



US009988683B2

(12) **United States Patent**
Adorjan et al.

(10) **Patent No.:** **US 9,988,683 B2**
(45) **Date of Patent:** ***Jun. 5, 2018**

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

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

(73) Assignee: **EPIGENOMICS AG**, Berlin (DE)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 717 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **12/794,449**

(22) Filed: **Jun. 4, 2010**

(65) **Prior Publication Data**

US 2011/0046005 A1 Feb. 24, 2011

Related U.S. Application Data

(63) Continuation of application No. 10/506,089, filed as application No. PCT/EP03/02034 on Feb. 27, 2003, now abandoned.

(30) **Foreign Application Priority Data**

Feb. 27, 2002 (EP) 02004551

(51) **Int. Cl.**
C12Q 1/68 (2018.01)

(52) **U.S. Cl.**
CPC **C12Q 1/6886** (2013.01); **C12Q 2600/154** (2013.01); **Y10T 436/143333** (2015.01)

(58) **Field of Classification Search**

None

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,565,552 A 10/1996 Magda et al.
5,567,810 A 10/1996 Weis et al.
5,574,142 A 11/1996 Meyer, Jr. et al.
5,580,722 A 12/1996 Foulkes et al.
5,585,481 A 12/1996 Arnold, Jr. et al.
5,587,371 A 12/1996 Sessler et al.
5,597,696 A 1/1997 Linn et al.
5,786,146 A 7/1998 Herman
5,837,832 A 11/1998 Chee et al.
5,958,773 A 9/1999 Monia
6,017,704 A 1/2000 Herman et al.
6,265,171 B1 7/2001 Herman et al.
6,331,393 B1 12/2001 Laird et al.
7,118,868 B2 10/2006 Berlin

FOREIGN PATENT DOCUMENTS

EP 02004551.4 2/2002
WO WO 1991/008230 6/1991

WO	WO 1995/000669	1/1995
WO	WO 1995/015373	6/1995
WO	WO 1997/046705	12/1997
WO	1998/032849 A2	7/1998
WO	WO 1999/028498	6/1999
WO	WO 2000/026401	5/2000
WO	WO 2001/007647	2/2001
WO	WO 2001/009183	2/2001
WO	WO 2001/068911	9/2001
WO	WO 2001/068912	9/2001
WO	2002/000926 A2	1/2002
WO	WO 2002/000928	1/2002
WO	WO 2002/074079	9/2002
WO	WO 2002/088381	11/2002
WO	WO 2003/064700	8/2003

OTHER PUBLICATIONS

Ushijima (Nature Reviews. 2005. 5: 223-231).*

Benner et al (Trends in Genetics (2001) vol. 17, pp. 414-418).*

Grunau et al (Nucleic Acids Research (2001) vol. 29, e65).*

Walsh et al teaches (Genes & Development (1999) vol. 13, pp. 26-36).*

Cottrell S.E., "Molecular diagnostic applications of DNA methylation technology", *Clin. Biochem.*, 37(7):595-604 (2004).

Database Biosis [Online], Biosciences Information Service, Philadelphia, PA, US; Mar. 1998 (Mar. 1998), Ho Coty et al., "Hypermethylation of the versican gene promoter in colorectal tumors", Database accession No. PREV199800197539 XP002265914 & *Proceedings of the American Association for Cancer Research Annual*, vol. 39, Mar. 1998 (Mar. 1998), pp. 539-540, 89th Annual Meeting of the American Association for Cancer Research, New Orleans, Louisiana, USA; Mar. 28-Apr. 1, 1998, Mar. 1998 ISSN: 0197-016X.

Ehrlich, M., "DNA methylation in cancer: too much, but also too little", *Oncogene*, 21(35):5400-5413 (2002).

Griffin and Smith, "Single-nucleotide polymorphism analysis by MALDI-TOF mass spectrometry", *Trends Biotechnol.*, 18(2):77-84 (2000).

Hiltunen et al., "Hypermethylation of the WT1 and Calcitonin Gene Promoter Regions at chromosome 11p in Human Colorectal Cancer", *British Journal of Cancer*, 76(9):1124-1130 (1997).

Kuismanen et al., "Epigenetic phenotypes distinguish microsatellite-stable and -unstable colorectal cancers", *PNAS U.S.A.*, 96(22):12661-12666 (1999).

Mills et al., "Increasing methylation of the calcitonin gene during disease progression in sequential samples from CML patients", *Leukemia Research*, 20(9):771-775 (1996).

NCBI Website. Genbank Accession No. AC026696, Dec. 12, 2001.

Olek and Walter, "The pre-implantation ontogeny of the H19 methylation imprint", *Nat. Genet.*, 17(3):275-276 (1997).

(Continued)

Primary Examiner — Katherine D Salmon

(74) *Attorney, Agent, or Firm* — Lathrop Gage LLP;
James H. Velema, Esq.; Sean M. Coughlin, Esq.

(57) **ABSTRACT**

Provided are 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. Further provided are novel nucleic acid sequences useful for the cell proliferative disorder specific analysis of said genes as well as methods, assays and kits thereof.

35 Claims, 23 Drawing Sheets

(56)

