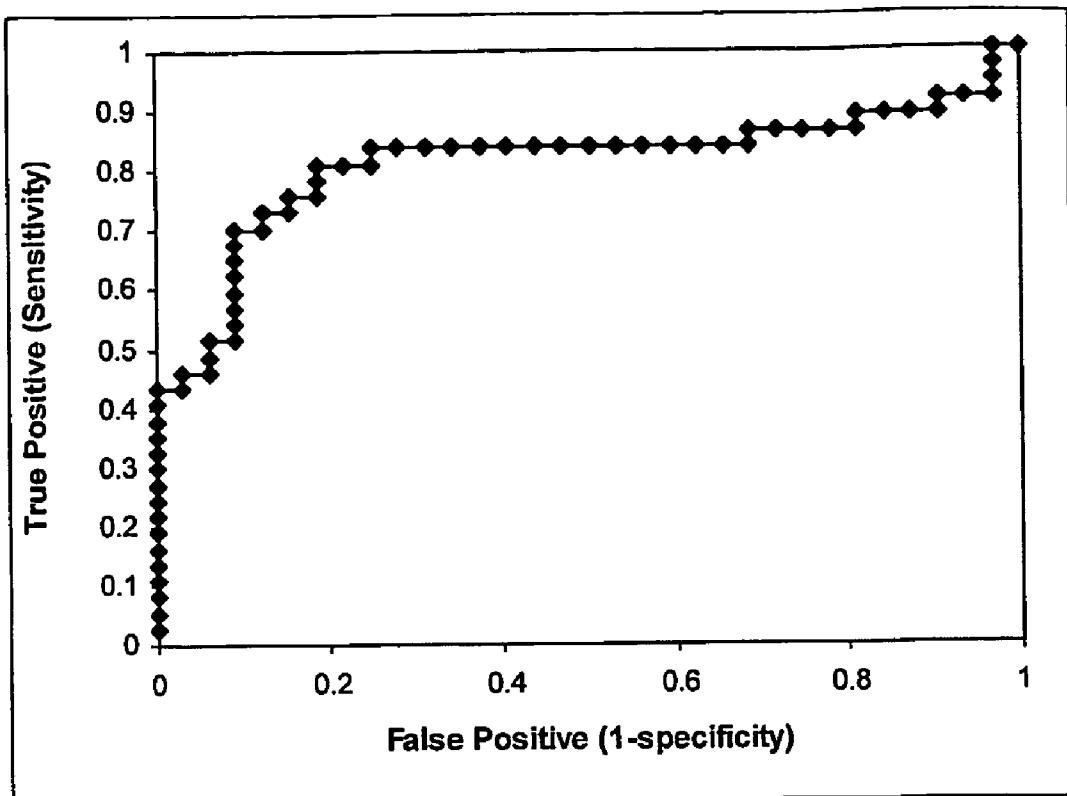
**FIGURE 5**

FIGURE 6

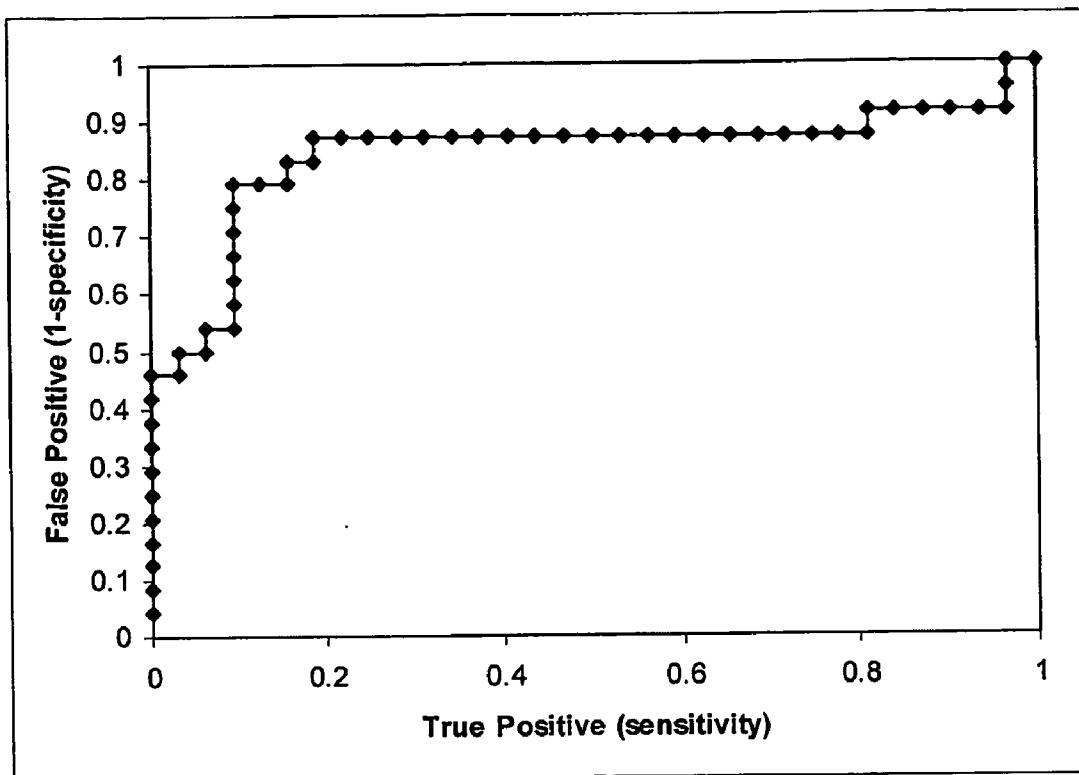


FIGURE 7

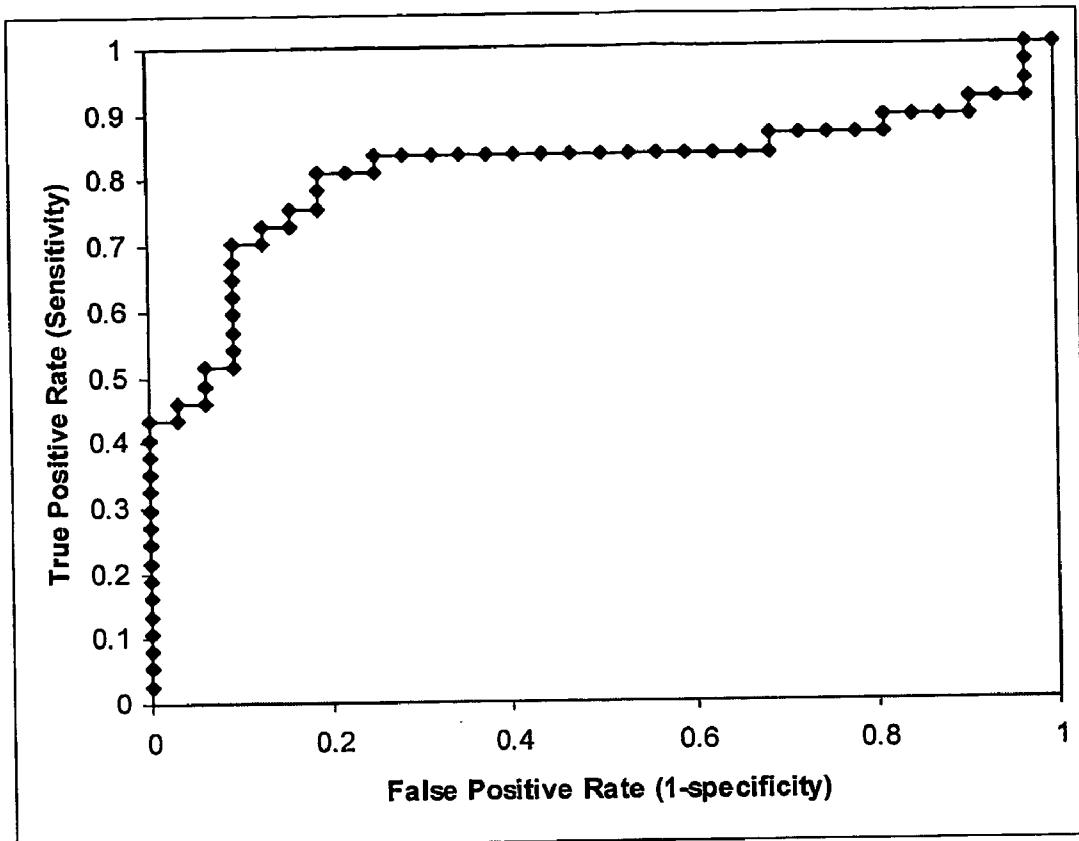


FIGURE 8

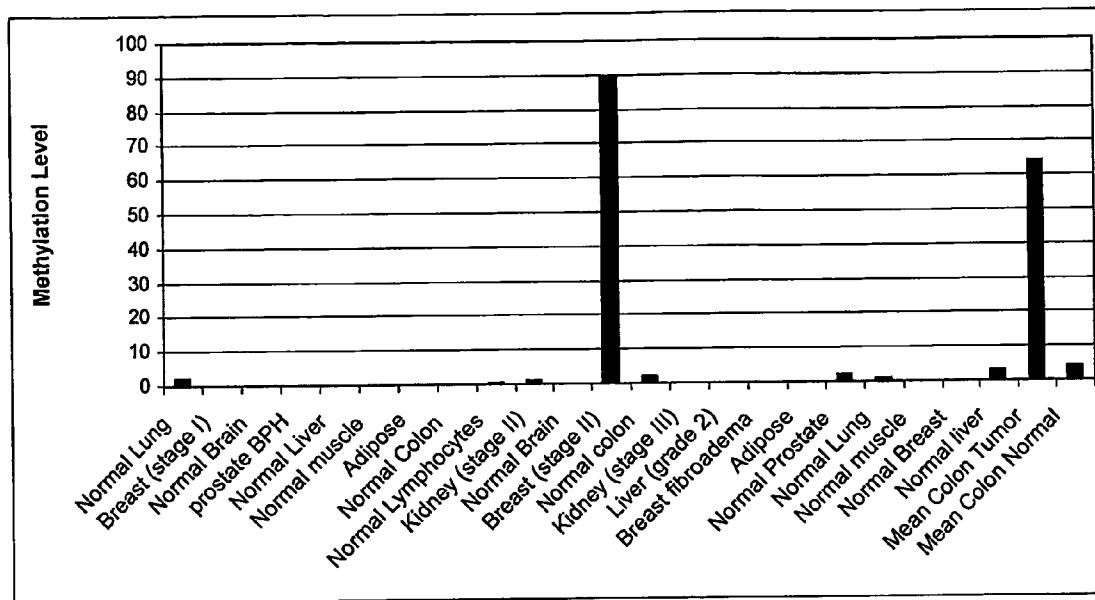


FIGURE 9

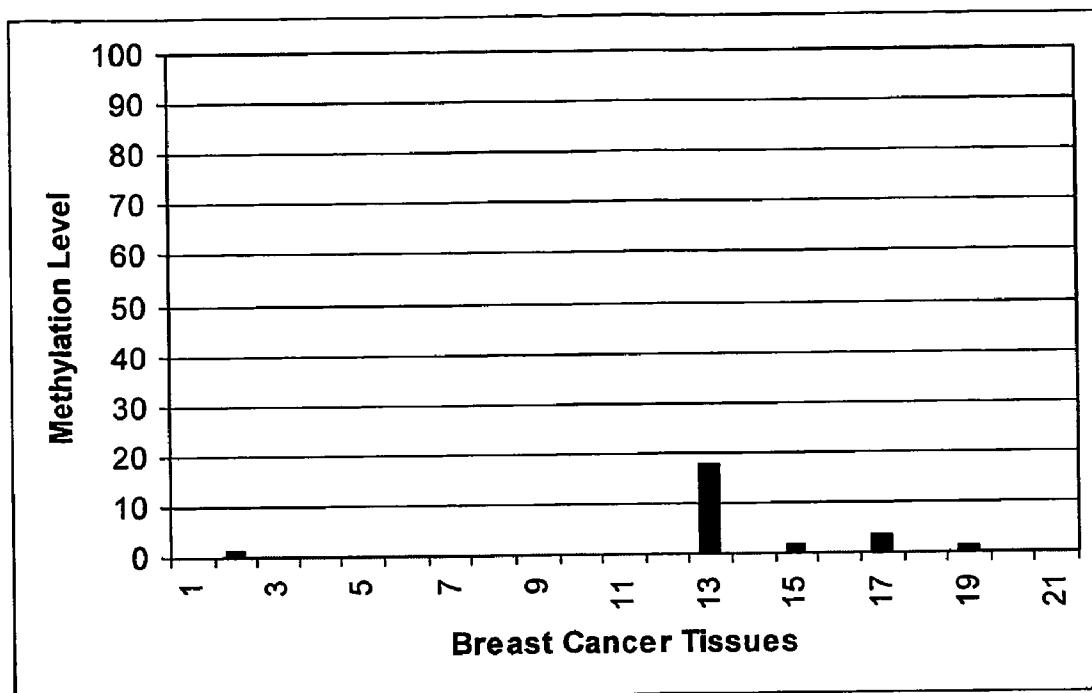


FIGURE 10

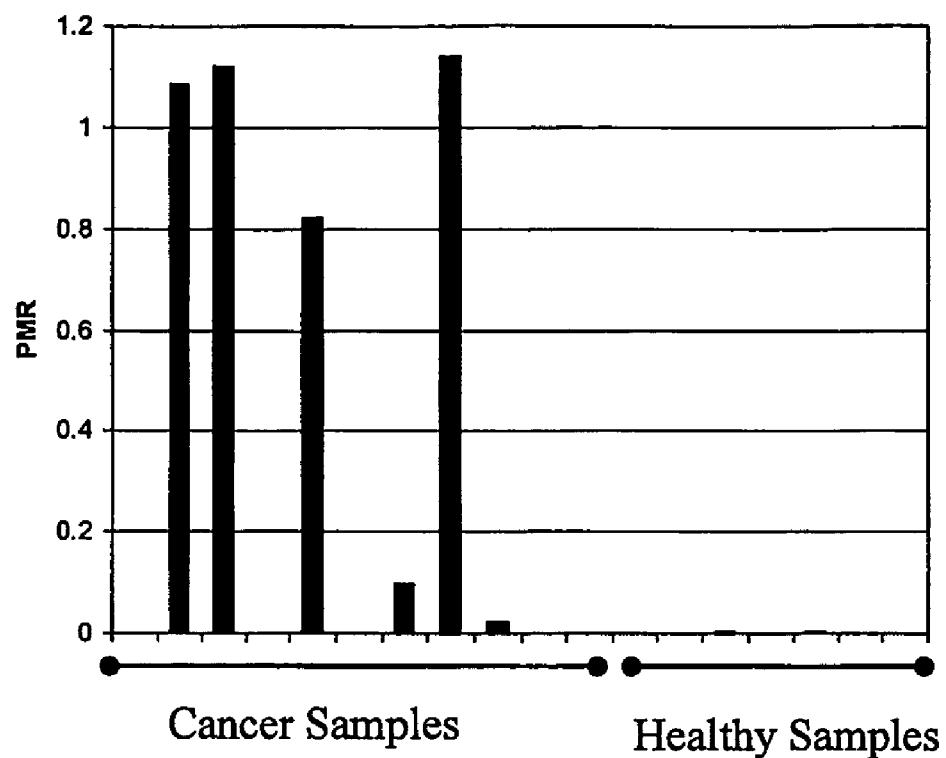


FIGURE 11

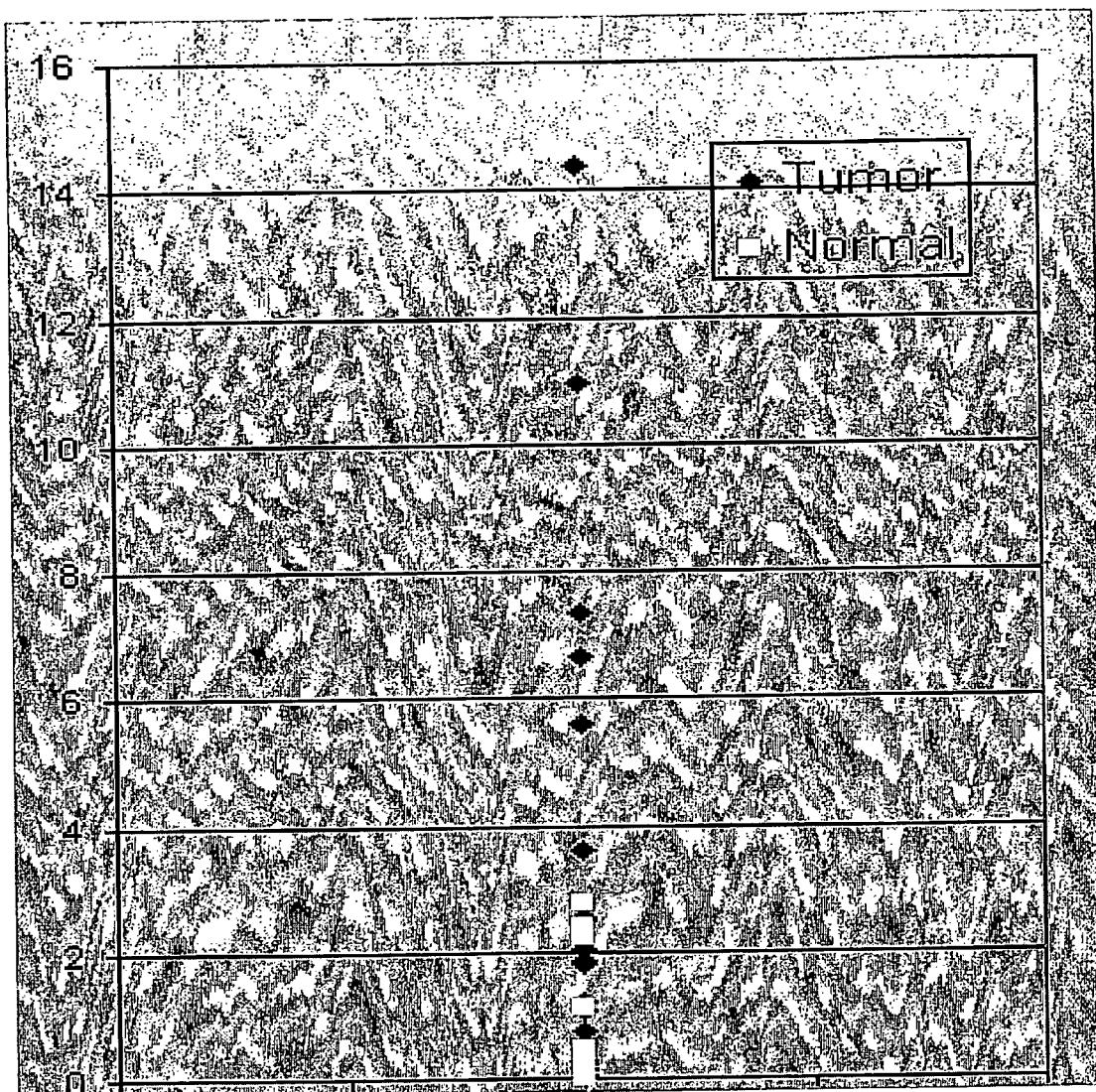


FIGURE 12

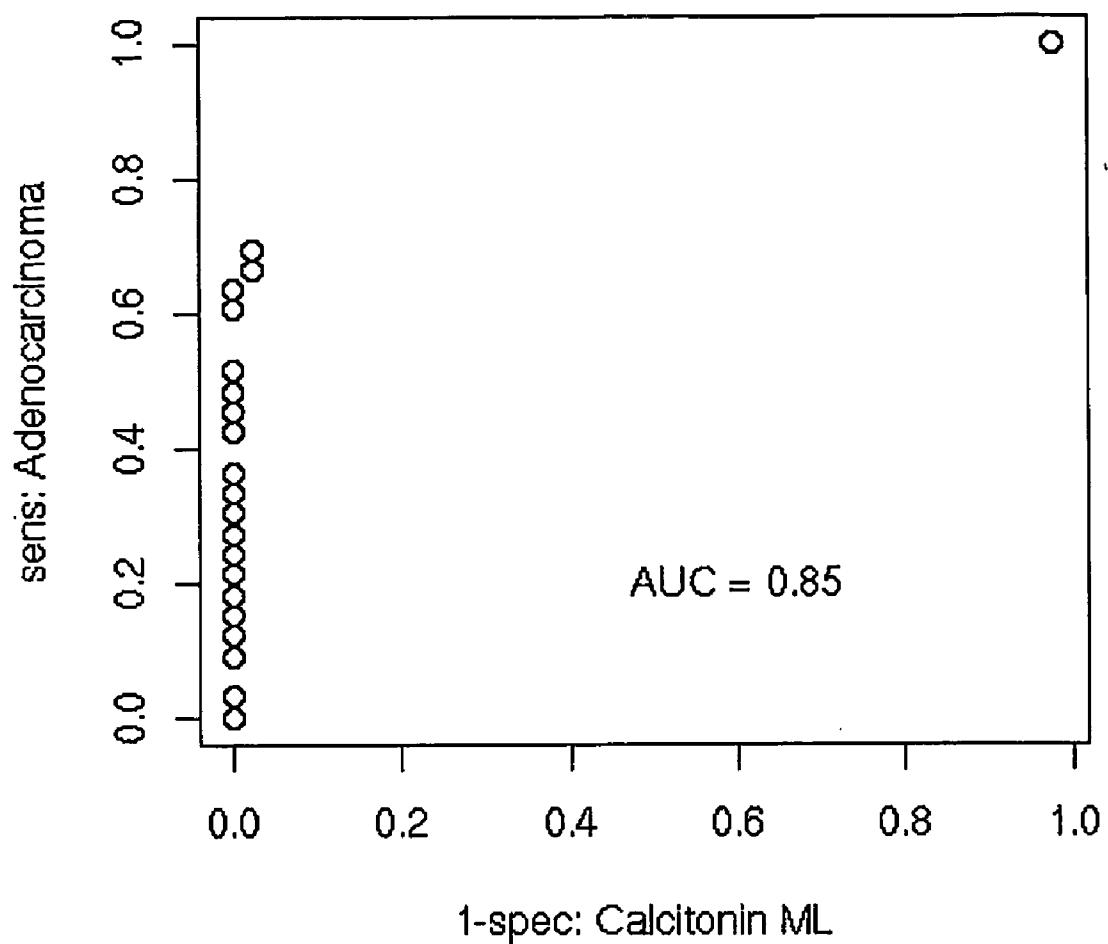


FIGURE 13

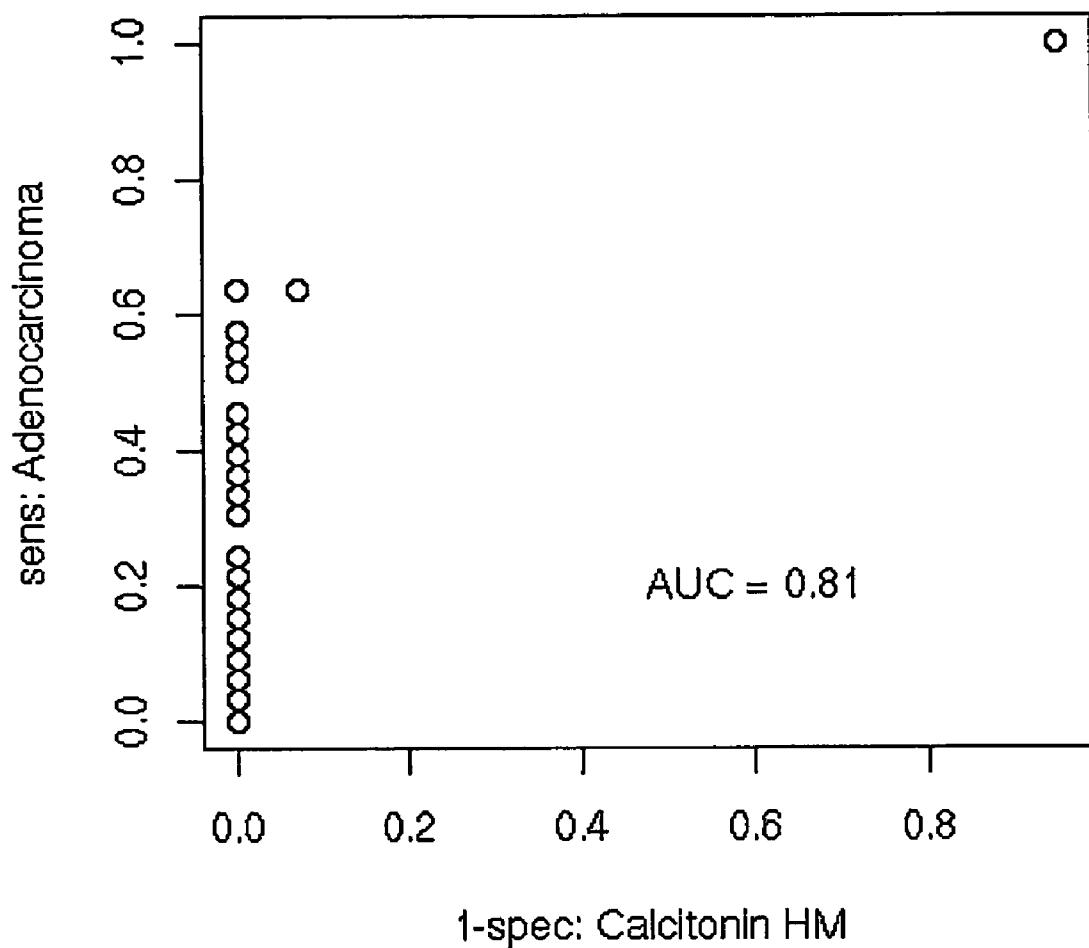


Figure 14

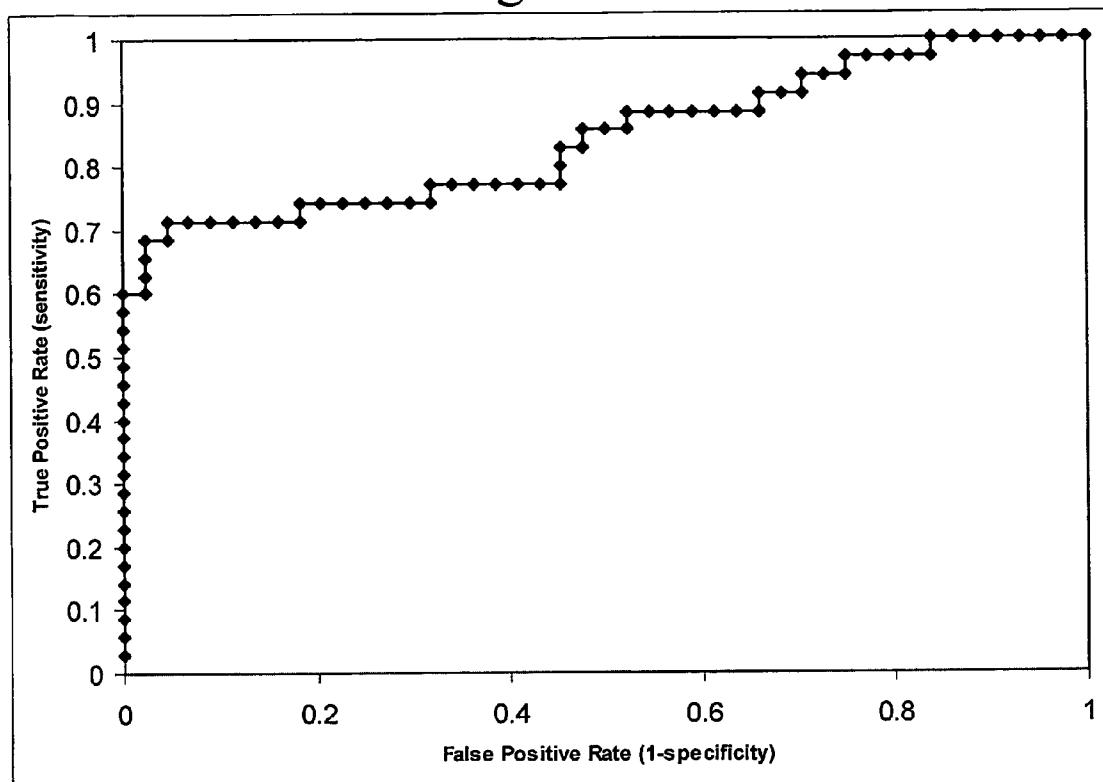


Figure 15

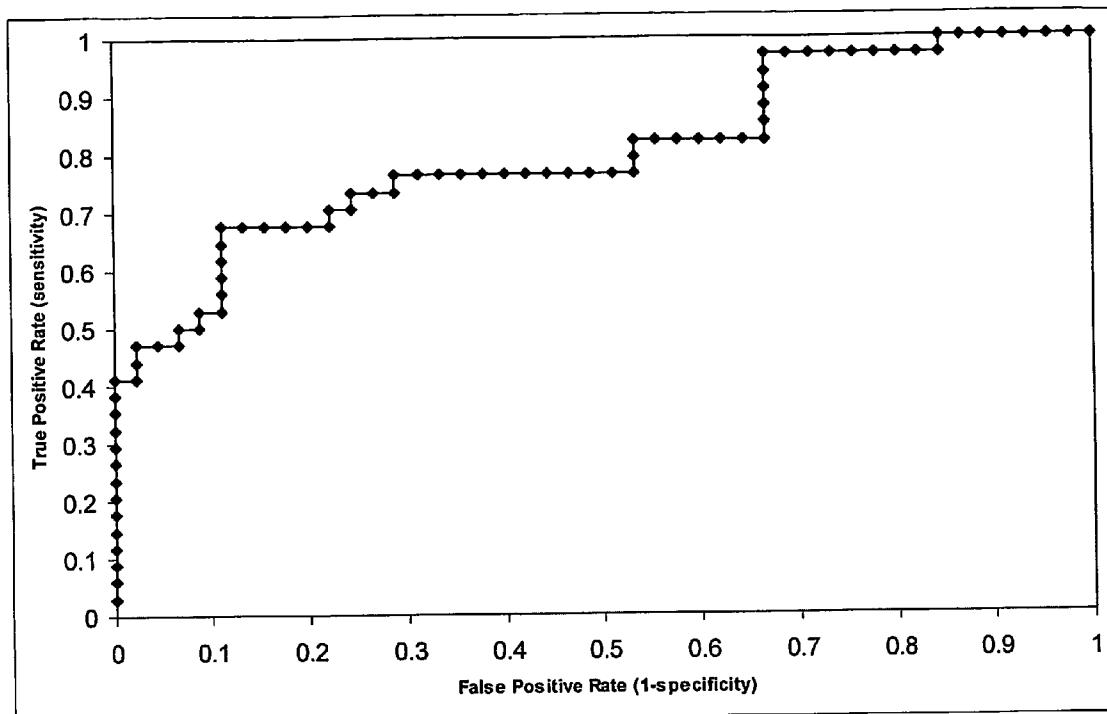


Figure 16

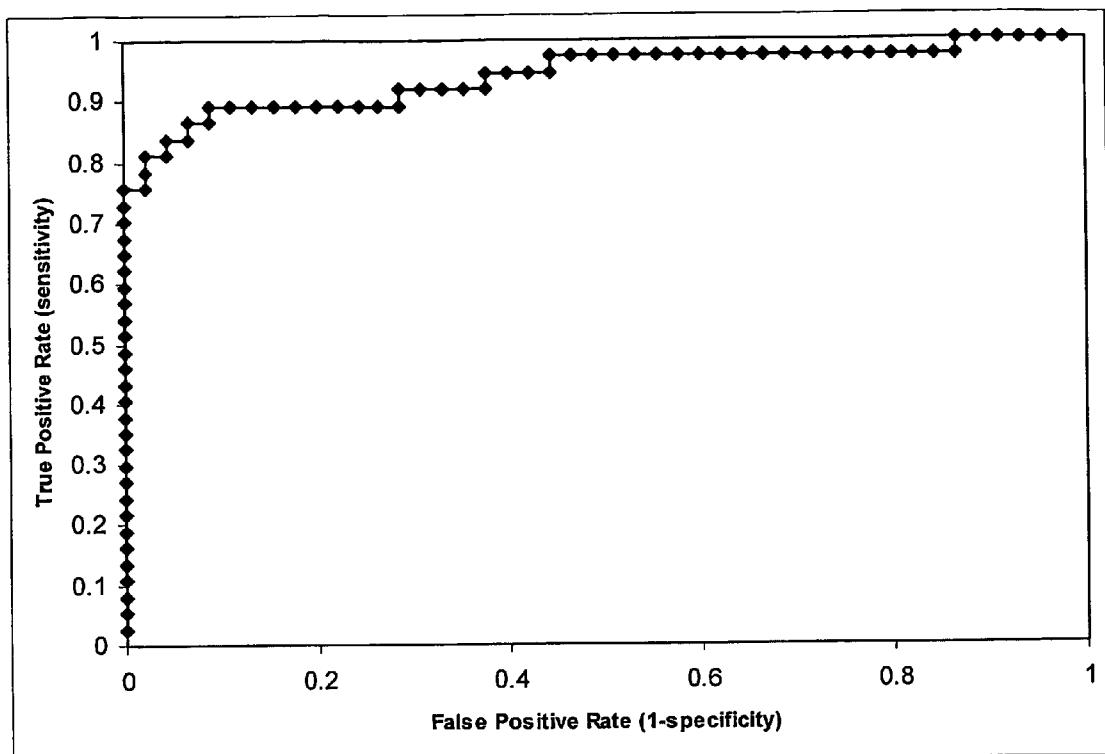


FIGURE 17

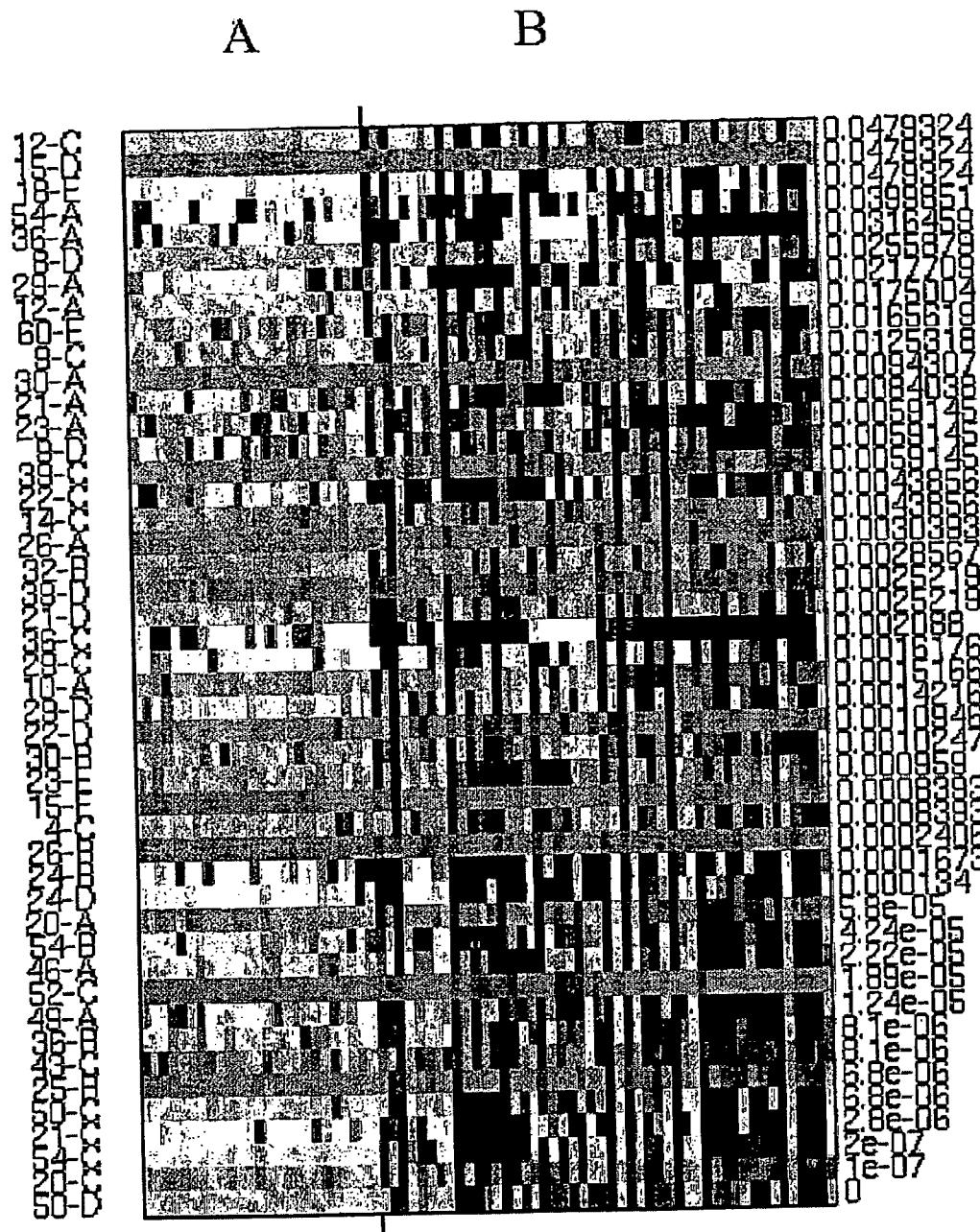


FIGURE 18

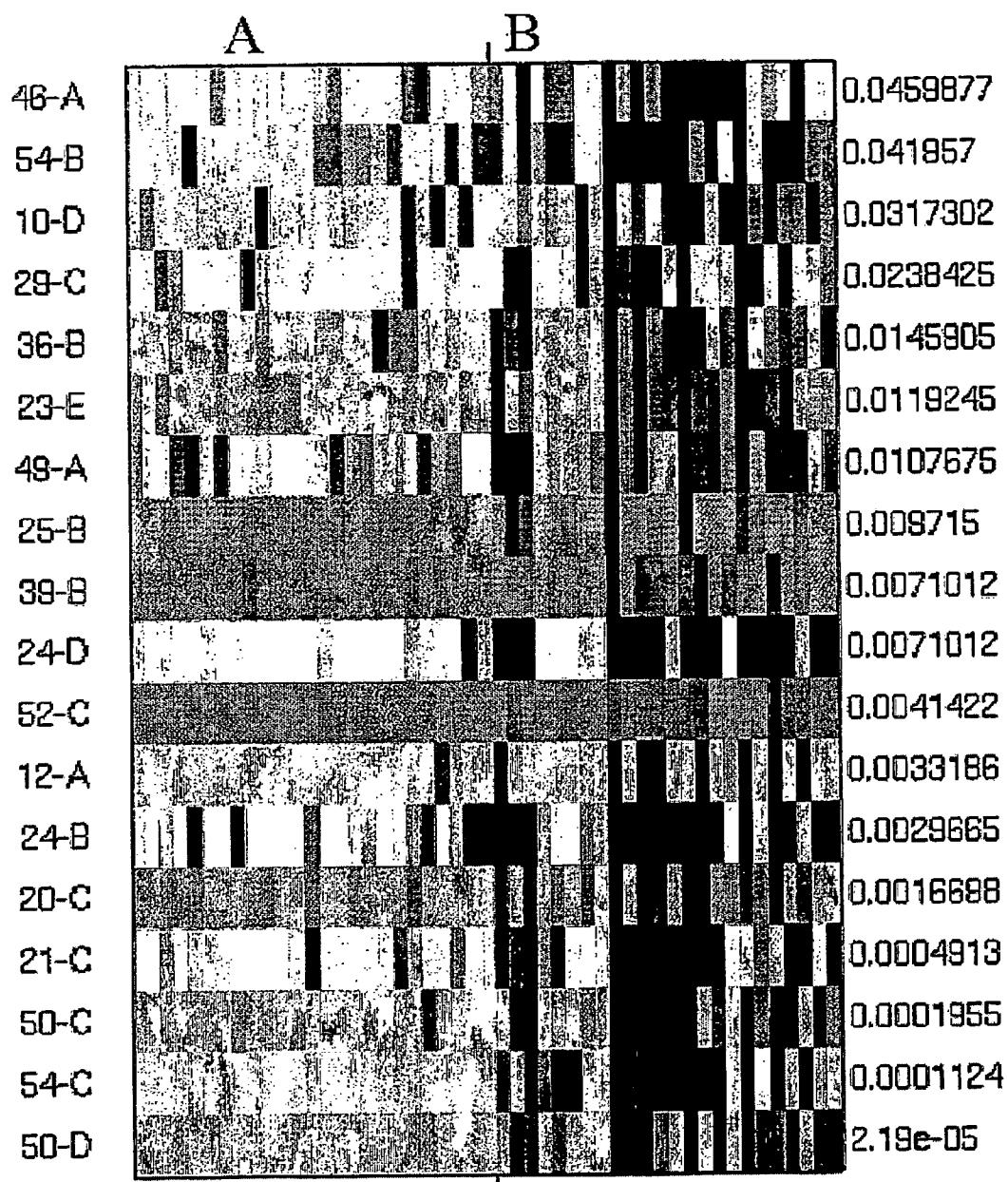


FIGURE 19

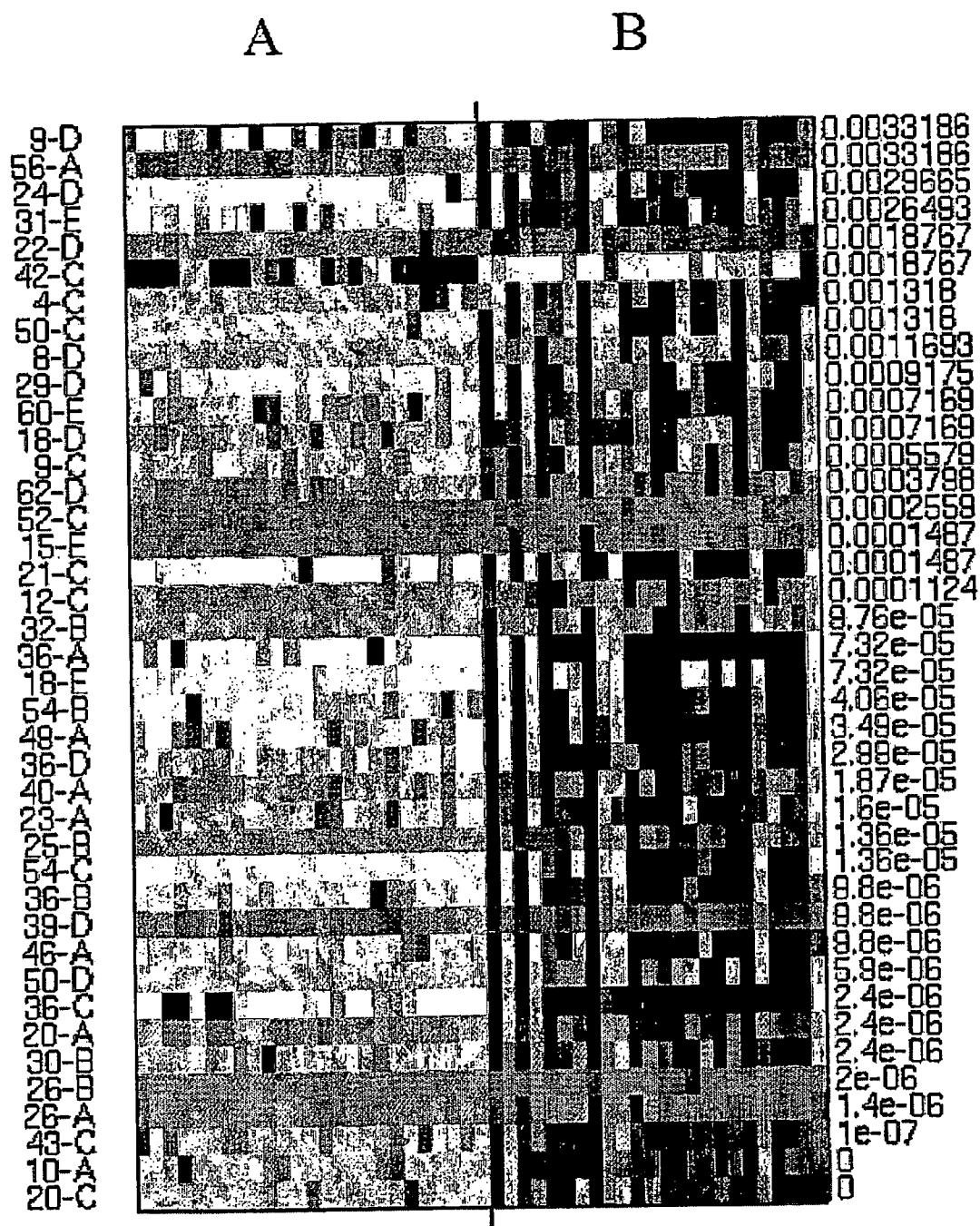


FIGURE 20

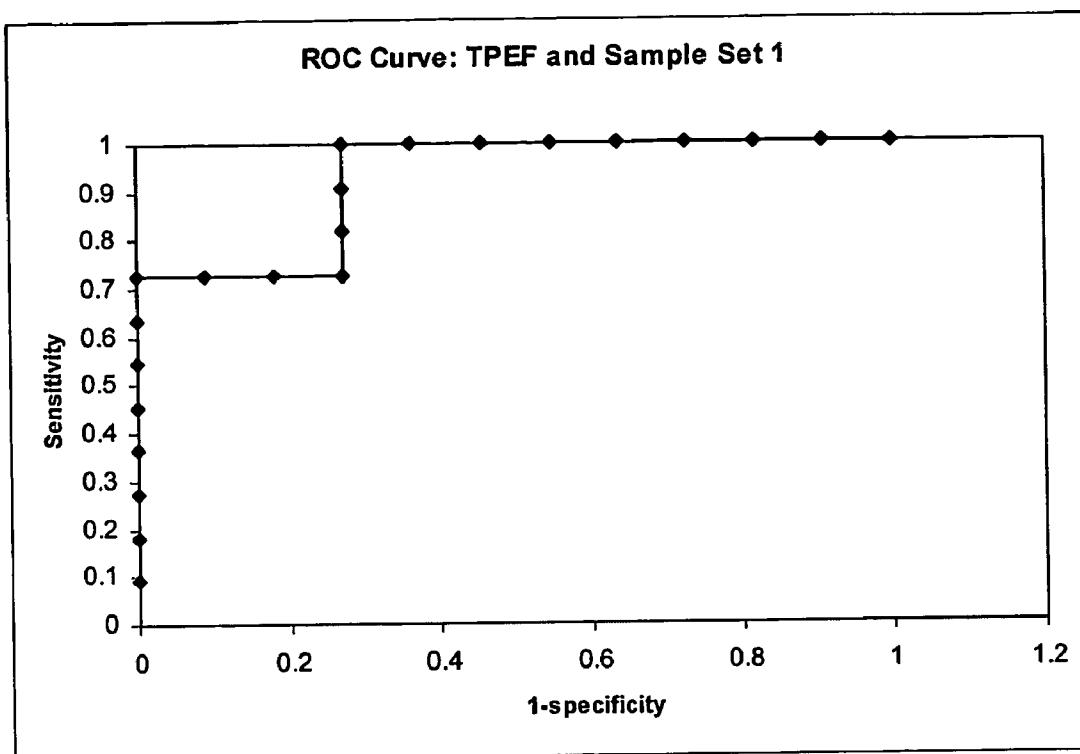


FIGURE 21

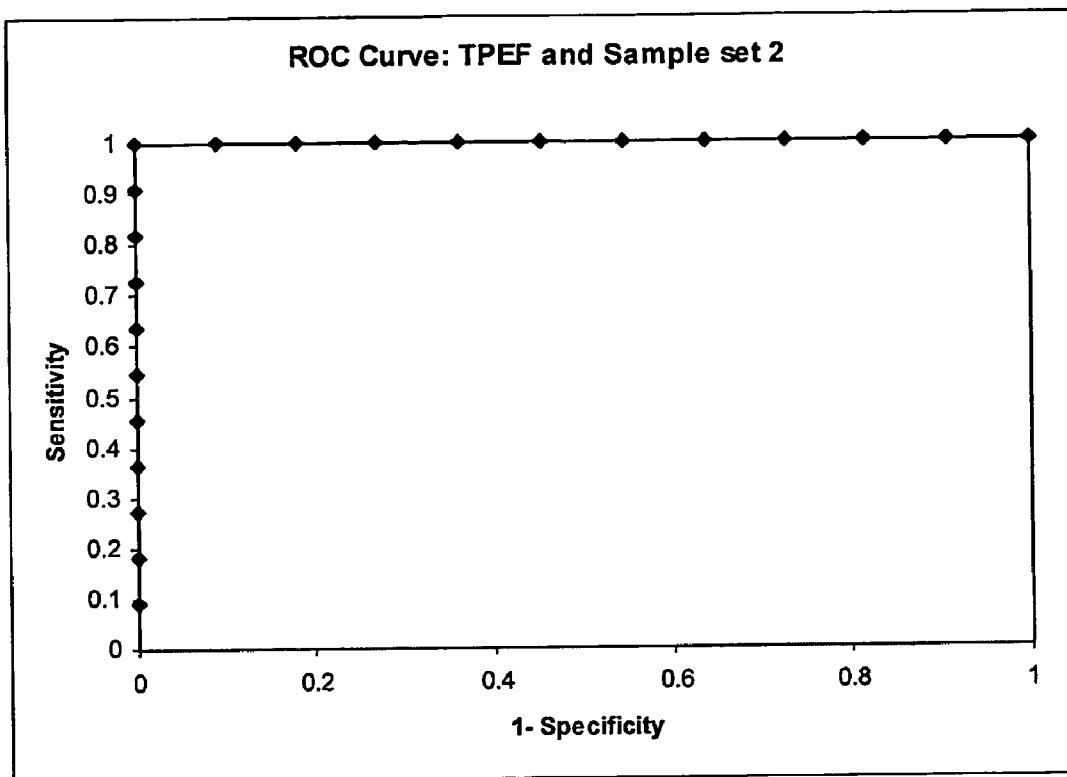
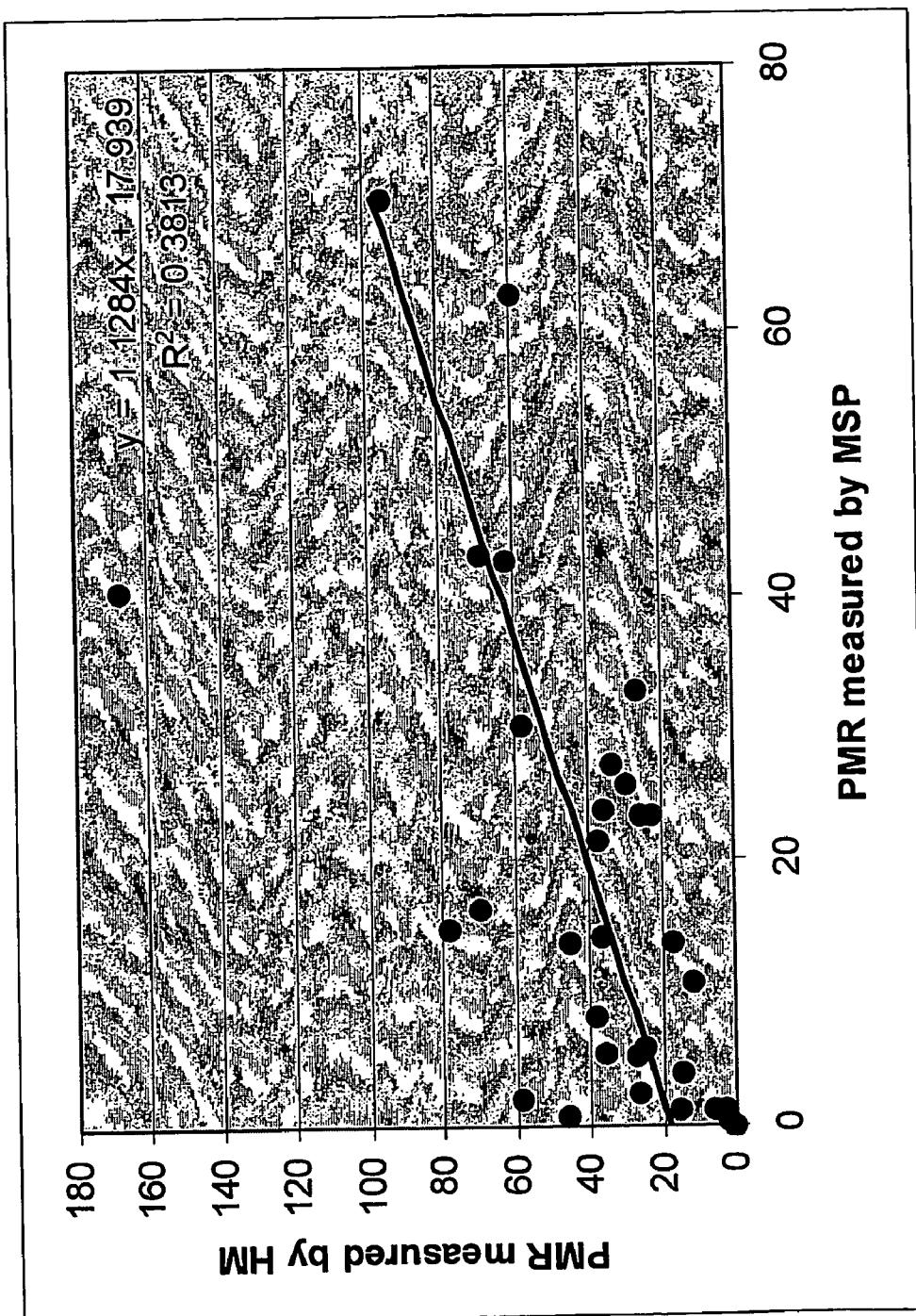
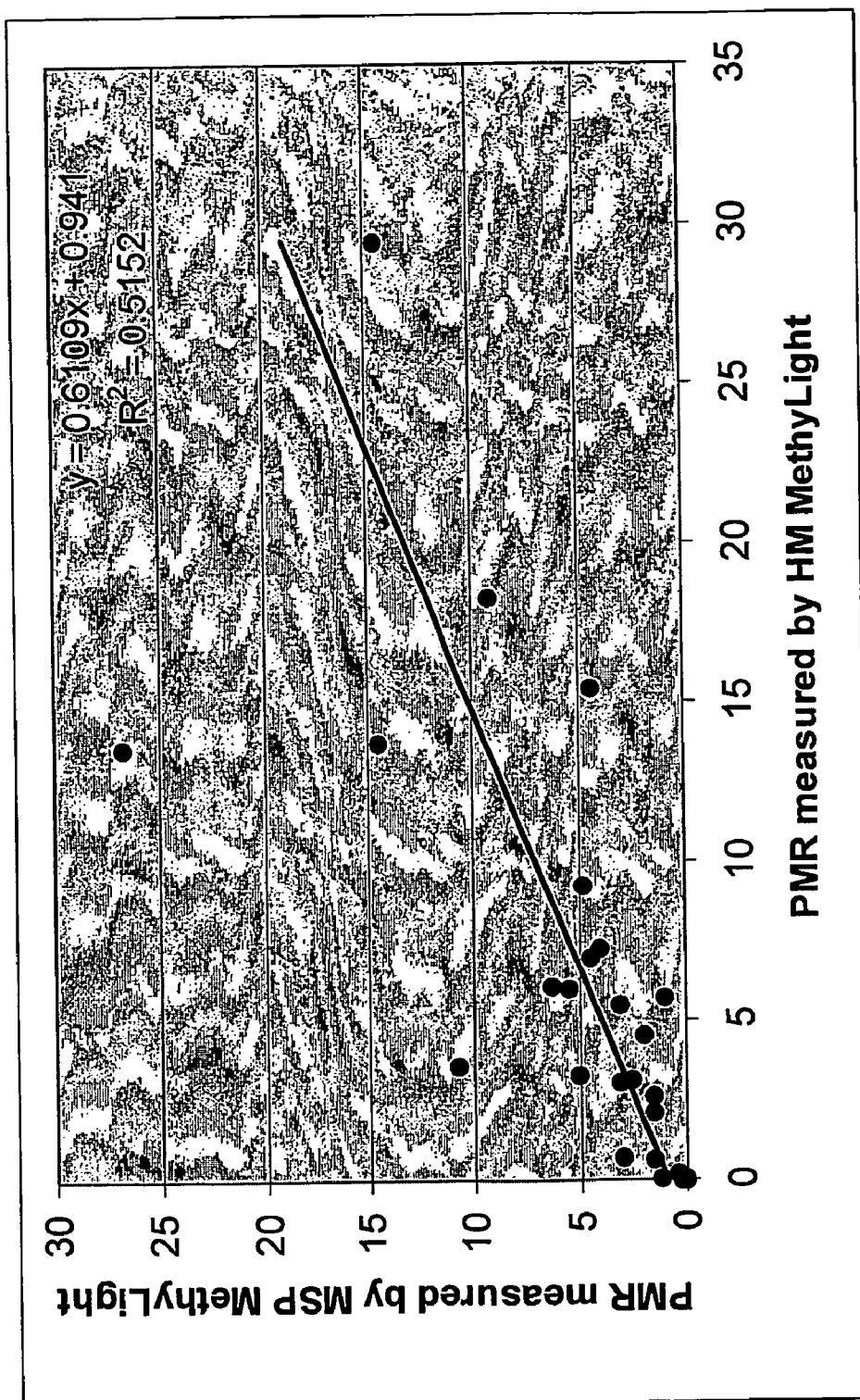


Figure 22



**Figure 23**



## 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 Iozzo 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 observation 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 disproportionately 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 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 from 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 adenine 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), palmitoyl 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 amplicate 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 amplicate is between 100 and 350 base pairs in length. It is particularly preferred that said amplicates 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 adenine). 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 disproportionately 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 amplicate 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 amplicates 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 amplicates were obtained by means of MSP amplification and/or blocking oligonucleotides, the presence or absence of an amplicate 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 amplicates 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 amplicates 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 amplicates 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 amplicates are then removed. The hybridised amplicates are detected. In this context, it is preferred that labels attached to the amplicates 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 amplicate 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 amplicates may carry suitable detectable labels as discussed above, namely fluorophore labels, radionuclides and mass labels.

[0091] In the final step the amplicates 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, MethylLight, 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); desulfonylation buffer; and DNA recovery components.

[0098] Typical reagents (e.g., as might be found in a typical MethylLight-based kit) for MethylLight 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 reagents 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<sup>M</sup>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-recognised 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 “MethylLight” 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 Adenomas 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 MethylLight 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 MethylLight 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-MethylLight and HeavyMethyl MethylLight assays. Please note that the term 'MethylLight' is used to describe real time PCR analysis of bisulfite treated DNA using probes of both the Taqman (single probe) and Lightcycler (dual probe) technologies.

#### Example 1

[0144] Analysis of methylation within colon cancer using an MSP-MethylLight 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-MethylLight 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: CGAAACCCTAAATATCCGAATAACGCCG (SEQ ID NO:81). The corresponding control assay was performed using the following primers and probes: Primer: TGGTGATGGAGGAGGTAG-TAAGT (SEQ ID NO:91); Primer: AACCAATAAACACTCTACTCCTCCCTTAA (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 MethylLight assay. The same DNA samples were also used to analyse methylation of the CpG island with a HeavyMethyl MethylLight (or HM MethylLight) assay, also referred to as the HeavyMethyl assay. The methylation status was determined with a HM MethylLight 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:	GGTGATTGTTATTGTTATGGTTG	(SEQ ID NO:44)
Reverse Primer:	CCCCTAACCTAAAAACTACAAAC	(SEQ ID NO:45)
Forward Blocker:	GTTATGGTTGTGATTTGTGTGGG	(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 MethylLight 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<sup>□</sup> 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-MethylLight Assay The colon cancer samples described in Example 1 were also analysed using a Calcitonin-MSP-MethylLight 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:**  
AGGTTATCGTCGTGCGAGTG; (SEQ ID NO:34)

**Primer:**  
TCACTCAAACGTATCCAAACCTA; (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 MethylLight assay. The same DNA samples were also used to analyse methylation of the Calcitonin-CpG island with a HeavyMethyl MethylLight 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:**  
TCTTGAGGTTATGTGTAATTGGGTGTGA. (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  $\mu$ M dNTPs, 5  $\mu$ M blocker; and 7  $\mu$ l of DNA, in a final reaction volume of 20  $\mu$ l);

[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 silico 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 MethylLight 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 MethylLight 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-MethylLight-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 CTCATTACACACACCCAAAC (SEQ ID NO:95). The resultant amplicate, 378 bp in length, contained an informative CpG at position 165. The amplicate 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 amplicate. 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 CTCATTACACACACCCAAAC (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  $\mu$ M of each dNTP, 1.5 mM MgCl<sub>2</sub> 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:  
TGGGATTAAGATTTCGGTTAGTTTC (SEQ ID No 36)

Reverse Primer:  
CACTACAACGCTACCGCGACTAAA (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  $\mu$ M dNTPs, 5  $\mu$ M blocker; and 7  $\mu$ l of DNA, in a final reaction volume of 20  $\mu$ l);

[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: TGGATAGGAGTTGGGATTAAGATTTC (SEQ ID NO:96)
Reverse Primer: CTTATTACAATTAAAAAAATTCACTACAA (SEQ ID NO:97)
Blocker: AAATTCACTACAAACACTACACAACTAAATTCAAC (SEQ ID NO:98)
          ATTAC
Probe: TTTTGTATTTTTTCGGTTATTACGTTT- (SEQ ID NO:99)
Fluor
Probe: LC640-ATGTGATTCTGTTGGTAACGTCGA- (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: TTTTTTTTCGGACGTCGTTG (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  $\mu$ M dNTPs, 5  $\mu$ M blocker; and 7  $\mu$ l of DNA, in a final reaction volume of 20  $\mu$ l);

[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: GTAGGGTTATTGTTGGGTTAACAAAT (SEQ ID NO:101)
Reverse Primer: TAAAAAAAAAAAAACTCCTCTACATAC (SEQ ID NO:102)
Blocker: AACTCCTCTACATACACCACAAATAATT (SEQ ID NO:103)
Probe: CGAAAAACATCGACCGAACAAACG-Fluor (SEQ ID NO:104)
Probe: LC640-GTCCGAAAAAAAAACGAACCTCC- (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/ $\mu$ l
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: GGACGTTTTTATCGAAGGCG	(SEQ ID NO 48)