References Cited

OTHER PUBLICATIONS

- Olek et al., "A modified and improved method for bisulphite based cytosine methylation analysis", *Nucleic Acids Res.*, 24(24):5064-5066 (1996).
- Pienkos et al., "Molybdenum cofactors from molybdoenzymes and in vitro reconstitution of nitrogenase and nitrate reductase", *Proc. Natl. Acad. Sci. U.S.A.*, 74(12):5468-71 (1977).
- Rein et al., "Identifying 5-methylcytosine and related modifications in DNA genomes", *Nucl. Acids Res.*, 26(10):2255-2264 (1998).
- Sanger et al., "DNA sequencing with chain-terminating inhibitors", *Proc. Natl. Acad. Sci. U.S.A.*, 74(12):5463-5467 (1977).
- Silverman et al., "Abnormal methylation of the calcitonin gene in human colonic neoplasms", *Cancer Res.*, 49(13):3468-3473 (1989).
- Toyota et al., "Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification", *Cancer Research*, 59(10): 2307-2312 (1999).
- van der Krol et al., "Modulation of eukaryotic gene expression by complementary RNA or DNA sequences", *BioTechniques*, 6(10):958-976 (1988).
- Wayne et al., "Mutations in the transcriptional activator EYA4 cause late-onset deafness at the DFNA10 locus", *Hum. Mol. Genet.*, 10(3):195-200 (2001).
- Xiong and Laird, "COBRA: a sensitive and quantitative DNA methylation assay", *Nucleic Acids Res.*, 25(12):2532-2534 (1997).
- Yates et al., "Methylational urinalysis: a prospective study of bladder cancer patients and age stratified benign controls", *Oncogene*, 25:1984-1988 (2006).
- Yu et al., "Specific inhibition of PCR by non-extendable oligonucleotides using a 5' to 3' exonuclease-deficient DNA polymerase", *Bio Techniques*, 23(4):714-720 (1997).
- Zeschnigk et al., "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.*, 5(2):94-98 (1997).
- Zeschnigk et al., "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.*, 6(3):387-395 (1997).
- Zon G., "Oligonucleotide analogues as potential chemotherapeutic agents", *Pharm. Res.*, 5(9):539-549 (1988).
- Adany et al. (1990) "Altered Methylation of Versican Proteoglycan Gene in Human Colon Carcinoma," *Biochemical and Biophysical Research Communications*. 171(3):1402-1413.
- Eads et al. (1999) "CpG Island Hypermethylation in Human Colorectal Tumors Is Not Associated with DNA Methyltransferase Overexpression," *Cancer Research*. 59:2302-2306.
- Eads et al. (2001) "Epigenetic Patterns in the Progression of Esophageal Adenocarcinoma," *Cancer Research*. 61:3410-3418.
- Feil et al. (1994) "Methylation analysis on individual chromosomes: improved protocol for bisulphite genomic sequencing," *Nucleic Acids Research*. 22(4):695-696.
- Gonzalzo et al. (1997) "Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE)," *Nucleic Acids Research*. 25(12):2529-2531.
- Gonzalzo et al. (1997) "Identification and Characterization of Differentially Methylated Regions of Genomic DNA by Methylation-sensitive Arbitrarily Primed PCR," *Cancer Research*. 57:594-599.
- Grigg et al. (1994) "Sequencing 5-Methylcytosine Residues in Genomic DNA," *BioEssays*. 16(6):431-436.
- Gut et al. (1995) "DNA and Matrix Assisted Laser Desorption Ionization Mass Spectrometry," In: *Molecular Biology: Current Innovations and Future Trends*. Horizon Scientific Press. Wymondham, United Kingdom. pp. 147-157.
- Gut et al. (1995) "A procedure for selection DNA alkylation and detection by mass spectrometry," *Nucleic Acids Research*. 23(8):1367-1373.
- Heid et al. (1996) "Real Time Quantitative PCR," *Genome Research*. 6:986-994.
- Herman et al. (1996) "Methylation-specific PCR: A novel PCR assay for methylation status of CpG islands," *Proc. Natl. Acad. Sci. USA* 93:9821-9826.
- Issa et al. (2001) "Accelerated Age-related CpG Island Methylation in Ulcerative Colitis," *Cancer Research*. 61:3573-3577.
- Karas et al. (1988) "Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons," *Analytical Chemistry*. 60(20):2299-2301.
- Liang et al. (2000) "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.
- Martin et al. (1995) "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*. 157:261-265.
- Model et al. (2001) "Feature selection for DNA methylation based cancer classification," *Bioinformatics*. 17(Suppl 1): S157-S164.
- Ausubel (1995) *Hybridization Analysis of DNA Blots*, Unit 2.10 in: *Current Protocols in Molecular Biology*. John Wiley & Sons. New York, New York. pp. 2.10.1-2.10.16.
- Basset, Jr. et al. (1999) "Gene expression informatics—its all in your mine," *Nat. Genetics*. 21(Suppl 1):51-55.
- Bowtell (1999) "Options available—from start to finish—for obtaining expression data by microarray," *Nat. Genetics*. 21(Suppl 1):25-32.
- Brown et al. (1999) "Exploring the new world of the genome with DNA microarrays," *Nat. Genetics*. 21(Suppl 1):33-37.
- Chakravarti (1999) "Population genetics-making sense out of sequence," *Nat. Genetics*. 21(Suppl 1):56-60.
- Cheung et al. (1999) "Making and reading microarrays," *Nat. Genetics*. 21(Suppl 1):15-19.
- Cole et al. (1999) "The genetics of cancer—a 3D model," *Nat. Genetics*. 21(Suppl 1):38-41.
- Collins (1999) "Microarrays and macroconsequences," *Nat. Genetics*. 21(Suppl 1):2.
- Debouck et al. (1999) "DNA microarrays in drug discovery and development" *Nat. Genetics*. 21(Suppl 1):48-50.
- Duggan et al. (1999) "Expression profiling using cDNA microarrays," *Nat. Genetics*. 21(Suppl 1):10-14.
- Genbank Database [online] (Jun. 25, 2016) "*Homo sapiens* versican (VCAN), transcript variant 1, mRNA," Accession No. NM_004385. Accessible on the Internet at URL: https://www.ncbi.nlm.nih.gov/nuccore/NM_004385. [Last Accessed Oct. 6, 2016].
- Genbank Database [online] (Sep. 5, 2016) "*Homo sapiens* transmembrane protein with EGF like and two follistatin like domains 2 (TMEFF2), transcript variant 1, mRNA," Accession No. NM_016192. Accessible on the Internet at URL: https://www.ncbi.nlm.nih.gov/nuccore/NM_016192. [Last Accessed Oct. 6, 2016].
- Genbank Database [online] (Sep. 9, 2016) "*Homo sapiens* cadherin 13 (CDH13), transcript variant 1, mRNA," Accession No. NM_001257. Accessible on the Internet at URL: https://www.ncbi.nlm.nih.gov/nuccore/NM_001257. [Last Accessed Oct. 6, 2016].
- Hacia (1999) "Resequencing and mutational analysis using oligonucleotide microarrays," *Nat. Genetics*. 21(Suppl 1):42-47.
- Lander (1999) "Array of hope," *Nat. Genetics*. 21(Suppl 1):3-4.
- Lipshutz et al. (1999) "High density synthetic oligonucleotide arrays," *Nat. Genetics*. 21(Suppl 1):20-24.
- Phimister (1999) "Going global," *Nat. Genetics*. 21(Suppl 1):1.
- Southern et al. (1999) "Molecular interactions on microarrays," *Nat. Genetics*. 21(Suppl 1):5-9.
- International Preliminary Report on Patentability corresponding to International Patent Application No. PCT/EP2003/002034, completed Jun. 9, 2004.