Reverse Primer: GCCACCCAACCGCGA	(SEQ ID NO 49)
Probe: ACCCGAAATCACGCGCAGAAA	(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  $\mu$ M dNTPs, 5  $\mu$ M blocker; and 7  $\mu$ l of DNA, in a final reaction volume of 20  $\mu$ l);

[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: GACGGATTTCACGTTTTC	(SEQ ID NO 42)
Reverse Primer: AAATAAAATACCACTCCGCGA	(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  $\mu$ M dNTPs, 5  $\mu$ M blocker; and 7  $\mu$ l of DNA, in a final reaction volume of 20  $\mu$ l);

[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: GTTAGTTAGTTAATTAAATAGATTAGTAG	(SEQ ID NO:106)
Reverse Primer: CAAAAAAACAAATAACCACCTCC	(SEQ ID NO:107)
Blocker: CCTCCACAAACTCACTCCTCACAAATAC	(SEQ ID NO:108)
Probe: red640 TTTCGTTTGATGGTAGATACGGGTGA-phosphate	(SEQ ID NO:109)
Probe: ATTAATGGTTTATAAGACGGATTTTTAAAC GT-fluorescein	(SEQ ID NO:110)

[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 $\mu$ M
Probes:	500 nM
Taq polymerase:	0.1 U/ $\mu$ l
dNTPs:	0.2 mM each
Magnesium Chloride:	4 mM
BSA:	0.25 mg/ml
Roche buffer with no MgCl <sub>2</sub> :	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 MgCl<sub>2</sub>

[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 dH<sub>2</sub>O). 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 dH<sub>2</sub>O 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)	GGATAGGAGTTGGGATTAA AGAT AAATCTTTTCAACACCA AAAT	414
EYA4 (SEQ ID NO 3)	GAT GGAAGAGGTGATTAATG CCCAAAAATCAAACAACA A	226
H-Cadherin (SEQ ID NO 4)	GGT TTTGTATTAGGTTGGAAGT CCAAATAAAATCAACAAAC AACAA	286
TPEF (SEQ ID NO 5)	GA TTGTTGGGTTAATAATG CTTCTCTCTCCCCCTCT C	295

[0247]

Table 2: Hybridisation oligonucleotides

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

**-continued****Table 2: Hybridisation oligonucleotides**

Gene	Oligomer sequence
Versican (SEQ ID NO 2)	AAGATTTTGGTTAGTT (SEQ ID NO 89)
Versican (SEQ ID NO 2)	ATGTGATTCTGTTGGTA (SEQ ID NO 50)
Versican (SEQ ID NO 2)	ATGTGATTCTGTTGGTA (SEQ ID NO 51)
Versican (SEQ ID NO 2)	GGGTAACGTCGAATTAG (SEQ ID NO 52)
Versican (SEQ ID NO 2)	GGGTAATGTTGAATTAG (SEQ ID NO 53)
Versican (SEQ ID NO 2)	AAAAAATTCGCGAGTTAG (SEQ ID NO 54)
Versican (SEQ ID NO 2)	AAAAAATTGTCGAGTTAG (SEQ ID NO 55)
EYA4 (SEQ ID NO 3)	TATATATACGTGCTGGTA (SEQ ID NO 56)
EYA4 (SEQ ID NO 3)	TATATATATGTCGTTGTA (SEQ ID NO 57)
EYA4 (SEQ ID NO 3)	AGTGTATGCGTAGAACAGT (SEQ ID NO 58)
EYA4 (SEQ ID NO 3)	AGTGTATGTCGAGAACAGT (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)	GAAGTGGTCGTTAGTTTT (SEQ ID NO 65)
H-Cadherin (SEQ ID NO 4)	TTGTTTAGCCTGATTTGT (SEQ ID NO 66)
H-Cadherin (SEQ ID NO 4)	TTGTTTAGCTGATTTGT (SEQ ID NO 67)
H-Cadherin (SEQ ID NO 4)	AAGGAATTCTGTTTGAA (SEQ ID NO 68)
H-Cadherin (SEQ ID NO 4)	AAGGAATTCTGTTTGAA (SEQ ID NO 69)
H-Cadherin (SEQ ID NO 4)	AATGTTTCGCTGATGTTG (SEQ ID NO 70)
H-Cadherin (SEQ ID NO 4)	AATGTTTCGCTGATGTTG (SEQ ID NO 71)
TPEF (SEQ ID NO 5)	ATTTGTTTCGATTAATT (SEQ ID NO 72)

**-continued****Table 2: Hybridisation oligonucleotides**

Gene	Oligomer sequence
TPEF (SEQ ID NO 5)	ATTTGTTTCGATTAATT (SEQ ID NO 73)
TPEF (SEQ ID NO 5)	ATAGGTTACGGTTGGAG (SEQ ID NO 74)
TPEF (SEQ ID NO 5)	ATAGGTTATGGGTTGGAG (SEQ ID NO 75)
TPEF (SEQ ID NO 5)	AATTGCGAACGTTGGG
TPEF (SEQ ID NO 5)	AATTGTCGAAATGTTGGG

**[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)	AATGTTTCGATGTTG (SEQ ID NO: 70)
H-Cadherin (SEQ ID NO 4)	AATGTTTTGATGTTG (SEQ ID NO: 71)
TPEF (SEQ ID NO 5)	AATTGCGAACGTTGGG (SEQ ID NO 76)
TPEF (SEQ ID NO 5)	AATTGTCGAAATGTTGGG (SEQ ID NO 77)
Versican (SEQ ID NO 2)	GGGTAACGTCGAATTAG (SEQ ID NO: 52)
Versican (SEQ ID NO 2)	GGGTAATGTTGAATTAG (SEQ ID NO: 53)
H-Cadherin (SEQ ID NO 4)	AAGGAATTCTGTTTGAA (SEQ ID NO 68)
H-Cadherin (SEQ ID NO 4)	AAGGAATTCTGTTTGAA (SEQ ID NO 69)
TPEF (SEQ ID NO 5)	ATAGGTTACGGTTGGAG (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)	AGTGTATGCGTAGAACAGT (SEQ ID NO 58)
EYA4 (SEQ ID NO 3)	AGTGTATGCGTAGAACAGT (SEQ ID NO 59)
Versican (SEQ ID NO 2)	AAAAAATTCGCGAGTTAG (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)	AAAAATTTGTGAGTTAG (SEQ ID NO 55)
Versican (SEQ ID NO 2)	AAGATTTCGGTTAGTT (SEQ ID NO 88)
Versican (SEQ ID NO 2)	AAGATTTGGTTAGTT (SEQ ID NO 89)
TPEF (SEQ ID NO 5)	ATTTGTTTCGATTAATT (SEQ ID NO 72)
TPEF (SEQ ID NO 5)	ATTTGTTTGATTAATT (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)	AATGTTTCGTGATGTTG (SEQ ID NO 70)
H-Cadherin (SEQ ID NO 4)	AATGTTTGTGATGTTG (SEQ ID NO 71)
TPEF (SEQ ID NO 5)	AATTGCGAACGTTGGG (SEQ ID NO 76)
TPEF (SEQ ID NO 5)	AATTGTGAATGTTGGG (SEQ ID NO 77)
H-Cadherin (SEQ ID NO 4)	AAGGAATTCTGTTTGAA (SEQ ID NO 68)
H-Cadherin (SEQ ID NO 4)	AAGGAATTCTGTTGTAA (SEQ ID NO 69)
Versican (SEQ ID NO 2)	GGGTAACGTCGAATTAG (SEQ ID NO: 52)
Versican (SEQ ID NO 2)	GGGTAATGTTGAATTAG (SEQ ID NO: 53)
EYA4 (SEQ ID NO 3)	AGTGTATGCGTAGAAGGT (SEQ ID NO 58)
EYA4 (SEQ ID NO 3)	AGTGTATGTTGAGAAGGT (SEQ ID NO 59)
EYA4 (SEQ ID NO 3)	AAGTAAGTCGTTGTGTT (SEQ ID NO 62)
EYA4 (SEQ ID NO 3)	AAGTAAGTTGTTGTGTT (SEQ ID NO 63)
TPEF (SEQ ID NO 5)	ATAGGTTACGGTTGGAG (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)	ATAGGTTATGGTTGGAG (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)	AATGTTTCGTGATGTTG (SEQ ID NO: 70)
H-Cadherin (SEQ ID NO 4)	AATGTTTGTGATGTTG (SEQ ID NO: 71)
TPEF (SEQ ID NO 5)	AATTGCGAACGTTGGG (SEQ ID NO 76)
TPEF (SEQ ID NO 5)	AATTGTGAATGTTGGG (SEQ ID NO 77)
TPEF (SEQ ID NO 5)	ATAGGTTACGGTTGGAG (SEQ ID NO 74)
TPEF (SEQ ID NO 5)	ATAGGTTATGGTTGGAG (SEQ ID NO 75)
Versican (SEQ ID NO 2)	GGGTAACGTCGAATTAG (SEQ ID NO: 52)
Versican (SEQ ID NO 2)	GGGTAATGTTGAATTAG (SEQ ID NO: 53)
H-Cadherin (SEQ ID NO 4)	AAGGAATTCTGTTTGAA (SEQ ID NO 68)
H-Cadherin (SEQ ID NO 4)	AAGGAATTCTGTTGTAA (SEQ ID NO 69)
H-Cadherin (SEQ ID NO 3)	AAGGAATTGTTGTAA (SEQ ID NO 69)
EYA4 (SEQ ID NO 3)	AGTAAAGTCGTTGTGTT (SEQ ID NO 62)
EYA4 (SEQ ID NO 3)	AGTAAAGTTGTTGTGTT (SEQ ID NO 63)
EYA4 (SEQ ID NO 3)	AGTGTATGCGTAGAAGGT (SEQ ID NO 58)
EYA4 (SEQ ID NO 3)	AGTGTATGTTGAGAAGGT (SEQ ID NO 59)
Versican (SEQ ID NO 2)	AAAAATTCCGCGAGTTAG (SEQ ID NO 54)
Versican (SEQ ID NO 2)	AAAAATTGTTGAGTTAG (SEQ ID NO 55)

[0251]

TABLE 6

Genes analysed according to FIGS. 16 to 18	
Number in Figures	Gene name
<u>Healthy vs Non-Healthy</u>	
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
<u>Healthy vs Carcinoma</u>	
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
<u>Healthy vs Adenoma</u>	
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
<u>Carcinoma vs Adenoma</u>	
18-B	CDKN2a
7-E	GPIb beta

[0252]

---

SEQUENCE LISTING

&lt;160&gt; NUMBER OF SEQ ID NOS: 110

&lt;210&gt; SEQ ID NO 1

&lt;211&gt; LENGTH: 965

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Homo Sapiens

&lt;400&gt; SEQUENCE: 1

gatcaattaa gggcatctta gaagtttaggc gttcccgctg ctccttga gcacggaggc 60  
caccaacccc taggggaaag agatgttagcg cgaggcaggg gtgtcggtgc aagaaatttc 120  
gacgcttctg gggactgagg acaaagggtgc ggacacgacc ccggggtacc tggagtccg 180  
tgactcgcbc caaggacggc acacctaggc gctaatttct gctctgcctc aaagaacctc 240  
aagcttagagt cttgcctcc gcccacagcc ccgggatgcc gctgctgcgc tcacccaca 300  
ggcagcgcggc ggaccggctg cagcagatcg cgccgtgcgc gttccaccgg gagatgggg 360  
agacgctgaa aagcttcttt ctggccactc tggacgctgt gggcgcaag cgcccttagtc 420  
cctacacctg cttagctgaa cgctcaggca cagtgaaact gaaacccgg tccctacctc 480  
tgctgagctg aacgctcagg cacagtggaa ctgaaacccgg gttctgggg atgtgagagc 540  
tgttgagggtc acgcgtaatt gggtgtgtat gaggggcgctt gttcgatgt tggtcaggtt 600  
tgatgcaagc aggtcatacg tggcgatgt tggtggatgc accggccgg agactcgag 660  
gcaggcgttgg gacacgtttg agtgaacacc tcaggataact cttctggcca gtatctgttt 720  
tttagtgtct gtgattcaga gtggggcacat gttggggagac agtaatgggt ttgggtgtgt 780  
gtaaatgagt gtgaccggaa gcgagtgta gcttgatcta ggcaggggacc acacagcact 840  
gtcacacctg cctgctcttt agtagaggac tgaagtgcgg ggggggggg acggggccgg 900  
aatagaatgt ctctggaca tcttggcaaa cagcagccgg aagcaaaggg gcagctgtgc 960  
aaacg 965

&lt;210&gt; SEQ ID NO 2

&lt;211&gt; LENGTH: 16579

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Homo Sapiens

&lt;400&gt; SEQUENCE: 2

tgttttttaa aatgatgtct agtgaacca gccaatggag aaaaaatataat ttatataataa 60  
ttttcctcta agttcaactag tcattttata gatcagttat aactttggct actccaatga 120  
gttaatttct ttcttagagac attataaaact cttaaaagat gggaaaataat tgttactttt 180  
ttgaataatg cctactatct aaaacaatgc acataaaaagc tgctcaataa actttttgtt 240  
aattgagaga aatataaaaa aattgaaagc agtctgtttg taataatgtt ttgctaaatgt 300  
ttacagatct tctctttttt ttggaaattat tgcaacaaat atttgacctt tgattttag 360  
gattaaaatg tgatctca tattttataa ggagcaagca ttagtccgtc aattcttttt 420  
gtcaatagaa taaaaatgga atgtatgtt gctgcttcta catgccccggt gataccatg 480  
ctgaacactt ttcttacaac gcagagggtt cttgttttc tgcgtgcgt tctacctttt 540  
ctcaccacag agacattaat gatgccatgt gtcttttgc accatcagtg cagcagctt 600

**-continued**


---

cattatgctc cttgtctcg	tcccactgta	gtctaattaa	ataaatggaa	acgggtagtc	660	
actgaaaata	aacagcaact	gaaaaagttt	tgccctttaa	gaaaccttct	gcattgtatt	720
ccgcctgcc	tcattgattc	taagatctct	accgcttcct	tgtgttcaga	aaccagactc	780
ttcagttgt	cttgcttcag	tcagacaataa	gctttataaa	aatgaccat	atttttcaga	840
ctgtggggag	agaagaaaaa	gactttgaga	aagtgactaa	ggagggagta	agagtgatac	900
aatctttgat	tcaatttagt	tggactgtgt	tctttatgcc	ttcatctttt	tctttataaa	960
aaagatagcg	tttttgcgtc	tctgctccag	gcagcctccc	agggtgagtg	cagatgtgct	1020
gagccatttg	gctgcctga	ggaccggctc	agagagcctc	tcttctact	aatgaagtca	1080
tctcctctcg	gcagccggc	ctctccatc	actccatgtt	ggaacttgc	tttggottgc	1140
atagttctgt	cgtgatgttag	atggcttttc	agttaagttc	tagagaaatc	tgctcaacca	1200
tgaacaggac	atttacagca	caggagaaaa	aaaaatttat	taaaacaata	attttactgt	1260
aacatttga	tttgggggtt	ggagagcaaa	aactcaat	ctgatgtctg	ggaaaaatct	1320
aatgtttat	ggctttctaa	ggaaatttg	ctgtgtttaa	tttatcagag	caatttat	1380
acttcgggtt	aaagcacaat	gagattgcca	ccatca	gcctttgc	agatttctt	1440
ttaagaata	aaaagaaggc	aataaagacc	ttatatttct	gccttccta	gaaagccaaa	1500
gagaattcaa	agtgcgtggat	gaaatgtctg	tccctaattt	gcagtgtttt	atttaaaaaa	1560
aaaattatga	cacaatctta	gaggataggc	ttccttttc	cagggcttt	gaaagttaaa	1620
agtacattt	taatctgttag	cattactgta	tcatccagtg	aggataattt	gcttagttgt	1680
gccttaggacc	agttagatca	aggagttgaa	actaagctct	ttctgaattt	gaatggact	1740
ggaatcttgt	catccctaga	gcagcaatag	atgatgccat	gtctaagaac	acttcacctg	1800
atcaataaaa	aaggagaagt	ctgaaatgta	atcctaaga	acctcagaag	gaggttttaga	1860
tgaagataat	aaaattccct	ggagacacat	ctgcccac	cacctttcct	ctccaagtca	1920
ttggcttgct	gccccctgagc	tcagccaaat	tcccaaagac	aagatcaaag	caaggaagca	1980
aggcacatag	cttctggctt	atggtttctt	ttagacacaa	gcattgtgaa	tgaagatctt	2040
gctttatcta	tacctggat	cttggagttt	tcacaaagaa	tatcccagt	acaaggatgt	2100
gatttaaagg	gggactcaca	gaataatattt	ttaattactg	tttgattcat	acttgagtgt	2160
aatggacaaa	taaatacaat	tttagttata	aaacctgaaa	aatatatattt	ttcccattca	2220
aaaactctg	atagttttgt	caggcattgg	agaatcagaa	ataaatgctc	tctcatttc	2280
tttctcttga	tgaaggaaag	tcattaaaaa	cagttacata	gaaataataa	tacaactgta	2340
gactctattc	taaataatagc	tggtatgttt	tccttttct	tttttatcca	ttctccattt	2400
tagacagcta	gaggggaaag	tctgccc	tagatgaaag	aagatatcca	tgggtataa	2460
tttcatttta	gttatctgat	agcctgaaaga	agtat	tttctgataa	ataatgttac	2520
aagtcaaaga	agagaaaaaga	caggcatttt	aaagctttc	aacattgtgc	atattctaat	2580
tccttaagtc	agccctgaga	aagtgcctt	gtcctgtaga	agtatttgg	caagtgttgt	2640
gaagtctatt	ggcattaaaa	gaaactaaaa	attttggagg	attagagaat	caaatttggc	2700
aacataaaatg	aaaacatctc	agaatgtaaa	aagtatctaa	cccatcactt	tcaatcatgt	2760
tccattatata	acagatggca	ttat	ttat	ttat	ttccctaaaa	2820
atgggcatta	atctgctatt	agttgcttgc	ttctgatcta	gttctcc	ttt gtaaagg	2880

-continued

**-continued**


---

ccaccacgct tcctatgtga cccgcctggg caacgccgaa cccagtcgcg cagcgctgca	5220
gtgaattttc ccccccaaact gcaataagcc gccttccaag gtaatcacgt ttcttttgtt	5280
ccccccctaa aaacacaaaaa caaaaaactt atagaaaaaa acccgcgagc tttagaaaaaa	5340
gaagcaattg gtagaaggct ttaattaagg caaagagctg taaggcgaag ttaagaaaaat	5400
gtaggcactt aaaaaatgca ggtaactttc ataaaggctt ttggggagag gcatacagag	5460
ggaccttgtt gttaaaaag attcagacaa aagaaaccca ggggggggtt ggggtaaaat	5520
gactaacgga attggggaa gggagggaaat aaattgtaaa gaaatcatag aaaagtggtg	5580
gttcttgag ctggagagaa gagagggacc tttggactt tgatTTTTT gtttgttgg	5640
ttcttaacac gctcgaggca aaagttgaa tggggactac caagacttgc cacagacaag	5700
tccccgaagc cgccttgtt caggccaccc ggtttccag cccctgggtt gtggtcagtg	5760
cctgggtttc ctggaaagcc actcccgccc agtcctgac agtgcgaccc ggcgcacaag	5820
cagectggga ctttgcgggg acctgacccc ttcaagaccgc aggcaagtctg ggaggaggc	5880
cgcccgaaaaa aggtgcagga tccccccgt gtcttttga cgacttgggg actgtcacgg	5940
ttctctcccg gcgccttggg gttttttgtt cctgcacgcg gtgcgaaggg gccagcagg	6000
aaggagcaga ggtatgggggg tggggttgtt ggagccccgc ggaggctctgg gaggccccctg	6060
ggcggggaaaaa gcctgttctg aatcggcagg gatgtgcaac aatttttctc accctgaaga	6120
gtgaaatagg gtctgtcgct cccatctcaa caagcaaaacc ggcacccaga ggcactgca	6180
gacaaagggtg gctcggggac ccgaatcagg ggcctctggg tcagtcctct cgcccaagact	6240
acaggagtc tccgtttct tacatattcc ccaccctctc ctatgttcc gctctagttt	6300
agcaacttta ctggcagggtc tagcggcagg ggcgcgcgcg tggccgggaa gccccgggg	6360
acgtcctcgg ctggagcggcc ccaccgttcc tcagaggcggtt gggcgctgtt gggcgacatc	6420
cctgggtcggt gcaagacccatggg ttgttctggc ccgcggctctg tgtactggg	6480
cgtggagcgg tctgggtgtt aggcaagggaa gacgggggtt gaaaaatagt gtcatagcct	6540
aaactatTTT ctctccaatc caactgcgtc tcggcaagtc ctgcatttc ctggagatg	6600
ctgcgggaag ggggagaaat ctacacggcg cctggagagt tgcctgcgcg cgccacacc	6660
cgcggcagag tctctctggg tcgcgttccatcttgc tctccctctg tctccatctt	6720
ctctgcctcc ctccccctct cccagcccttctt gttctctctc tctagcttct gctccccc	6780
ccccacagct gacaaatgaa tggctagtgt gaaatccctg cctttccctgt gctccaaggc	6840
agcagggagg gaggagcggag ggaggcggtt cgctccctcg gatctcccc cagttcagct	6900
cacaagactt gcagaaccta gatgtctagg aattgggagt tttgcggcg gtgtggcg	6960
ccctgtatgg agaagtcccg cacaggcggaa gaaaaacaag ccccccagac caagcgctgca	7020
tctttcacaa ctttgtgttc aaggatggaa ggcccttagtt ttcccttaag tcattcattt	7080
tgcactccac aatctgcggc gtacatcttag gatTTTggag gaccctgaaa aaaggctttt	7140
gtctgtgcaa agtgcagaga tgccttcgtc cgagtcggct ggaacgcacg cggcgccctg	7200
ggtcggatgtt gcccgtttagt ggacgaacgc acatggccgg ggtggagcc ggtttcgag	7260
gtcttagacgc gcccaccctgg gcgttttaaa gaaagataat acacataccg tgctctcaaa	7320
acccatgacc acttgaaccc aacggtaggg ccccgccccca ggatattgca aaagaaggc	7380
tcctaataccaa aacttaaac ttccctcaact tcagggcgccc ggttcggacgg gaaagtggag	7440

**-continued**


---

agaaggcggg cagtggagg aaaaaagaaa gggaaaggaag ggaagtggaa ggaaagagga	7500
ccgtggaggg gagaaaagg agaaaggac ggagatgggg cagatggcat tcaaacagcc	7560
aggaacggac cggaagggtt cgtttcgggg tttgggtctg aatttcgaga ctcgttttag	7620
gatactctt tcccttccc agcggcaaca aatgtgactg cagaggcgtg ccggatggaa	7680
agggtggaa agaactagac aagcggaggt ggtccagccc cgcccaaagg ggctgggtct	7740
ggggcccccc cccactcgct cacctcgccg caagggagga cccgcaacgg gcgcctcccg	7800
cagcggctgg gttagccagc cgccacgtgg gttggcgctt cggctccag ccgagacggg	7860
tgtgacttcc tgggctcaga gactgcgatg cagccgcaca ggcggccggc gcagccgagg	7920
cccccaccca ccgaggccaa tggggccggg gcggggaaagg cagggccag cttaaccgag	7980
agagccgagt ggcctcgccc ccgggattcg gggccggctg cttctggagc tgaccgag	8040
tccatccaga cagcgctgaa cactcttattt aaaaaggctat ttaagttta ttttatttttta	8100
ttttatttttt gcttgcttct tgggagtggt aggagggca tcgtgctttt gttttgaggc	8160
aggggagagg taaaattta ctccaagcaa actttttgtc agaaaaatga gactttatac	8220
cgtattccat ttgttactt tttaaataca ttcaaggcac tctaaagccc attcttagtc	8280
aattaacaga ttaatttcgt tccagactgt cataaagagt tcgcctggaa tggcaaatac	8340
tgccattcga caagactaat agaatttcat gggggcaaac atggaaggcgc tgcgagtgt	8400
actgaggttc ctgtgcctca atgaggaacc ttagcgggaa cctggcgggg actggaagct	8460
cagcacctat ttagggagtg actgagctgg aaacctggag aagcttgggtc cagacgagac	8520
gggtgcttctt agagctgcat ttaagtccctt cccatTTTCG tcgcataatgg aagaactgg	8580
ttttaaagat ctttccatc acctcagattt ctgttagctaa cagccacgtc tttttttttt	8640
tgtctgctga aagctgtgac ctaattttttttaaagcacttc ttgacaattt cgccaaaccc	8700
atccgtaaaa tattcacata aagtccctaaa ataggatgtt gtgagtgta tataaaaaata	8760
ggaaggtacc aatggaagca acaatttagtg ttctgagata gaagcgggttc ccagtttgta	8820
acttggaaat gttaaagagca gtacatcagg gagaaggctt ccaagagatct gacttttgt	8880
tctgagctctt gagtaaaagaa agaaatatgt tagaattgtt ttgtctggcc gcagtaat	8940
tgtaaagggt tccaatggat tccaatgttta ttgagcaattt acttcctcca gcaccaagtc	9000
cacagtcctt aatcttcttg tgagaagattt aggttaggaag cgggatgaga ttcaagacaa	9060
tcatttagga tttatgtcag aattcaaaatgtt ttgtattttt ttggttagata aaaacagagt	9120
tggttttaaa ttgtatTTCC tgctttcaaa aaaacgaggtt ttgaaaggaa tccccctaga	9180
tttacatagg ttacatagg taagtttctt ctttttccctt gtgtttgcaacttggaggaa	9240
gatgcctttt atattccaga atacctgtat ctatacactt gtaaattctg ctagggaaagg	9300
acctaaccacg ctaaccacaaa ggtgaccatg attaatgttta aagataactac ccattttct	9360
ctctttttttt gtattttagt gcttaaaatc tttggagttt atgattggtc acttaattgtt	9420
gtAACAGAGC CTTTGAGAGC GCAAAGATCA ACTAACTCTG TCTCCAAAG TCTGGCAACT	9480
AAGCCCTCCT TTCTGGTGA GATTTAAACA TTTTTGTGA CGTCATTCTG GGACAAACAT	9540
CGGTAACACAC TGTATTTAAA ATGGGATATT GCGCTGTGAT GGAGCAGAAC TTCTCAAGAC	9600
ACATGGTTCTT TTGAAAGAA AGAATTAATG ACTGTGTTAA GGTATTTTTT CACCACCCAC	9660
GTAGAGAAAC CAGTGGAGAC TGACAAACAC TCATTGTTGG GAGTTACTTG CTCTGGTCTG	9720

---

**-continued**


---

agaaaatatg	ctgattgtta	cataggctg	tatTTTACCA	gaaaacacaa	agcatgtact	9780
ctccctgatt	tggaatcaga	aaagggggat	tagatATGGA	tgtCTTGGCC	aagCTCATT	9840
tcaattatgt	aatTTTGCc	agTTTCTCT	TGTAGCCTAG	ACTTGGTAaa	ttACATATA	9900
ataaaAGTCTG	TGTGTGTG	TGGGGTCGCT	GtGTGTGT	ATGTATACAG	GGTTTTAAA	9960
cactgctaag	ttCTTACCAT	tGAATGTGTT	AGAGGTGGTT	GTATTCAGA	GGTCAGTACA	10020
aagaatCTTT	GTtTAATGTG	TCAgTTCTGT	ttCCCTGAAAC	ACACCTCTGG	Gtagattac	10080
caactataaa	aggatacAAA	TGTATATACT	ttCCACATCT	GTATACATCC	TACCTCATT	10140
ttCTAAGGAT	TTTAAGAGGT	GTtTATGTCT	ATAAAATGAG	ATTACATATT	ttACAATAAA	10200
catATTTGAT	AAATTATAAT	CTTTTATTA	TTAAATTATT	GATTGATAT	TACTGAGGT	10260
atAGTGGAA	GAGCACTAGA	CttGGAATCA	gaaAGTTGA	GTtGGACTCA	CTAGCTGAGA	10320
aaACCAAGCA	AAGATAATTAC	TTTTGGACCT	TTAGTGTCT	AATCTATTA	ATGAATATGC	10380
AGTTATAAAG	ATCAAAGAG	TAAGTACTTT	gaaaaactat	ACAAAAAAGG	GATGATAATA	10440
tattatgtct	TGGGATTCT	ATTGATCTAA	ttCTATAATT	TTAATAATAA	CTGCATTACC	10500
caAGCCATGT	AGTTCTACT	GAATCTATAA	AATAAATCAG	AGAGGGCAT	ATTGTTAA	10560
AAATTGTTA	GTtTCATTAA	TcAGAAgaca	GATTACAGAA	ATTGACTCCA	ATGAAGTTG	10620
acATTCACTG	gAAAGGAAGA	ATTATTCAAT	AAATGGTTc	AGGACAATTG	ATGGGTCCAT	10680
gCAAATAAG	AATCAGTCTC	TACCTGATT	CTCATCCTAA	AAGAAATATA	AGATGAATCA	10740
aatATGAAAT	GTAGGGAATA	ACTTTTACA	AAATGATAAA	AGAAAATCAT	GGGATTAAAA	10800
AAAATAAGT	AGGAATAGAG	AAACCTGTTc	TAAGCAAAAC	ATAAAACCCA	GTGCCATTCT	10860
AGGAATAACT	GATAGGTCTG	ACTAAGTAA	AATTACACAA	AAGCAAATAA	TATATTGGA	10920
AAACATCTAC	GACATTGAG	AGACAAAAAG	CTAATAACTC	TTGATATTCC	ATGGACTTT	10980
ATGAATAACT	TTGAAAAGTC	CACATCTATC	AAAATATGAG	CAAAAATTAT	GTACAGGTAG	11040
TGGCAGAAAG	AGACATAGAG	TTTAAAATT	ATAGAAAAAA	TCGCTCAGCC	TCACTGTAA	11100
TTAAATAAAT	AGTAATTAAA	ACACCAATGA	TATGCTACAC	CATTTTCA	CCAATCAGAG	11160
GAGTAACAAA	AAAAATAAT	TTTTTCTTT	TTTTTTTTT	TTGAGATGGA	GTCTGTACT	11220
CAGGTTGGAG	TGCACTGGCA	CGATCTCAGC	TCACTGCAAC	CTCCGCCCTC	CAGGTTCAAG	11280
CAATTCTCCT	GTGTCAGCCT	CCCTAGTAGC	TGGGATTATG	GGCACCCACC	ACCATGTTG	11340
GCTAAATTTC	GTGTTTTAG	TAGAGTGGGG	TTTCACCATG	TTGGCCATGG	CAGTTGATCT	11400
TGAACCTCTG	ACCTCAAGTG	ATTCACCCAC	CTCAGCCTCC	CAAAGTGCTG	GGATTACAGG	11460
CATGAGCCAC	TGTGCTTGGC	AAAAATGAAA	TTAAAAAGTT	TGATGCAAGT	TTGTGTTGGC	11520
CAGGTTGTGG	AGAAAACCTGGG	TCTCTCATAT	ACTGTTGTAT	GAAGTAAAT	AGGTTCTCT	11580
ACTTCAAAAA	ATAATTGGC	AATCACTTT	TTTGTATTCT	TTATCAAAAG	TTTTAAAAG	11640
ATGTATATAC	TCTTGTATTG	AGCAATTCCA	CTTCTCGACA	TTATTATATG	GATGGATATT	11700
GGTACCAATA	TCACCCAAGA	CAAATGGACA	AGGATGTTCT	TTGCAAGCATT	GCTTCTAAGA	11760
GCAAATATCT	GGAAACAGTG	TAACAGGCTC	ATCATAGGGA	ACCATTAAA	TAAGAAGTTGC	11820
TGTATCCACC	TAATGGAAAA	CTATGAGCT	GTCAAAAATA	ACATGGAAAA	ATTGTATTTG	11880
CTGATTGGA	ACATTCTCCA	AAATATATTG	TATTCTAAGT	GAAGAAATTG	ATAGATAAAA	11940
CTGAGTATAT	ATTATACATC	TAATTATGTA	AAATGGGTGT	ATATTCTTAC	TTTAACAAA	12000

**-continued**


---

tatttactga	gtgttctagg	cattgtggat	acaggagtga	acaaaacatc	aaagactgct	12060
ctgttggAAC	ttttactatg	gtaggtagAG	aaaggcaATT	cccaatcaca	tatgcAAAT	12120
ggtgatATgt	gctaggatGA	aaatagAGCA	tgggAAAGGG	aatAGAGTAC	agtATgtCTT	12180
aggtagAGTA	atgagAAAGG	agttcaccAG	gcagagaACa	gaggAGTGT	aaggGGCGAC	12240
aatgtgAAA	agcttagCCA	agtccAGGTT	caggaATTCA	catgtgCTA	ttacttGTA	12300
gaagtaAAATT	catgAAATTt	cgcacctaAA	tcacgtGATT	tctctctGCT	gtacaATGAT	12360
tctttctGTG	agaaAGAACCC	agtgagCAAT	atgatATGT	atgctcATTa	catccACTT	12420
gtagctGAAG	tatCCTAGAA	agttggTTG	ctagggTACC	taattAGCAG	aatggTGGAC	12480
aaaaccACCG	tgtctctGTC	cttaAGTCAC	cccttGTCAG	ctgagCTTC	ccAAATCAA	12540
gggagaatac	taaataATCTA	ttaatAGAA	tcttggCTT	ctctcgGAGG	ctgcccATCC	12600
ctggTGTGTC	ccttatCCTC	cctgcAGACA	tttctCATGG	ataggAGTTC	cattgACCTT	12660
gtcttttATG	tctcccAGGC	aattAAACAG	tctcAGTACCC	ctgtacCTGG	attgttCAAT	12720
gactttCTTC	ccttctCATT	gacttcACTG	gagtagAAGG	ctaaatATAG	gcattcACTC	12780
tcataatCTCT	caggTCTTTC	catgcAGACA	gagttgttAT	cactAGTAGA	cggTGTtAT	12840
aagagagACA	tggAAAAGTG	gaagaATGGA	aatAGGGATT	tggagtgGTG	ctaaaACAA	12900
agaaAGAGGC	tctgaaAGAGC	tttccATTtA	atATGATCAA	acatAGAAGA	caAAAGATGT	12960
aacttaATGA	ccaaggATAG	ggacatAGCC	ttgagAAAT	tacAGAAGTG	gaatGAACGT	13020
cttggactCC	aaagaAGAGG	gaatggAAAT	gcagttCTGA	ggcagCTAA	gtcaataAAAG	13080
ccctctGTATA	tatttactGT	agaATAAAT	gatcatCTT	taattcaAGG	tttgcaAAACT	13140
taggtGACTA	caggggcGAG	gcatggAGT	gaaatggGTA	agagACTAAG	tggggcAGT	13200
aatggcCTTg	agcacacAGA	cttaAGGGGA	gagactGCTG	cttggtaAT	gctcatGATE	13260
agccagCTGA	aatatAGACT	tagtGTTGT	agacCTTCTG	atTTTTGAA	agaAGTTAGA	13320
gattttATAT	ttagtGTAAT	cagtCTAATT	tttaAAATTGA	gtAACTTATT	caaACATATT	13380
tatGCAACTT	attcaAAATGT	atttagAGCA	tgtctGATCT	cctcCTTGTa	ttcttACCTA	13440
gtcttagGGG	taatCTTAAG	gaaagtGATG	cgtataCTG	ttgtacCTAT	gtAGATGGAG	13500
gaccattCTC	ctctccccAA	caaAGACAC	ttttATAAGG	attgagtgTG	gcacAAAGAG	13560
ttcgattttt	agtatttAA	tttagtGAGGA	aactCAAATA	cagaAGATCC	tttcctGGGT	13620
cacacCTTtG	gttttttAA	atgagATTG	actctCTAGA	tattgacAA	catcatATTa	13680
ttgggtCATG	gtAAATCTC	aagtggACTG	attcaAAAGT	gttcaAAAGT	atttcAGGTG	13740
ttaAGGAATT	gttggcacGG	agtttcaaAG	gtgtttCTGc	ataAGAAATA	gaggcGTTGA	13800
tcatgtATT	gatctacAA	cattcCTCAA	ggatataCC	tgaacCTGGC	aatGAAGAAA	13860
tattgtGGGA	attattaAGT	tgaataAAAGT	actagCTG	ccttccAGAA	gcctactGTT	13920
ctacAGAGGA	gattAGACAT	gtttacAGAT	aatTTGAAGT	acatAGCAAT	aaattatAGG	13980
atgattttAT	ccaAGCATAA	agtgtatATT	gtctactATAG	tggtaACCTA	ggttataATT	14040
ttcaAAATGTA	agAGCCCCAA	atactATTCT	ttttactCTT	agtGATATGA	aactATGTTA	14100
gaacatCTGC	agAAAGTCTG	aattAGTAAT	ttaataATAA	aatattGAGT	actGATATAT	14160
tgtatATCAC	tGAGATTAT	tttatataCA	ttattttATC	aattcCTATG	atcactTCGG	14220
taagtactAA	ttttcacaAC	taagtagtGA	cagAAATGA	ggtttagAGA	gatataGTAT	14280

**-continued**


---

tttacccaaa gtcacagaat ataagtagag aatgcagaac ttgcattctg ggctattga 14340  
 cagtggagcc cagacttttta aaatttaggc tataaagatt tcctggaaaa aaagaaaaat 14400  
 aaaaaataga gaacaaattt tcctgaagaa aagtaggaaa ttattattta taagtataat 14460  
 taaagagaga tccttaattt cttttttttt ttttttttg agacggagtc tcgctctgtc 14520  
 gcccaggcgc aatctcagct cactgcaacc tccacccccc gggttcatgc cataaagaaa 14580  
 gtttctattt gattggcgc acctgattgt ctactagaac ataggcttc cagtagggtt 14640  
 aacaagtcat aacagttcta gtgatataag ataaatacaa actaataata aggttaattt 14700  
 agaattatca acatatgttt ttcatgtga attatcagg tggctataga tggaaaatat 14760  
 aagatatata atgttattagc ttaaaaaatg ctcttcagag agtgtctaga attatagaat 14820  
 ttaaaaaattt tggtttgaaa aggtctaatt ctatattcaa aatatataaa gaaatcaaatt 14880  
 gcaaaaaaat taagagaaaa ggtcttaattt aattaaatgc tggtaactc caagccctat 14940  
 ttatgttcaa gatcgtttt gtaatttatg actaatttca gcattagctt gtttcaatc 15000  
 taggtttgct aaataatata gaatttctt tccaatgaag ttttcaggg tcccgctga 15060  
 aaaatataata ccagtttagga tgcttcacac tgtaaggaac aaaaagtca gatTTTATTG 15120  
 ggtttaaaca ataatgacat ttataatctc atataaaatg aagttcaaaag gcatggtcca 15180  
 ctcttaggtct aatccttaag gacttacgcc ttgccttgc tgatttttc tcagttcat 15240  
 ttccagaggt atctcactta ctcatggatg caaaatccct acaacagttc cagcagaaag 15300  
 attccagcat gataatgtcc agggaaagaaa actgaccacc ttttccttgc ttctttctc 15360  
 aggatgaatc tatctttccc agggaccctt cactggagga agagatcccc tcacatctca 15420  
 ctggccaaaa ctatattata tactgttcca aatccaatca ttggaaaagg gaatggtacc 15480  
 tctataattt acttaggata tcaggatcat cccttggct ggggtttgac ccaggattct 15540  
 ccaaagtatg tgactgaggg taggcactag aatggaacca gagttgttgg agaaaggaga 15600  
 aacgcatggc tattatagaa tagttctaaa tgctaccgag gagggcacaa ctgtaaaaac 15660  
 caaataatct ttgcatttc tttgaagtt ggcactttga tttcttagatg gttcccaac 15720  
 acaggttctt ctccctctca tcattatagc tgccatttca tttgagctgg aaggaaacat 15780  
 tctgagaccc agattgttaa atgtcttttca aaaaatcttgc caataaatta aatggcaaaag 15840  
 ccagggcagt ttcttgcgtc agtacagggt attttcttgc attctttact cttgagactt 15900  
 tagaactgtt ggtactgttt taaaattcat ggcaagaact ggtcactttt gtaattaaca 15960  
 cttccattata atacatttttgc tttgtttgtt tagccagctt gaaactacat ggagtctgt 16020  
 cttaaaaaag cctgccgaag tccttatttgc ctgtttttgtt attatgttca tgaaccacca 16080  
 attgggttctt tctcacccat cacttgcgtt agatgcctt ctttcaacat ctttcttat 16140  
 tgctcccat cttctcttgc tctatattatg atcagcttc tggtttctaa tagacttgc 16200  
 ggtcaccat ttctttttgtt gccagcttgc atccacttagt tacttgcact gttggctctca 16260  
 ttgttcttca tacagcctac tcacttgcgtt ttgtcacacaca catccacatt gttatgtt 16320  
 cccaaactgtt attctggat gatTTTGGTA atggctgcatt tggatggatg tcaaaactaat 16380  
 aattaaagca ttgagatagc ttctgttattt acaagtttac tcctgtttt atagtccaaag 16440  
 aggagctact tcttctactc ttattactta atgcttaata ctacccttat tacatacaat 16500

-continued

gcacaaaag catgtgattt atgaccact ttaaacctga atgcttgta ttactttgt 16560  
gttttcttc atttctagg 16579

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

<400> SEQUENCE: 3

acatTTTaaat tgcatTTTT gCTTGTaATA tattacatAT taaaAGAAat tactatTTT 60  
ctGAATACCA ttaataCTAA taatAGCAtC acAGTgACTA agAGTggACT taaATAAAGG 120  
ctggCTTgAA actcAGGgTCtT gCCATTtAtt AGCAAGCTtC taaaATTCT gAGCCtTAG 180  
ttttCTCaCA tGTgAAATGG agAAAATAAA tCTGCAAAT tAATTAAATC tGTCCATTCA 240  
tATTTTCTC tCCAGTgTAT attAACTgGC ATTCCTCGT aggCCAGAAt GTGCTCTCAA 300  
ccATGCTCCA aATCCGCTTT gtGCCAAccc CACTGCCAGA ACCCTTCTA CCTTGAGAAC 360  
cAGAAAAGGA AACATTATGC CTGGCAATGC CTACACCCtC CAAAATAAAT CTGAGGAAA 420  
GAACACCCAG TAAGTgATGA gAGCAGCAAC GACTGCCtTt ATCAATTAA ATTACAACA 480  
CCACCTTTtC tagAGCCTCT taAGCATTGT AGATAATTCC CCACTCATTAA AAAAATAAAT 540  
tGTAACCATA AGTATTCCAGG GTTgATACTG CTTTGAATT AGACAGTgCT CATATCAGTT 600  
GCATAAGACC AACCTAAAGT AGAGGATGAA ATCTTTTtC tGAACCTTT TCAGAACGTA 660  
ACTTAGTgAA TATATTAAAAA CTTAAACTTtC tttGAATGGG AGTAATTCT ACGGATTAA 720  
CTGTAATCTC ttagACCACA CCTAAGGTAA TGTAGAGGTT GTTgATCAT AGGCTTGTG 780  
ATTAGAGACC ACTGGATTtG TCTTGGAAA AGTCACTCAT GATTCTTGG GCTTTGGTT 840  
CCTCATCTT TATACCGGtC TAATAATGAC CACCGTAGAG TTGTCATGGA AGTTAAATGA 900  
ACTCATGTAC CATAAGTgAC CAATAACATG ATTGACACTC AGTgATTtAT CAATAAAATT 960  
AAACATTtAT TATCAATATG ACAGAGAAGG TGCCGCTAA AtAGACAAATA GGTTTTGG 1020  
AGAGGTGATT AAATGGATGC AAAATTtATG GATTGTTtAT TCCGCTTACtC tttGCTGTG 1080  
CCCTGGTTG TGGCATAACAC ACgtGTGGGT AtAAAATCgT AAATCCTATG TAGTCGCGTA 1140  
GTGCTATGCG AGAAGGCTTA GACACGAAAT GTCATTTCAg CAATGTGCT AGAGAAGCTC 1200  
TGACGCCGCC TTGGAAgTAA GTGCTTGTG CTCGACCTTt GGGCGTCTGG GACGGATGCC 1260  
TATACCTGCA CCCAGCAGCA CTGGAAGGGG CCCAGGGCCtC TCGCAGCACA GCCTATCCCC 1320  
AGACCGCTTA GTCCTTCATA ACATATATCT CCACGGAAAA GGtTATTCC TCCCCTCAGA 1380  
AAAAGCGCCC CAGTCTGGTC TGGGTTGGTT TTTATTTCAc GTTgTTGCAA GTAGGCGAAG 1440  
TCCCTTCTGT CTCCCTCCCTT GGGGTAAGTG GAAAGGAGTC CGGCGAGGGG CCCCAGTGG 1500  
CTGCAcAGG AGGAActGGGT AGCAGAGAG ATTCAGGCAa TTCCGGGGGtC TGCCCCACAG 1560  
AAGCAGGTGG GGATCGACAG TGGCTCTCCG GCCCAGGGAG GAGAGCGCGG TCGCGGGTCC 1620  
CTCCCTCTAG CCTGGGGCTG GCAGCCGCTC GAGTCGGCCC GGGTGGGGGtC GGGTGGGGG 1680  
CGGCGCGGAG GGCACGGAGA TTACGGCGGC GECACCCGGG ACATCCAGGG CCCCAGGAGC 1740  
CTGGCGGGTC CCCACGCGAG ATCGAAACCC ATGACAATAG GCAgTCACCC GAGGTCAAAT 1800  
AAAAACGGAG TGGGTCCCCC GCGCGCCGtC GCCCCCGCG TCCCTGGCGG CCTCCCCCGA 1860  
GGCCCCCGGC GGCCTCACGA GCCCGAGTA GCCGGTGGC AGTGCCTCCCC CGCCCCACCT 1920

**-continued**


---

ccctgcgcaa	gtgcgaggct	gccggcagcg	cggcgcacgc	tccggccgtt	cccgggttcc	1980
gcgcaaaact	tccatcctgt	ccacgtgaag	ttgtcgctgc	cttagagagg	gggaaagagc	2040
tgcgggaaaa	gcccgggagt	gacgactgcg	gcggctggc	gcgcctctc	attttcttt	2100
cttctccctt	ccccctgtc	gcagtccgga	gttttgctc	ctctcccttc	ctcctcccccc	2160
tcggagccgg	cttctccctc	cgcccccgtt	ctcccccgct	tgtgtacgct	atttgttgt	2220
gggtggccga	aggggatgtc	ctgtttcac	cagaggcaca	gcgcgaaggg	gaaacttcga	2280
cactggaagg	aacgagaata	aatacttaat	tacggacgca	ctgaaccgcg	gctgggacag	2340
acacttcggg	aacctcgaggc	ggaccggggc	acgagggtgag	tgacccttc	ttccaacccc	2400
cgccccaggg	ctcccggggg	agcctgagtt	gagagaaccc	ccaaacttcc	cgggaaagtg	2460
cgcgaggctc	cgccggggac	gccgagcgct	gggtactgag	gacgcgcagc	tggacggtgc	2520
gtgggcgcct	gcgtccccgg	ggggcgcttg	gaggccgggt	gccccacgcc	tgagggcccg	2580
ggccgcgtcg	accgcagcg	tgatctctgc	cctagaagac	gtccccaagc	cccaagggtc	2640
cttccgagc	ctgcctgtcc	cttccgggt	oggcgcggag	cctgcgcgtt	acggagttca	2700
tccagcagtc	cagcgcgcgg	cttctacctg	caccccgct	ccacctggca	gaggcgcgag	2760
catcggggtc	tccccacat	cttcttatg	acgtgtat	cttctgtatg	acccctaga	2820
tggtccaggc	gcgaggatgc	tgaccagag	tccttcggag	ggtcacaggc	gcctggctt	2880
tcccggtgcc	gggtgcgtgt	gtactttaaa	ggctcgcggt	ctaattccca	ggcactgatc	2940
gggttttca	actgcggcga	tcccaactta	atagtttta	tgtggcggtt	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

agtttcagnc	atca	tnca	tctttatcaa	tgnccttggg	nccacagggn	gn	gncc	ttct	60
gnccaaacag	cagccaaatgn	agnatctt	aaaggcacat	cactttat	ttattc	agt	gt	ttca	120
gttatgnat	cacttaatcc	tcaccacaac	cttgcngnt	cagtacaatt	tgtgn	ttca			180
ttttatctgt	gaacaatgg	aggcagnnnt	ttgngtaatt	tacaagtaag	tggc	agatcc			240
gggggctgtg	cctagacaat	gtggctccag	actccttgc	cctccctg	ttct	gtgt	ca		300
tcttgattga	gtctagccat	ctatttttt	aacacacact	aattgcgttc	tccat	gttcc			360
aggccctgtg	ccagatgtca	gaaatacactc	cccatccact	ggatcttggc	atgg	tccatt			420
gctcagtgtc	ctgctatgtt	acgtgttaca	taatgcattt	attccggat	ttag	ccat			480
gaaaataatg	ctgaaagaca	ttgtatgcat	gtttctacaa	caaaactatg	taactt	atgt			540
ttcttttatt	gcttggctca	tca	ggagact	cttatccaa	tca	gagacaa	ta	cat	600
tttgggcttgc	tcat	ttcagt	ttgtgtact	tttccttc	gccc	agat	ttt	gg	660