* cited by examiner

FIGURE 1

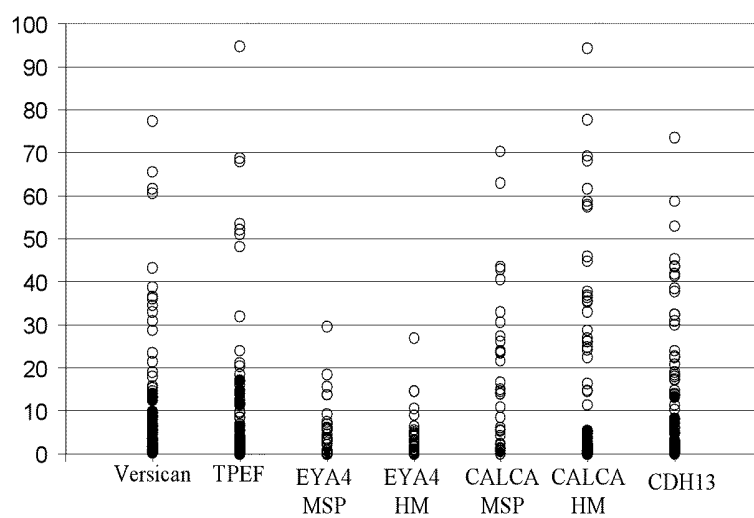


FIGURE 2

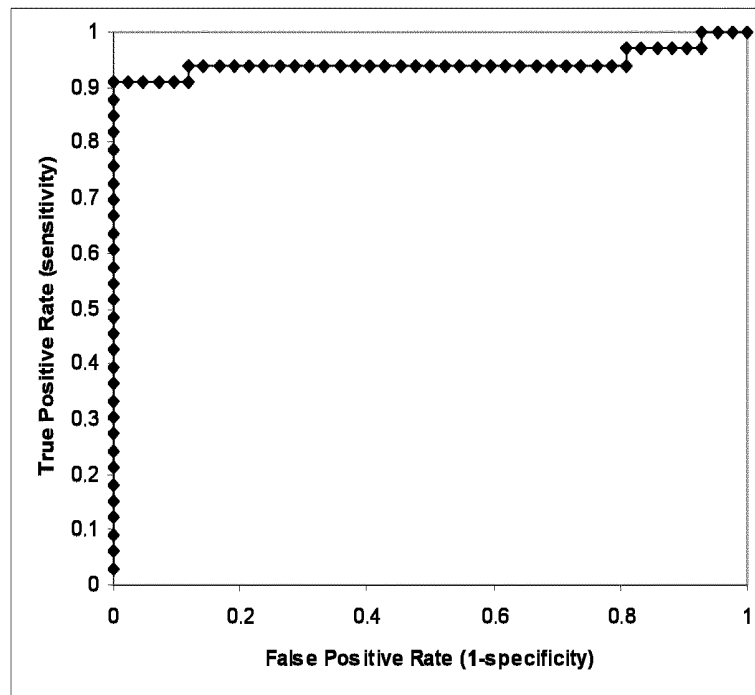


FIGURE 3

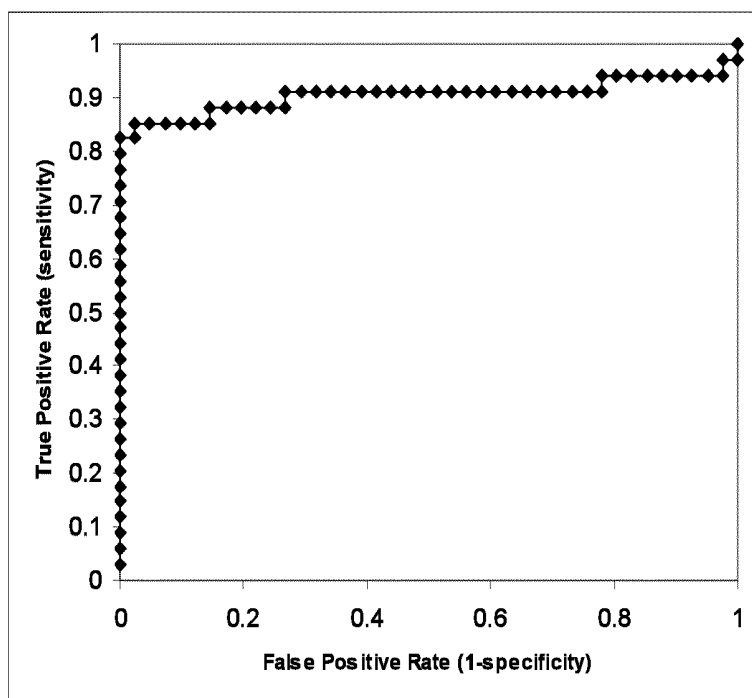


FIGURE 4