-continued

ttcataaaat ctattgcaca aatgtccctt attgtaaaaac accttaaatt ctttctaagg	720
gaaggctgca tggaaatgtat acagtaaggt cttccctgca ttttcttaga ttccatttag	780
ggaaggcaga cctatgttc cctcttaccc tcagtcacca agccccccat ttataaaaatc	840
ccttaagcag tgactactgc tggtctgagt acctggaggt agttcagagc ttctgaaagg	900
taaaatccat atacaaaaag aagttcccttc cgatccagat cccagcttgc gtgccagatg	960
cacatttgag gagtaggtgg ctatgcacgc tctcacctga gcagttata aaatctattt	1020
cccccttaatg gaattttttc tgcaagctcg aattgtatctg tcatcttgc gattttgttag	1080
atggcaggga agcaccaaac accatcatga ctggggccac agtggtggaa aaaaaggaaa	1140
aaagaaaaaa aaaatccact gccaagcctt gccaggcgta naaaggctg gaactgctgg	1200
ggccatTTTA tctgatttat tggaaataga gtggatctt ttaacatTTT aataaagaga	1260
atcttttgc actaggctgg aagtggccgc cagtcggccctg tgcaatttcca ttctgtggaa	1320
aagtggaaatc agctggcatt gcccagcgta atttgtgagg ctgagccccaa acagtccaaa	1380
gaagcaaatg ggatgccacc tccgcggggc tgcgtccctcg cgggtgtctc accccgtatc	1440
tgccatgcaa aacgaggagg cgtaggaag gaatccgtct tgtaaagcca ttggcttgg	1500
tcatcagcct ctacccaaatg ctggctgtat gctgctgtat ctatTTTgg gaagtggct	1560
ggctggcgag gcagagccctc tcctcaaagc ctggctccca cggaaaatat gtcagtgc	1620
gccgcgtgca tgaatgaaaaa cggccggggc cgcttctagt cggacaaaat gcagccgaga	1680
actccgtctcg ttctgtgcgt tctcctgtcc caggttagga agaggggctg ccggccgc	1740
tctgcgtcccc gtttctgtcat tcggatgcgc cggcacggggc agggttaggg ggcttgcgg	1800
gggtcggggc ctccggcgc ggcggcgaaag acagatcggtt gctcggttagg gaggtcattc	1860
cgagccaga gatccttaggc acccccccaca cacaggctcc cactctggcg tgctgcgtg	1920
tgtgtgtgtg tgggtgtgtg tgggtgtgtg tgggtgtgtg tgggtacgtt cgtaacggg	1980
aqqa	1984

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

<400> SEQUENCE: 5

gtctttgggtg agatatgtgt tttacaagtt ttaatggaga aaaatgtaag tattttacct 60  
cctgaaacct ggctatttga gtaatggaaa aatagtcact ttccccagga cagtggttct 120  
caatcatggc tatgtgtttc tccaggaaaa cttaaaaaat atatatatac caatgcttct 180  
gtgtcacttc tagggattcc aagtcttga atacgaactc tgcatcagta ttcttaatt 240  
atccaggtga ttgtgtatgtg aaatcatgac tgagccccac tgctctaaga tgaaataaac 300  
tttcctcagc actgaaatca ccaaactaaa ctaccaaaat taattaaggg catggaaatc 360  
aataaggcat agggaaagctt ttacattata aaattatttc tttaaatcac agctcattgt 420  
ttatatgtta ttgcatttgc tagaaaaggg tgaaaaaata gcaaatttaa ttactctcag 480  
tttggaaaat tatccagaaaa tgaagatgac gactctgaaa cattgtcaat atcatttgc 540  
ctataaataa tgttctaata catttactac acactgatag atacttttc atatgaatata 600  
tatacattaa aactaaggca ataatgcatt tagaacattc tatctatatac tatgtatctt 660

-continued

**-continued**


---

ttggcagctc tatctaattgg gtattctgtt tccttaaaa atttagctgc tctgtctaga	3000
agccgatttt ctgatgcctc caacgtctgg tctaaattgtat ctgttttaat ggagtcttcg	3060
tccgttagga gcgagatgcc accgactaga atgctggat ctgctgccta attgccagga	3120
gtgagagaca ctgagattca gaaatctttt gaggtggag gggagaggga cagtctcgga	3180
cggaggcgga gatgttaagat aaaggatgg atttcacaca gaaaaaaa aaagatttcg	3240
ttgaggact gaggtgctgc acgatcacat ctctcaaagg agaagttaaa aagcaaggaa	3300
gtgggaggag gttggaggat aaagtactta aaaggattac tcgggtacaa tttgttttc	3360
tgctgggtgc tgcaaaggat agatagtccc gtttcaaag tataatgaatg cctctttaa	3420
gtgattggga atggacacta attgcctgtt aaatgttatac aaatgctctc ctaaatttcg	3480
gggacacaga aagaggggca caaaaggaga atttaaatag aaaaaggag gatccggagg	3540
cttttgaag cggggggaga agaaggagga gggataacag agaggaatag agaaggagag	3600
cggagagaag ataaacaaaa acaaaaacag gaatcactga ataatcacac accaaaaaga	3660
aagctcttcc ctaggggca tccaaaacac tgagactgca atagtaccc cggtcatgga	3720
agaaagatgt tcctctccac ctttgcctt gaaagctttt ggtcccttta ctggcacta	3780
aaattccatt aggctaaaga gtgtgtctaa ctgcctgaag aatgcacgac acggaaggcg	3840
gttcccgcta tgccgtttgc ctttccccgtt ggagagaatg aaagaaaacgc gcagagccag	3900
agactcctgc cgagtttagac cttctctcgat cgccccaggt caccggccat ccggaaaga	3960
cccgagtaag gaacgcaggg tcaactgcctg ggccaacaaa tggagcccg tctcccttc	4020
ccggacgccc ctggccggcc gatgctcccg gcaacccacc cgcggcgat gcagaggagc	4080
ctttctcttt ctctcagacc acttgtcccg accaatctga ctttccaaac acatctgacc	4140
gcaccccttca ggtggacaca ctaataggct acgggctgga gaggagcggg tgatgaggag	4200
aggattcaa acctgcgaac gcttgggctt ggtcgagat gccccggcc tgggaggaga	4260
gaggggagaa gagagaagga aggagagcgc ctgcccggat ggctgagctg cctcggcgag	4320
cagccttggg gttgcacgct ctttgtggat atgctgctgt tgcttccagg tcggcaagag	4380
cggttctaac accatcgcc ttcaccctct ttccctgtaaa tccctagaga aacgtccctg	4440
gcctctccgc cgcgacattc ccagcctgca tccccctaca gcctaggcgg cgcgctcccg	4500
cacgctggag cgccggcgc cagcaggacg ccctctcccg cgccgactcg cccctctctg	4560
ccctgctgct gctgctccctc tgacacctcc gccccccacca tctccagctc ggagagacgc	4620
cacccagccg cggcccgacac tgcggccccc gggtcacgcg cgaaagagg ggcgtactcc	4680
ggaccccgcc ttccgttaggg ggcgtctgg agcggagatg gaggcgaatg gtatatgagt	4740
gtgcgggtat cccaccctga agcccgatct tctcaattga gccatcccc gcctagcccc	4800
actcggggca ggcgtggcg agcggaccca tctgtggctt ccgcggccgc ctctctcttgc	4860
catccttgca cctactcgatc gacccctccc tcccgggacc tgcacccctgc tccaccaatc	4920
agagcccgac tgcctttcc cacgtgaccc cggcgccggct gaggacatgc tgcttccaa	4980
acggccagagg gatgcggcg gcaagatcg agaggcggtt gccggcgatc ggggcgcctt	5040
gactctccctt ccaccctgcc tccctgggctt ccactcgatc gccccctggac tccctgtctcc	5100
tcctgtcctc cggcttccca gagctccctc cttatggcag cagcttcccg cgtctccggc	5160
gcagcttctc agcggacgac cctctcgatc cggggctgag cccagtcctt ggtatgttgc	5220

**-continued**


---

gaaaactctcg agatcatgcg cgggtttggc tgctgcttcc ccgcgggtg ccactgccac	5280
cggccgcgcc tctgctgcgg cggccgcgg gatgctca gtgcacactg agcccgcgc ccggccccgg	5340
cgtatcctgtg ttccctcgaa gccgtttgtc gctgcagagt tgacacactt agtcatggtg	5400
ctgtgggagt ccccgccggca gtgcagcgc tggacactt gcgagggtt ttgctggctg	5460
ctgctgctgc ccgtcatgtc actcatcgta gcccgcggc tgaagctcgc tgcttccct	5520
acacctttaa gtgactgcca aacgcccacc ggctggatt gctctggtaa gtccagaacc	5580
cccgccccgg accctttaac tccgcagaag aacacgcgtc tccagcacag accagctac	5640
cctagcgcgc ctcctcagcc ctcacccatc tactgccta gacccctaatt accaccacc	5700
tctatccaga gaaacaaggaa gaactgttgc aggcccggg gtgaggggtg gttctggat	5760
gggcagaaag tgcaggtgta gcaggaaacc tttgcattgtc tgccgttaca ttggagctgc	5820
gaggattttg agaaatatta aacgggatgg ttttctgggt tcactgtttt gaaagagcac	5880
caatccttagg gaaacactg aaacagaagc tttgtcatca ttaaaagaaaa aagtcttact	5940
aggatgagga agaaataact ttatgagaaa gaatgagcga gaaagcaata aatcaaattgg	6000
tgactgcagg ggaatcgctg attcctggca aaggtccat gaggtcgac tggctcccg	6060
ttgaagacca ggtcacacag attctagagg agctgggtt caatagaatt tctctctctc	6120
tctctctctc tctctctctc tctctctatc tatctatctc tctctctctc	6180
tcattccctt ctctcctagg cggcaaaaga cattgggtt gcagtcaga tatgccttc	6240
tctttgcctc cctaagcttc aaggtagtac agggaggtt agaaaaagaa cactttcggt	6300
gtctcccaagg ccggagtggtt catgactgag gctggtcagg ctccatgttgc gcgagccgg	6360
ggcggaaaccg acttcagtgg ggcgtgactc ctccatttctt ggacaggctt ctgtggagtg	6420
ggtcaggcac tcttcttgcgat cgtcgggtt ctttcagatt ctgacggcga acgcttggca	6480
ggcttcgcgc tgctgaagct tcctaattaa ataggccag agatggggat tggctgcact	6540
cctagctggc atagcattcg gtttgcacgc ctgttagata gggtgtatgt aattttcat	6600
cttctgtgaa tataattttt ctgttagttaa atctggctct gaataaaatgt tctttcaaag	6660
atgtatataa gctgaagtgt atgtactttt agagaggagg gaatgaccaaa ctgtactca	6720
gggtgaaagc ctgtatagtt cctagtttactgtatgtt atgccaatggaa gaaaattattt	6780
atgcatttcattt ctaatttttc ctttacaaag acaagtttgcgat atatgcaccc ctattagatt	6840
tgggtcaata gattgttctc ttttttggca gtttctaaat ttggcattttt aataaaactc	6900
aacatgtttc tataacttctc tgattcatgc gtacatgtgtt gtttttttttggaa aagaataag	6960
tttcaatttttgc tattgccta atcaattttt agatgttttta ttatggtaat aattatgagc	7020
ctgcaaaaac aatttttggaa aatgttgcgtt gtttttttttgcgttgcgtt aataaggttgc	7080
tcattccctag cccttcattt gtttttaggaa ataactaact taaatgtgaa gttgcattt	7140
gcaatcaaga aattacatata ttaccagata ttttaaagggg gactgcataaa actaaagaga	7200
ataaaactggt tttgcagata gttgtcaag aacttggcac cccgcatttcccccctgtttaa	7260
cttagaggttgcgat atcaatcttc atttgcgcgc aacagaccat cacagaaac actgtgcctg	7320
tttatcttta ttatggggc tttgtttccctt ctttgcgttgcgat atacatttca aataaggggt	7380
tgtttcagtc gttgtcaag aagaacaattt aaagatgggg aaatgtttaaa agggatttca	7440
gagatcatca ctagctttt tccaaaatgtt ggtttttgtt ggtcataaaat attgtccacc	7500

-continued

taatgagcaa aaaataaaaa taaaaaaaaa acaggaagca aatgttaagc tttcattcac	7560
cactgtcagt attaacgc aa gctttaaaaa atagcactat cagaaaaagg tactaaagg	7620
gaattgacta gaaaagaatt gtggaaaatg gaaacgaata ttgatcaactt aactagattt	7680
tgagggttatac agtagacagt gacccgtcag tacagctata gttgttggat ttaaaaattta	7740
ggacaagttat tttaaagctt caaagtagtg ct tttttttt 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 gggtattttt gaagtttaggc gttttcggtt ttttttttga gtacggagg 60  
tattaatttt tagggggaaag agatgttagcg cgaggttaggg gtgtcggtt aagaatttc 120  
gacgtttttgc gggatttgagg ataaagggtgc ggatacgtt tcggggattt tgagttcg 180  
tgatttcgcgt tacggacggt atattttaggg gtaattttt gttttttttt aaagaatttt 240  
aagtttagagt tttttttttc gtttatagtt tcgggatgtc gttgttgcgt ttatcgata 300  
ggtagcgttc ggatcggtt tagtagatcg cgcgttgcgc gttttatcgg gagatggtgg 360  
agacgttggaa aagttttttt tttgttattt tggacgttgtt gggcggttaag cgttttagtt 420  
tttatttttgc ttgagttgaa cgtttaggtt tagtggaaattt gaaattcggt tttttttttt 480  
tgtttagtttgc aacgttttagg tatagtggaa ttgaaatttcg gttttcgccc atgttagagt 540  
tgtttaggtt acgcgttaattt ggggtgtatg gagggcggtt gttcggtatg tttgttaggtt 600  
tcatgttaagt aggtttagtgcgt cgtcgagtg tttggatgcg atcggttcgag agattccgg 660  
gttaggtttgg gatacggttgc agtgaatattt ttaggatattt tttttggtaa gtattttttt 720  
tttagtgcgtt gtgatttgcgtt gttggatata gttggagat agtaatgggtt ttgggtgtgt 780  
gtaaatgagt gtgatcgaa gctgagttgtca gtttgcgtt ggttagggattt atatagtattt 840  
gtttatattttt tttgtttttt agtagaggat tgaagttgcgg ggggtgggggtt acgggggtcg 900  
aatagaatgt ttttggata ttttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt gtttgcgtt 960  
aaacq