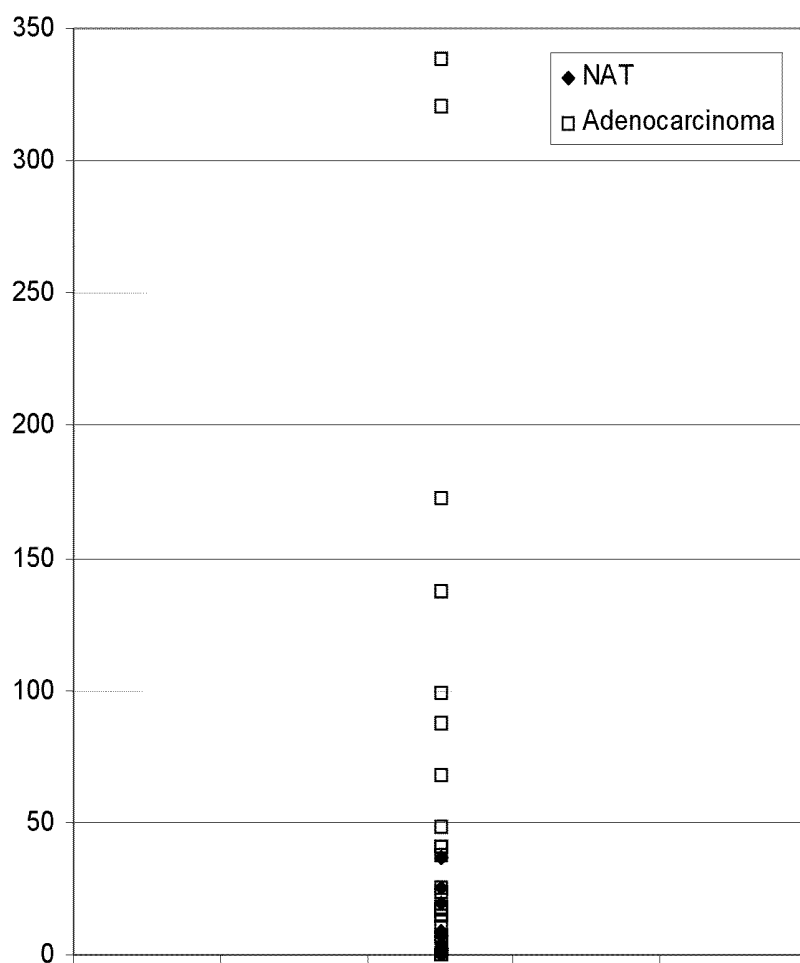


FIGURE 5

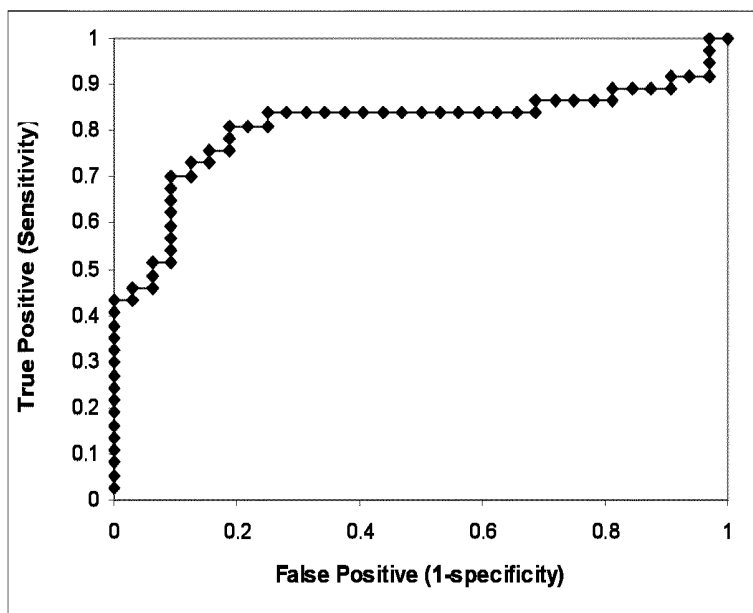


FIGURE 6

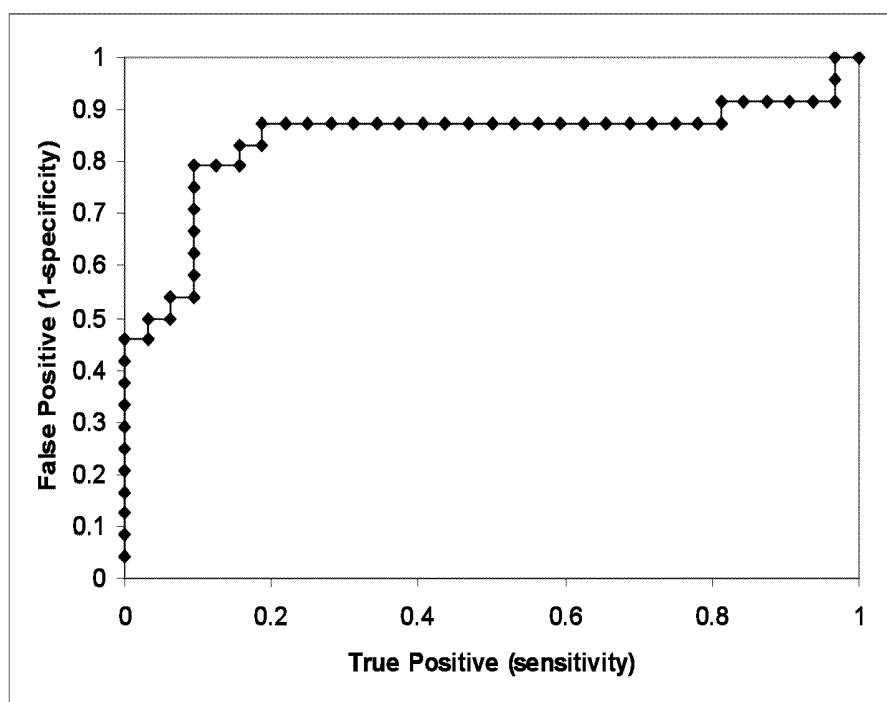


FIGURE 7

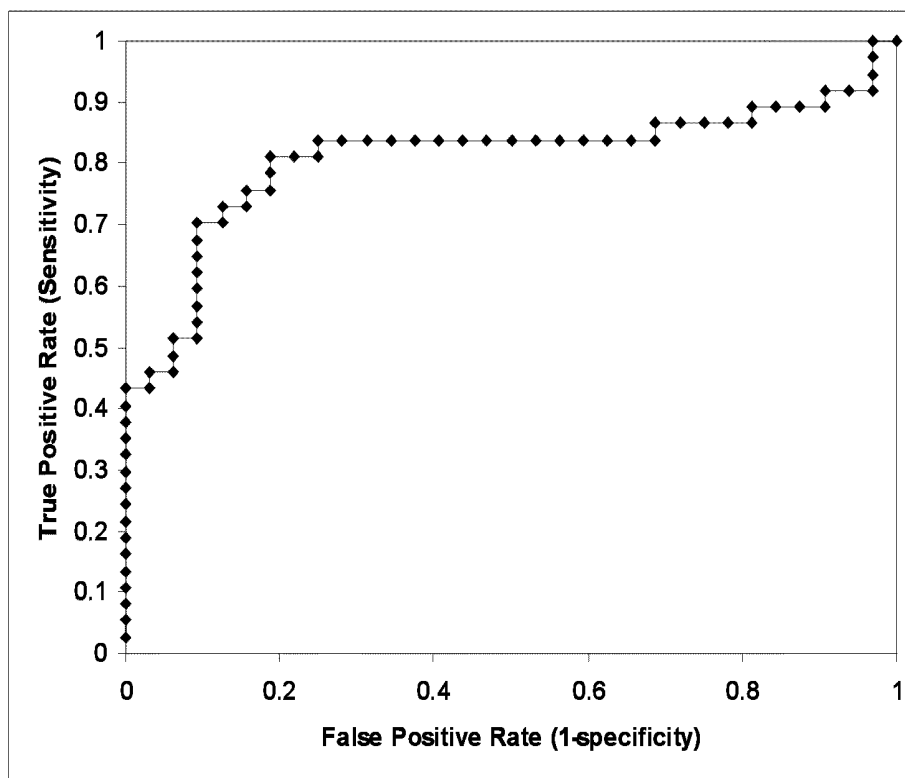