<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

cgtttgcata gttgtttttt tggtttcggt tgggttttgt taagatgtt tagagatatt 60
ttatccggt ttctgtttttt tttttcgta ttttagttttt ttatcaaaga gtaggttaggt 120
gtgtatgtgt tgggtgggtt ttgttttagat taagtttata ttctgtttcg gtttatattta 180
tttatatatata tttaaaattta ttattgtttt ttaatatgtg ttatatttga attatagata 240
ttaaaaaaaata gatattgggtt agaagagtat ttggagggtt ttatcaaac gtgttttaag 300
```

-continued

ttttttttcgg	cgccgtcgat	ttatatatttc	gtacgacgt	gatttttttgc	360	
tattaaattt	gtatataattt	cgaataggcg	tttttttatta	tatthaatttca	cgctgtgat	420
taatagtttt	tatatttctgt	agaatcggtt	tttagtttttta	ttgtgttttgaa	gcgttttagtt	480
tagtagaggt	agggaaatcggt	gttttagttt	tattgtgtttt	gagcgtttag	tttagtagag	540
gtagggatttta	aggcgtttgt	cgttatagc	gtttagagtg	gtaagaaaaga	agttttttag	600
cgtttttattt	atttttcggt	ggAACCGCta	gcgcgcgattt	tgttgttagtc	ggttcccccg	660
ttgttttgc	ggtagcgat	gtacggat	ttcgggggtt	tggggggagg	taaggat	720
agttttaggt	ttttttaggt	agagtagaaaa	ttagtttttta	ggtagtgcgt	tcgtggcg	780
agttacggaa	tttttaggtat	ttcggggtgc	tgttgcgtatt	tttgcgttta	gttttttagaa	840
gcgtcgaaat	tttttaggtac	gatatttttgc	tttcgcgtta	tatttttttt	tttttaggggt	900
ttttttttttt	cgtgtttaaa	ggaggtagcg	ggAACGTTA	attttttaaga	tgtttttat	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

tgttttttaa aatgatgtt agttgaatta gtaatggag ataaaaatat ttatataaa 60  
ttttttta a gtttttag ttatattata gattagttt aattttggtt attttaatga 120  
gttaattttt ttttagagat attataaatt ttttaaagat ggaaaataat tgtttatttt 180  
ttgaataatg tttattttt aaaataatgt atataaaagt tgtttaataa atttttgtt 240  
aattgagaga aaatataaaa aattgaaagt agtttgggtt taataagtgt ttgtaagtt 300  
ttatagattt tttttttttt ttggaattt t gtaataaat atttgatttt tgatttgtt 360  
gattaaaatg tgattttta tatttaataa ggagtaagta ttatgttgtt aatttttttt 420  
gttaatagaa taaaatgga atgtatgtt gttttttta tatgtcggtt gatattttatg 480  
ttgaatattt tttttataaac gttagggttt ttttgggttt tgttttgcgt tttatttttt 540  
tttattataag agatattaaat gatgttatgt ttatgttgtt attatttagt tagtagttt 600  
tattatgtttt ttgtttcgg ttttattgtt gtttaattaa ataaatggaa acgggttagtt 660  
attgaaaata aatagtaattt gtaaaagttt tgtttttaa gaaattttt gtattgtatt 720  
tcgtttgtt ttattgattt taagttttt atcggtttt tggtttttaga aatttagattt 780  
tttagttgtt ttgttttagt ttgataaaat gttttataaa aatgattttt attttttaga 840  
ttgtggggag agaagaaaaa gatgttggaa aagtgtttaa ggagggagta agagtgtat 900  
aattttgtt ttaatttagt tggattgtt tttttatgtt ttttattttt tttttataaa 960  
aaagatagcg tttttgtatgt tttgttttagt gtagttttt agggtgagtg tagatgtttt 1020  
gagttattttt gttgttttga ggatcggtt agagagttt tttttttattt aatgaagttt 1080  
ttttttttcg ttagttcggt tttttttattt attttatgtt ggaattttgtt tttgtttgtt 1140  
atagttttgtt cgtgtatgtt atggttttt agttaaagttt tagagaaattt tgtttaatttt 1200  
tgaataggat atttataatgtt taggagaaaa aaaaattttt taaaataaaat atttttttgtt 1260

**-continued**


---

aatattttga tttgggggtt ggagagtaaa aatthaatat ttgatgttg ggaaaaattt	1320
aatgttttat ggtttttaa gcaaatttg ttgtgtttaa ttatttagag taatttat	1380
attcgggtt aaagtataat gagattgttta ttatttattt gtttttgta agatttttt	1440
ttaagaata aaaagaaggt aataaagatt ttatattttt gtttttttaa gaaagtaaa	1500
gagaatttaa agtgttggat gaaatgttg ttttaattt gtatgtttt atttaaaaaa	1560
aaaattatga tataatttaa gaggatagg ttttttttt tagggttttt gaaagtaaa	1620
agtatatttt taattttagt tattattgtt ttatattgtt aggataattt gtttagttgt	1680
gtttaggatt agtagatattttaa aggatgtttaa attaagttt ttttgaattt gaatggatt	1740
ggaattttgt tatttttaga gtagtaatag atgatgtt gtttaagaat attttat	1800
attaataaaa aaggagaagt ttgaaatgtt attttaaaga attttagaa gaggtttaga	1860
tgaagataat aaaattttttt ggagatataat ttgttattta tattttttt ttttaagtt	1920
ttggtttggttt gtttttgagt ttagttaat ttttaaagat aagattaaag taaggaagta	1980
aggtatataag ttttgggtt atgggtttttt ttagatataa gtattgtgaa tgaagattt	2040
gttttattta tattttggat tttggagttt ttataaagaa tatttttagt ataaggatgt	2100
gatttaaagg gggatttata gaataatattt ttaatttattt tttgattttt attttagt	2160
aatggataaa taaatataat tttatgttataa aaattttgaaa aatataatattt ttttattta	2220
aaaatttttgc atagttttgt taggtatttttgg agaatttagaa ataaatgtt ttttatttt	2280
tttttttga tgaaggaaag ttattaaaaa tagttatataa gaaataataa tataattgtt	2340
gatttttattt taaatataatgt tggtatgtt tttttttttt tttttattta ttttttattt	2400
tagatgttta gggggaaag tttgttttagt agatagaaag aagatatttaa tgggtataa	2460
tttttatttaa gttatgtttagt agtttgaaga agtatttaat tttttgataa ataatgtt	2520
aagttaaaga agagaaaaaaga taggtatgtt aaagttttt aatattgtt atattttat	2580
tttttaagttt agtttggata aagtgtttttt gttttgtttaa agtattttgtt taatgtt	2640
gaagtttattt ggtattaaaaa gaaattttttttaa attttggagg attagagaat taaatttgg	2700
aatataaaatg aaaatattttt agaattgtttaa aagtatttaat ttttattttt ttaattatgt	2760
tttattttat atagatgttta ttatattatgt tgaatgtttaa ttattttat ttttttttt	2820
atgggttattt atttggattt agttgtttttt gttttttttt gttgaaaggt	2880
tttattttat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	2940
tagattttttt attttaaaaaa tggttttttttt taggttattt gatttttttt gaaattttttt	3000
agattttattt aaattttttaa agtggataat aaattttat tttttttttt gttttagt	3060
ggtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	3120
tttatttttga attaagtaat tttttttttt tttttttttt tttttttttt tttttttttt	3180
ttggattttt atttggattt gttttttttt tttttttttt tttttttttt tttttttttt	3240
gttattttatg aaggttgaag tttttttttt tttttttttt tttttttttt tttttttttt	3300
aaattttttttaat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	3360
ttgtttttttt atttggattt gttttttttt tttttttttt tttttttttt tttttttttt	3420
gataaaataat aaaagaaaatg aaagatgtat atggtgagaa ttgaaattttt tttttttttt	3480
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	3540

**-continued**


---

ataaaaagaaa	ttttataaaat	gcgtatgtat	aggagtttaa	gttattaaaa	gttttaaagt	3600
ataagtttaa	attaaattaa	ttaaagaagt	tgagaggaaa	aattgggttt	tatTTTtaat	3660
tattattgtt	tttagggttt	atgttaata	taattttta	agtagaggtt	tttagagagaa	3720
gagttgtgag	gatattttta	tatttgtga	gaaggaaaag	tttggttttt	atTTTtagtat	3780
tttttagtgtt	atattgtatgt	gtatTTTgga	tttattttgt	ttttatttgta	taaatttata	3840
tttgattttta	aagaaaagga	aaattttaaag	ttttttttt	ttaaggggat	agaaaattttt	3900
tgtgttaatt	gtttgatttt	tttttttgta	aggTTTTatt	ggaaattttt	tgtaatataa	3960
tgttagggat	ttttttatgt	gttgatgtt	tttataatgt	gggggtgggtt	tgattgaaga	4020
aaaaaaatcg	tatatacgt	tgaagatta	tggTTTTatt	ttcggaaagt	atgaaaggt	4080
attgatattt	ttaagaagtt	tttggatttt	aggaaaattta	ttaaatattt	tatTTtagaga	4140
tatTTggaaa	gattgaagga	aaggaagaac	gaagaaagta	gaatttagat	ttatgtgggg	4200
agagattttgt	ggtagaggaa	aagtattttt	tttgaattcg	ataaggattt	tgTTTgggg	4260
aattttttgt	ttagTTTTT	attattaggg	ttttttgaag	tcgggTTTT	tattgggtag	4320
ttttttggga	gtgttagtggg	gaatttttat	atTTTTTTT	tagTTTTcg	aaggatttcg	4380
tttttttagt	tttttttta	ggttggtagg	agTTTTgagt	ttgatatttt	tttttgatgg	4440
gataggttaag	ttttgtgggc	gcgttaaatac	gttgaatta	agTTTTTgt	tgattttata	4500
tttttgcgt	tttgcgagaa	gaagtgtatcg	tatTTaatg	tttattgtt	gtttgtttt	4560
taagagttt	ggggTTTTT	tttttaatt	tagaattatgt	tgtacggggg	gcggggaaagt	4620
gggggtgggg	aaggagtggg	agggttagtgg	ttttcgcgag	tagacgcgt	ttattgagt	4680
agTTTTGaa	tggggagcgt	tgtgtttttt	aagtgcgtt	gtattttttt	ttaggaagaa	4740
acgttaagag	gtgggagtgt	ttggggaggg	aggtaggcgg	tttttatcgt	aggcgcgggg	4800
agttgttttt	tcgttttttc	gtttgtttt	taagtttga	tttttaggag	tggttgaagt	4860
tgcggagcgt	ttttggagtt	tgtgaatgaa	ttttttttt	ttttttttt	ttttttttcg	4920
ttgagTTTT	ttttcggtt	tgacggata	gtgatataat	gatgtatgg	gttataattc	4980
gtatttgaat	ttttaggcga	gttgtttcga	gtttttttgg	ggaagaattt	taggcgtgc	5040
gacgttaatag	tcgagaatat	taggtgtt	ggataggagt	tgggattaag	atTTTcggtt	5100
agttcgtat	ttttcgtat	tttttagtat	cgtttgcgtat	ttttcgtatt	ttttttcggg	5160
ttattacgtt	ttttatgtga	ttcgtttggg	taacgtcgaa	tttagtcgc	tagcgtgt	5220
gtgaattttt	tttttaaatt	gtaataagtc	gttttttaag	gtaattacgt	ttttttgtt	5280
tttttttaa	aaaataaaaa	taaaaaattt	atagaaaaaa	attcgcgagt	ttagaaaaaa	5340
gaagtaattt	gtagaaggtt	ttaattaagg	taaagagt	taaggcgaag	ttaagaaaat	5400
gtaggtattt	aaaaaatgt	ggttaattttt	ataaaggttt	ttggggagag	gtatataagag	5460
ggatttttgt	gttggaaaag	atttagataa	aagaaatttta	gggtgggggt	ggggtaaaat	5520
gattaacgga	attggggaa	gggaggaaat	aaattgtaaa	gaaattatag	aaaagtgggt	5580
gttttttgag	ttggagagaa	gagaggattt	tttgggtattt	tgattttttt	gttggtttg	5640
tttttaatac	gttcgaggta	aaagtttga	tggggattat	taagatttgt	tatagataag	5700
ttttcgaagt	cgtttgggt	taggttattt	gttttttag	tttttgggt	gtggttagtg	5760
tttgggtttt	ttggaaagtt	atTTTcggtt	agtttttgat	agtgcgtt	ggcggttaag	5820

**-continued**


---

tagtttggga	ttttgcgcgg	atttgatttt	tttagatcgt	aggtagttg	ggaggaggtt	5880
cggtcggggg	agggtgttagga	tttgcgtcg	gttttttga	cgatttgggg	attgttacgg	5940
tttttttcg	gcgtttttgg	gtttttttgt	tttgcgtcg	gtgcgaaggg	gttagtaggg	6000
aaggagtaga	ggatgggggg	tgggttgg	ggagttcgc	ggaggtttgg	gaggttttg	6060
ggcgggaaaa	gtttgtttt	aatcggttagg	gatgtgtaat	aattttttt	atttgaaaga	6120
gtgaaatagg	gtttgtcg	tttattttaa	taagtaaaatc	ggtatttaga	gcgtattgt	6180
gataaagggt	gttcggggat	tcgaatttagg	ggtttttgg	ttagttttt	cgttttagatt	6240
ataggagttt	ttcggttttt	tatataaaaa	ttatttttt	ttagttttt	gttttagtt	6300
agtaatttt	ttggtaggtt	tagcggtac	ggcgccgcg	tgtgtcgga	gtttcggggg	6360
acgttttcgg	ttggagcgtt	ttatcg	tttttttttt	gttagggcgt	ggcggttta	6420
ttttgtgcgt	gttagatttag	ggtattcgg	ttgttttgg	tcgcgtttt	tgttatgg	6480
cgtggagcgg	tttgggtgtt	aggtagagga	gacggggta	aaaaatagt	gttatagtt	6540
aaattattt	tttttaatt	taattgcgtt	tcggtaagtt	ttgtatgtt	ttggagatg	6600
ttgcgggaag	ggggagaaat	ttatacgcc	tttggagagt	ttttttcg	cgcgtatatt	6660
cgcggtagag	tttttttgg	tcgcgtttt	tatttttgt	ttttttttt	tttttatttt	6720
ttttgtttt	ttttttttt	tttagttttt	ttttttttt	tttagttttt	gcgtttttt	6780
ttttatagtt	gataaatgaa	ttgttagtgt	gaaattttt	ttttttcgt	gttttaaggt	6840
agttagggagg	gaggagcgt	ggaggcgg	cgtttttcg	gatttgcgtt	tagtttagtt	6900
tataagattt	gtagaattt	gatgttttag	aattgggagt	tttgcggcgg	gtgtggcgt	6960
ttttgtatgg	agaagttcg	tataggcgga	aaaaataag	tttttttagat	taagcgtgt	7020
tttttataa	ttttgtgtt	aaggatggaa	gttttttagt	tttttttaag	ttattttattt	7080
tgtattttat	aatttgcggc	gtatatttag	gagtttggag	gattttgaaa	aaaggtttt	7140
gtttgtgtaa	agtgttagaga	tttttttcg	cgagtcgtt	gaaacgtac	cggcggttt	7200
gttttagttc	gtcgttttagt	ggacgaacgt	atatggtcg	ggtggagtt	gttttccgag	7260
gttttagacgc	gttattttgg	gcgtttttaa	gaaagataat	atatatatcg	tgtttttaaa	7320
atttatgatt	atttgaattt	aacggtaggg	tttcgttttta	ggatattgt	aaagaagggt	7380
ttttaatttta	aaattttaaat	tttttaattt	tttagggcg	cgtcgacgg	gaaagtggag	7440
agaaggcggg	tagtgggagg	aaaaagaaa	ggaaaggaag	gaaagtggaa	ggaaagagga	7500
tcgtggaggg	gaggaaaagg	agaaaggaaac	ggagatgggg	tagatggat	ttaaatagtt	7560
aggaacggat	cggaagggtt	cgtttcgggg	tttgggttt	aatttcgaga	ttcggtttag	7620
gatatttttt	ttttttttt	agcggtataa	aatgtgattt	tagaggcgt	tcggatgg	7680
agggtggaa	agaatttagat	aagcggtat	gttttagtt	cgtttaaagg	gtttgggtt	7740
gggggttcgtt	tttattcg	tatccgt	taagggagga	ttcgtaacgg	gcgtttttcg	7800
tagcggttgg	gttagtttagt	cgttacgtgg	gttggcg	cggtttttag	tcgagacgg	7860
tgtgattttt	tggtttttaga	gattgttagt	tagtcgtata	ggcggtcggc	gtgtcgagg	7920
tttttatttt	tcgaggtaa	tggggcg	gcggggaaagg	taggttttag	ttaatcgag	7980
agagtgcagt	ggttcgttt	tcgggattcg	gggtcggtt	tttttggagt	tgattcgagt	8040
tttatttaga	tagcggtgaa	tatttttatt	aaaaggttat	ttaagttta	tttttatttt	8100

**-continued**


---

ttttattttt gtttgaaaa tgggaggtggt aggaggggta tcgtgtttt gtttgagggt	8160
agggggagagg taaaatata tttaaagttaa attttttgtt agaaaatgaa gattttat	8220
cgtatttat ttgttgattt tttaaatata tttaggttat tttaaagttt atttttagtt	8280
aattaataga ttaatttcgt tttagattgt tataaagagt tcgtttgaa tggtaaatat	8340
tgttattcga taagattaat agaattttat ggggtaaat atggaagggt tgcgagtgt	8400
attgaggttt ttgtgttttta atgaggaatt ttagccccca tttggccgggg attggaagtt	8460
tagtattttat tttagggatgt attgagttgg aaatttggag aagtttgggt tagacgagac	8520
ggtgtttttt agagttgtat tttaagtttt ttatattttgt tttatattttg aagaattgga	8580
ttttaagat ttttttattt attttagatt ttgttagttaa tagttacgtt tttttttttt	8640
tgtttgttga aagttgtat ttaattttttt taagtatttt ttgataattt cgtaaattt	8700
atttgaaaaa tattttatata aagttttaaa ataggatgtt gtgagttaa tataaaaata	8760
gaaaggattt aatgaaagta ataattttgtt ttttgagata gaagccgttt ttagttgt	8820
atttggaaat gtaagagta gtatattttg gagaagggtt ttaagagttt gatttttgt	8880
ttttagttttt gagtaaagaa agaaatatgt tagaatttgc ttgttttttc gtagtaat	8940
tgtaaagggt ttaatggat ttaatgttgc tttagtattt attttttttta gtattttttt	9000
tatagttttt aatttttttg tgagaagattt aggttaggaag cgggatgaga tttaagataa	9060
ttattnnntttaa ttgtatttttt tggtttttaa aaaacgaggt ttgaaaggaa ttttttttta	9120
tttataatagg tttatagatg taagttttttt tttttttttt gtgttttttgc atttggaggaa	9180
gtatgtttttt atatttttaga atatttgtat ttatataattt gtaaatttttgc ttagggaaagg	9240
atttaaatag ttaaattttaa ggtgattatg attaatgttta aagatattttt ttattttttt	9300
ttttttgtt gtatttttag gtttaaatat tttggagttt atgattttgtt attttaatttgc	9360
gtaatagatg ttttgagagc gtaaagatata attaatttttgc ttttttttttgc ttgttttttgc	9420
aagttttttt tttttggta gattttaataa ttttttttttgc cgtttttttttgc ggataaataat	9480
cggtaaatat ttttttttttgc atggatattt gtttttttttgc ggatgttttttgc ttttttttttgc	9540
atatgtttttt tttgaaagaa agaattttatg attgttagtttttgc ggtattttttgc ttttttttttgc	9600
gttagagaaat tagtttttttgc tgataaataat ttttttttttgc gagtttttttgc ttttttttttgc	9660
agaaaaatatg ttgttttttttgc ttttttttttgc ttttttttttgc gtttttttttgc ttttttttttgc	9720
tttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	9780
tttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	9840
ttaatttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	9900
ataaaatgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	9960
tatttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	10020
aagaatttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	10080
taatttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	10140
tttttaaggat ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	10200
tatatttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	10260
atagtggaaat ggttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	10320
aaattaagta aagatatttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	10380

-continued

**-continued**


---

gttttttatg	tttttaggt	aattaaatag	ttttagtatt	ttgtatttg	attgttaat	12720
gatTTTTTTT	tttttttatt	gatTTTATTG	gagttagaagg	ttaaatatAG	gtatTTATTT	12780
ttatATTTTT	tagTTTTTT	tatgtagata	gagTTTtat	tatttagata	cggTgtttat	12840
aagagagata	tggaaaAGTG	gaagaatggA	aataggGATT	tggagTggTg	ttaaaaATAA	12900
agaaagaggt	tttgaaAGt	tttttattta	atatgattA	atatagaAGA	taaaagatGT	12960
aatTAATGA	ttaaggatAG	ggatataGTT	ttgagAAAAT	tatagaAGtG	gaatgaacGT	13020
tttggatTTT	aaagaagagg	gaatggAAAT	gtagTTTGA	ggtagTTAA	gttaataaAG	13080
tttttGATA	tatttattGT	agaaataAAAT	gattatTTT	taatTTAAGG	tttgtaaATT	13140
taggtGATTA	tagggcGAG	gtatggAGtG	gaaatgggTA	agagattaAG	tggggTAGt	13200
aatGGTTTG	agtatataGA	ttaaggggA	gagattgTTG	tttggtaAT	gtttatGATT	13260
gatTTTATAT	ttagtgtAAA	tagTTTAA	tttaaattGA	gtaatttATT	taaatatATT	13380
tatgtAATT	atTTAAATGT	atTTAGAGtA	tgtttGATT	ttttttGTA	tttttattTA	13440
gttttagGGG	taatTTAAG	gaaagtGATG	cgtatattG	ttgtatttat	gtagatGGAG	13500
gattatTTT	ttttttTAA	taaaagatAT	ttttataAGG	attgagtGtG	gtataaAGAG	13560
ttcgatTTT	agtatTTAA	tttagtagGA	aatttaAAATA	tagaagATT	ttttttGGGT	13620
tatattttG	ttttttTAA	atgagattG	atTTTTAGA	tattgataAA	tattatATT	13680
ttgggttatG	gtAAattTTT	aagtggattG	atTTAAAGT	gtttaaaAGT	atTTAGGT	13740
ttaagGAATT	gttggTACGG	agtttAAAG	gtgtttttGT	ataagaATA	gaggcgttGA	13800
ttatgtATT	gatttataAA	tatTTTTAA	ggatataATT	tgaatttGgt	aatgaagAAA	13860
tattgtGGG	attatTAATG	tgaataAAAGT	attgttttG	tttttagAA	gtttattGTT	13920
ttatagAGGA	gattagataAT	gtttatAGAT	aatttGAATG	atatgataAT	aaattatAGG	13980
atgatttATT	ttaagtataAA	agtgtatATT	gttattatAG	tggtaattTA	ggttataATT	14040
ttaaatGTA	agAGTTTAA	atattatTTT	ttttatTTT	agtgatATGA	aattatGTTA	14100
gaatatttG	agaaAGTTG	aatttagAAT	ttaataATAA	aatattGAGT	attgatataAT	14160
tgtatattat	ttagatttat	tttatataATA	ttatTTTATT	aattttatG	attatttcGG	14220
taagtattAA	tttttataAT	taagttagtGA	tagAAATGA	ggtttagAGA	gatatagtAT	14280
tttatttAA	gttatAGAAT	ataagtagAG	aatgtagaAT	ttgtattttG	ggttatttGA	14340
tagtgGAGT	tagTTTTA	aaatttagGT	tataaAGATT	ttttggaaaa	aaagaaaaAT	14400
aaaaatAGA	gaataAAATT	tttgaAGAA	aagttagGAA	ttattatTTA	taagtataAT	14460
taaAGAGAGA	tttttaATT	ttttttttt	tttttttttG	agacGGAGt	tcgtttGTC	14520
gtttaggcgt	aatTTTGT	tattgtAA	tttattttC	gggttatGt	tataaAGAAA	14580
gtttttATT	gattggTTG	atttgattGT	ttatttagA	ataggTTTT	tagtagGGTT	14640
aataagTTA	aatagTTTA	gtgatataAG	ataaaATAA	attaataATA	aggtaattt	14700
agaatttATA	atatatGTT	tttattGtGA	attattAGT	ggtttatAGA	tggAAAAT	14760
aagatataAT	atgtattAGT	ttaaaaaATG	tttttagAG	agtgtttAGA	attatAGAAT	14820
ttaaaaATT	tgtttGAAA	aggttaATT	tttatttAA	aatatataAA	gaaatTAAT	14880
gtaaaaAAAT	taagagAAAA	ggtttaATT	aattaaATGT	tgtataATT	taagTTTAT	14940

---

-continued

---

ttaggttaa gattagttt gtaatttatg attaattttt gtatttagttt gtttttaattt	15000
taggtttgtt aaataatata gaattttttt tttaatgaag ttttttaggg ttttagttga	15060
aaaatataata ttagtttagga tgtttttaat tgtaaggaaat aaaaagtac gatittattt	15120
gtttaaata ataatgatat ttataatttt atataaaatg aagtttaaag gtatggtttt	15180
tttttaggtt aatttttaag gatttacgtt ttgtttttgt tgattttttt ttagttttat	15240
tttttagaggt attttattta ttatggatg taaaattttt ataatagttt tagtagaaag	15300
attttagtat gataatgtt agggaaagaa attgattttt ttttttttg tttttttttt	15360
aggatgaatt tattttttt agggattttt tattggagga agagattttt ttatattttt	15420
ttggtaaaaa ttatattata tattgttttta aatttaattta ttggaaaagg gaatggattt	15480
tttataattt gattttttttt ttttgggtt ggggtttgtt ttaggattttt	15540
ttaaagtatg tgatttgggg taggttattttt aatggaaatg gagttgtt agaaaggaga	15600
aacgtatggt tattatagaa tagttttaaa tgttatcgag gagggtataa ttgtaaaaat	15660
taaataattt tttgttattttt ttttgaagttt ggtatttttga ttttttagatg gttttttat	15720
ataggtttttt tttttttttt ttattatagt tgttttgaaa ttgtttgtt aagggaaatatt	15780
tttgagattt agattgttaa atgtttttttt taaagttatg taataaattta aatggtaaag	15840
ttaggttagt tttttgtt agtataagggtt atttttttt attttttttattttt tttgagattt	15900
tagaattgtt ggtattgttt taaaattttt ggtaagaattt ggttattttt gtaattaata	15960
ttttttata atatattttgt ttgtttgtt tagttatgtt gaaattttat gtagttgtt	16020
ttttttttttt tttttttttt ttgttcgaag tttttttttt ttgtttttgtt attatgttga tgaatttata	16080
attgggtttttt ttttttttttta tattttgtatg agatgtttttt tttttttat tttttttat	16140
tgtttttttat tttttttttgtt tttttttatgtt attagttgtt tgttttttaaa tagattttgt	16200
gtttttttat tttttttttgtt gttttttttt atttatttttatttttgaattt gtttttttttta	16260
ttgtttttttta tatagtttat ttatgttgg ttgtttatata tattttatattt gttatgtt	16320
tttaaattttt attttggat gttttttttt gttttttttt gttttttttt gtttttttttta	16380
aattaaagta ttgagatgtt ttttttttttta ataaatgtt tttttttttt atagtttaag	16440
aggagttttt ttttttttttta ttttttttttta atgttttaata ttatttttat tataatataat	16500
gtataaaaaaaat tatgttattttt atgatttttta tttaaattttga atgtttttgtt tttttttttt	16560
gtttttttttt atttttttagg	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

ttagaaatg aagaaaaata taaagtaaat tataagtatt tagttttaaa gtgggttata	60
aattatgtt tttttgttga ttgtatgtta taagggttgtt attaagtattt aagtaataag	120
agttagaaatg gtagtttttt ttggattataaaaataggaa taaattttgtt atagtagaaat	180
ttatattttat tttttttat ttagttttttttttaa atgtttttttt tttttttttt	240
tttttagaaata tagttttttttttaa atatataata atgtggatgt gtgtgataat tataagttag	300

**-continued**


---

taggttgtat gaagaataat agagattata gtttaagtaa ttagtgata ggagttgtat	360
taaaaagaaa tgggtgatta taaagtttat ttagaaatag atagttgatt ataaatagag	420
taagagaaga tggggagtaa tagagaaaga tggtaaaaga aaggtatTTT tattaagtgt	480
aaggtagaaaa ggaattaatt ggtggTTTt gtatataata taaaataga gaataaggat	540
tccggtaggt ttttaaagt atagattta tggtagTTTt agttggtaa gtaaataaaa	600
taaatgtatt ataaggaggt gtttaattata aaagtgatta gttttgtta tgaattttaa	660
agtagtatta atagtttaa agtttaaga gtaaagaatg aaagaaaata ttttgttattg	720
agttaaagaaa ttgttttgg tttgttattt aattttatgt atgattttgg aaaagatatt	780
taataatttgg gtttttagaa tgttttttt tagtttaat tttaaggtag ttataatgtat	840
gaggaggaga agaatttgg tttgggaaattt atttagaaat taaagtgtta attttaaaag	900
aatggtaaaa gattatttgg tttttatagt tggtttttt tcggtagtat tttagattat	960
tttataatag ttatgcgtt ttttttttta atataattt gttttttt tagtgtttat	1020
tttttagttat atattttggaa gaattttggg ttaaattttta gtttaaggga tgattttgat	1080
atttttaagtt aattatagag gtattttttt ttttttaat gattggattt ggaatagtat	1140
ataatatagt tttggtagt gagatgtgag gggattttt ttttttagtga agggttttg	1200
gaaagatag atttttttg agaagagaat aaaggaagag gtggtagtt ttttttttg	1260
gatattattttt ttttggaaattt tttttgttgg aattgttga gggattttgt attttatgt	1320
aagttagata ttttggaaa tgaagttgag aaagaatttag taaaggtaag gcgtaaattt	1380
ttaaggatta gatttagtgg ggattatgtt tttgaattttt attttatatg agattataaa	1440
tgttattttt gtttaaattt aataaaatcg tgatTTTt ttttttagt ttgaaagtat	1500
ttaattttgtt atatattttt tagttggat tttgaaaaat tttattggaa aagaattttt	1560
atattttta gtaaatttttag attgaaaata agttaatgtt gaaatttagt ataaattata	1620
aaatttgattt tgaatataaa tagggTTTgg agttgtatag tattttat attaagattt	1680
tttttttaa tttttttgtt tttgattttt ttatattttt tgaataaga attagattt	1740
tttaaaataa aatttttaaa ttttataatt ttagatattt tttgaagagt atttttaag	1800
ttaatattttt atatattttt tttttttat ttatagatattt attgataatt tataatgaaa	1860
aatatatgtt gataatttttta aattaattttt attattttatgtt tttttttttt ttatatttt	1920
agaattttgtt tgattttgtt attttattgg aaagttttatg ttttagttaga taattttgtc	1980
ggattaatttta aatagaaattt tttttttatgg tatgaatttcg ggaggtggag gttttagtga	2040
gtttagatttgg cgtttggcg atagagcgag atttcgtttt aaaaaaaaaa aaaaaaaaaa	2100
aattaaagat ttttttttaa ttatattttttaaa aaataataat tttttttttt ttttttaggaa	2160
aattttgtttt ttattttttaa tttttttttt ttttttaggaa attttttataa gtttaattttt	2220
aaaagttttgg gttttattgt taaatagttt agaatgttaag tttttgttattt tttttttata	2280
ttttgtgatt ttgggtaaaa tattttatattt ttttaaattt tttttttgtt tatttttttag	2340
ttgtgaaaaat tagattttat cgaagtgttattt ataggaatttgg ataaaataat gtatataaaa	2400
taaatttttag tggatataaa tatattttat ttttaatattt attttttttttaa ttattttttt	2460
agatttttttgg tagatgtttt aatatagtttt tatattttttaa agagttttttttttaa gaatgttattt	2520
tggggttttt atattttttttttaa attataattt attttttttttttaa aggttattttat tttttttttttttttaa atatataattt	2580

-continued

tatgttttga ataaattttt ttataattta ttgttatgt ttttaaatta tttgttaata	2640
tgttaattt tttttgtaga atatgggtt ttggaaagg agagtttaga ttatattttaa	2700
tttaataatt ttataatat ttttttattt ttaggtttag ggtatattt tgagaatgt	2760
tttagattt aatatatgt taacgtttt attttttatg tagaaatattt tttgaaattt	2820
cgtgttaata attttttat atttgaataa ttttgaaata ttttgaaattt agtttattt	2880
aagattttt atgatattt aatatgtgt ttgttaat ttagagagtt aaattttt	2940
taaaaaattt aaagggtgtga tttaggaaag gatttttgtt attttagttt tttatattt	3000
ttaaatattt aaaatcgaaat ttttggttt atatatttattt ttataaaaag tgttttt	3060
tggggagagg agaatggttt ttatatttata taggtataat aagtatacgtt attttttt	3120
ttaagattttt tttaagattt aggtttagaaat ataaggagga gatttagatattt gttttaaata	3180
tatattgaaata agttgtataa atatgttga ataagtattt taattttaaa attagattgt	3240
ttatattttaa tataaaattt ttaattttttt ttaaaaaattt agaaggttt gtaatattt	3300
gtttatattt tagttggttt attatgagta ttaattaatg agtagttttt ttttttaagt	3360
ttgtgtgtttt aaggtttattt ttgtttttt ttagttttttt attttttttt attttatgtt	3420
tcgtttttgtt agttatattttaa gtttgtaaat tttgaattaa aagatgattttt ttttttttta	3480
tagttaatattt attagaggggtt tttattgtt ttagttttttt tagaattgtt tttttttttt	3540
ttttttttttt gagtttaaga cgtttttttt attttgttaa ttttttttaag gttatgtttt	3600
tatttttgtt tattttttttt tttttttttt tgattttttt tgattttattt aatggaaag	3660
tttttagaaag tttttttttt tttttttttt tagttttttt attttttttt attttttttt ttatttttttt	3720
attttttttt gttttttttt taatgtatcgttt ttattttttt tagaattttttt ttatttttttt	3780
aaagatttga gagatatgag agtgaatgtt tatattttttt tagttttttt agtgaagttt	3840
atgagaagggg aagaaaatgtt ttgataattt taggtatagg gtattttttt tagttttttt	3900
tttggggagat ataaaagata aggttaatgg aatttttttattt tatgagaat tttttttttt	3960
aggataagggg atatattttttt gatgggtttt tttcgagaga aagttaatgtt tttttttttt	4020
agatattttttt tttttttttt tgattttttt gaaatgtttt tagttttttt tgattttttt	4080
atagagatac ggtgggtttt tttttttttt tgattttttt tgattttttt taattttttt	4140
tttaggatattt tttagtttata aagtggatgtt aatgtttttt attatattttttt ttgtttttttt	4200
gtttttttttt atagaaagaa ttattttttt tagttttttt gtagagagaa attacgtgtt tttaggtgcga	4260
atttttttttta atttttttttta tataagtttta agtattttttt tagttttttt gaatttttttga attttggattt	4320
ggtttaatgtt tttttttttt tttttttttt tgattttttt gttttttttt tggtttttttt	4380
ttttttttttt atttttttttta agatattttttt tttttttttt tttttttttt tgttttttttt	4440
tatatttttttta tttttttttt tagttttttt ttgtttttttt tgattttttt ttatattttttt	4500
atagtaaaag tttaataga gtagttttttt atgtttttttt tagttttttt tttataatgtt	4560
tttagaaatattt tagtttttttta tttttttttt aatgtttttt tagttttttt tttataatgtt	4620
atgtataata tttttttttt aatgtttttt tagttttttt tagttttttt tttataatgtt	4680
ggagaatgtt tttttttttt aatataattttt tttttttttt tagttttttt tagttttttt	4740
ttttttttttt tagttttttt tagttttttt tagttttttt tagttttttt tagttttttt	4800
ttttttttttt tagttttttt tagttttttt tagttttttt tagttttttt tagttttttt	4860

-continued

**-continued**


---

ttttaaagg ttttgttatt taattaagt attaattatt aatttaaag tatttaagtt	7200
taataatata ataaaagaga gaataatggg tagtattttt aatattaatt atggtttattt	7260
ttggtttagt tggttagtt ttttttagt agaattata agtgtataga tataggtatt	7320
ttgaaatatt aaaggatttt ttttttagt tgtaaatata gggaaaaaga agaaatttat	7380
atttgtaaat ttatgtaaat ttaggggat ttttttaaa tttcgtttt ttgaaagtag	7440
gaaatataat taaaattaa ttttgtttt atttattaaa agaattaaa ttttgaattt	7500
tgtatataat tttaaatgtat tggttagt tttatccgt tttttatata atttttttat	7560
aagaagatta aggattgtgg atttgggtt ggaggaagta attgttaat tatattggaa	7620
tttattggaa tttttataa tattattgcg gtagataat ataatttaa tatattttt	7680
tttttattta gagtttagaa taagaagtta gatttttgg gattttttt ttgatgtatt	7740
gttttaataa ttttaagtt ataaattggg aatcggtttt attttagataat attaattgtt	7800
gtttttattt gtattttttt atttttatata tatatttata atattttattt ttaggatttt	7860
atgtgaatat tttaggatg gtttggcga aattgttaag aagtgttaa aataattagg	7920
ttatagtttt tagtagataa aaagaaaaag acgtgggtt tagttataga atttgaggtt	7980
atggaaaaga tttttaaat ttagttttt taatatgtat taaaatggg gatattaaa	8040
tgtattttaa agaagtatcg ttccgtttt attagttttt tttaggtttt tagtttagtt	8100
attttttaaa taggtgttga gtttttagt ttccgttagt ttccgttaag gttttttt	8160
gaggtatagg aatttttagt atattcgtat ttttttatgt ttgtttttt tgaaattttt	8220
ttagttttgt cgaatggtag tatttggat tttaggcga ttttttatgtat gttttggaa	8280
cgaatataat ttgttaattt attaagaatg ggttttagag tgattttat gtatattaaa	8340
agttaataaa tggaaatacgg tataaagttt tattttttt ataaaaagtt tgtttgagtt	8400
aaatttttaat ttttttttg ttttaaaataa aagtacat gtttttttta ttatattttaa	8460
gaagtaagta aaaataaaat aaaataaaat aaaattttaa tagtttttta ataaagatgt	8520
ttagcgtttt ttggatggaa ttccgggtt tagttttttt agtcgggttca gatttcggg	8580
ggcgagggtt ttccggtttt tcgggttaatg tgggtttttt ttttttcgtt tcgtttttat	8640
tggtttcgtt gggtgggggg ttccgggttgc tcgggtcgat gtgcgggtt attgtatgtt	8700
ttaggtttttag gaagttatata tcgtttcggt tgggagtcga agcgttaatt tacgtggcgg	8760
ttaggttaatt tagtcgttgc ggggagcgtt cggtcggtt ttttttttgg cgcgagggtga	8820
gcgagtgggg gcggttttta gatttgtttt ttttggccgg ggttggatta ttttcgtttt	8880
ttagttttt ttttattttt ttatccgtt acgtttttgt agttatattt gttgtcggtt	8940
ggaaaggaa aagagtattt taaaacgagt ttccaaattt agatattttt ttccaaacgaa	9000
atttttccg ttccgttttg gttgtttgaa tttttttttt tttatccgtt tttttttttt	9060
ttttttttt ttttacggt tttttttttt tttatccgtt tttttttttt tttttttttt	9120
tttttattgt tcgtttttt tttatccgtt cggtcgacgt tcgttttggaa gttgaggaag	9180
tttaagttttt ggatttaggat tttttttttt gtaatattt ggggcggggg tttatccgtt	9240
gtttaagtg gttatgggtt ttgagagttac ggtatgtat taaagacgtt	9300
tagggtggcg cggttagatt ttccaaattt gttttttttt cggtatgttgc cggtcggtt	9360
ttaagcggcg aattggattt agggcgatcg gtcgtttta gtcgattcgc gagaaggtat	9420

-continued

ttttgtatTTT tggatTTT aatTTTaaatTTT tttagggTTT tttaaatttt tagatgtacg 9480  
tcgtatTTT tggatgtAA aatgaatgtat ttaaggaaaa attagggTTT ttatTTTtg 9540  
aatataaaggat tggatTTT tttttttttt tcgttgc 9600  
gggatTTT tattaggggg cgTTTtatTTT cgTCgtaaaa ttTTTaaattt tttagatattt 9660  
aggtttgtA agttttgtA gttgaattgg ggtagattc gaaggagcgt atcgTTTttt 9720  
tcgtttttt tttttgttT tttggagta cgggaaaggT agggattttt tattagttat 9780  
ttatTTTtta gttgtgggg aggggacgta gaagtttagag agagagaata gaggttggA 9840  
gaggggaagg gaggttagaga agatggagat agaggagat ataaagatga ggaacgcgt 9900  
ttagagagat ttTgtcgCgg gtgtgcgcgc ggttagataa ttTTTtaggc gtcgtgtaga 9960  
ttttttttt ttttgcgtat attttttagg gatatgttagg atttgtcga acgtatTTTgg 10020  
atTTTggatTTT aaatagtTTT ggttatgata ttatTTTttt atttcgTTT tttttgttta 10080  
atatTTTtagat cgTTTtacgt tttagtgcata agatcgCggg tttagaataat tcggatgttt 10140  
taggtttgtA cgtattaggg atgtcgTTT atagcgTTTtta cgTTTgtA ggacgggtggg 10200  
gcgttttagt cgaggacgtt tttcggggtt ttcggatatac gcgcgcgtcg ttgtcgttag 10260  
atTTTgttagt aaagttgttT aatttagagcg gagaatttagg agagggtggg gaatatgtaa 10320  
ggaaacggag gatTTTgtA gtttggcga gaggattgtt ttagaggttt ttgattcggg 10380  
ttttcgagtt atTTTgttt ttagtgcgtt ttgggtgtcg gtttggTTTtg tgagatgggA 10440  
gcgtatTTTtta ttttaggggt agaaaatttT ttgtatTTTtta ttgtcgtattt 10500  
agaataggtt ttttgcgttT aggggtttt tagatTTTcg cgggtttta ataattttat 10560  
tttttatttt ttgttttttt tttgttggTT tttcgtatTC gcgtgttagga taaaagaatt 10620  
tagggcgtc gggagagaat cgtgatagtt tttaagtgcgt taaagagata cggcggggat 10680  
tttgcgttT tttcggcgttT atTTTtttT agattgttg cggTTTgaag gggtaggtt 10740  
cgcgtaaaggT tttaggttTt tttggcgtcg ggtcgtatttT ttagaggtt ttcgggaggt 10800  
gttttttagt aatatttagt attgattata tattaggggt tgggaaatttA ggtggTTTgt 10860  
attaaggcgg tttcgggat ttgttggtaaTTTtttgg tagTTTtttTttaaattttt 10920  
gtttcggcgtc tggtaagaat aataataata aaaaattaa agtggtaaaag gttttttttt 10980  
tttttttagt ttaagaattt attatTTTttt tatgattttt ttataattta tttttttttt 11040  
tttttaattt tcgttagttt ttttattttt atTTTattttt gggTTTtttT tttttgtt 11100  
tttttaataa ttaaggTTTttt tttgtatgtt ttttttaaa agTTTtttTtga aagttattt 11160  
gtatTTTttta agtggTTTata ttttttaat ttcgtttttt agTTTtttTgt ttaattttaa 11220  
gtttttttt aattgttttt ttttttaag ttcgcgggtt ttttttata agTTTtttTgt 11280  
ttttgttttt taagggggaa ataaaagaaa cgtgattatt ttggaaaggcg gttttttgt 11340  
gtttgggggg aaaattttattt gtacgttgc gctgattgggt tcggcgttgc tttaggggggt 11400  
tatataaggaa gcgtgggtt tcggggaaagg atgcggggg tgccggacgg tggtaaga 11460  
tgcggggagga tgcgggggttgc gtcgaagatt ttggTTTtag tttttgttta taatattaa 11520  
tgttttcgtt tggtaatgcgtt tcgttgcgttgc gttttttttt tagaaagggtt cggggtaggt 11580  
cgTTTgttaag tttaaatgcgtt gttgtgtata tttattattt ttatattattt gtatcgTTT 11640  
agtcgaggag gagatTTTtagc gagaagaagg aggagggaga ggaggagggtt ttattttat 11700

**-continued**


---

gttttaaaag cgtttcgtag ttttagttat ttttaagagt ttaggtttgg aaagtaggcg 11760  
 gaggggcgga aaggtagttt ttgcgcgtt cggtagggat cgtttgcgtt tttttttagg 11820  
 tatttttatt ttttgcgtt ttttttgcgtt aagaagtatt aatcggtttg gggatagtag 11880  
 cgtttttat ttagggattt attagtaat atcggtttgt tcgcggaaat tattgtttt 11940  
 ttatttttt tttatttta tttttcgtt ttgcgttag ttaggtttgg gtttagggaa 12000  
 aggagttttt aggttttttag ggggttaggtt agtaatagat aatttgatc gattatttt 12060  
 ttgcggagt atataaaattt gtaaaattttag taaagaattt ggtagtacgc tgtttacgcg 12120  
 tttatagagt ttgtttgttt tattaaaggg aagtgttagg ttttaggttt ttgttaattt 12180  
 gaaagagata ttgagaaaac gagattttc ggggatttag agggaaagtg taagaattt 12240  
 ttattgtatt ttttaggaat tgtttaatgg ggagttcggt tttaaaagat ttggtaata 12300  
 aaagggttggaa taggaaattt ttttagttaa attttttgc ggatttaaag agaatattt 12360  
 tttttgtta taaatttttt tttatataag ttttagattt gttttttcg tttttttttt 12420  
 ttttagttt ttttaagtat ttttgatgtag aatatttgc aatttttttgc agtaatagg 12480  
 attttttgcgaa agtattaattt atttttatgtt tttttcgaa ataagattat aattttttat 12540  
 gcgtatatgc gattttttt ttttagtttgc gtttatttttgc ttgtttaat agtattaata 12600  
 tatggaaagag tttttgttat tttgtttaataa aagatttttgc ataggattttt atagagaaaa 12660  
 gggtaaataa gttgatataa aggattttttgc ttttttttgc aaagaggat tttggatttt 12720  
 tttttttttt gaagtttaatgtt atgatgttataat ttttttttttgc ttttttttttgc 12780  
 tatttagtata atattaggaa tatttgcgtt ttttttttttgc ttttttttttgc 12840  
 gaaagtattt ttataattttt ttttttttttgc gtttatttttgc agaaaattat attaaatata 12900  
 ggattttaaa atagtagtga ttaaagatga aagttaattt ttttttttttgc ttttttttttgc 12960  
 tagtttgcgtt taaatttatgtt ttttttttttgc gtttatttttgc ttttttttttgc 13020  
 ttttataatgtt ttttttttttgc gtttatttttgc ttttttttttgc ttttttttttgc 13080  
 agtgcgttattt aggttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc 13140  
 attttttttttgc ttttttttttgc gtttatttttgc ttttttttttgc ttttttttttgc 13200  
 aattttatgtt ttttttttttgc gtttatttttgc ttttttttttgc ttttttttttgc 13260  
 aatattttatgtt ttttttttttgc gtttatttttgc ttttttttttgc ttttttttttgc 13320  
 tttaatttttttgc ttttttttttgc gtttatttttgc ttttttttttgc ttttttttttgc 13380  
 tatataatgtt ttttttttttgc gtttatttttgc ttttttttttgc ttttttttttgc 13440  
 ttgttttgcgtt taggttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 13500  
 ttttttttttgc ttttttttttgc gtttatttttgc ttttttttttgc ttttttttttgc 13560  
 tttaatttttgc ttttttttttgc gtttatttttgc ttttttttttgc ttttttttttgc 13620  
 ttttttttttgc ttttttttttgc gtttatttttgc ttttttttttgc ttttttttttgc 13680  
 tatgtatataatgtt ttttttttttgc gtttatttttgc ttttttttttgc ttttttttttgc 13740  
 atttttatgtt ttttttttttgc gtttatttttgc ttttttttttgc ttttttttttgc 13800  
 gtttatttttgc ttttttttttgc gtttatttttgc ttttttttttgc ttttttttttgc 13860  
 agatgttttttgc ttttttttttgc gtttatttttgc ttttttttttgc ttttttttttgc 13920  
 tttaatgtt ttttttttttgc gtttatttttgc ttttttttttgc ttttttttttgc 13980

**-continued**


---

ttagggttg	atthaaggaa	ttagaatatg	tataatgtt	aaaagttta	aattgttgt	14040
ttttttttt	tttgatttg	taatatttt	attttagaaaa	ttaaatttt	tttaggtta	14100
ttagataatt	aaaatgaaat	tatattttat	ggatatttt	ttttatttt	taggttagat	14160
ttttttttt	agttgtttaa	aatggagaat	ggataaaaaa	aaaaaggaa	aatatattag	14220
ttatatttag	aatagagttt	atagttgtat	tattattttt	atgttaattgt	tttaatgtat	14280
tttttttat	taagagaaag	aaaatgagag	agtattttt	tttgattttt	taatgttga	14340
taaaattatt	agaagttttt	gaatggaaa	atataatattt	tttaggtttt	ataattaaaa	14400
ttgtatTTT	ttgttattta	tatTTtagt	tgaattaaat	agtaattaaa	atattatttt	14460
gtgagttttt	ttttaaatta	tatTTTgt	ttggggat	ttttgtgaa	aatttaaga	14520
tttaggtat	aaataaaatgta	agattttat	ttataatgtt	tgtgtttaaa	agaaattata	14580
agttagaagt	tatgtgtttt	gtttttttgt	tttgattttt	tttttggaa	tttggtttag	14640
tttaggggta	gtaagttaat	gatttggaga	ggaaagggtgt	agatggtaga	tgtgtttta	14700
ggaaattttt	ttatTTTtat	ttaaattttt	tttgaggtt	tttaggatt	atattttaga	14760
ttttttttt	tttatttgatt	aggtaagtg	tttttagata	tggattattt	tattgtgtt	14820
ttagggatga	taagattttt	gttttattta	aatttagaaa	gagtttagtt	ttaatttttt	14880
gatttgattt	gttttaggtt	taatTTtagt	aattttttt	attggatgt	atagtaatgt	14940
tatagattaa	aaatgttattt	ttaatttttta	aagattttgg	aaaaaggaaag	tttatttttt	15000
aagattgtgt	tataattttt	tttttaataa	aaatattgt	aatttagggat	agatatttttta	15060
ttagtattt	tgaattttttt	ttggTTTTT	aggaaaggta	gaaatataag	tttttattt	15120
ttttttttt	atTTTTaaaa	aagaaattttt	gtaaaaggta	aagtgtatgt	ggtaattttta	15180
ttgtgtttta	attcgaagta	tataaatttt	tttgataaaat	taaatataga	taaatttttt	15240
tagaaagtta	taaaatattta	gatttttttt	agatattttaga	tattgagttt	ttgtttttta	15300
atTTTTaaat	taaaatgtt	tagtaaaata	tttgTTTTaa	taaattttttt	ttttttttgt	15360
gttgtaaatg	ttttgtttat	ggttgatgt	atTTTTtag	aattttatgg	aaagattttt	15420
tatattacga	tagattttat	taagttaaag	ttaagtttta	atatggatgt	atgggagagg	15480
tcgggttgc	gagaggagat	gattttattta	gtgagaagag	agttttttt	agtcggttt	15540
taaggtagtt	aatggttta	gtatattttgt	atttattttt	ggagggttgtt	tggagttag	15600
tatTTAAAC	gttattttttt	ttataaagaa	aaagatgaag	gtataaagaa	tatagttaa	15660
ttaaattttt	ttaaagattt	tatttttttt	atTTTTTTT	tagttatttt	tttaaagtttt	15720
ttttttttt	ttttatagt	ttgaaaata	tagttattt	ttataaagt	tttggttgt	15780
tgaagtaaga	taaattttttt	ttgaatataa	ggaagcggtt	gagatTTTT	15840	
aattttatgt	gtagggcgga	atataatgt	gaaggttttt	taagaggtaa	aatttttgt	15900
gttggTTTT	atTTTTtagt	attttcgtt	tttattttt	taatttagatt	atagtgggt	15960
cgagataagg	agtataatgt	aagttgttgt	attgtggtg	taaaaagata	tatggtatta	16020
ttaatgtttt	tgtggtgagt	aaaggttagaa	cgtaagatag	aaaataagga	aatttttgcg	16080
ttgttaagaaa	agtgttttagt	atgggttattt	atcggtatgt	agaagttagta	agtatatttt	16140
ttatTTTTat	tttattttt	aaaagaattt	taggattaat	tttgTTTTT	attttttttat	16200
gagaaattat	atTTTTattt	tataaattaa	aggtaaata	tttgTTTTaa	taattttaaa	16260

**-continued**


---

aaaaagagaa gatttgaag attagtaaa tatttattat aaatagattg ttttaattt	16320
tttatatttt ttttaattt aataaaagt tattgatgt ttttatgtg tattgtttt	16380
gatagtaggt attattnaa aaagtaataa ttattnnta ttttaagga gtttataatg	16440
tttttagaaa gaaattaatt tattggatg gttaaagtta taattgattt ataaaatgt	16500
tagtgaattt agaggaaaat tatataaaa tattttgtt tttattgggtt ggtttagtta	16560
gatattattt taaaaataa	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 gtttgaata tattatatat taaaagaat tattattnn	60
ttgaatatta ttaatattaa taatagtatt atagtgatta agagtggatt taaataaagg	120
ttggtttgaa attaggttt gttattnattt agtaagttt taaaattnn gagtttttag	180
ttnnnntata tgtgaaatgg agaaaataaa ttgtaaaat taattaaattt tgtnntttt	240
tattnntttt tttagtgtat attaatttgtt attttcgtt aggttagaat gttnntttaa	300
tatgttnna aattcgtnn gtgttaattt tattgttaga attttnntt ttttggaaat	360
tagaaaagga aatattatgt ttggtaatgt ttattnntt taaaataaaat ttgttaggaaa	420
gaatatttag taagtgtatg gagtagtaac gattgtttt attattttaa atttataata	480
ttattnnttt tagagttttt taagtattgt agataatttt ttattnntta aaaaataaaat	540
tgttaattata agtattttgg gttgatattg ttttgaattt agatagtgtt tatatttagtt	600
gtataagatt aatttaaagt agaggatgaa atttttttt tgaattttt ttagaacgta	660
atttatgtgaa tatataaaa ttaaattttt ttgaatggg agtaattttt acggattaaat	720
ttgttaatttt ttagattata tttaaggtaa tgtagaggtt gttgtattat aggttttg	780
attagagatt attggatttg ttttggaaa agttattttt gattnnttgg gttttgggtt	840
ttttattnntt tatatcggtt taataatgtat ttcgttagag ttgttatgaa agttaaatgt	900
atttatgtat tataagtgtat taatataatgt attgatattt agtgattttt taataaattt	960
aaatattttat tattaatatgt atagagaagg tgctgtaaa atagataataa ggtttttgaa	1020
agaggtgattt aatggatgtt aatattttat gattgtttt ttcgttttattt ttgttgtgt	1080
tttttgggtt tggtatatac acgtgtgggtt ataaaatcgta aatattttatg tagtcgcgt	1140
gtgtatgcgtt agaagggtta gatacgaat gttattnntt taatgtgtttt agagaagttt	1200
tgacgtcggtt ttggaaagtaa gtcgttgggtt tttgattnntt gggcggttgg gacggatgtt	1260
tatatttgta tttagtagta ttggaaaggggg tttaggtttt tcgtgtataa gtttattttt	1320
agatcggtta gtttttata atatataattt ttacggaaaa ggttattttt ttgcgttaga	1380
aaaagcggtt tagtttgggtt tgggttgggtt tttattnnttac gttgtgtttaa gtagggcaag	1440
ttttttttgtt tttttttttt ggggttaagtg gaaaggagtt cggttaggggg ttcgttagtgg	1500
tttgatatttttgg ggaattgggtt agcgagagag tttaggttaa ttccgggggtt tgtnnttata	1560
aagtaggtgg ggatcgatag tggtttttcg gtttagggag gagagcgccgg tcgcgggtt	1620

**-continued**


---

ttttttttag tttggagggtt gtagtcgttc gagtcggttc gggtgggggc ggggtggggg	1680
cggcgccggag ggtacggaga ttacggccgc gttattcggg atattttaggg tttcagggtt	1740
ttgggcgggtt ttacgcgag atcgtaaatt atgataatacg tagtttattc gaggtaaat	1800
aaaaacggag tgggttttc gecgcgtcgc gttttcgcg ttttggcg ttttttcga	1860
gttttcggc ggttttacga gttcgttaga gtcgggtggcg acgtcgttt cgttttattt	1920
ttttgcgtaa gtgcgagggtt gtcggtagc cggcgtacgt ttcggcgtt ttccgtttt	1980
gcgtaaaatt ttattttgtt tacgtgaag ttgtcggtt ttttagagagg gggaaagagt	2040
tcggggaaaaa gtcggggagt gacgattgcg cgggttgggc gcgtttttt atttttttt	2100
ttttttttt ttttttgcg tagttcggg gttttggttt tttttttttt ttttttttt	2160
tcggagtcgg tttttttttt cgtttcgtt tttttcgtt tgtgtacgtt atttgcgtt	2220
gggtggtcga agggatgtt ttgttttat tagaggata ggcgaaagg gaaatttcga	2280
tattgaaagg aacgagaata aatatttaat tacggacgtt ttgaatcgcg gttggatag	2340
atatttcggg aattcggggc ggatcggggc acgagggttag tgattttttt ttttatttt	2400
cgttttaggg ttttcggggg agttttaggtt gagagaattt ttaaattttt cggaaagt	2460
cgcgagggtt cgtcggggac gtcgagcgtt gggattttag gacgcgtagt tggacggcgt	2520
gtggcgtttt gcttttcggg gggcggtttt gaggtcgggtt gttttacgtt tgagggttcg	2580
ggtcgttcgg atcgtacggg tgtttttgtt tttagaaagac gtttttaagt tttaagggtt	2640
tttttcgagt ttgtttgttt ttttcgggtt cggcgcggag tttgcgcgtt acggagttt	2700
tttagtagtt tagcgcgcggg tttttatgtt ttttcgtt ttatggta gaggcgcgag	2760
tatcggggtt ttttttatat tttttttatg acgtgtatata ttttttgatg atttttttaga	2820
tggtttaggc gcgaggatgt tgatttttagt ttttcggag gtttataggc gttttgggtt	2880
tttcgggttc ggggtgcgtgt gtattttaaa gtttcgcgtt ttaattttta ggtattgtac	2940
gggtttttta attgcggcga ttttatttttta 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

gatatttttagt ttacgttata taaaaattat taaaatggga tcgtcgtagt tgaaaat	60
gattagtgtt tggagattag aacgcgagggtt tttaaagtat atacgttattc ggtatcggg	120
aagtttaggc gttttgtgatt tttcgaagga ttttgggtt gtttttcgcg gttttggatta	180
ttaggggggtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	240
tttcgcgtttt tgtaggtgg aggccccccc taggtttttttt tttttttttt tttttttttt	300
tgaatttcgt tacgcgttggg tttcgcgtcg atttcggaaatggataggtag gttcggaaagg	360
gattttttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	420
cgggttttttta ggcgtgggggtt attcggtttt taagcggtttt tcggggacgtt aggcgttttac	480
gtatcggttttta gttgcgcgtt ttttagtattt aacgtttttttt gttttttttt gttttttttt	540
tatTTTTTGGG gaaagtttgg gggttttttt aatttaggtt ttttcggag tttttggggcg	600

**-continued**


---

ggggttggaa gaaggggta tttatccgt cggtcggttc gtttcggggtt ttcaagtgt	660
tgttttagt cgcgggttag tgcgttcgtt attaagtatt tatttcgtt ttttttagtg	720
tcaagtttt ttttcggt tgcgtttttt gtgaaaatag gatattttt tcggttattt	780
tataataaat aecgtatata agcggggag aagcggggcg gagggagaag tcggttcga	840
gggggaggag gaaaggagag gagttaaaat ttccggattgc gataggggg aaaggagaag	900
aaaagaaaaat gagagagcgc gtttgcgtcg ctgtacgtt attttcgtt tttttcgtt	960
gtttttttt ttttttaag gtagcgataa ttttacgtgg ataggatgga agtttgcgc	1020
ggaagtccgg aacggtcgga gcgtgcgtcg ctgtgtcggt agtttgcgtat ttgcgttaggg	1080
agggtggggcg ggggcgacgt ctgtatcggt tatttcgggt tcgtgaggc gtcgggggtt	1140
tcggggggagg tcgttaggga cggggggggc ggccgcgcgc gggggattta ttccgtttt	1200
atttgatttc ggggtgattgt tatttcgtt gttttcgat ttccgcgtgg gatcgttttag	1260
gttttcgggg ttttggatgt ttccgggtggc gtcgtcgtaa ttttcgtgtt ttccgcgtcg	1320
tttttatttc gtttttattc gggtcgattc gagcgggtgt agtttttagg ttgagggag	1380
ggattcgcga tcgcgtttt tttttgggtt cggagagtta ttgcgtattt ttatccgttt	1440
ttgtggggta gttttcgaa ttgtttggaa tttttcgtt atttagttt tttgtgttagg	1500
ttattgcggg tttttgtcg gatttttttt tatttattttt aaggagggag atagaaggaa	1560
tttcgtttat ttgtataaac gtgaaataaa attaattta gattagattt gggcggtttt	1620
tttgacggga ggaatattt ttttcgtgg agatatatgt tatgaaggat taacgggttt	1680
ggggataggt tggttgtcgaa agggtttggg ttttttttag tggttgtgg tgtaggtata	1740
ggtattcggtt ttagacgttt aaaggtttagg tagtaacgtt ttatccgtt ggcggcgat	1800
gagttttttt aggtatattt ttgaaatgtt atttcgtgtt taagttttt gctgtatgtat	1860
tacgcgatta tataggattt acgattttat atttatacggt gtgtatgtt taatttaggg	1920
atatagtaaa ggtagacgga ataataattt tataaattttt gtattttttt aattttttt	1980
ttaaaaattt tattttttat tttagcggta tttttttgtt tatttgcgtt attaaatgttt	2040
taatttattt attaattttt ggtgttaat tatttgcgtt tttttttgtt tatttgcgtt attaaatgttt	2100
ttatatttattt ttatgcgtt attttttttt tttttttttt tttttttttt tttttttttt	2160
aaattaaagt ttaaaaattt atgagtgttattt tttttttttt tttttttttt tttttttttt	2220
tataaagtttt atgatataat aattttttata ttatccgtt tttttttttt tttttttttt	2280
attaattcgat agaaattttt tttttttttt tttttttttt tttttttttt tttttttttt	2340
tacgttttga aaaagggtta gaaaaagat tttttttttt tttttttttt gttttttttt	2400
aatttgcgtt agtattttttt aattttttttt tttttttttt tttttttttt tttttttttt	2460
attttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	2520
tgttgtaaat ttaaaaatgtt aaaggtagtc gtttgtgttt ttattttttt ttgggtgttt	2580
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	2640
gtttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	2700
ttgaggtat attttgggttt aacgaggtttt gttgtttttt tttttttttt tttttttttt	2760
tgttgtttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	2820
ttaagggtttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	2880

**-continued**


---

tttttattta agtttatttt tagttattgt gatgttatta ttagtattaa tggtattnag	2940
aaaaatagta attttttaa 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 attatnnnta tttttattaa tgnntttggg nttataggnn gngnntttt	60
gnttaaatag tagtaatgn aggnatttaa aaaggatat tattnntat ttattnatgt	120
gttatgntat tatttaattt ttattataat ttgtgnngnt tagtataatt tgcgnnttta	180
ttttattngt gaataaatgg aggtagnnnt ttgnntaatt tataagtaag tggtagattc	240
gggggttggt tttagataat gtgggttttag atttttgtt tttttttgtt ttttgggttga	300
ttttgattga gtttagttat ttattnntt aatataattt aattgttgtt tttatgtttt	360
agggtttgtt ttagatgtta gaaatataattt ttattnntt ggatttttgtt atggtttatt	420
gttttagtgtt ttgttatgtt acgtgttata taatgtattt attcggat ttatgtttag	480
gaaaataatg ttgaaagata ttgtatgtat gttttataaa taaaattatg taattnatgt	540
tttttttattt gttttggta ttaggaagtt ttattnntaa ttagagtaaa tatattagaa	600
tttgggtttt gttattnatgtt ttgttgaatt tttttttttt gtttagat ttattnnttgg	660
tttataaaattt ttattnatgtata aatgtttttt attgtaaaat attttaaattt ttttttaagg	720
gaagggtgtt tggaaatgtt atagtaaggt tttttttgtt tttttttttaga ttttttttag	780
ggaaggtaga ttttagatgtt tttttttattt ttagttttaa agttttttat ttataaaattt	840
ttttaagttag tgattttgtt tttttttttt gttttttttt gttttttttt atttttttttt	900
taaaattnat atataaaaag aagttttttt cgattnatgt ttttagttttt gttttttttt	960
tatatttgag gttttttttt gttttttttt tttttttttt gttttttttt gttttttttt	1020
ttttttatgtt gttttttttt tttttttttt gttttttttt gttttttttt gttttttttt	1080
atggtagggaa agtattaaat attttatgtt tttttttttt gttttttttt gttttttttt	1140
aaagaaaaaaa aaaattnatgtt gttttttttt gttttttttt gttttttttt gttttttttt	1200
gtttattnatgtt tttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	1260
attnatgtt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	1320
aagtggaaattt agtgggtttt gttttttttt gttttttttt gttttttttt gttttttttt	1380
gaagttttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	1440
tgtttatgtt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	1500
tttattttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	1560
ggttggcgag gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	1620

-continued

gtcgctgtat	tgaataaaa	cgtcgctggg	cgttttttagt	cggataaaat	gtagtcgaga	1680
atttcgttcg	ttttgtgcgt	ttttttgttt	taggttaggaa	agaggggttg	tcgggcgcgt	1740
tttgcgtttc	ttttttgtat	tcggatcgtt	cggtaacgggt	agggttaggg	ggttttcggg	1800
gggtcgggggt	tttcggtcgc	ggcggcgaag	atagatcggtt	gttcggtagg	gaggttattt	1860
cgagtttaga	gatttttagt	atttttata	tataggtttt	tattttggcg	tgcgctgcgt	1920
tgtgtgtgt	tgtgtgtgt	tgtgtgtgt	tgtgtgtgt	tgtgtacgtt	cgttaacggg	1980
agqa						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
```

-continued

tttttttag aaagaattta aggtgttta taataaagga tatttgtta atagaattta	1320
tgaattaaaa tagaaaattt gggtagaaag gaaaagtta gtaaaattgaa atgataagtt	1380
taatatttaa tgtatgtt ttgatggta taagagttt ttgatgtta aagtaataaa	1440
agaaatataa gttatatagt tttgttgtag aaatatgtat ataatgttt ttgtatttat	1500
tttttatggg ttaaaatatcg gaataaatgt attatgtat acgtaatata gtaggatatt	1560
gagtaatgga ttatgttaag atttagtggta tggggatgtat ttttagtat ttgttatagg	1620
gtttgaaata tggagatagt aatttagtgtg tgttaaagaa atagatggtt agattaatt	1680
aagatgtagt agaaggtagg gaggggtaag gagttggag ttatattgtt taggtatagt	1740
tttcggattt gtttatttt tgtaaattat ntaaannntt gtttttattt gtttatagat	1800
aaaatgaaan tataaatttg attgantnta taaggtgtg gtgaggattt agtcatgnta	1860
taattattgtt ataaatataa agtcatgtgt ttttaagat nttnattt gttgtgttt	1920
ggntagaagg ntntnnnttgg ntnttttaag gntattgata aagatgnnag tcatgnttga	1980
aqtt	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  
  
gtttttgggt agatatgtgt tttataagtt ttaatggaga aaaatgttaag tatttttattt 60  
tttgaatattt ggtttatttga gtaatgagaa aatagttatt ttttttagga tagtggtttt 120  
taattatggt tatgtgtttt tttagaaaaa tttaaaaat atatatatat taatgttttt 180  
gtgttatttt tagggatttt aagtttttga atacgaattt tgatttagta ttttttaattt 240  
attttagtga ttgtgtatgt aaattatgtat tgagttttat tgtttaaga tgaataaaat 300  
tttttttagt attgaaattt taaattttaa ttattaaaat taattaaggg tatggaaattt 360  
aataaggat agggaaagttt ttatattata aaattttttt tttaaatttat agtttattgt 420  
ttatatgtta ttgttattt tagaaaaggg tgaaaaaataa gttaattttaa ttatTTTtag 480  
ttgaaaaat tatttagaaa tgaagatgac gatTTTgaaa tattgttaat attatttgat 540  
ttataaataa tgtttaataa tatttattat atattgatag atatTTTTt atatgaataat 600  
tatataattaa aattaaggta ataatgtatt tagaatattt tatttattat tatgtttttt 660  
aagttagtta gaaattaaga tatgatgttta taagtatgag atgttaaggt gtggggtttag 720  
aaattatattt gtatTTTTt attaataattt aatataattt ttaatattat atatattttaa 780  
tttaatTTTt tataTTTTt attatTTTta attatgtgtta taaatataag tataatattat 840  
tttatgtattt tatttattta tattttattt tatttattta tataggggat ttttttaat 900  
ttatttattat taaattatattt tattttattt taatttttag aataagtttt ggaggttagt 960  
attgttattttt tttatattttt ataaatgagg aaattgttta tagttataaa gttattgtgt 1020  
tagatataattt agaagtttaa tatataattt gtgaatataat gtataaaaaat agagagatag 1080  
atatgtataa tagtttattttt ttatattgag taaaagtttt taatttggtt tagaaatttt 1140  
tttgtaaaaa ttgagtaaaa atcgaggat ttttttattt gttatattagg tataagggtt 1200

---

-continued

---

attttatttt ttaataaagg atgaatattt aaatgtggat tttaagggtt aattttagat	1260
ttttgaatt tttgatagtg ggatttgaa tttgtttatt gttttaaagt ttttaagga	1320
atttatatga ttaatttagt ttagaaatta ttggattta ttgtcgaagt ttgagaatta	1380
aagtttgggt tttatgcgg tttatagaa aggtaaatg aagtattatg gatagaattt	1440
atacgaaaa agtagtttt tttttttaga agttaatagg tagtaatata gtagaaatta	1500
gtgatttatg ttttgtttt tgaagttagg tagaatttta tagagttttt gtagtgttat	1560
tgacgagatt tgaaaaatgg ggtaagttgt ttgatgtttt taaagttata tttttttat	1620
ataaaatgag ataataatttt ttgtttata ggggtgttt aaagattaaa taaaataat	1680
atgttttatt ttatatggta taatgtttga tatttaagaa gtaaaggata tattttattt	1740
ttattgaagt aatttagaaag tatgaaatta tgaaggagat aagagtttg attggtagt	1800
tattttattt ttttaggttt atttattttt tttaaattat ttttgttga gaataatttt	1860
taagttttt atttaagtgt tgagtaattt tattttttat aatgatgtt tttttatgag	1920
aaaaaaaaat gtttttaagt ttttggaga aaatataattt gtattttttt tattgaaaaaa	1980
ttaataattt ggatttgtt ttttgttatt aatttttagag tttatgtt ataaataaaag	2040
tgttttagtt taagaagattt gaaagtaaat atggtagt attttaaaat aagaattttt	2100
taaatatatg gtatgattgtt gttatattat tagtaattat atgatacgtt atgtaaagta	2160
tagtttagt atttaaattt aattttataa agtaaatttga tttgttttg ttggggaaaa	2220
gttaaagtat taatttattt gttatgttag ttttgtttat tttttgttta tttagtgata	2280
agtttaaataa atgtatataat ttttattttt atatttagta atataatttt ttgtttatgt	2340
agtgtatgtt ttttgttattt tgggtgtt ttttagttt agaattttgtt ttttgtgtt	2400
attataatata taagtataga gtaagtgtaa taaaatttga gtattttttt tgaaaaggtt	2460
ttgttttaaa ttgtttataa atttaaaggaa tttttgttga agtaatcgta ttgtttaatt	2520
agttataattt agtaatttaat tttttggag tttaattttt tttttgtttaa aacgttttag	2580
gaagagtata tattttttaga aagtatgtt aaaaattttt tagtagaaaa tttaaaaata	2640
gtttttttttt gtttaaggtt tttttaaaat ttattttata tagttaaattt ttgaaattttt	2700
agttagttttt gttttattat tataattttt gtataaaat tttaaggat ttgtttttta	2760
gttttaagta tgattttattt ttataagttt gatttagttt tatatttagt ttgtttagga	2820
aatgtatgt tttttttttt ttgtttagta gttgttaat ttgttattt attttatgtt	2880
ttttttttttt tgaaagttt tagcgaaaga aatttttaat tttttgtttt gtaatatttt	2940
ttgttagttt tttaatgg gtattttgtt tttttaaaga atttagttgtt ttgtttaga	3000
agtcgatttt ttgtatgttt taacgtttgg tttaattgtt ttgttttaat ggagtttcg	3060
tcggtgagga gcgagatgtt atcgatttga atgttggat ttgttgttta attgttagga	3120
gtgagagata ttgagattt gaaatttttgg gaggtggag gggagaggaa tagttcg	3180
cggaggcggaa gatgtaaat aaaggatgtt aattttatata gaaaaaaa aaagatttcg	3240
ttgaggtatt gaggtgtgtt acgatttat ttttaaagg agaagttaaa aagtaaggaa	3300
ttggggaggag ttggagggtt aaagttttaa aaggattat tcgggtataa ttgtttttt	3360
ttgttgggtt tgtaaaggat agatgttcc gtttttaaag tatatgtat tttttttaa	3420
gtgattggaa atggatattt attgtttgtt aaatgttattt aaatgtttt tttaaatttt	3480

-continued

gggatataaga aagaggggta taaaaggaga atttaaatag aaaaaggggag gattcggagg 3540  
tttttgaaaag cggggggaga agaaggagga gggataatag agaggaatag agaaggagag 3600  
cgagagaag ataaaataaa ataaaaatag gaattattga ataattatat attaaaaaga 3660  
aagtttttt ttatgggta tttaaaatat tgagattgt aatgtgattt cggttatgga 3720  
agaaagatgt tttttttat tttgttttc gaaagtttt gtttcgtta ttggcgatta 3780  
aaattttatt aggttaaaga gtgtgtttaa ttgtttgaag aatgtatgt acggaaggcg 3840  
ggtttcgtta tgtcgtttgt tttttcggt ggagagaatg aaagaaacgc gttaggttag 3900  
agattttgt cgagttagat ttttttcgt cgttttaggt tatcggttat tcgtaaaga 3960  
ttcagactaag gaacgttaggg ttattgttg ggttaataaa tggagttcgt ttttttttt 4020  
tcggacgtcg ttgttcggtc gatgtttcg gtaatttatt cgccgcgtat gttagaggat 4080  
ttttttttt ttttagatt atttgttgc attaattgtt ttttttaat atattgtatc 4140  
gtatTTTTTA ggtggatata ttaatagggtt acgggttggg gaggagcggg ttagaggag 4200  
agggatttaa atttgcgaac gtttgggtt ggtcggagtt gcgggggggtt tgggaggaga 4260  
gagggggagaa gagagaagga aggagagcgt ttgtcgggat ggtttaggtt ttcggcggag 4320  
tagttttggg gttgtacgtt ttgtgggag atgttgggt ttttttaggt tcggtaagag 4380  
cggttttaat attatcgttt ttttttttt tttttagataa ttttagaga aacgtttttg 4440  
ttttttcgt cgcgatattt ttagttgtt tttttttata gtttaggcgg cgcttttcg 4500  
tacgttggag cgtcggcgt tagaggacg tttttttcg cgtcgattcg ttttttttg 4560  
ttttgttgg ttttttttt ttagattttgc gttttttata ttttagttc ggagagacgt 4620  
tattttgtcg cggttcgtat tcgcggttcg gggttacgcg cggaaagaggg gcgttagttc 4680  
ggatTTTGTG ttccgttaggg ggcgtttgg agcggagagt gaggcgaatg gtatatgagt 4740  
gtgcgggttag tttttttgtt agttcgagtt ttttattgtt gttatTTTTG gtttagttt 4800  
attcgggtta gcgttggcg agcgatTTTGC tttgtggtt tcgcggcgt ttttttttg 4860  
tattttgttca tttattcggtt gattttttt tttcgggatt tttttagttt ttttattttt 4920  
agagttcgat ttttttttt tacgtgattt cggcgggtt gaggattgt tttttttttaa 4980  
acgttagagg gatgcggcg cgtagatgtcg agaggcgggtt gtcgggttc gggcggtttt 5040  
gattttttttt ttatTTTTGGGTTT tttcgggtt ttatcggtt gttttttggat tttttagttt 5100  
ttttttttttt cgggtttttttaa gagttttttt tttatggtag tagtttttcg cgtttttcg 5160  
gtatTTTTGGGTTT agcggacgtat ttttcgtt cgggtttggat ttttagttt ggtatgtt 5220  
gaaatTTTTGGGTTT agattatgcg cgggtttggat tttttagttt tttttagttt 5280  
cgtcgtcgat tttgttgcgt tcgttcgcgg gatgttttagt agttcggttgc tttttagttt 5340  
cgatTTTTGGGTTT ttttccggaa gtcgtttggat tttttaggtt gttttaggtt gttttaggtt 5400  
ttgtggagtt tttcgggttgcgt tttttaggtt gttttaggtt gttttaggtt 5460  
ttgtggagtt tttcgggttgcgt tttttaggtt gttttaggtt gttttaggtt 5520  
atTTTTGGGTTT tttttaggtt gttttaggtt gttttaggtt gttttaggtt 5580  
gtgattgttca aacgtttatc gtttggaaat tttttaggtt gttttaggtt 5640  
ttcgttttcg atTTTTGGGTTT tttttaggtt gttttaggtt gttttaggtt 5700  
tttagegcgt tttttaggtt gttttaggtt gttttaggtt gttttaggtt 5760

---

-continued

---

gggtagaaaag ttaggtgtta gtaggaaatt tttgtatgtt tgcgtttata ttggagttgc	5820
gaggattttgg agaaatatta aacgggatgg tttttgggt ttattgtttt gaaagagtat	5880
taatttttagg ggaatatttg aaatagaagt tttgttattta ttaaagaaaa aagttttatt	5940
aggatgagga agaaataattt ttatgagaaa gaatgagcga gaaagtaata attaaatgg	6000
tgattttagg ggaatcggtt attttggta aagggtttt gaggtcgat tggttttcg	6060
ttgaagatta ggttatatacg attttagagg agttgggtt taatagaattt tttttttttt	6120
tttttttttt tttttttttt tttttttttt tttttttattt tttttttttt tttttttttt	6180
ttatTTTTTTT tttttttttt cggtaaaaga tattggttt gtatTTTtta tatgttttttt	6240
ttttttttttt ttaagttttt aaggtgtat aggggagttg agaaaaagaa tattttcggt	6300
gttttttagg tcggagtggtt tatgattgag gttgggttagg ttttatgttag gcgagtcgag	6360
ggcggaatcg attttagtgg gcgttgattt tttttttttt ggataggttt ttgtggagtg	6420
gtttaggtat tttttttttt cgttcgggtt ttttttagatt ttgacggcga acgtttttttt	6480
gttttcgttt ttgttaagtt tttaattaa atagggttagg aggtggagttt ttgttttatt	6540
tttagtttgtt atagtattcg gtttgatagt ttgttagata gggtgtatgtt aatttttttt	6600
tttttgtgaa tataattttt ttgttagttaa atttggttttt gaataaaatgtt ttttttaaag	6660
atgtatataaa gttgaagtgtt atgtatattt agagaggaggaa gatgtttttttt ttgttaattt	6720
gggtgaaatgtt ttgtatagttt ttttagttttt attgtatgtt atgttaaaag gaaaattttt	6780
atgtattttt ttaattttttt ttttataaag ataaatgtttagt atatgttattt ttattttttt	6840
tgggttaataa gattttttttt tttttttttt gttttttttt ttgtttttttt aataaaattttt	6900
aatatgtttttt tataatttttt tgatgttgc gtatgttgtt gttttttttt aaagaataaag	6960
ttttttttttt ttattttttt attttttttt agatgttttta ttatgttaat aattatgtt	7020
ttgtaaaaat aattttttttt gatgttgcgtt gttttttttt ttaatataaga ttgtttttttt	7080
ttttttttttt tttttttttt gttttttttt gttttttttt ttaatataaga ttgtttttttt	7140
gtatgtttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7200
ataaaatgtt tttttttttt gttttttttt tttttttttt tttttttttt tttttttttt	7260
tttagaggtgtt attatattttt attttagttt aatagattat tatagaaaaat attgtttttt	7320
tttattttttt ttattttttt tttttttttt tttttttttt tttttttttt tttttttttt	7380
tgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7440
gagattttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7500
taatgtttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7560
tattgtttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7620
gaattttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7680
tgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7740
ggataatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7800
ttaatgtttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7833

&lt;210&gt; SEQ ID NO 15

&lt;211&gt; LENGTH: 7833

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial Sequence

-continued

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

<400> SEQUENCE: 15

-continued

---

-continued

---

aacgggatta	tttatTTTTt	gtagatatta	gtagaaaaat	aaattgtatt	cgagtaattt	4500
tTTTAAGTAT	ttaatTTTTt	aTTTTTTTt	tATTTTTTtG	tTTTTtaAtt	tTTTTTTtGA	4560
gagatgtgat	cgtgttagtat	tttagtgTTT	taacgaaAtt	tTTTTTTTt	tTTTGTGTGA	4620
aatttatTTTt	tttattttat	atTTTCGTTt	tcgttCgaga	ttgttttttT	tTTTTTTtAT	4680
ttttaaagat	tTTTGAATTt	tagtgTTTTt	tATTTTGtG	aATTAAGTAG	tagATTTAG	4740
tATTTAGTC	GGTGGTATTt	cgttttttat	CGACGAAGAT	tTTTAAAAA	tagATAATT	4800
agattagacg	ttggaggatAT	tagaaaATCG	gtttttAGAT	AGAGTAGTtA	aATTTTAA	4860
gaaatAGAA	tatttattAG	atAGAGTTG	taATTAATAT	tGTAAATAAA	ggaattAGAA	4920
atTTTTTCG	ttatAGGTTT	ttagtagAGA	AGGTAAATA	AATATAGATT	aAGATTAAAT	4980
aATTTATAG	tagAGAATGA	GAATATGTTA	tTTTTATAG	taAGGTGtG	GTGGTAATT	5040
attaggttA	tGAAATAAG	ttatGTTGA	aATTAAGGT	AAAGTTTTA	AAAGTGTtA	5100
tGTGTAATT	ATGATAATGA	aatAGGATT	GTTAGGATT	TAGAGTTGG	ttatGTAAGT	5160
AGAATTTAG	AGAATTTTT	AGTAGAGGAA	AATTGTTTT	GAATTTTG	tTAAGTAAAT	5220
tTTTGTATA	tTTTTAATA	ATATATGTT	tTTTAAAGAC	GTTTGTAA	AAGTAAGT	5280
AAATTTAAA	GGAGTTAATT	ATTGGTTGA	ATTGGTTAAT	AAATGCGGTT	GTTTTATAG	5340
AGGTTTTA	AATTATTTAA	TAGTTGAAG	TAAGTTTT	TAAATGGAA	TGTTGTAA	5400
ttgttgtatt	tATTTGTAT	ttagtGTTAT	AGTGTtATT	AGAAATAAA	TTGAAATTG	5460
gtAAGTATT	ttaAGTGGTA	GAAGAATATT	ATTTATTGAG	TAGAGAATTG	TATTATTGAA	5520
TATGTAATA	AAAATATATA	TATTATTTAG	ATTTGTTATT	AGGTATTTAA	GAAGTAGATA	5580
AGATTGTT	AGTAATTGGA	TTAGTGTtT	AATTTTTTT	TAGTAAGGTA	AAATTAGTT	5640
ATTTATTAGA	ATTAATTTA	AGTTTATGAA	TTGTATTtG	TATTGCGTAT	TATATGATTG	5700
Ttagtaatat	GATAATAATT	TATTATGTAT	TTGTAAATT	TTTATTTAA	AATATTATAT	5760
TATATTTATT	TTTAATTTT	TTGAGTTAGA	ATATTTATT	TGTGGTATAT	ATATTTAGA	5820
ATTGATGTAG	AGGAGTAGAG	TTAGTGTtG	AGATTTTTA	GTAGAAATAG	TGTAGATATA	5880
TTTTTTTAG	AAAATTTAAG	AATTTTTT	TTTTTTATGG	AAAGAAATT	ATTATAAAGT	5940
GTGAGATTAT	TTATAGTTA	AGTAGGGGGT	TTGGGAGTA	TTTTTAATA	AGAATAGTT	6000
AAGATAAAATA	AATGAATTG	GGAAAATAAG	ATATATTGTT	AATTAGAATT	TTTATTTTT	6060
TTATGTTTT	ATATTTTTG	ATTGTTTA	AAAGGTAAG	ATGTATTTT	TGTTTTTAG	6120
GTGTTAGGTA	TTGTGTATG	TaggataGAA	TATGTTATT	TTATTAATT	TTAAAGAATAT	6180
TTTTATGAGA	TAAGAAATAT	TATTTTATT	TATATAAAAG	GAATATGTT	TTGAAAGTAT	6240
TaggtaATT	TTTTAAGAA	ATAAATTCG	TTAGTGTAT	TGTGGGATT	TTGTGAAATT	6300
TTGTTGATT	TTAGAGTATA	AGATATAAGT	TATTAATT	TGTGTATTG	TTGTTGTtA	6360
TTTTTGAGA	GGGGAAATTA	ATTGGGAACG	TATTAGTTT	GTTTATGATA	TTTTATTGt	6420
TTTTTTGTG	GAgtcgtAGT	AAGGTTAAA	TTTTAATT	TAATTCG	TAATAAGATT	6480
TAGTGATT	TGAATTTGGT	TGATTATATG	AATTTTGAA	GAATTTGA	AATAATAGAT	6540
AAATTTAAG	TTTTTATT	AGGGATTAG	AAAATTGGA	GTGTTTTT	GGGATTATA	6600
TTTTAATATT	TATTTTGTT	GGAGAAGTAA	GGTATTAT	ATATTTAT	GATAAATGAA	6660
AGGATATT	CATTTTGTT	TagTTTTAT	AGAGAGGTT	TTGAGATAGG	TAAAGTTT	6720

---

-continued

---

ttathtagt taaagatgag ttgttgtata tgttgtttt ttgtttta tgtatatgtt	6780
tattaaatat gtattaagtt ttaaatatgt ttgatatagt aattttgta ttgtagataa	6840
ttttttatt tgtaaaatgt gagtaataat aatattgtt ttttgggtt gtttaaaga	6900
ttaaaataaa aatgtatggt ttaatggtag tggatttggg gggattttt atataaataa	6960
gtgaatggag gtatgaataa ataaatataat aaagatgtgt gtatttatat ttatataat	7020
aattaaaaat agttaaagat gtataaatta aagttaaatg tatgtgatat tgaagtat	7080
gttgattatt gataatgaag tatagtataa ttttaattt tatattttaa tattttat	7140
ttaataaattt atattttaat ttttagtta ttaagatat atagatatag atagaatgtt	7200
ttaaatgtat tattgtttt gtttaatgt ataatatttta tatgaaaaag tatttttag	7260
tgtgttagtaa atgttattaga atattattta taggttaaat gatattgata atgttttaga	7320
gtcgttattt ttattttgg ataattttt aaattgagag taattaaatt tgttttttt	7380
ttatttttt ttataatgtt aaataatata taaataatga gttgtgat ttaaaataa	7440
ttttataatg taaaagttt tttatgtttt attgattttt atgttttaa ttaattttgg	7500
tagtttaagt ttgtgatttt agtggtaggg aaagttttt ttattttaga gtagtgggt	7560
ttagttatga ttttatatta taatttattt gataattaaa gaatattgat gtagagttcg	7620
tattnaaaga tttgaaattt ttagaaagtga tatagaagta ttggatata tataattttt	7680
aagttttttt ggagaaatat atagttatga ttgagaattt ttgttttgg gaaagtgatt	7740
attttttat tattnaaataa gttaagttt aggaggtaaa atatttatat tttttttat	7800
taaaatttgt aaaatataa 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 gggatttttta gaagtttagt gttttgttg ttttttttga gtatggaggt	60
tattaatttt taggggaaag agatgttagt tgaggttaggg gtgttgtt aagaaatttt	120
gatgttttg gggattgagg ataaagggtt ggatatgatt ttgggttatt tggagtttt	180
tgatttgtt tatggatgtt atattaggg gttatattt gttttttttt aaagaatttt	240
aagtttagt tttttttttt gtttatagtt ttggatgtt gttgttgtt ttattgtata	300
ggtatgtttt ggattgggtt tagtagatt tggatgtt gttttttttt gggatgttgg	360
agatgttgaa aagttttttt tttgttattt tggatgtt gggatgttgg tggatgttgg	420
tttttttttt ttgagttgaa tggatgtt gttttttttt gaaattttgtt tttttttttt	480
tgttgagttt aatgttttagg tatagtggaa ttgaaattt gttttgttgg atgtgagagt	540
tgttgaggtt atgtgttattt gggatgtt gggatgtt gttttgttgg tggatgttgg	600
tgtatgttgg gatatgtttt ggtatgtt ggtatgtt gttttgttgg tggatgttgg	660
tttagtgtttt ttgatgtt gttttttttt gttttttttt gttttttttt gttttttttt	720
gtttttttttt ttgatgtt gttttttttt gttttttttt gttttttttt gttttttttt	780
gtttttttttt ttgatgtt gttttttttt gttttttttt gttttttttt gttttttttt	840

-continued

<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

gttttgtata	gttgggggtt	tgtttttgtt	tgttgggtt	taagatgttt	tagagatatt	60
ttatttgggt	tttgatatttt	tatTTTGTa	tttaggtttt	ttattaaaga	gtaggtaggt	120
gtgatagtgt	tgtgtgggtt	ttgttagat	taagtttata	tttgggggtt	gttatattta	180
tttatata	tttaaattta	ttattgtttt	ttaatatgtg	tttatttga	attatagata	240
ttaaaaaata	gatattgggtt	agaagagtt	tttgagggtt	ttatTTAAAT	gtgttttaag	300
tttgggggtt	agttttttgg	gtgggtgtat	ttatata	gtatgtat	gattttttgtt	360
tatTTAAATT	gtatata	tgaatagggt	ttttttata	tatTTAATT	tgtgtgtattt	420
taatagtttt	tatTTTTGt	agaattgggtt	tttagttta	ttgtgtttga	gtgttttagtt	480
tagtagaggt	agggaaattgg	gttttagttt	tattgtgtt	gagtgttttag	tttagtagag	540
gtagggattta	agggttttgtt	tgtttatagt	gtttagagtg	gtaaaaaga	agtttttttag	600
tgtttttattt	atTTTTGGT	ggaatgtgt	gtgtgtgatt	tgtttagttt	ggTTTGGGTT	660
ttgtttgtgt	ggtgagtgta	gtagtggat	tttgggggtt	tgggtggagg	taaggatttt	720
agttttaggtt	ttttttaggtt	agagtagaaa	ttagttttta	ggtgtgtt	ttgtgtgtgt	780
agttatggaa	tttaggtat	tttgggggtt	tgtttgtatt	tttgggttttta	gttttttagaa	840
gtgttgaat	tttttaggtat	gatatttttgg	tttgtgttta	tatTTTTTTT	tttttaggggtt	900
tgggtggtttt	tgtgtttaaa	ggaggttagtg	ggaatgtttta	atTTTTAAGA	tgttttttaat	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

tgttttttaa aatgatgtt agtgtaaatta gttaatggag ataaaaataat ttatataaa	60
tttttttta agtttattag ttatTTTATA gattagttat aatttgggtt atttaatga	120
gttaattttt ttttagagat attataaatt ttttaaagat ggaaaaataat tgTTTttttt	180
ttaaataatg tttattttt aaaataatgt atataaaagt tgTTTAAATA atTTTgttg	240
aattgagaga aaatataaaa aattgaaagt agtttgggtg taataagtgt ttgttaagtt	300
ttatagattt tttttttttt ttggaattat tgtaataat atttgatttt tgattttag	360
gattaaaatg tgatTTTTTA tatttaataa ggagtaagta tttagtttgt aattttttt	420
gttaatgaa taaaatgga atgtatgttt gtgttttta tatgttgggtt gatattttag	480

---

-continued

---

ttgaaatattt ttttataat gtagagggtt ttttgtttt tggtttgtgt ttttattttt	540
tttattatag agatattaat gatgttatgt gtttttttgtt attattatgtg tagtagttt	600
tattatgtt tttgtttgg ttttattgtt gtttaattaa ataaatggaa atgggtatgtt	660
attgaaaata aatagaattt gtaaaagttt tgttttttaa gaaattttttt gtattgtattt	720
ttgtttgtt ttattgtttaa taagattttt attgtttttt tttgtttttaga aatttagattt	780
ttagtttgtt tttgttttag ttagataat gttttataaa aatgatttat attttttttaga	840
ttgtggggag agaagaaaaa gatttgaga aagtgattaa ggagggagta agagtgatatt	900
aattttgtat ttaattttatgt tggattgtgt tttttatgtt ttttattttt tttttataaa	960
aaagatagtg ttttgatgt tttgttttag gtagttttt agggtgagtg tagatgttt	1020
gagtttttg tttgtttga ggattgggtt agagagttt tttttttattt aatgaatttta	1080
tttttttttgg ttagtttggtt tttttttattt attttatgtt ggaatttgggtt tttggttgtt	1140
atagtttgtt tttgtatgtt atgggtttttt agtttaattt tagagaaattt ttttttaattt	1200
tgaataggat atttataatgta taggagaaaa aaaaattttat taaaataat attttattgt	1260
aatattttga tttgggggtt ggagagtaaa aatttaatattt ttgatgttttggaaaaattt	1320
aatgttttat gtttttttaa ggaattttgtt ttgtgtttaa ttttatttagag taattttat	1380
attttgggtt aaagtataat gagattgtta ttattttttt gttttttgtt agattttttt	1440
ttaagaata aaaaagaaggt aataaaagatt ttatattttt gtttttttaa gaaagttaaa	1500
gagaattttaa agtgttggat gaaatgtttt tttttttttt gtagtgtttt attttaaaaaa	1560
aaaattatga tataattttta gaggataggt tttttttttt tagggttttt gaaagttaaa	1620
agtatattttt taattttgtat tattttgttta ttatattttttagt aggataattt gtttagttgt	1680
gtttaggattt agtagatgttta aggagttgaa attaagttt ttttgaattt gaatggattt	1740
gaaatttgtt tattttttaga gtagtaatag atgatgtttt gtttaagaat atttttttg	1800
attaataaaaa aaggagaagt ttgaaatgttta attttaaaga atttttagaag gaggtttttaga	1860
tgaagataat aaaattttttt ggagatataat ttgttattta tttttttttt ttttaatttta	1920
ttggtttggtt gttttttgagt ttatgttaat ttttaaagat aagattaaag taaggaatgtt	1980
aggatataatg tttttgggtt atgggtttttt ttagatataat gtagtggaa tgaagatttt	2040
gttttattta tattttggat tttggagttt ttataaagaa tattttttttt attaaggatgt	2100
gattttaaagg gggattttata gaataatattt ttaattttttt tttgattttt attttgtt	2160
aatggataaa taaatataat tttagttataa aaattttgaaa aatataatattt tttttttttt	2220
aaaatttttg atagttttgtt taggtatttttgg agaatttagaa ataaatgtttt tttttttttt	2280
tttttttttga tgaaggaaag ttataaaaaa tagttatataa gaaataatataa tataattgtt	2340
gattttattttt taaatataatgtt ttgttattttt tttttttttt tttttttttt tttttttttt	2400
tagatgttta gaggggaaag tttgttttagt agatagaaag aagatatttttta tgggttataa	2460
tttttttttta gttatgtttagt agtttgaaga agtattttat ttttttttttta ataattttat	2520
aagttaaaga aagaaaaaaga taggttagttt aaagttttttt aatattgtgtt atattttat	2580
tttttaagttt agtttggaga aagtgtttttt gttttgtttaga agtattttgtt taagtgtt	2640
gaagtttttattt ggtataaaaaa gaaataaaaaa attttggagg attagagaat taaattttgtt	2700
aatataaaatg aaaatattttt agaatgtaaa aagtattttaa ttttattttt ttaattttatgtt	2760

---

-continued

---

tttattataat atagatggta ttatTTTatag tgaatgttaa ttatTAATat ttttttaaaa	2820
atgggttatta atttGTTTatt agttGTTTgt ttttgattta gTTTTTTTt gttgaaaggt	2880
tttatttaat tgtattataa aggataaaaa gggtAAAAGA taataagttt tttagattgg	2940
tagatttGTTt atttataaaaa tggTTTTGGG taggttattt gaTTTTTG aaatttaatt	3000
agatttattt aaatttaaga atgagaatat aaatttatAT tttgaagtgt ttatAGAAA	3060
gTTTTTTTt aatGTTTGAa gTATATTTTt TAATGAATAT tTATTTTATT TTATTTTTT	3120
ttatTTTGA attaAGTAAT tttGAATTa aAGTTGTTat GATTGATT gAAAGATTa	3180
ttggattattt aattGtGTGA tttGGGATA gtaATTTTTt GTATTTAGT ttGTTTATAT	3240
gttataATG aAGGTTGAAG tttGATTTG tttGtGATT ATTATTTAA ATATTGATG	3300
aaatttaattt ttagtGTTG gaATGGTAGt ATAATAAATT TATTAAGAAt AAATAATTa	3360
ttgtAAAAATt ATATTGATTt ttaAAATGATG TAATTGATAG TTATATTATTt gtagAGGGTT	3420
gataAAAtAAT AAAAGAAATG AAAGATGTAT ATGGTGAAGA TTGAAATTAT TTGATAAGT	3480
tttttatttG ttttattttt AAAATTAAATG ATTATGTTG ATGTTATAA ATTATAAAAT	3540
atAAAAGAAA ttttATAATt GTGTATGTAT AGGAGTTAA gTTTAAAGAA GTTTAAAGT	3600
ataAGTTAA AttaAAAttAA ttaAAAGAAGT TGAGAGGAA AATTGGTTT ttttTAAT	3660
tattattGTTt tTGAGGTTTt ATGTTTAATA TAATTTTTA AGTAGAGGTT ttagAGAGAA	3720
gAGTTGTGAG gATTTTTA tatttGtGTA GAAGGAAAAG tttGTTATTt ATTtTAGtAT	3780
ttttAGTGTt ATATTGATGT GTATTTGGA ttttATTTGT ttttATtGTA TAAATTATA	3840
tttGATTtTA aAGAAAAGGA AAATTAAAG ttttttttT ttaAGGGGAT AGAAATTtTT	3900
tgtGTTAAtt GTTGTATTt tttttttGTA AGGTTTATTt GGAAATTtTTt TGTAATAAA	3960
tgtAGGGGAT ttttttatGT GTTGATGTTG ttttATAAGT GGGGTGGGTT TGATTGAGA	4020
aaaaAAATTG tATATATGTA tGAAAGATTa TGTTTTTtTTtGGAAAGT ATGAAAGGTG	4080
attGATATTt ttaAGAAAGTT tttGTTATTt AGGAAAATTa ttaAAATTTt TATTTAGAGA	4140
tatttGAAA GATTGAAGGA AAGGAAGAAT GAAGAAAGTA GAATTAGAT TTATGTGGGG	4200
AGAGATTGTt GGTAGAGGAA AAGTATTtTTt tttGAATTtG ATAAGGGATT TGTTGGGG	4260
AATTtTTGTt ttagTTTTt ATTATTAGGG ttttttGAA tttttttttt tattGGGTAG	4320
ttttttGGGA GTGTAGTGGG GAATTtTTt ATTtTTTTt TAGTTTTG AAGGATTtTG	4380
ttttttTAGt GTTTTTTA GTTGTGTTAGG AGTTTTGAGT TTGATATTtTTt ttttGATGG	4440
GATAGGTAAg ttttGtGGGT GTGTAATAAT GTGTAATAA AGTTTTGTt TGATTtATA	4500
gttttGtGTG ttttTGAGAA GAAGTGTATTG TATTAATTG ttttATGTTG ttttGTTTT	4560
taAGAGTTG GGGGTTTTt ttttTAATTt TAGAATTAGT TGATGGGG GTGGGAAGT	4620
GGGGGTGGGG AAGGAGTGGG AGGGTAGTGG ttttGtGAG tagAGTGTAT ttttGAGtG	4680
AGTTTTGAA tGGGGAGTGT ttttGTTTTt AAGTTGATTG GTATTTTTG TTAGGAAGAA	4740
ATGTTAAGAG GTGGGAGTGT ttttGGGGAGG AGGTAGGTGG ttttATGTtG AGGTGTGGGG	4800
AGTTGTTTTt ttttGTTTTt GTTGTGTTT TAAGTTGGA tttttagGAG tGGTGTGAGT	4860
TGTGGAGTGT ttttGAGTT GTGTAATGAA ttttTTTTt ttttTTTTt ttttTTTTG	4920
TTGAGTTTTt ttttGGTTt tGATGGTATA GTGATAATAAT GATGATGGGT GTTATAATTt	4980
GTATTGAAAT tttGAGGTGA GTTGTGTTG A GTTTTTGG GGAAGAATTt TAGGTGTGTG	5040