FIGURE 8

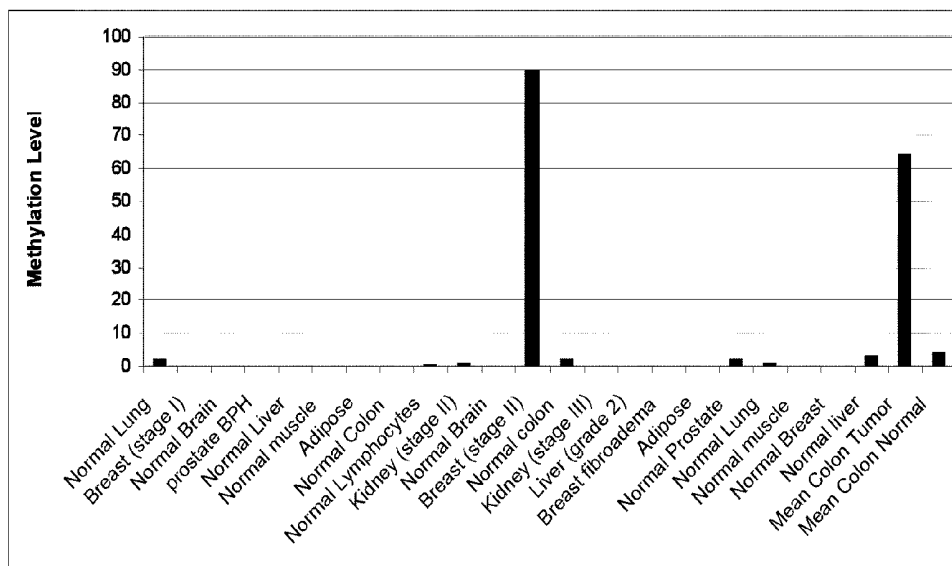


FIGURE 9

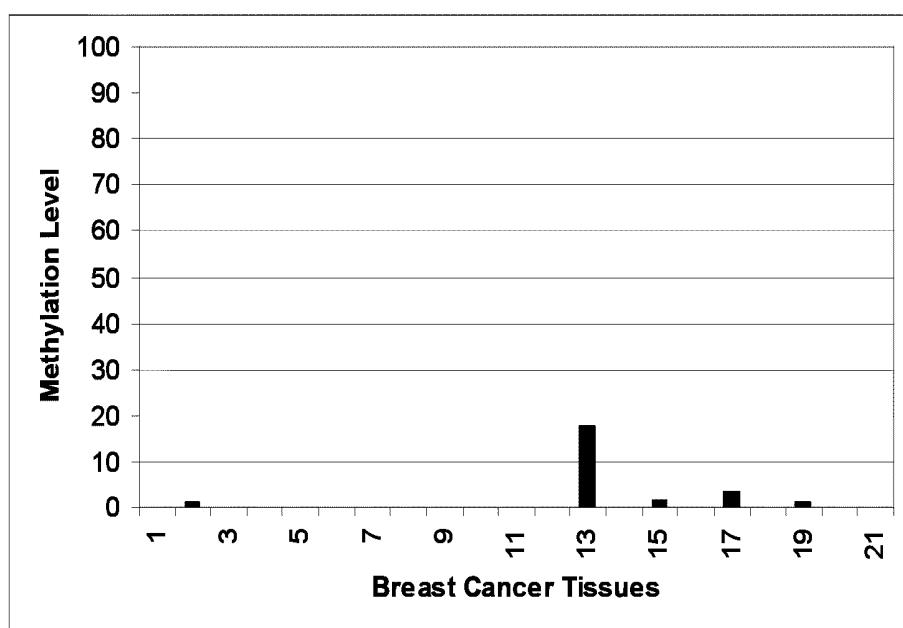


FIGURE 10

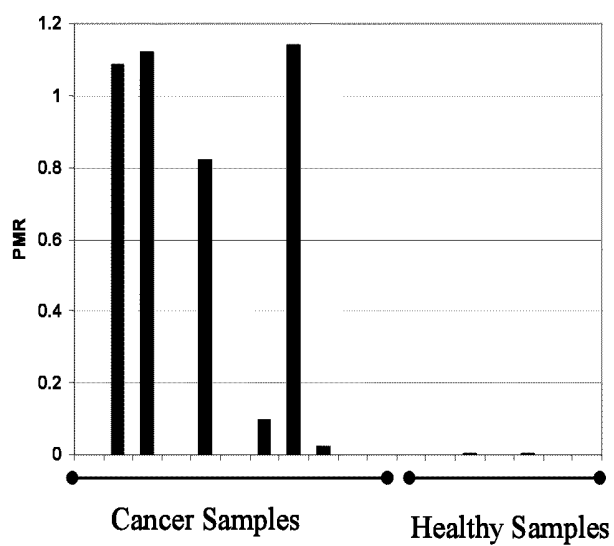


FIGURE 11

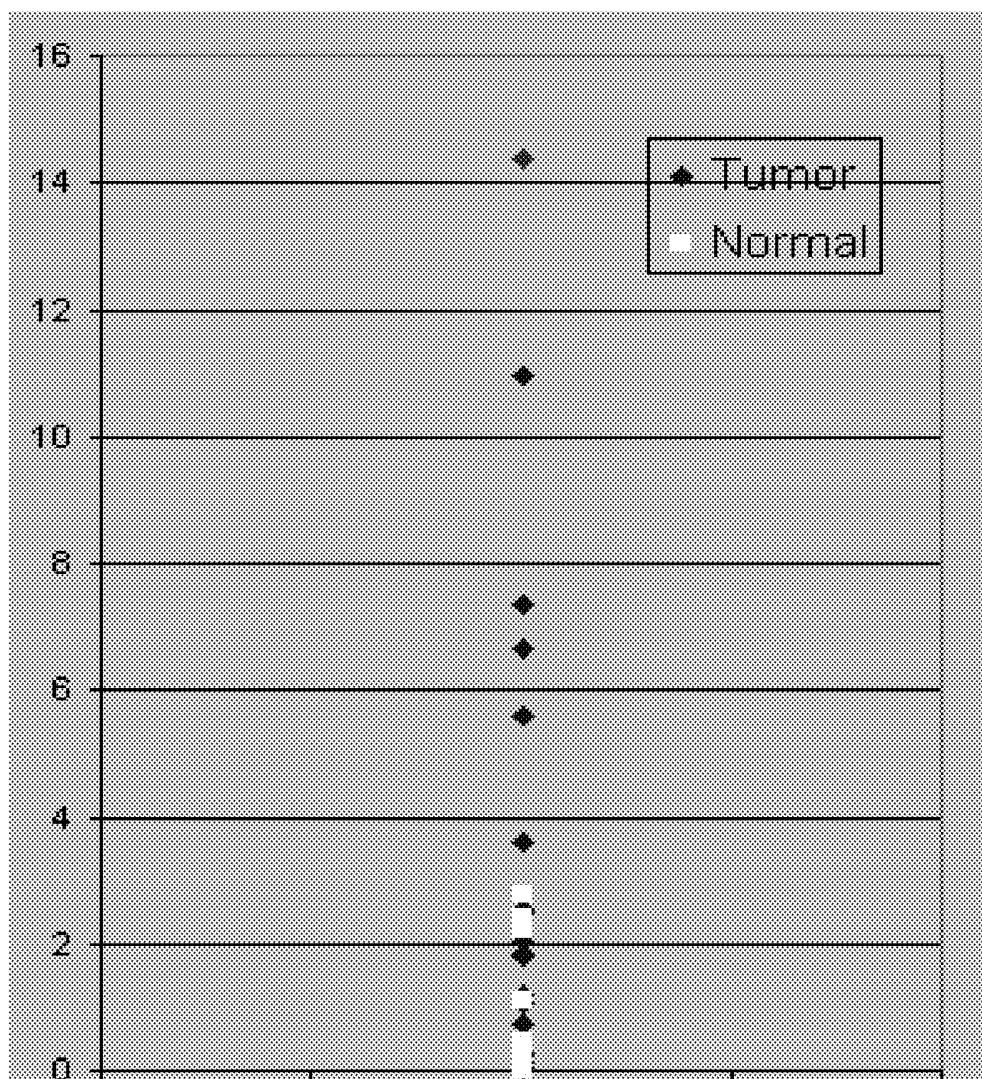


FIGURE 12

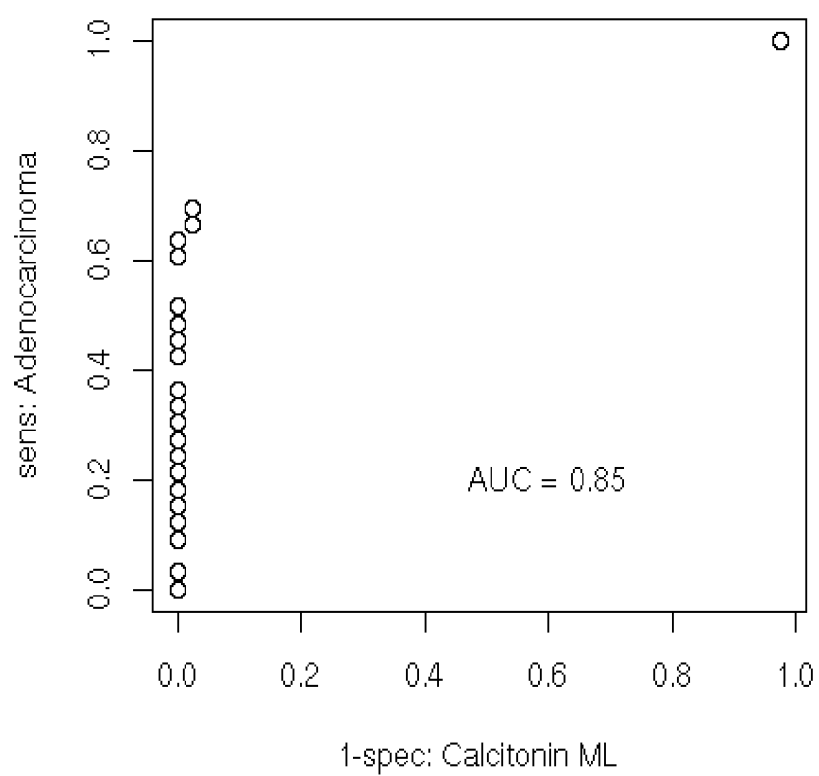


FIGURE 13

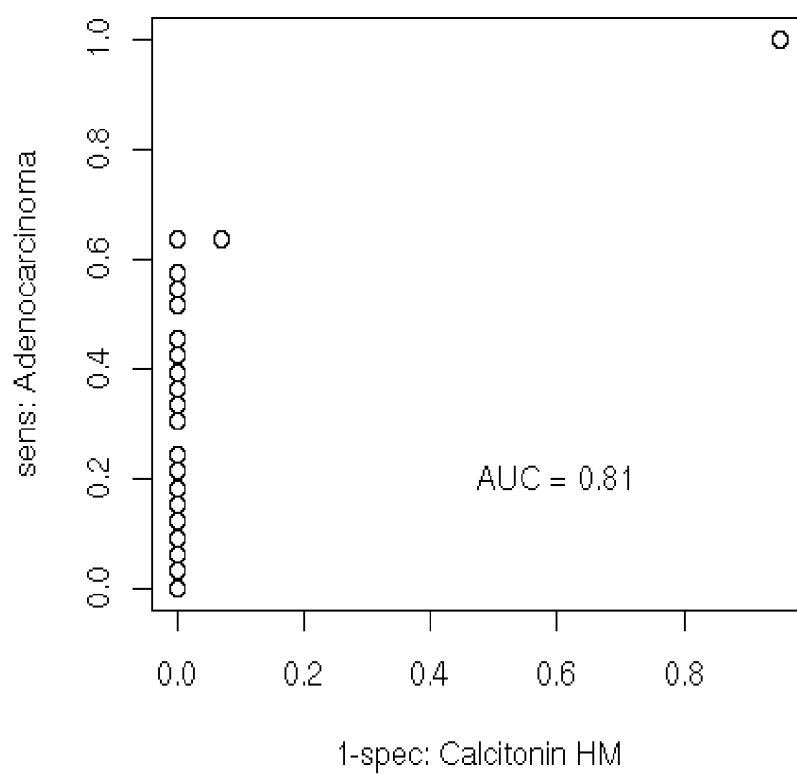


Figure 14

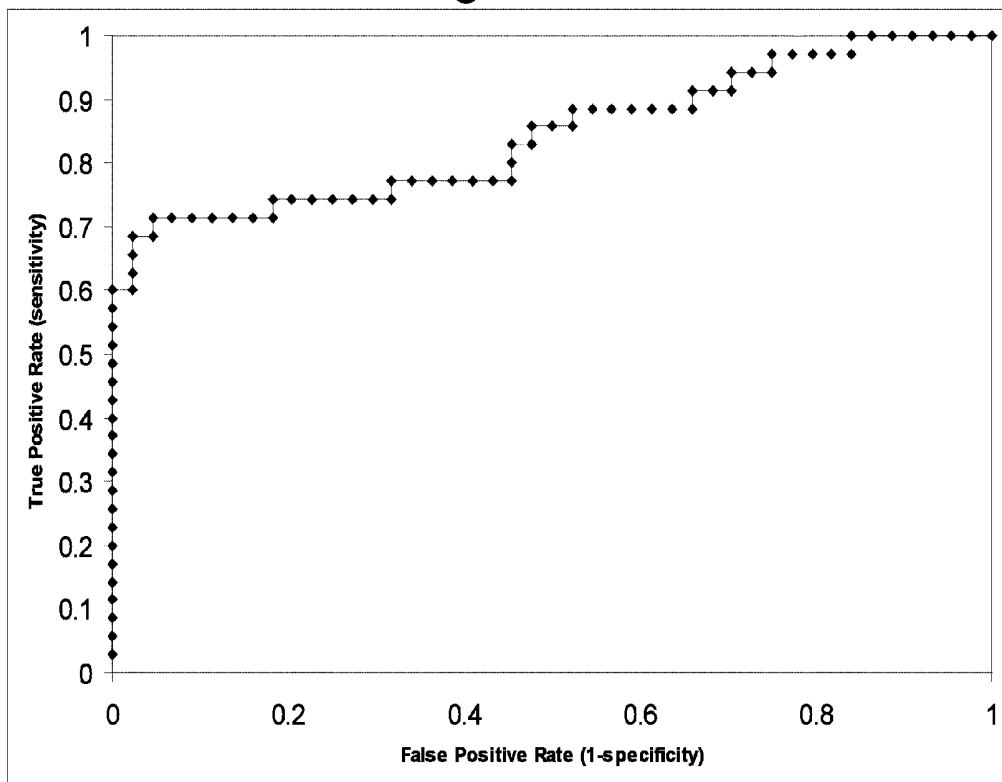