-continued

**-continued**


---

atttatgatt atttgaattt aatggtaggg ttttgttta ggatattgt aagaagggt	7380
ttaattta aaattnaat tttttaattt ttaggggtgg tggtggatgg gaaagtggag	7440
agaagggtgg tagtgggagg aaaaagaaa gggaaaggaag ggaagtggaa ggaaagagga	7500
ttgtggaggg gaggaaaagg agaaaggat ggagatgggg tagatggat tttaatagtt	7560
aggaatggat tggaaagggtt tggtttgggg tttgggttt aatttgaga ttgttttag	7620
gatattttt ttttttttt agtggtaata aatgtgattt tagaggtgt tggtggatgg	7680
agggtggaa agaatttagat aagtggaggt ggtttagtt tgtaaagg ggttgggtt	7740
ggggtttgg tttatgtt tattttgtgt taagggagga ttgttaatgg gtgtttttg	7800
tagtggttgg gttagtttagt tggtatgtgg gttgggttt tggttttag ttgagatgg	7860
tgtattttt tgggtttaga gattgttaga tagttgtata ggtgggttgt gtagttgagg	7920
tttttattta ttgaggtaa tgggggtggg gtggggaaagg tagggtttag tttaatttag	7980
agagttgagt gggtttgtt tggggatttg gggttgggtg tttttggagt tgatttgagt	8040
tttattttaga tagtggtaa tatttttattt aaaaggttat ttaagttta ttttatttta	8100
tttttattttt gtttgggttt tgggagtggtt aggagggttta ttgtgtttt gtttggaggt	8160
aggggagagg taaaattta tttaaagtaa attttttgtt agaaaaatgt gattttatat	8220
tgtattttat ttgtgttattt tttaatata tttaggttat tttaagttt atttttagtt	8280
aattaataga ttaattttgt tttagattgt tataaagagt ttgtttggaa tgtaaatat	8340
tgtattttga taagattaat agaattttat ggggttaaat atggaaagggt tggagtgta	8400
attgagggtt ttgtgtttta atgaggaatt ttagtggggta ttgtgtgggg attggaagtt	8460
tagtattttat tttagggagtg attgagttgg aaattttggag aagttttgggt tagatgagat	8520
gggtttttt agagttgtat tttaagttt tttatgtgt tgataatgg aagaattggaa	8580
ttttaaagat tttttttattt atttttagatt ttgttagtaa tagttatgtt tttttttttt	8640
tgtttgttaa aagttgtat ttaattttttt taagtatttt ttgataattt tgtaaattt	8700
attttgaaaa tattttata aagttttaaa ataggatgtt gtgagtgtaa tataaaaata	8760
ggaaggattt aatggaaagta ataatttagt ttttgagata gaagtgggtt ttagttgt	8820
attttgaaaat gtaagagta gtatattagg gagaagggtt ttaagagttt gatttttgt	8880
tttgagtttt gagtaaagaa agaaatatgt tagaattgtt ttgtttttt gtagtaatat	8940
tgttaagggtt ttaatggat tttaatgtgtt tttagtaattt atttttttaa gtattaagtt	9000
tatagttttt aatttttttg tgagaagattt aggttaggaag tgggtgaga tttaagataa	9060
ttatatttaga ttatgttagt aattttaaat ttgtgttttt ttgttagataaaaatagat	9120
tggttttaaa ttgtatTTTT ttgttttaaa aaaatgaggt ttgaaaggaa tttttttaga	9180
tttatatttagt ttatagatgt taagttttt tttttttttt gtgtttgtttaaatttgggg	9240
gatgtttttt atatttttaga atattttgtat ttatattttt gtaaattttt ttaggaaagg	9300
atttaaatag ttaaattttaaat ggtgattatg attaatgtta aagatattat ttattttttt	9360
ttttttgtt gtattatttag gtttaaatat ttggagttttaa atgattgggtt atttaattgt	9420
gtaatagat ttttgaggtt gtaaagatata attaattttt gtttttaaag ttgtgttatt	9480
aagttttttt tttttgggtga gatttaataa tttttgtta tggttattttt ggtataat	9540
tggtaaataat tttttttttttaaa atggatattt gttttgtat ggagttagat tttaaagat	9600

---

-continued

---

atatggttt tttgaaagaa agaattaatg attgttagaa ggtatTTTT tattatttat 9660  
gtagagaaat tagttaggat tgataaatat ttattgttgg gagttatTTT ttttggTTT 9720  
agaaaatATG ttgattgtta tataGGTTtG tattttatTA gaaaatATAA agtATgtATT 9780  
tttttGATT tggAAattAGA aaaggGGGAt tagatATGGA tGTTTGGTT aagtTTTATT 9840  
ttaattAGtG aattttGTT agTTTTTTT tGtAGTTAG atTTGGTAA ttatATAAT 9900  
ataaaAGTTG tGtGTGtGTG tGGGGTGTGT GtGTGTGTG atGTatATAg GGTTTTAAA 9960  
tattgttaAG ttttattAT tGAatGTGTT agAGGTGGTT GTatTTAGA GGTTAGTATA 10020  
agaattttt GTTTAATGTG ttagTTTGT ttttGAAAT atTTTTGG GTAGATTAT 10080  
taattataAA aggatataAA tGTatATAtt ttttataTTT GTatATAttT tattttATT 10140  
tttaaggat ttaaAGAGGT GTTTATGTTT ataaaaATGAG attatATAttT ttataATAAA 10200  
tatatttgat aaattataAT ttttttattA ttaaatttattT gatttgatAT tatttgaggT 10260  
atagtGGAAT gaggattAGA tttGGAATTa gaaAGTTGA GTTGGATTa TTAGTTGAGA 10320  
aaattaAGTA aagatattAT tttGGATTt ttagtGTTT aattttatTA atGAatATGT 10380  
agttaataAG ataaaaAGAG taagtATTTT gaaaaattAT ataaaaAAAG gatGATAATA 10440  
tattatGTT tGGGATTtTt atGATTtAA ttttataAtt ttaataATAA ttGTattTT 10500  
taagttatGT agTTTTTATT gaatttataA aataaattAG agAGGGGTt ATTGTTTAA 10560  
aaattGTTtA GTTTATTAA ttagAAGATA gattatAGA attGATTtA atGAAGTTGG 10620  
atatttagtG gaaAGGAAGA attatttAAT aaatGGTTT aggATAATTG ATGGTTTAT 10680  
gtaaaataAG aattAGTTT tatttGATTt tttatTTAA aagaaaATATA agatGAATTa 10740  
aatatGAAAT gtAGGGAATA attTTTATAA aatGATAAA agaaaATTAT gggattAAAA 10800  
aaaataAGTT aggaatAGAG aaattGTTT taagtaAAAT ataaaATTA GTGTTTTT 10860  
aggaaATATT gatAGGTTG AttaAGTAA aattatATAA aagtaAAATA tataTTGGA 10920  
aaatatttAT gatatttGAG agataAAAG ttaataATTt ttGATATTt AtGGATTtTT 10980  
atGAATAATT ttGAAAGTT tatatttATTt AAAATATGAG taaaAATTAT GTATAGGTG 11040  
tGGTAGAAAG agatATAGAG ttttaAAATTt atGAAAAAA ttGTTAGTT ttatttGTA 11100  
ttaaataAAAT agtaatTTAA AttaATGTA tATGTTATt tatttttTA ttaattAGAG 11160  
gagtaataAA aaaaATAAT ttttttttt ttttttttt ttGAGATGGA GTTGTtTT 11220  
taggttggAG tGtGTTGGA tGATTTAGT ttattGTAAT ttttGTTTT tagGTTAAG 11280  
taattttttt GTGTTAGTT ttttagtGAGT tGGGATTATG GGTATTtTTtTtTtTtT 11340  
gttaattttt GTGTTTTAG tagAGTGGGG ttttattATG ttGTTATGG tagGTGTTT 11400  
tgaatttttG attttaAGtG AttttatttAt ttttagTTT taaAGtGTTG ggattatAGG 11460  
tatGAGTTt tGtGTTGGT AAAAATGAA taaaaAGTT tGATGTAAGT ttGtGTTGGT 11520  
taggttGtGG agaaATTGGG tttttatATt AtGTTGtAT gaaAGTAAT AGGTTTTTT 11580  
attttaaaaaA AttaATTGtG Attttattttt ttttGATTtTt ttaattAAAG tttttAAAG 11640  
atGtATATAAt ttttGATTt AgtaatTTA tttttGATA ttaattATG gatGGATtT 11700  
gttattAAATA ttaattTAAGA taaATGGATA agGATGTTT ttGtAGTATT GTTTTAAGA 11760  
gttaatATAAtt GGAATAGTG taatAGGTTt AttaTAGGGA AttttAAAG taaaAGTTGT 11820  
tgttattttt TAATGGAAAA ttatGtAGTT GttaaaaATA AtatGGAAAA AtGtATTG 11880

-continued

---

-continued

---

tgtatattat tgagatttat tttatatata ttattttatt aatttttatg attatttgg 14220  
taagtattaa ttttataat taagtagtga tagaaaatga ggtttagaga gatatagtat 14280  
tttatttaaa gttatagaat ataagtagag aatgtagaat ttgtatTTTg ggTTTattga 14340  
tagtgagtt tagattttta aaatttaggtt tataaagatt ttTTggaaaa aaagaaaaat 14400  
aaaaaataga gaataaattt tttgaagaa aagtaggaaa ttattatTTta taagtataat 14460  
taaagagaga ttttaatTTt tttttttttt tttttttttg agatggagtt ttgtttgtt 14520  
gtttaggtgt aatTTtagtt tattgttaatt tttatTTTTt ggTTTtatgt tataaagaaa 14580  
gtttttattt gattggTTTg atttgattgt ttattagaat ataggTTTTt tagtagggTT 14640  
aataagttat aatagtTTta gtgatataag ataaatataa attaataata aggttaattt 14700  
agaattatta atatatgttt ttattgtga attattagtt ggTTTataga tggaaaaatAT 14760  
aagatataata atgtatttagt ttaaaaaatg ttttttagag agtgtttaga attatagaat 14820  
ttaaaaattt tgTTTgaaa aggttaattt ttatTTtaa aatataaaa gaaatttaaat 14880  
gtaaaaaaaaat taagagaaaa ggTTTtaattt aattaaatgt tttataattt taagTTTtat 14940  
ttatgtttaa gattagTTTTt gtaatTTatg attaattttta gtattagTTTt gtttttaattt 15000  
tagTTTgtt aaataatata gaattttttt tttaatgaag ttttttaggg ttTTtaggtga 15060  
aaaatataata ttatTTtagga tgTTTtaat ttttaaggaat aaaaagttagt gattttattg 15120  
ggTTTaaata ataatgatata ttataattttt atataaaaatg aagtTTaaag gtatggTTta 15180  
tttttaggtt aatTTtaaag gattttagtt ttgtTTTTgt tgattttttt ttagTTTtat 15240  
tttttagaggt atTTTattta ttatTTgatg taaaattttt ataatagTTTt tagtagaaag 15300  
atTTTtagt gataatgtttt agggattttt tattggagga agagattttt ttatTTTTta 15360  
aggatgaattt tattttttt agggattttt tattggagga agagattttt ttatTTTTta 15420  
ttggTTaaaa ttatTTata tattgtTTta aatttaatTTt ttggaaaaagg gaatggattt 15480  
tttataattt attaggata ttaggattat tttttggTTt gggTTTgtt ttaggatttt 15540  
ttaaagtatg tgattgaggg taggtattag aatggaaatTTt gagTTTgtt agaaaggaga 15600  
aatgtatggt tattatagaa tagTTTaaa tgTTTattgag gagggTataa ttgtaaaaat 15660  
taaataattt ttgttattttt ttgttaaggTt ggtattttta ttttagatg tttttttat 15720  
atagTTTTTTt ttTTTTTTta ttattatagt tgTTTTgaaa ttgagTTgg aaggaaatAT 15780  
ttttagattt agattgttaa atgtTTTTTTt taaagtatg taataaattt aatggtaaag 15840  
ttagggttagt ttTTTTgattt agtatagggt atTTTTTTTTt atTTTTTTattt ttgagattt 15900  
tagaattgtt ggtattgttt taaaattttt ggtaaagatt ggTTTattttt gtaatttaata 15960  
ttttttata atatTTTTgt ttgtTTTgtt tagtagTTta gaaattatAT ggagTTTgt 16020  
ttttaaaaag ttgttgaag ttTTTTTTta ttgtTTTTgtt attatgtta tgaatttata 16080  
attggTTTTTTt ttTTTTTTta tattttagtga agatTTTTTTt ttttaatAT ttTTTTTTat 16140  
tgTTTTTTat ttTTTTTTgt tttatTTTTat attagTTTgtt tgTTTTTTaa tagattttgt 16200  
ggttattttat ttTTTTTTgtt gttagTTTTTTt atttatttagt tatttgaattt gtggTTTTta 16260  
ttgtTTTTta tatagtTTTat ttatTTTgtgg ttgttataata tatttataattt gttatTTgtt 16320  
tttaaatttGTt attttggaaat gattttggta atggTTTgtat tggatgagat ttaaatttaat 16380  
aattaaagta ttgagatagt ttTTTTTTat ataagtTTTat ttTTTTTTt atagTTtaag 16440

---

-continued

---

aggagttt ttttttattt ttatttattt atgtttaata ttatTTTtat tatataataat 16500  
gtataaaaag tatgtgattt atgatttattt ttaaatttga atgtttgtga tttatTTTgt 16560  
gtttttttt 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

ttagaaatg aaaaaata taaagtaaat tataagtatt taggtttaaa gtgggtata 60  
aattatatgt ttttgtgtta ttgtatgtaa taagggttgtt attaagtattt aagaataaag 120  
agtagaaagaa gtagttttt ttggattata aaaataggag taaattttaa atagtagaaag 180  
ttatTTTtat ttttaatta tttagttttaa ttgtatTTTt attaaaaattt 240  
ttttagaata tagtttggaa atatataata atgtggatgt gtgtgataat tataagttag 300  
taggttgtat gaagaataat agagattata gtttaagtaa tttagtgata ggagttggta 360  
taaaaagaaa tgggtgatta taaagttttaat ttagaaatag atagttgattt ataaatagag 420  
taagagaaga tggggagtaa tagagaaaga tggtaaaga aaggtatTTTt tattaagtgt 480  
aaggtgagaa ggaattaattt ggtggTTTtat gtatataata ttaaaataga gaataaggat 540  
tttggtaggt ttttaaagt atagattttt tttttttttt agttggtaa gtaaataaaaa 600  
taaatgtattt ataaggaggt gtttaattata aaagtgatta gtttttggta tgaattttaa 660  
agtagtattttaa atgTTTtaa agtttaaga gtaaagaatg aaagaaaata ttttggatTTG 720  
agtttaagaaa tttttttttt tttttttttt aattttatgtt atgattttgg aaaaagatattt 780  
taataatttgg gttttttttt tttttttttt tagtttaat ttttaaggtag ttataatgtat 840  
gaggaggaga agaatttgg tttttttttt tttttttttt tttttttttt tttttttttt 900  
aatggtaaaa gattatttgg tttttttttt tttttttttt tttttttttt tttttttttt 960  
tttataatag ttatgtgttt tttttttttt atataattttt gttttttttt tagtggTTTtat 1020  
tttttagttt atatTTTgaa gaattttttgg ttaaatttttta gtttaaggga tgattttgtat 1080  
attttaagttt aattatagat gtattttttt tttttttttt gttttttttt gttttttttt 1140  
ataatataatgt tttttttttt gagatgttagt gggatttttt tttttttttt tttttttttt 1200  
gaaagatag attttttttt agaagagaat aaaggaagag gtggtagtt tttttttttt 1260  
gatattttttaa tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1320  
aagttagata tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1380  
ttaaggatata gttttttttt gttttttttt tttttttttt tttttttttt tttttttttt 1440  
tggtagttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1500  
tttaattttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1560  
atattttttttaa gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1620  
aaattttttttaa tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1680  
ttttttttttaa tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1740  
tttaaaataa aatttttttaaa tttttttttt tttttttttt tttttttttt tttttttttt 1800

-continued

-continued

---

-continued

---

ttataattta ttaaatatgt ttattgtaaa atatgtaatt ttattttata gatataaata	6420
ttttttaaaa ttttagaag aatgaggtag gatgtatata gatgtggaaa gtatatatat	6480
ttgtatttt ttatagttgg taaatttatt tagaggtgtg ttttaggaaa tagaatttgat	6540
atattaaata aagatttttt gtattgattt ttgaaatata attatttta atatatttaa	6600
tggtaagaat ttagtagtgt taaaaattt tgtatataata tatataatata tgtatttat	6660
atataatata agatttata tatatgtaat ttattaagtt taggttataa gggaaattg	6720
gtaaaaattt attaattgaa atggagttt gtttagat ttatatttaa tttttttt	6780
ttgattttaa attagggaga gtatgttt tggtttttt ggtaaaatat agatttatgt	6840
aataattagt atatttttt agatttagt aagtaatttt taataatgag tgttttag	6900
tttaatttg tttttttag tgggtggta aaaaatattt tattataggtt attaattttt	6960
ttttttaaaa gaattatgtg ttttgagaag ttttgttta ttataggta atatttatt	7020
ttaaatatag tggttattga tggttttta tgaatgatgt taataaaaat gtttaaattt	7080
tattaggaaa ggagggtta gtttagat tttggagat agagttagtt gatgggtg	7140
tttttaagg tttgttatt taattaatgtt attaattatt aatttttaag tatttaagtt	7200
taataatata ataaaagaga gaataatggg tagtattttt aatattaattt atggttattt	7260
ttggtttagt tggttagtt ttttttagt agaatttata agtgtataga tataatgtt	7320
ttggaatattt aaaggtattt ttttttagt tgtaaatata gggaaaaaga agaaatttat	7380
atttgtaaat ttatgtaaat ttagggggat ttttttaaa ttttgttttt ttgaaatgt	7440
gaaatataat taaaattaa ttttgttttt atttattaaa agaattaaaa ttttgaattt	7500
tgtatataat ttaaatgt ttttttgaat tttatgtt tttttatata atttttttat	7560
aagaagatta aggattgtgg atttgggtgt ggaggaagta attgttaat tataatggaa	7620
tttattggaa tttttataa tattattgtg gtagataat ataattttaa tataattttt	7680
tttttattta gagtttagaa taagaagtta gatgggaa gatggggat ttgatgtatt	7740
gttttaataa ttttaagtt ataaattggg aattgtttt attttagaat attaattgtt	7800
gttttattt gtagttttt atttttatat tataattata atattttattt ttaggatttt	7860
atgtgaatattt taggtttagt ggtttggta aattgttaag aagtgtttaa aataatagg	7920
ttatagtttt tagtagataa aaagaaaaag atgtgggtgt tagttataga atttgagggt	7980
atggaaaaga tttttaaat ttagttttt taatatgttag taaaatgggaa ggatttaaaa	8040
tgtattttaa agaagtattt tttgtttgtt attaagtttt ttttagtttt tagtttagtt	8100
attttttaaa taggtgttga gtttttagt tttgttagt tttgttaag gttttttat	8160
gaggtagtattt aatttttagt atatttttagt tttttttatg tttgtttta taaaatttta	8220
ttagttttgt tgaatggtag tattttgtt ttttaggtgaa tttttatgtg tagttggaa	8280
tgaatattat ttgttaattt attaagaatg ggttttagag tgattgtat gtattnaaaa	8340
agttataaa tggatatgg tataaagttt tattttttt ataaaaatgtt tggtggagt	8400
aaattttat ttttttttg ttttaaaaata aagatgtat gttttttta ttattnaa	8460
gaagtaagta aaaataaaat aaaataaaat aaaattttaa tagttttta ataagagtgt	8520
ttagtgttgtt tggatggaa tttgggttag ttttagaagt agttgggtt gaattttgg	8580
ggtaggttta tttgggtttt ttgggttaagt tgggtttgt tttttttgtt ttgtttttat	8640

-continued

---

-continued

---

gttttgagtg tggtaagaat aataataata aaaaaattaa agtgttaaag gttttttttt 10980  
 ttttttagt ttaagaattt attattttt tatgattttt ttataattt tttttttttt 11040  
 tttttaatt ttgttagtta ttttattttt attttattttt gggttttttt tggttgaattt 11100  
 tttttaata ttaaggtttt tttgtatgtt tttttttaaa agttttatg aaagttattt 11160  
 gtattttta agtgtttata ttttttaat tttgttttat agtttttgtt ttttaattaaa 11220  
 gtttttattt aattgttttt ttttttaaag ttgtggggtt ttttttata agtttttgtt 11280  
 ttttgggttt taagggggga ataaaagaaa tgtgattatt ttggaagggtg gtttattgt 11340  
 gtttgggggg aaaattttattt gtatgtgtt gtgattgggt ttgggtgtgt ttaggtgggt 11400  
 tatataggaa gtgtgggtgtt ttggggaaagg atgtggaggg tggggatgg tggtaaga 11460  
 tggggagga tgggggggtt gttgaagattt ttgggttttag tttttgttta taatatttaa 11520  
 tgggggtgtt tggtgtgtt gtatgtttgg agtttttttt tagaaagggtt tggggtagtt 11580  
 tggggtaag tttaaatgtg ggtgtgtata tttatttattt ttatatttattt gtattgttag 11640  
 agttgaggag gagatttagt gagaagaagg aggaggagga ggaggagggtt ttattttag 11700  
 gttttaaaag tgggggttag ttttagttat ttttaaggtt ttaggtttgg aaagtaggtg 11760  
 gaggggtgga aaggtagttt ttgtgttttgg tggtagggat tgggtttttt ttttttagg 11820  
 ttttttattt tttgggtgtt tttttttgtat aagaagtattt aattgggttg gggatagtag 11880  
 tgggggtttat ttagggattt atttagtaat attgtttgtt ttgtggaaat tattgttttt 11940  
 ttatttttttt tttattttttta tttttttgtt tttgtgttag ttaggtttgg gtttagggaa 12000  
 aggagttttt aggttttttag ggggttaggtt agtaatagat aattgaggtt gatttttttt 12060  
 tttttggagt atataaaattt gtaaaatttag taaagaattt ggttatagtg tggttatgt 12120  
 tttatagagt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 12180  
 gaaagagata ttgagaaaat gagattttttt ggggattttttag agggaaagtg taagaatttt 12240  
 ttattgtattt ttttagggat tggtaatgg ggagttttgtt tttaaaagat tttggtaata 12300  
 aaagggtgga taggaaattt ttttaggtta atttttttttgg ggtttaaag agaatatttt 12360  
 tttttgtta taaattttttt tttatataag ttttagattt gtttttttttgg tttttttttt 12420  
 ttttttagttt ttttaagttat tttttaggttag aatattttgtat aatttttttttgg agtaataggg 12480  
 atttttttggaa agtattaattt atttttttagt tttttttggaa ataagattat aattttttat 12540  
 gtgtatgtt gattttttttt ttttaggttag gtttatttttta ttgtgttaat agtattaata 12600  
 tatggaaagag tttttttgtat tgggttataa aagatttttta ataggattttt atagagaaaa 12660  
 gggtttaataa gttgtatataa aggattttttt tttttttttaga aaagagggtt tttggatttt 12720  
 tttttttttt gaagtttaagt atgagtttat ataataggaa taaaataaaat ttaaggtgt 12780  
 tattagtata atattaggaa tattagaatg gatggtaat tttttttttt atataaataat 12840  
 gaaagtattt ttataattttt tttttttgaa gttttttttttagt agaaaattat attaaatata 12900  
 gatattttaaa atagtagtga taaaagatga aagttaattt tttttttttaa tttttttgtat 12960  
 tagttgggtt taaattttatg ttttaaaattt ttttagtaattt tagatttttg tatatgtgt 13020  
 tttataagat tttttttata ttttgtat ttttaggtattt tagtatgtt attgattttt 13080  
 agtgataat aggtagaaga tttgttagga taatttttagt ttttattttatg tttttttttt 13140  
 atttttttttgg tttttttttttaa gttttttgtat gtaatataat ttttagttat attattttgg 13200

---

-continued

---

aattaatgtg tttttgtagt gaattattta tttttagtaa atttattgtt ttattatttt 13260  
 aaatattgaa atttgatttt attagatgtt tagaatgata gttatagagt agaatttagat 13320  
 ttaattttt atgtataata tgtaaataaa ttaagggtga gaaagttatt gttttaaagt 13380  
 tatataatta atagtttagt gttttttta gtgttaatta tagtaatttt agatthaaga 13440  
 tttttgatt taggaatgga gagaaataaa ataaaatgaa tttttattga aatataattt 13500  
 ttagatatta agatagattt ttttgtgaaa tatttaaga tttggatttg tttttttt 13560  
 tttaaatttg agtaaattta gttgagttt aaagagttt agtaatttg ttaaggttat 13620  
 tttgttaagtg atagatttg tagtttgaaa agtttattat tttttttttt tttttttt 13680  
 tatgatataag ttaagtaaaa ttttttaata aaggagaatt agattagaaa taagtaatta 13740  
 atagtagattt aatgtttattt ttttagggaaa ttttaatattt taatatttttataaataat 13800  
 gtttttttta tataatggaa tatgattgaa agtgatgggt tagatatttt ttatattttt 13860  
 agatgtttttt atttatgtt ttaaatttttga ttttttaattt tttttttttt tttttttt 13920  
 tttatgtta atagattttta taatattttga ttaaataattt ttataggata aaggtatttt 13980  
 ttttagggttt atttaaggaa tttagaatatg tataatgtt aaaaagtttta aattgtttgt 14040  
 tttttttttt ttttgatttg taatattttt atttagaaaaa ttaaataattt ttttaggtta 14100  
 ttagataattt aaaaatgaaat tatattttat ggatattttt tttttttttt tagggtagat 14160  
 tttttttttt agttgttttaa aatggagaat ggataaaaaaa gaaaaaggaa aatataatttt 14220  
 ttatatttttag aatagagttt atagttgtat tattttttt atgttaatttg ttttaatgtat 14280  
 tttttttttt taagagaaaag aaaaatgagag agtattttt tttgattttt taatgtttga 14340  
 taaaatttttta agaagttttt gaatggggaaa atatataattt ttttaggtttt ataatttttt 14400  
 ttgttattttt tttttttttta tatttaagta tgaattttttaat agtaattttttttaatattttt 14460  
 gtgagttttttt ttttaattttta tatttttttta tttttttttttaat tttttttttttaatattttt 14520  
 ttttaggttat aaataaaagta agattttttat ttataatgtt tttttttttttaatattttt 14580  
 agtttagaagt tatgtttttt gttttttttt gttttttttt tttttttttttaat tttttttttt 14640  
 ttttaggggtta gtaagttaat gttttttttttaat gttttttttttaat tttttttttttaat 14700  
 gggaaatttttta ttatattttttaat tttttttttttaat tttttttttttaat tttttttttttaat 14760  
 tttttttttttt tttttttttttaat tttttttttttaat tttttttttttaat tttttttttttaat 14820  
 ttagggatgtta taagtttttaat tttttttttttaat tttttttttttaat tttttttttttaat 14880  
 gatttgattt gttttttttttaat tttttttttttaat tttttttttttaat tttttttttttaat 14940  
 tatagatttaaa aatgttattttttaat tttttttttttaat tttttttttttaat tttttttttttaat 15000  
 aagattgtgtt tataattttttttaat tttttttttttaat tttttttttttaat tttttttttttaat 15060  
 ttttagtattttt tttttttttttaat tttttttttttaat tttttttttttaat tttttttttttaat 15120  
 tttttttttttt tttttttttttaat tttttttttttaat tttttttttttaat tttttttttttaat 15180  
 ttgttgggtta atttgtttttttaat tttttttttttaat tttttttttttaat tttttttttttaat 15240  
 tagaaagttta taaaatatttttaat gttttttttttaat tttttttttttaat tttttttttttaat 15300  
 atttttttttttaat taaaatatttttaat gttttttttttaat tttttttttttaat tttttttttttaat 15360  
 gttgttaatg tttttttttttaat gttttttttttaat tttttttttttaat tttttttttttaat 15420  
 tatattttatgtta tagaaatttttaat gttttttttttaat tttttttttttaat tttttttttttaat 15480

-continued

ttgggttgtt gagaggagat gatTTTatta gtgagaagag aggttttttg agttgggttt	15540
taaggtagtt aaatggttt gtatatttgt atttattttg ggagggtgtt tggagtagag	15600
tattaaaaat gtttttttt ttataaagaa aaagatagaag gtataaagaa tatagtttaa	15660
ttaaattgaa ttaaagattt tattttttt atttttttt tagttatTTT tttaaagtTTT	15720
ttttttttt tttttatagt ttgaaaata taggttattt ttataaagta tttgtttgat	15780
tgaagtaaga taaattgaag agttggTTT ttgaatataa ggaagtgta gagatttttag	15840
aattaatgag gtaggggtgga atataatgtt gaaggTTTT taagaggtaa aattttgtt	15900
gttggTgtttt atttttagt atttttgtt ttatttttt taatttagatt atagtgggat	15960
tgagataagg agtataatgtt aagttgttgtt attgtatggtg taaaaagata tatggtttta	16020
ttaatgtttt tggtgtgagt aaaggtagaa tgtaagatag aaaataagga aattttgtt	16080
ttgttaagaaa agtgtttagt atgggttattt atgggtatgtt agaagtagta agtatataattt	16140
ttatTTTTT tttatttgata aaaagaattt taggattaaat gtttggTTTT attaaaatat	16200
gagaaaattt atttttaattt tataaattaa aggttaaataa ttgttgtaa taattttaaa	16260
aaaaagagaa gatttgtaag atttagttaaa tattttattt aaatagattt tttttttttt	16320
tttatTTTTT ttttttaattt aataaaagtt tattgtatgtt tttttttttt tattgtttta	16380
gatagttaggtt attttttaaa aaagtaataa ttatTTTTt ttttttaagga gtttataatgt	16440
tttttagaaa gaaattaattt tattggatgtt gttttttttttaa aggtttttttttaatgt	16500
tagtgaattt agaggaaaat tataatataaa tatttttgtt tttttttttt ggtttatgtt	16560
qatattttt taaaaataa	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	gtttgaata	tattatata	ttaaaagaat	tattat	ttt	60	
ttgaatatta	ttaatattaa	taatagtatt	atagtgatta	agagtggatt	taaataaagg		120	
ttggtttcaa	attttaggtt	gttattttatt	agtaagttt	taaaaatttt	gagtttttag		180	
ttttttata	tgtgaaatgg	agaaaataaa	tttgtaaaaat	taattaaatt	tgtttat	ta	240	
tat	tttagtgtat	attaattgg	at	ttttttgtt	aggtagaaat	gtgttttaa	300	
ttatgtttta	aattttgttt	gtgttaat	ttt	ttttagttttaga	at	ttttttta	ttttgagaat	360
tagaaaagga	aatattatgt	ttggtaatgt	ttat	tttttttt	taaaataaaat	ttgttaggaaa	420	
gaatatttag	taagtgtat	gagtagtaat	gatt	tttttttt	attat	tttaaaat	atttataata	480
ttat	tttttttt	tagagtttt	taagtattgt	agataat	ttat	tttatta	aaaataaaat	540
tgtat	ttata	ttata	tttttttt	tttttttt	tttttttt	tttttttt	tttttttt	600
gtataagatt	aatttaaagt	agaggat	gaa	at	tttttttt	tgaat	ttttagt	660
at	tttagt	gaa	tat	ttttaat	ttt	ttttag	ttttagt	720
ttgtat	ttttagatt	ata	ttttagt	aa	tgt	ttttagt	ttttagt	780
attagagatt	atttggattt	tttttggaaa	agttattt	tttttgg	tttttggttt		840	

**-continued**


---

ttttattttt tataattgggtt taataatgat tattgttagag ttgttatgga agttaaatga	900
atttatgtat tataagtgtat taatataatg attgatattt agtgatttat taataaatta	960
aaatattttat tattaatatg atagagaagg tggtgttaaa atagataata ggtttttgga	1020
agaggtgatt aaatggatgt aaaattttatg gattgtttat tttgtttatt ttgttgtgt	1080
tttttgggtt tggtatatat atgtgtgggt ataaaattgt aaattttatg tagttgtgt	1140
gtgtatgtgt agaagggtta gatatgaata gttatTTTaa taatgtgtt agagaagttt	1200
tgtatgtgtt ttggaaagtaa gttgtgttg tttgatTTTt ggggtttgg gatggatgtt	1260
tatattttgtat tttagtagta ttggaaagggg ttttaggtttt tgtagtata gtttattttt	1320
agattgttta gttttttata atatataattt ttatggaaaa gggtattttt ttttggtaga	1380
aaaagtgttt tagtttgggtt tgggttgggtt tttatTTTat gttgtgtaa gtaggtgaag	1440
ttttttttgtt tttttttttt ggggttaagtg gaaaggagtt tggtaggggg tttgttagtgg	1500
tttgatagg ggaattgggtt agtgagagag ttttaggtaa ttttgggggt tttttatag	1560
aagttaggtgg ggattgtatag tggttttttt gtttagggag gagagtgtgg ttgtgggttt	1620
tttttttttag tttggaggtt gtagttgtt gatgtgtttt ggggtgggggt ggggtgggggg	1680
tgggtggag ggtatggaga ttatgggtgtt gttatTTTggg atatttaggg ttttggagtt	1740
ttgggtgggtt ttatgttag gttgtaaattt atgataatag gtagttttt gaggtaaat	1800
aaaaatggag tgggtttttt gtgtgtgtt gttttttgt gttttgggtt ttttttttga	1860
ggtttttgtt ggttttatga gttgtgttaa gttgggtgggt atgtttttt tttttttat	1920
ttttgtgtaa gtgtgaggtt gttggtagtg tggtgtatgt tttgggtgtt tttgggtttt	1980
gtgtaaaattt ttatTTTgtt ttatgtgaag tggtgtgtt ttttagagagg gggaaagagt	2040
tgtggaaaaa gttggggagt gatgtgtgt gttgggtgggt gttttttttt attttttttt	2100
ttttttttttt tttttttgtt gtagttgtt gttttttgtt tttttttttt tttttttttt	2160
tggaggtgg tttttttttt tttttttgtt tttttttgtt tggtgtatgtt atttggatgt	2220
gggtgggtga aggggatgtt ttgtttttat tagaggtaa gttgtgaagg gaaattttga	2280
tattggaaagg aatgagaata aatatttaat tatggatgtt ttgaattgtg gttggatag	2340
atattttggg aatttgaggt ggattgggtt atgagggtgag tgattttttt ttttaatttt	2400
tgttttaggg tttttggggg agtttgagtt gagagaattt ttaaattttt tggaaagtgt	2460
tgtgaggtttt tggtggggat gttgagtgtt ggggtattgag gatgtgtatgt tggtgtgt	2520
gtgggtgttt tggtttttgg ggggtgtttt gagggtgggt gttttatgtt tgagggttt	2580
ggttgggtttt gttgtgtgtt gttgtgtgtt ttttagaagat gtttttaagt ttttaagggtt	2640
ttttttgtgtt ttgtgtgttt tttttggggg tggtgtggag tttgtgtgtaa atggagttt	2700
ttagtagtt tagtggtgtt tttttatTTTt tttttgtttt ttatTTTgtt gagggtgt	2760
tattgggggtt ttttttatat tttttttatg atgtgtatTTTt tttttgtatg atttttttaga	2820
tggtttaggt gtgaggatgt tgatTTTgtt gttttggag ggttataagggt gttttgggtt	2880
ttttgggtgtt ggggtgtgtt gtatTTTaaa ggtttgtgtt ttaattttta ggttattgtt	2940
gggtttttta atttgggtga ttttattttta atagttttta tttgtgtgtt attgtgtgtt	3000