Figure 15

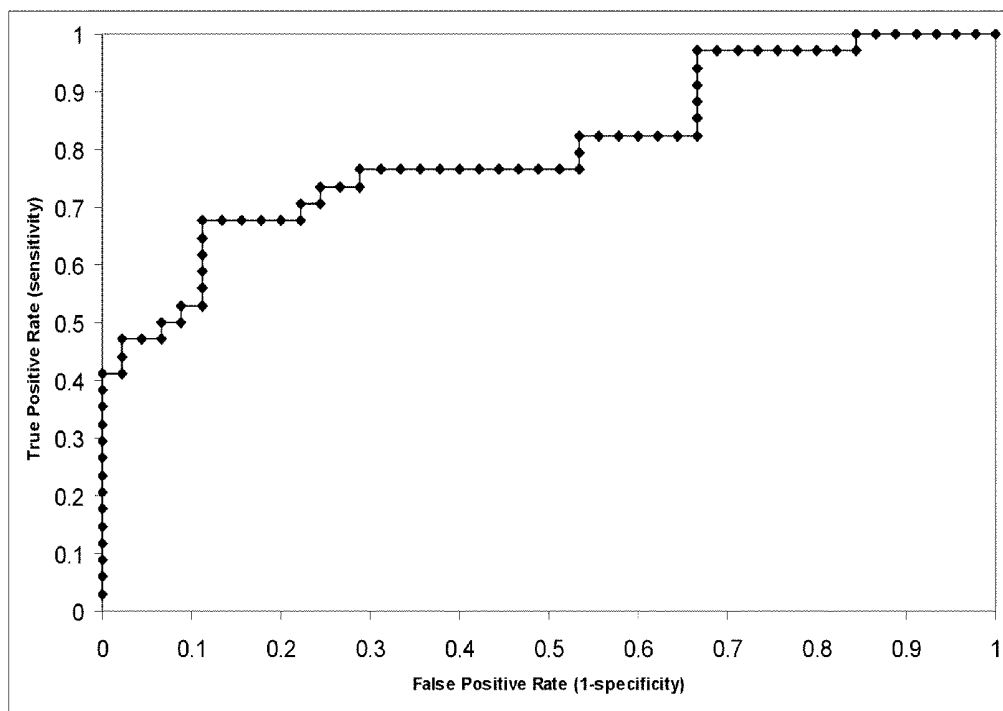


Figure 16

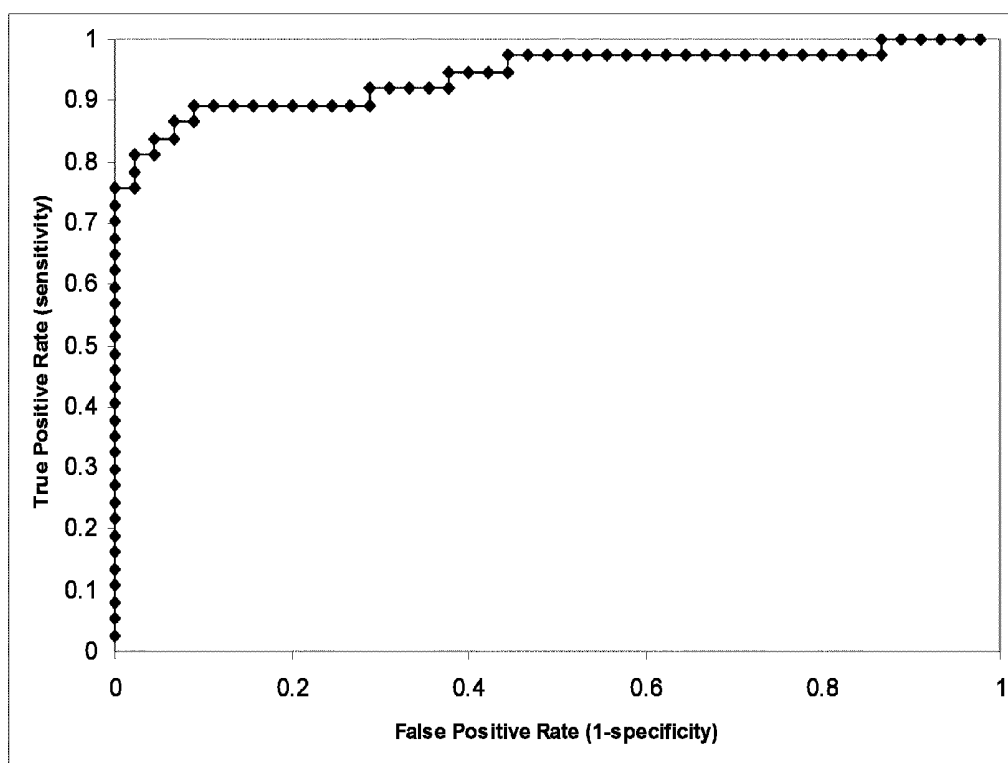


FIGURE 17

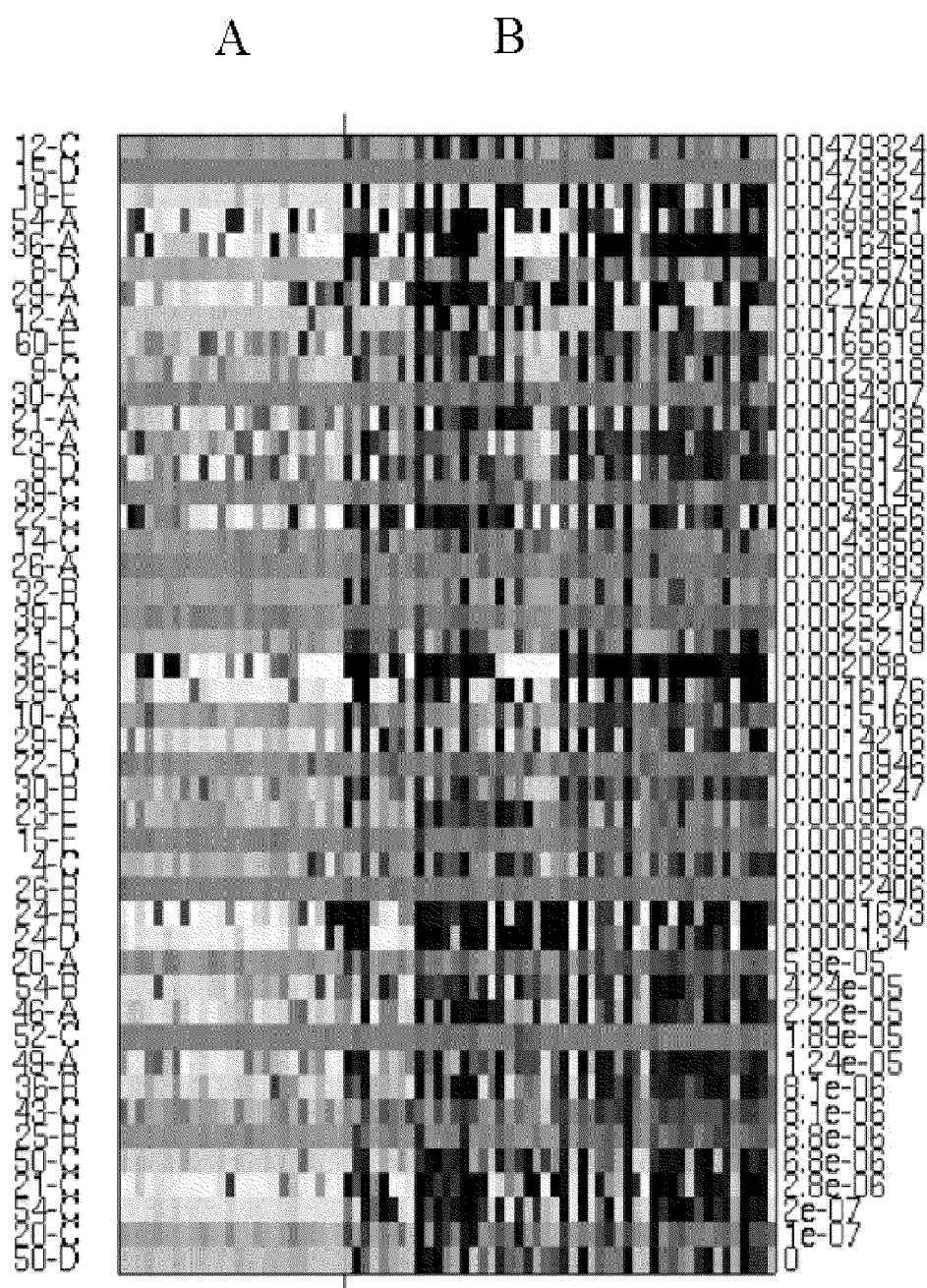


FIGURE 18