-continued

---

-continued

---

ttatthaatt tttatgataa ttttatgggt gttattatta agttggata aaagatgagg	2160
aaatthaagt ttaaagaatt atgagtgatt tttttaaaga taaattttagt ggtttttaat	2220
tataaagttt atgatataat aattttata ttattttagg tgtggttaa gagattatag	2280
attaattttgt agaaatttattt ttatttaaa gaaagtttag ttttaatata tttttaagt	2340
tatgtttga aaaaggtttta gaaaaagat tttttttt atttttagttt ggttttatgt	2400
aattgatatg agtattgttt aattttaaag tagtattaat tttgaatatt tatggttata	2460
attttttttt taatgagtgg ggaatttattt ataatgttta agaggtttta gaaaagggtgg	2520
tgttgtaaat ttaaaatgtt aaaggtagtt gttgttggtt ttattttttt ttgggtgttt	2580
ttttttgttag atttttttg gagggtgttag gtattgttag gtataatgtt ttttttttg	2640
gttttttaagg tagaaagggt tttggtagtg ggggtggat aaagtggatt tggagtagtgg	2700
ttgagagtagt attttggttt aatgaggaat gttagttat atatattgga gagaaaaata	2760
tgaatggata gatttaatta attttgtaga tttttttttt ttattttata tttgagaaaa	2820
ttaagggtttt agaattttta gaagtttgtt aataaatgtt agatttgagt tttaagtttag	2880
ttttttttta agtttatttt tagttattgtt gatgttattttt ttatgtttaa tggtagtttt	2940
aaaaatagta atttttttaa atatgtataa tattataagt taaaatgttta attttaaatgt	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 ttttttattaa tgnttttggg nttataggnn gngntttttt    60
gnttaaatag tagttaatgn aggnattta aaaggtatata tttttttat ttatttttagt    120
gttatgntat tatttaattt ttatttataat tttgtgnngnt tagtataatt tttgntttta    180
ttttttttgt gaataaatgg aggtagnnnt ttgngtaatt tataagtaag tggtagattt    240
gggggttggg tttagataat gtggttttag atttttttttt tttttttttt ttttgggtgtt    300
ttttgattga gtttagttat ttatttttttt aatataatatt aattttttttt ttttattttt    360
aggttttgtg ttagatgtta gaaatataattt tttttttttt ggattttttgtt atggttttt    420
gttttagtgg tttgttattttt atgtgttata taatgttattt atttttttttt ttagttttt    480
gaaaataatg ttgaaagata ttgtatgtat gtttttataa taaaattatgt taattttatgt    540
ttttttttttt gttttgggttta ttaggaattttt tttttttttt ttaggtttttt ttttattttt    600
tttgggttttgg tttttttttt ttaggtttttt tttttttttt gttttttttt ttttattttt    660
tttataaaattt ttatgttata aatgtttttt attgtaaaat tttttttttt tttttttttt    720
gaagggttggta tggaaatgtat atgtttttttt tttttttttt tttttttttt tttttttttt    780

```

**-continued**


---

ggaaggtaga ttttagatgtt ttttttattt ttagtttaa agtttttat ttataaaatt	840
ttttaagtat tgattattgt tggggaggt atttggaggt agtttagagt ttttgaagg	900
taaaatttat atataaaaag aagtttttt tgatttagat ttttagtttg gtgttagatg	960
tatatttgag gagtaggtgg ttagttagat ttttatttga gtagttaata aaatttattg	1020
tttttaatg gaatttttt tgtaagttt aattgattt ttttttgtt gatttgttag	1080
atggtaggga agtattaaat attattatga tttgggttat agtggggaa aaaaaggaaa	1140
aaagaaaaaa aaaattttt gtttaagttt gtttaggtgt naaagggtt gatttgtgg	1200
gtttatTTTA tttgatTTT tgaaataga gtggatttt ttaatatttt aataaagaga	1260
attttttgtt attaggttgg aagtgggtt tagtttttg ttttttgtt ttttttgtt	1320
aagtggaaatt agtgggtatt gtttagtgtt atttgtgagg ttgagtttta atagttaaa	1380
gaagtaaatg ggtatgttatt tttgtgggt ttttttttt tgagggtttt attttgttatt	1440
tgttatgtaa aatgagggag ttttagaaag gaattttgtt ttttttgtt ttttttgtt	1500
ttatttagttt ttatTTTAatg ttttttgtt gtttttgtt atttttgtt gtttttgtt	1560
gttttgttgg gtagagtttt tttttaaagt ttttttgtt ttttttgtt gtttttgtt	1620
gttttgttgg ttttttgtt gtttttgtt ttttttgtt ttttttgtt gtttttgtt	1680
attttgttgg ttttttgtt ttttttgtt taggttagga agaggggtt ttttttgtt	1740
tttttgttgg ttttttgtt ttttttgtt ttttttgtt ttttttgtt ttttttgtt	1800
gggtttgggt ttttttgtt gtttttgtt ttttttgtt ttttttgtt ttttttgtt	1860
ttagtttaga gtttttgtt ttttttgtt ttttttgtt ttttttgtt ttttttgtt	1920
tgttgtgtgt ttttttgtt ttttttgtt ttttttgtt ttttttgtt ttttttgtt	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

ttttttgtt aatgaatgtatatatata tatatatata tatatatata tatatatata 60
tatatatatgtat tttttttttt gttttttttt tttttttttt tttttttttt tttttttttt 120
tttggaaatgtat tttttttttt gttttttttt tttttttttt tttttttttt tttttttttt 180
atttttttttt gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 240
tagttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 300
gagttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 360
tggttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 420

```

**-continued**


---

agttagttaa	ttttttaaat	agattagtag	tagtattatg	aaagtattgg	gtagaggttg	480
atgattagga	ttaatggttt	tataagatgg	atttttttt	aatgtttttt	tgttttgtat	540
ggtagatatg	gggtgagtat	tttgtgagga	gtgagtttg	tggagggtgt	atttatattg	600
tttttttgg	ttgttgggtt	ttagtttat	aaattatgtt	gggtaatgtt	agttgatttt	660
atttttttat	agaatggaat	tgtatgggg	attgggtgg	attttagttt	tagtgtaaaa	720
agattttttt	tattaaaatg	ttaataagat	ttatttttt	ttaataaat	tagataaaaat	780
gttttagta	gttttagttt	ttntatgtt	tggtaaggtt	tggtagtgg	tttttttttt	840
tttttttttt	ttttttttat	tattgtgg	taagttatga	tggtgittgg	tgtttttttg	900
ttattttata	aattataaag	atgatagatt	aatttgagtt	tgtagaaaaa	attttattaa	960
ggggtaatag	attttattaa	ttgtttaggt	gagagtttga	ttagttttt	attttttaaa	1020
tgtgtatttg	gtattaaagt	tgggatttg	attggaagga	atttttttt	gtatatggat	1080
tttatttttt	agaagttttg	aattttttt	aggtatttag	aatagtagta	gttattgttt	1140
aagggatttt	ataaatgggg	gtttttggg	ttgagggtaa	gaggatatt	tagtttgg	1200
tttttaata	ggaatttaag	aaaatgtagg	gaagatttt	ttgtattatt	tttatgtat	1260
ttttttttat	aaagaattha	aggtgtttt	taataaagga	tattgtgt	atagaattta	1320
tgaattaaaa	tagaaaattt	gggttagaa	gaaaagttt	gtaaattgaa	atgataagtt	1380
taaattttaa	tgtattgtt	ttgatttgg	taagagttt	ttgatagtt	aagtaataaa	1440
agaaatataa	gttatatagt	tttggtag	aaatatgtat	ataatgtttt	ttagtattat	1500
tttttatggg	ttaaatattg	gaataatgt	attatgtat	atgtaatata	gtaggatatt	1560
gagtaatgg	ttatgttaag	atttgtgg	tggggatgt	tttttagtat	ttggtagttag	1620
gtttggaata	tggagatgt	attagtgt	tgttaaagaa	atagatgtt	agatttaatt	1680
aagatgtat	agaaggtagg	gagggttaag	gagtttggag	ttatattgtt	taggtatagt	1740
ttttggattt	gttattttt	tgtaaattat	ntaaannnn	gtttttttt	gtttatagat	1800
aaaatgaaan	tataaattgt	attgantnta	taaggttg	gtgaggatta	agtgtgnnta	1860
taattattga	ataaatataa	agtgtatgt	tttttaagat	ntttntattt	gttgggtttt	1920
ggntagaagg	ntntnnttg	tggnttaag	gntattgata	aagatgnnag	tgtgnntga	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

gtttttgg	tg agatatgtgt	tttataagtt	ttaatggaga	aaaatgtaa	tattttattt	60
tttgaatatt	gttttatttga	gtaatgagaa	aatagtattt	tttttttagga	tagtggttt	120
taattatgtt	tatgtgtttt	tttaggaaaa	ttttaaaaat	atatatata	taatgttttt	180
gtgttatttt	tagggatttt	aagtttttga	atatgtattt	tgtatttagta	tttttttaattt	240
attttaggtga	ttgtgtatgt	aaattatgt	ttagtttttt	tgttttaaga	tgaaataaaat	300
tttttttagt	attgaaatta	taaatttaaa	ttattaaaat	taattaagg	tatggaaattt	360

**-continued**


---

aataaggtat	aggaaagttt	ttatattata	aaattatttt	tttaaattat	agtttattgt	420	
ttatatgtta	tttgttattg	tagaaaagg	tgaaaaaata	gtaaaattaa	ttattttag	480	
tttggaaaat	tathtagaaa	tgaagatgt	gatttgaaa	tattgttaat	attattgtat	540	
ttataaataa	tgttttaata	tattttatt	atattgtat	atatttttt	atataaat	600	
tatataattaa	aattaaggta	ataatgtatt	tagaatattt	tatttatatt	tatgtat	660	
aagtaggtta	gaaattaaga	tatgagttt	taagtatgt	atgttaagg	gtggggtag	720	
aaattatatt	gtat	tttatttatt	attaataatt	aatataattt	ttaatattat	780	
ttttaatttg	tatattttta	attat	tgtta	taaataaag	tataatatt	840	
tttatgtatt	tat	tttatttta	tat	tttattta	tatagggat	900	
tttattattat	taaattat	at	tttatttta	taat	ttttaggt	960	
attgttattt	tttat	at	aaat	atg	ttttaggt	1020	
tagatattat	agaagttaa	tat	at	at	gtataaaat	1080	
atatgtataa	tagtttattt	ttat	attt	ttttag	aataagg	1140	
tttgcggaaa	ttgagtaaaa	att	gagg	ttt	tttat	1200	
attttatttt	ttaataagg	at	gaat	attt	ttttag	1260	
ttttgaaatt	tttgatagtg	ggat	tttgg	ttt	ttttaag	1320	
atttatatga	ttaatttagt	ttag	aaat	tttta	ttgt	1380	
aagttgggt	tttattgtgg	ttt	tat	aaat	gtattat	1440	
atatgttttt	atgtat	ttt	ttttag	ttttag	ttttag	1500	
gtgat	tttgcgtt	tgaat	ttttag	tat	ttttag	1560	
tgtat	ttttttgg	gtat	tttgcgtt	tttgcgtt	tttgcgtt	1620	
ataaaatgag	ataatattt	ttgtt	tttata	gggtgttt	aaat	aaaaataat	1680
atgtttattt	tttatatgtta	taat	gttt	tttata	gtat	tttttattt	1740
ttatgtat	aatt	at	tttgcgtt	tttgcgtt	tttgcgtt	1800	
tat	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	1860	
taagttttt	at	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	1920	
aaaaaaaat	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	1980	
ttaataatt	ggat	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2040	
tgttttagtt	taagaagatt	gaa	atgtat	atgtat	atgtat	2100	
taaataat	gtat	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2160	
tagttat	at	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2220	
gttaaagat	taat	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2280	
agtttaata	atgtat	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2340	
agtgtat	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2400	
attataat	taat	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2460	
ttgtttaaa	ttgtttaata	at	tttgcgtt	tttgcgtt	tttgcgtt	2520	
agtataatt	agtaat	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2580	
gaagagtata	tattttaga	aag	atgtat	tttgcgtt	tttgcgtt	2640	

**-continued**


---

gtttttttt gtttaagaggt tttttaaaat tttatttata tagttaaattt ttgaaatttt	2700
agttaggtttt gttttattat tataattattt gtataaaatat ttttaaggat tttgttttta	2760
gttttaagta tgatttattt ttataagttt gatttagttat tatattagtt ttgttatgga	2820
aatatgatatg tttttatttt ttgttgtaga gttgttaaat tttgatttat atttatgttg	2880
tttttttgt tgaaagttt tagtgaaga aatttttaat ttttgtttt gtaatattag	2940
ttggtagttt tatttaatgg gtattttgtt ttttaaaga atttatgtt tttgtttaga	3000
agttgatttt ttgatgttt taatgtttgg tttattgtat ttgttttaat ggagtttttgg	3060
ttggtgagga gtgagatgtt attgattaga atgttggat ttgttggttta attgttagga	3120
gtgagagata ttgagattt aaaaattttt gagggtggag gggagaggga tagtttgga	3180
tggaggtgga gatgtaagat aaaggatgg attttatata gaaaaaaa aagattttgg	3240
ttgaggtattt gaggtgtgtt atgatttat ttttaaagg agaagttaaa aagtaaggaa	3300
gtggggaggag gttggagggtt aaagtattttaa aaggattat ttgggtataa tttgtttttt	3360
tgttgggttt tgtaaaggat agatagttt gtttttaaag tataatgtt ttttttttaa	3420
gtgattggga atggatatttta attgtttgtt aaatgttattt aatgtttttt ttaaatttttag	3480
gggatataaga aagaggggta taaaaggaga atttaaatag aaaaaggag gatttggagg	3540
ttttgaaag tggggggaga agaaggagga gggataatag agaggaatag agaaggagag	3600
tggagagaag ataaataaaa ataaaaatag gaattattga ataatttat attaaaaaga	3660
aagttttttt ttatgggtt tttaaaatattt ttagattgtt atagtgtt tggttatgga	3720
agaaagatgt ttttttttat ttttgtttt gaaagttttt ggttttttta ttgggttatta	3780
aaattttattt aggttaaaga gtgtgtttaa ttgtttgaag aatgttagtag atggaaaggt	3840
ggttttttta tggtgtttgt tttttttgtt ggagagaatg aaagaaatgt ttagagttag	3900
agatttttgt ttagtttagat tttttttgtt tggttttaggt tattgttatt ttggtaaaga	3960
ttttagttaag gaatgttaggg ttattgtttt ggttaataaa tggagttttgtt tttttttttt	4020
ttggatgttg ttgtttgggtt gatgtttttt gtaattttt tttgtgtat gtagaggagt	4080
tttttttttt ttttttagattt atttgttttgg attaatttga ttttttaat atatttattt	4140
gtatttttta ggtggatata ttaataggtt atgggttgg gaggagtg gtagaggag	4200
agggatttaa atttgtgaat gtttgggtt ggttggaggt gtgggggggtt tgggaggaga	4260
gaggggagaa gagagaagga aggagagtgt ttgttggat ggtttagttt ttttggtag	4320
tagttttggg gttgtatgtt ttgttggag atgttggat ttttttttagg ttggtaaaga	4380
tggtttttaat attattgttt tttttttttt tttttgtaaa ttttttagaga aatgtttttt	4440
gtttttttgt tttgtatattt ttagttgtt tttttttata gtttaggtgg tttgtttttt	4500
tatgttggag tttgtgtgtt tagtaggtt tttttttttt tttgtgtttt tttttttttt	4560
ttttgttgggtt tttgtttttt tttatattttt gtttttattt ttttttagttt ggagagatgt	4620
tattttagttt tttgtgtat tttgtgtttt ggtttagtgg tggaaagagg gtttagttt	4680
ggatttttgtt tttgttaggg ggtgtttttgg agtggagagt gaggtgaatg gtatatgt	4740
gtgtgggttag tttatatttga agttttagttt tttttatttga gttatgtttt gtttagttt	4800
atttgggtta tttgtttgggtt agtggattttta tttgtgtttt tttgtgtgtt tttttttttt	4860
tatttttgtt tttatatttgg tttttttttt ttttggatt tttatatttgg tttatatttattt	4920

**-continued**


---

agagtttgc tatgtgattt tgggtgggtt gaggattgt tgtttttaa	4980
atgttagagg gatgtgggtg gtagagttt agaggtgggtt gttgggttgtt ggggtgtttt	5040
gattttttt ttattttgtt tttttgggtt ttattingttt gtttttgat ttttgttttt	5100
ttttgttttt tggttttta gagttttttt tttatggtag tagttttttt tgttttttgt	5160
gtagttttttt agtggatgtat tttttgtttt tgggggttag ttttagttttt ggatgttgtt	5220
gaaatttttt agattatgtg tgggtttgtt tgggtttttt tgggtttgtt ttattgttat	5280
tgttgggtttt tttgttgtt tttgttgtt gatgttttagt agtttgttgtt ttgggtttttt	5340
tgattttgtt tttttttggaa gttgttttgtt gttgttagt gttatgaatt agttatgggt	5400
ttgtgggagt ttttggta gtgttagt gttatgtt gttgggtttt ttgttgtt	5460
ttgttgttgtt ttgttatgtt atttattgtt gttgtttttt tgaagttgtt tgttttttt	5520
atttttttaa gtgattgtta aatgtttttt gtttggaaatt gttttttggaa gttttagaattt	5580
tttgggtttt attttttaat tttgtagaag aatatgttta ttttagtata tagtttttatt	5640
ttagtgtgt ttttttagtt tttttttttt tattgtttta gatttttaat attattttttt	5700
tttattttaga gaaataaggg gaattgttgtt aggtttgggg tggaggggtt gttttggat	5760
gggttagaaag ttaggtgtta gtaggaaattt tttgtatgtt tttgtttata ttggagttgt	5820
gaggattttt agaaatatttta aatgggtatgg tttttttttt ttattgtttt gaaagatgtat	5880
taatttttagg gaaatatttta aatagaagt tttgttattta ttaaagaaaa aagttttattt	5940
aggatgagga agaaataatttta ttatgagaaa gaatgagtgta gaaagtaata aattaaatgg	6000
tgattgttagg ggaattgttg atttttgtta aaggtgttat gagggtgtat tgggtttttt	6060
ttgaagatta ggttatatacg attttagagg agttgggttt taatagaattt tttttttttt	6120
tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	6180
ttatttttttt ttttttttagg tggtaaaaga tattggttttt gtagtttaga tatgtttttt	6240
tttttttttt ttaagttttt aaggtgtat aggggagttg agaaaaagaa tattttgtgg	6300
gttttttttagg ttggagtggtt tagtattttt gttttttttt tttttttttt tttttttttt	6360
ggttggaaattt attttaggtt gttttttttt ggttggatgtt ttgtggagtt	6420
gtttaggtat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	6480
gtttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	6540
ttagtgggtt atgtatattt gttttagtata ggggtgtatgtt aatttttttat	6600
ttttttgttata tataatttttt tttttttttt tttttttttt tttttttttt tttttttttt	6660
atgtatataaa gttgaagtgtt atgtatattt agagaggagg gaatgattaa ttgtatattt	6720
gggtgaaagt tttgtatattt ttttagtattt attgtatgtt atgtatattt gaaaattttt	6780
atgtattattt ttaattttttt ttttataaaat tttttagtattt attgtatgtt atgtatattt ttattttttt	6840
tgggttataa gttgtttttt tttttttttt tttttttttt tttttttttt tttttttttt	6900
aatatgtttt tataatttttt tttttttttt tttttttttt tttttttttt tttttttttt	6960
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7020
ttgtaaaaat aatttttttt gttgtatgtt gttttttttt tttttttttt tttttttttt	7080
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7140
gtaatattttt aattttttttt ttattttttt tttttttttt tttttttttt tttttttttt	7200

-continued

ataaatttgtt tttgttagata gggtgttaag aatgggttat ttgtttttta tttttgttaa	7260
tttagagggtt attaattttt atttgagttt aatagattat tatagaaaat attgtgttg	7320
tttattttta ttattgaggt ttgtttttt tttgttttg atataittta aataagggtt	7380
tgttttagtt gttgaagtaa aagaataatt aaagatgggg aaatggtaaa agggtattta	7440
gagattatta ttagttttt tttaaaatgt ggagttttgt ggttataaat attgtttatt	7500
taatgagtaa aaaataaaaa taaaaaaaaaa ataggaagta aatgttaagt ttttattttat	7560
tattgttagt attaatgtta gttttaaaaa atagttttat tagaaaagga tattaaagga	7620
gaattgatta gaaaagaatt gtggaaaatg gaaatgaata ttgatttattt aattagattt	7680
tgaggttatt agtagatagt gatttgttag tatagttata gttgttggat tttaaaattta	7740
ggataagttat tttaaagttt taaagtagtg ttttttttg tttaaaaattt gtaagatgtt	7800
ttaatgattt gagtggtttt ttgttaattt 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 taataaaaaaa 60  
aagtattatt ttgaagtttt aaaatattt gttttaaattt taaatatta aattatagtt 120  
gtattgttaag gttattgtt attgataatt taaaatttt gtttaagtgtat taatattttt 180  
ttttatTTT tataatTTT ttttagttaa tttttttttt gtatTTTTT ttgtatgtgt 240  
tatttttaa agtttgtttt aatattgata gttggtaatg aaagtttaat atttgttttt 300  
tgttttttt ttatTTTT tttttgtta ttaggtggat aatattttt attataaaaat 360  
tttatTTTT ggaaaagagt tagtgatgtat ttttgaatTTT ttttttattt tttttttattt 420  
tttaattgtt tttttgtttt aatattgtt aataattttt atttggaaatg tattttgata 480  
aagaggaaat aaagttttaa taataaagat aaataggat agtgtttttt gtgtggttt 540  
gtttggtttta aatgaagattt gattttttt aagtttaatg ggggtggaaatg ggggtgttaa 600  
gtttttgata attttttgtt aaaaattttt tattttttt attttatgtt gtttttttttta 660  
aaatattttg taaatatgtt atttttgtt tgtaaatgtt aattttatTTT ttaagttatgt 720  
tattttttaa aataatgtt aaaaattttt gggtagggaaatg tttttttttt tagtttgtt 780  
agttttaat attttttaaa attttttttt taggttttata atttttattt taataaaatgt 840  
tttaaaaatgtt gatttagttt tagtaaatgtt aatTTTTTTT ttttaaaaat aatatatTTT 900  
tatgtatgtt gttttttttt aataatgtt aataatgtt attttttttt attttttttt 960  
aattttttttt aatggaaat aatTTTTTTT tttttttttt aatTTTTTTT tttttttttt 1020  
tttttttttttta aaggataaat tagaaatgtt tttttttttt tttttttttt tttttttttt 1080  
agtaataattt aggaattttt taggtttttt tttttttttt taggtttttt tttttttttt 1140  
tttaaaatgtt tatatatTTT agtttataat tttttttttt tttttttttt tttttttttt 1200  
gattttttttt aatTTTTTTT tttttttttt aatTTTTTTT tttttttttt tttttttttt 1260  
taggttttttta aattttttttt aatTTTTTTT tttttttttt aatTTTTTTT tttttttttt 1320

**-continued**


---

tatthaatta ggaagttta gtagagtcaa gtttgttaag tgtttggtgt tagaatttga	1380
aggaaatttga gtgagtaaga agagtgtttt atttatttta tagaagtttg tttagaaatgt	1440
gaggagttag tgtttatgtaa agttggttt gtttttgtt tgtttatatg gagtttgatt	1500
agtttttagtt atgtttattt tggtttggaa gatttgtaaa gtgtttttt ttttaatttt	1560
tttgtattat tttgaagttt agggaaagtaa agagaggggt atatttggat tgtaaaattta	1620
atgttttttgg ttttttagga gagaaggaa tgagagagag agagagatag atagatagag	1680
agagagagag agagagagag agagagagag agaaatttta ttgaaattta	1740
gttttttttag aatttgtgt atttggttt taatggaga ttgtgtat tttatgtat	1800
ttttgttagg aattagtat tttttgtat ttattattt atttattgtt tttttgtta	1860
ttttttttta taaagttatt tttttttat ttttagtaaga tttttttttt taatgtat	1920
aaagtttttg ttttagtattt ttttttagga ttgggtgtttt tttaaaatag tgaatttaga	1980
aaattattttt gtttaatatt tttaaaattt tttgtatgtt taatgtatgt gtaagtatgt	2040
aaaggttttt ttttatattt gtattttttt tttatgtat aattattttt tttttttgg	2100
ttttagtataat tttttttgt tttttttgtt agaggtgggt ggtttaggg gtttagggta	2160
gttaggagggtg aggggtttagg gaggtgtgtt agggtaggtt ggtttgtgtt ggatatgtgt	2220
gtttttttgtt ggagttaaag gggtggggat ggggttttg gatttatgtt agtaatttta	2280
gttgggtgggt gttgggtatg tatttaagga ggttagggaaa gtagttagttt ttattgggt	2340
gtttatgtatg agtagtatga tgggtatgtatg tagtagttttag taaaatgtt tgtaaagtgt	2400
ttagttgtt tattttgtt gggattttta tagtattatg attagttgtt gtaattttgt	2460
agtagtaaat gtttttttagg gaatataggaa ttgtgggggt tgggtatgtt gttattgtat	2520
attttgttggaa tgggtgtatg agaggtgggt gttgggtatgtt tgggtatgtt tggggaaatgt	2580
gttagttaaat ttgtgtatgtt ttttaggtt ttttaggtt ggggttttagt	2640
ttggagttttagg aggggtttagg tggagggaga gttaaagggtt ttttaggtt ggttagttttag	2700
aaggaggggag ttttggaaag ttggaggata ggaggagatg ggagttttagg ggttagatgt	2760
tggagttttagg ggaggttaggg tggagggaga gttaaagggtt ttttaggtt ggttagttttag	2820
tttttaggtt tgggtttgtt atttttttttgg tgggtggaa gtagtaggtt ttttagttgt	2880
ttgggggtttagt gttggaaagag gtagttgggtt ttttaggtt ggttaggtat gtagttttgt	2940
ggagggaggg gttgtatgtt aggtgtatgtt atgtaaaggag ggttgggtt gttaaatgt	3000
agatgggtttt gtttttttagg tgggttttttgg agtgggttta ggtgggtat gttttaaatgt	3060
agaagttttgg gtttttaggtt gggttatttg tatattttata tattttttgtt tttttttttt	3120
gttttttaggtt gttttttttttaa gaaagggtttttt gttttttttt tttgtgtgtat	3180
ttttgggttttgg ttaggtgtggg ttgtgggtgg gttgggtttt ttttaggttgg agatgggtgg	3240
ggtggagggatg ttagaggatg agtagtagta gggtagagag gggtaggtt ggtgtggaga	3300
gggtgttttttgg tgggtgatgtt gtttttttagt gttgtggatg gttgtgttta gttgttaggg	3360
ggatgttaggt tgggtatgtt gttgggtgg gtttaggtt gtttttttagt ggattttatgt	3420
gaaagagggtt gtaggtgtatg ggtgttagaa ttgtttttgtt tgattttggaa gtaatagtag	3480
tattttttat aagagtgtgtt aattttaaagg ttttttttttgg aggtatgttta gtttttttgg	3540
taggtgtttt tttttttttt tttttttttt tttttttttt ttaggtttttt tttttttttt	3600

**-continued**


---

attttagtta agtgttgta ggtttgaatt tttttttta ttatgggtt ttttttagtt	3660
tgttagttat tagtgtgtt atttgggagg tgggtttaga tgggtttgaa aggttagatt	3720
ggttgggata agtgggttga gagaaagaga aagggtttt tggatgttt gtgggtgggt	3780
tgtgggagt attgggtgg tagtgggtt tggaaagggg aggtgggtt ttatgggtt	3840
gtttaggtag tgatgggtt ggggtttgt tggatgggtt gtgatgggtt	3900
gtgatgagag aagggttaat ttggtaggag ttttgggtt tggatgtttt tttttagttt	3960
tttagtgaaa agggtaatgt gtatagtggg atttgggtt tggatgtgtt attttaggtt	4020
tagttagata ttttttttag tttatggaa ttttaggtt tagtaatggg attaagagtt	4080
tttgggata aggggtggaga ggaatatttt tttttatga ttgggttat tttttagtt	4140
ttagtgggtt ggtgtttta taggaaagag tttttttttt ggtgtgtat tttttagtga	4200
ttttttttt tttttttttt ttgtttttaa aagttttgg atttttttttt tttttagtta	4260
aatttttttt ttgtgtttt tttttgtgt tttttagatt taggagagta tttgataata	4320
ttaataggt aatttagtgtt ttttttaat ttttaaaag aggtattttt atatggaa	4380
aatgggatta tttttttttt gtagatatta gtagaaaaat aaattgtatt tgagtaattt	4440
ttttaagtat tttatgggtt aattttttttt tttttttttt tttttttttt tttttttttt	4500
gagatgtat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4560
aatttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4620
ttttaaagat tttttagttt tagtgggtt tttttttttt tttttttttt tttttttttt	4680
tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4740
ttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4800
agatttagatg ttggaggtat tagaaaattt gtttttagat aggttagttt aattttttaa	4860
gaaatagaa tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4920
atttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4980
aattttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5040
atttagttta tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5100
tgttagtaatt atgataatga aataggattt gtttagattt tagatgtttt ttatgttaat	5160
agaattttttt agaattttttt agtagaggaa aattttttttt gaattttttt tttttttttt	5220
ttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5280
aaattttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5340
aggttttttt aattttttttt tagtttgaag taaagttttt tttttttttt tttttttttt	5400
ttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5460
gtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5520
tatgtttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5580
agattttttt agtaattttt tttttttttt tttttttttt tttttttttt tttttttttt	5640
atttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5700
ttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5760
tatatttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5820
attttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5880

**-continued**


---

ttttttttag aaaatttaag aatattttt ttttttatgg aaagaatatt attataaagt	5940
gtgagattat ttatagttta agtagggggtt ttgggagttt ttttttaata agaatagttt	6000
aagataaaata aatgaatttg ggaaaataag atatattgtt aattagaatt tttttttttt	6060
ttatgatttt atattttttg attgttttaa taaaggtaag atgtatTTT tgTTTTtag	6120
gtgttaggtt ttgtgttatg taggatagaa tatgttattt ttatTTTattt tttaaaatAT	6180
ttttatgaga taaagaatAT tattttattt tatataaaag gaatATgtt ttgaaagtat	6240
taggttaattt gtttaagaa ataaattttg ttagtgcatt tgTTGGGATT ttgtgaaatt	6300
ttgtttgatt tttagtata agatataagt tattaattt tgTTgttattg ttgtttgttA	6360
gtttttgaga gggaaatTA attggaaatg tattgtttt gtttatgata ttTTTTTGT	6420
ttttttgtg gagttgttgt aaggTTAA tttaatttt taaattttgg taataagatt	6480
tagtgatTTT tgaattttgt tgattatAT aatttttga gaaatttga aataatAGAT	6540
aaatttttaag ttTTTatttT aggatttag AAAATTGGA gttgggTTT gggatttATA	6600
tttaatattt tattttgtt ggagaagtAA ggtattattt atatttatAT gataaATgAA	6660
aggatattttt gattttgtt tagTTTTTAT agagaggTTT ttgagatAGG ttAAAAGTTT	6720
ttatTTtagtg taaagatgag ttgtgtata tgTTTGTtT ttgttttta tgTatATgtt	6780
tattaaatAT gtatTAAGTT tttaatATgt tTGatATgt aattttgtA ttgttagataA	6840
tttttttatt tgtaaaatgt gagtaataAT aatattttt ttTTGGGTTT gttttaaAGA	6900
ttaaaataaa aatgtatggT ttaatggtag tggtttggg gggattttt atataaataA	6960
gtgaatggag gatGAATAA ataaatataA aagatgtgt gtatTTTAT ttatATatAT	7020
aattaaaaat agttaaAGAT gtataaATTA aagttaaATg tatgtgcatt tgaagtatAT	7080
gttgattattt gataatgaaA tatagtataA ttTTTAattt tataTTTAA tattttatAT	7140
ttaataattt atattttat ttttagTTA tttaagatAT atagatATAG atagaATgtt	7200
ttaaatgtat tatttttta gtttaatgt ataaatatttA tatgaaaaAG tatttttagA	7260
tgtgtgtAA atgttattAGA atattatttA taggttaatA gatattgata atgttttagA	7320
gttggTTTTT ttatttttgg ataaattttt aaatttgagAG taattaaATT tgTTTTTTT	7380
ttatTTTTT ttataatggT aaataatataA taaataatgA gttgtgattt aaAGAAataA	7440
ttttataatg taaaAGTTTTT ttatgtttt attgattttt atgttttAA ttatTTTgg	7500
tagTTtaatg ttgtgatttt agtggTgagg AAAGTTTAtt ttatTTtagA gtagTggggt	7560
ttagttatgA ttTTTatttA taatttatttG gataattAAA gaatattgat gtagagTTTg	7620
tatTTAAAGA ttggaaATTt ttAgAAgtgA tataAAggtA ttggTatATA tataTTTTA	7680
aagtTTTTT ggagaaatAT atagttatgA ttggaaATTA ttgtttggg gaaAGTgatt	7740
atTTTTTAT tatttaaATAA gttaaGTTT aggaggtaAA atatttatAT ttTTTTTAT	7800
taaaattttgtt aaaatataA ttTTTattAAA 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: 0002044:2462022

<400> SEQUENCE: 27

aatctttt 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: 0002064:1224019

<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

tttgtttag gtttggaaatgt gt 22

<210> SEQ ID NO 31  
<211> LENGTH: 22  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: 0002383:1529022

<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

cttctcttctt cccccctc 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

aggttatcg t cgtgcgagt 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

tcaactcaaac gtatccaaa 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

tgggatataag 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

tttttttttc ggacgtcggtt 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 cggcgcaat 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 ttttaacgt ttttc 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  
aaataaaaata 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 tattgttag 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
 <pre>&lt;210&gt; SEQ ID NO 46 &lt;211&gt; LENGTH: 23 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: HM PRIMER CALCITONIN</pre>	
 <pre>&lt;400&gt; SEQUENCE: 46</pre>	
ggatgtgaga gttgttgagg tta	23
 <pre>&lt;210&gt; SEQ ID NO 47 &lt;211&gt; LENGTH: 24 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: HM PRIMER CALCITONIN</pre>	
 <pre>&lt;400&gt; SEQUENCE: 47</pre>	
acacacccaa acccattact atct	24
 <pre>&lt;210&gt; SEQ ID NO 48 &lt;211&gt; LENGTH: 21 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: TPEF FORWARD PRIMER</pre>	
 <pre>&lt;400&gt; SEQUENCE: 48</pre>	
ggacgtttt tatcgaaggc g	21
 <pre>&lt;210&gt; SEQ ID NO 49 &lt;211&gt; LENGTH: 15 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: TPEF BACKWARD PRIMER</pre>	
 <pre>&lt;400&gt; SEQUENCE: 49</pre>	
gccaccccaa cgcga	15
 <pre>&lt;210&gt; SEQ ID NO 50 &lt;211&gt; LENGTH: 18 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: 0002044:2183A188</pre>	
 <pre>&lt;400&gt; SEQUENCE: 50</pre>	
atgtgattcg tttgggta	18
 <pre>&lt;210&gt; SEQ ID NO 51 &lt;211&gt; LENGTH: 18 &lt;212&gt; TYPE: DNA &lt;213&gt; ORGANISM: Artificial Sequence &lt;220&gt; FEATURE: &lt;223&gt; OTHER INFORMATION: 0002044:2183B188</pre>	
 <pre>&lt;400&gt; SEQUENCE: 51</pre>	
atgtgatttg tttgggta	18
 <pre>&lt;210&gt; SEQ ID NO 52</pre>	

---

-continued

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

<400> SEQUENCE: 52
ggtaacgtc 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
ggtaatgtt 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
aaaatttgt 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
tatataacg tgtggta 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 tgtggta 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 tagaagg 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 tagaagg 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 ttgttg 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  
aagtaagttt ttgttg 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

---

gaagtggctcg ttagttt 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

gaagtggcttg 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

ttgttttagcg 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

ttgttttagtg 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

aatgttttg ttagtgg

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 ggtaggag

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

ataggttatc ggtaggag

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 gaacgatcgca atccca 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 aatatccccga ataacgccc 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 aaataactcac 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

tgttgagggtt atgtgttaatt 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 cactcaaatac 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 gtgggg 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 gtttagttt 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 gtttagttt 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 ggaggttag 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
aaccataaaa 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 aacacacaaat 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	
ctttattacaa tttaaaaaaaaaa 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	
aaaaaaaaaaaaaaa aaaaaaaaaactc ctctacataac	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
aactccctcta catacaccac aaataaaatt 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
gttagtttagt 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
cctccacaaaa actcactcct cacaataac 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 acgggggtga

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 gattttttt 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 adenine 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 amplicates, or is not amplified;

d) determining, based on the presence or absence of, or on a property of said amplicate, 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 amplicates 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 from 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 amplicate nucleic acid molecule carrying a detectable labels; and combinations thereof.

**22.** The method of claim 21, wherein the detectable amplicate label is selected from the label group consisting of: fluorescent labels; radionuclides or radiolabels; amplicate mass labels detectable in a mass spectrometer; detachable amplicate fragment mass labels detectable in a mass spectrometer; amplicate, and detachable amplicate 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 amplicate, or detachable amplicate 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 amplicates 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.

\* \* \* \* \*