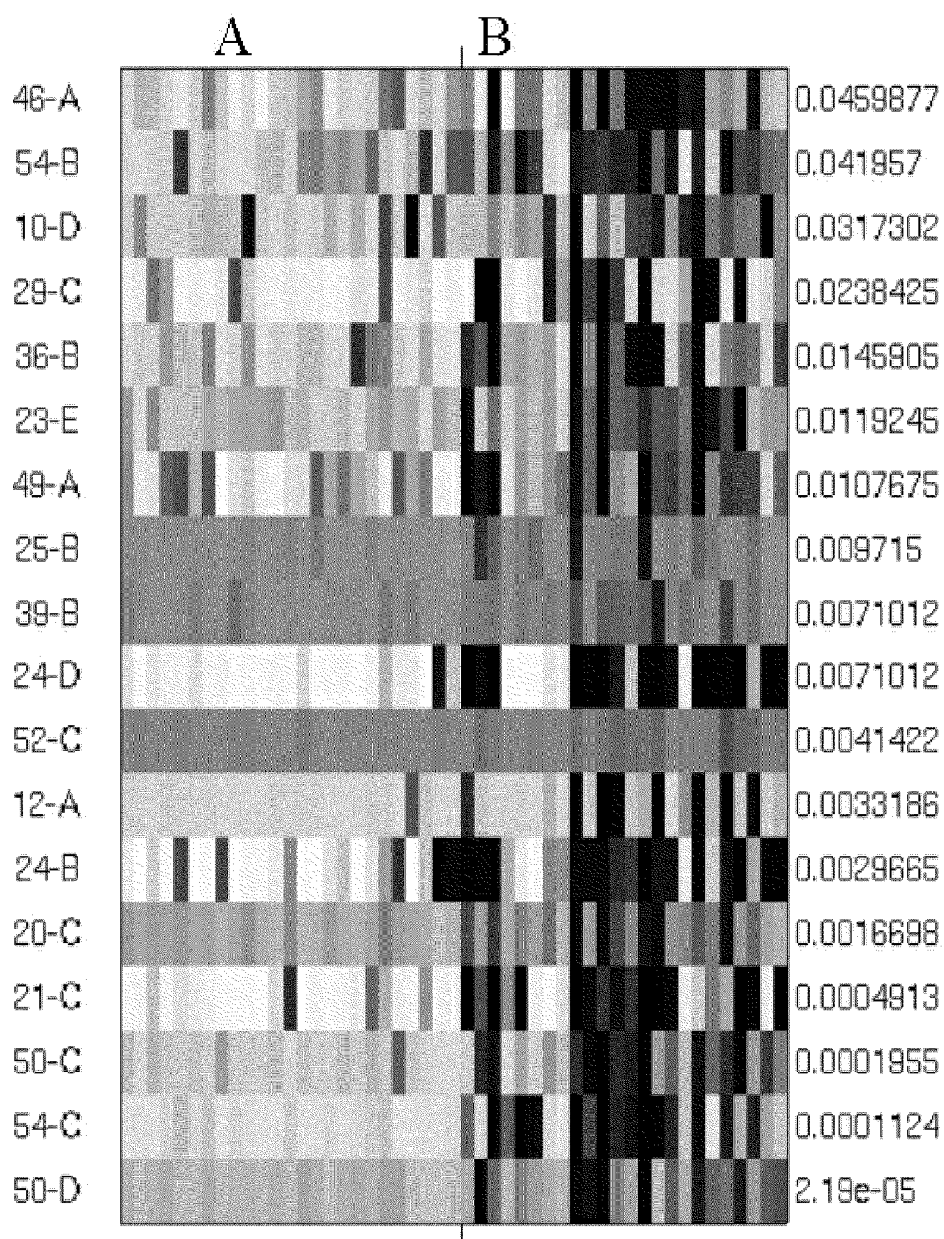


FIGURE 19

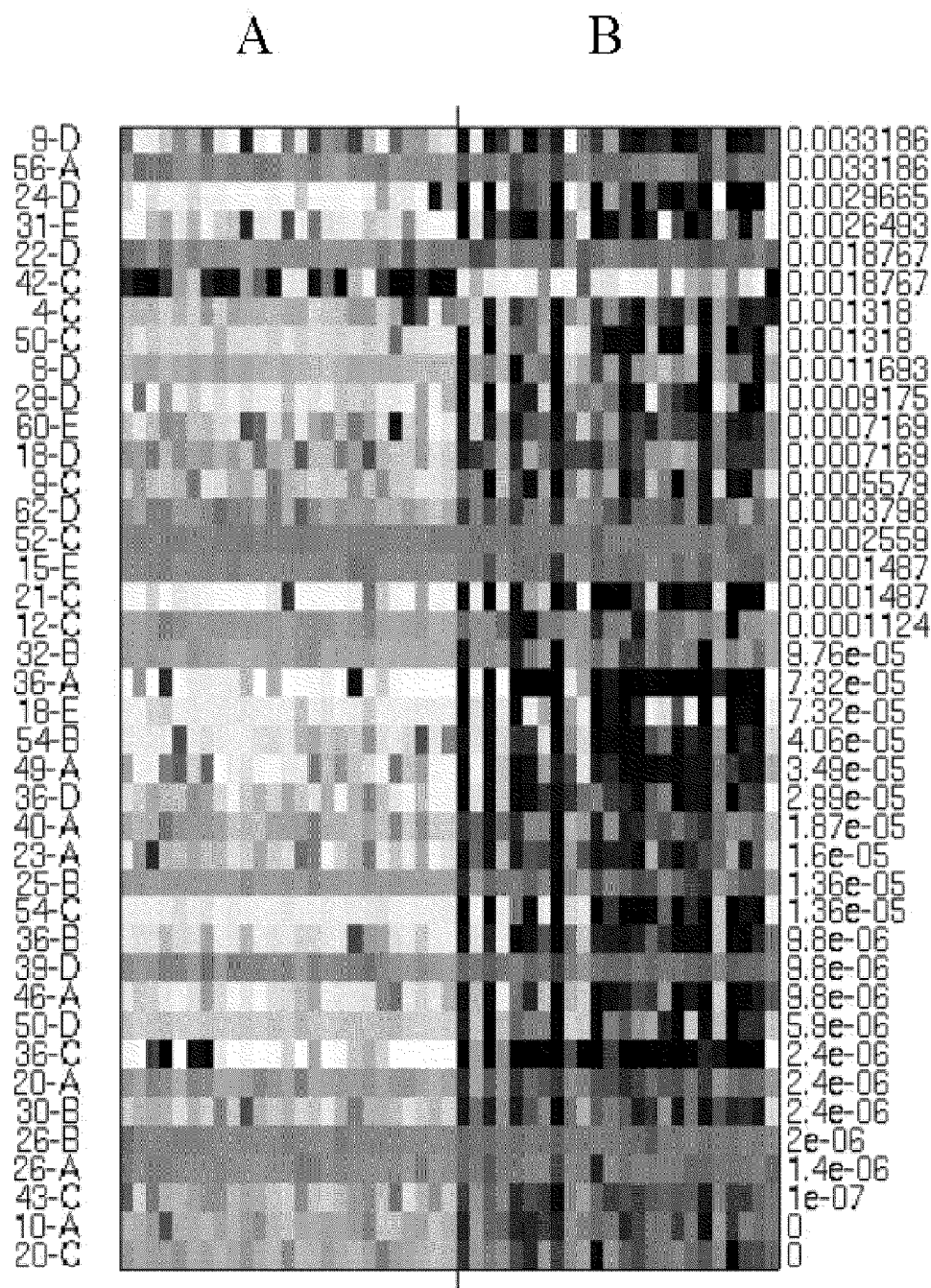


FIGURE 20

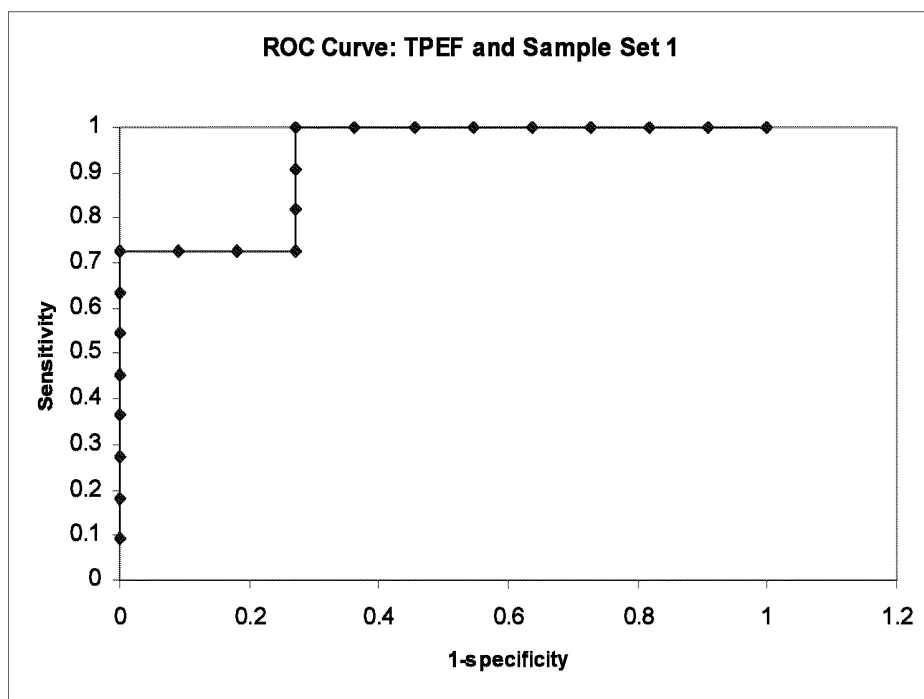


FIGURE 21

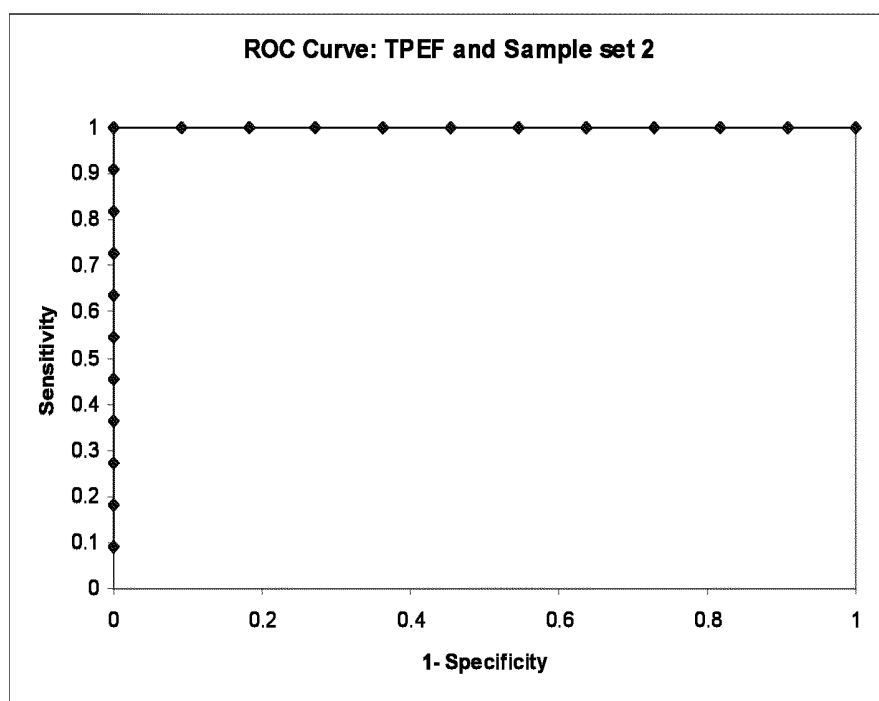


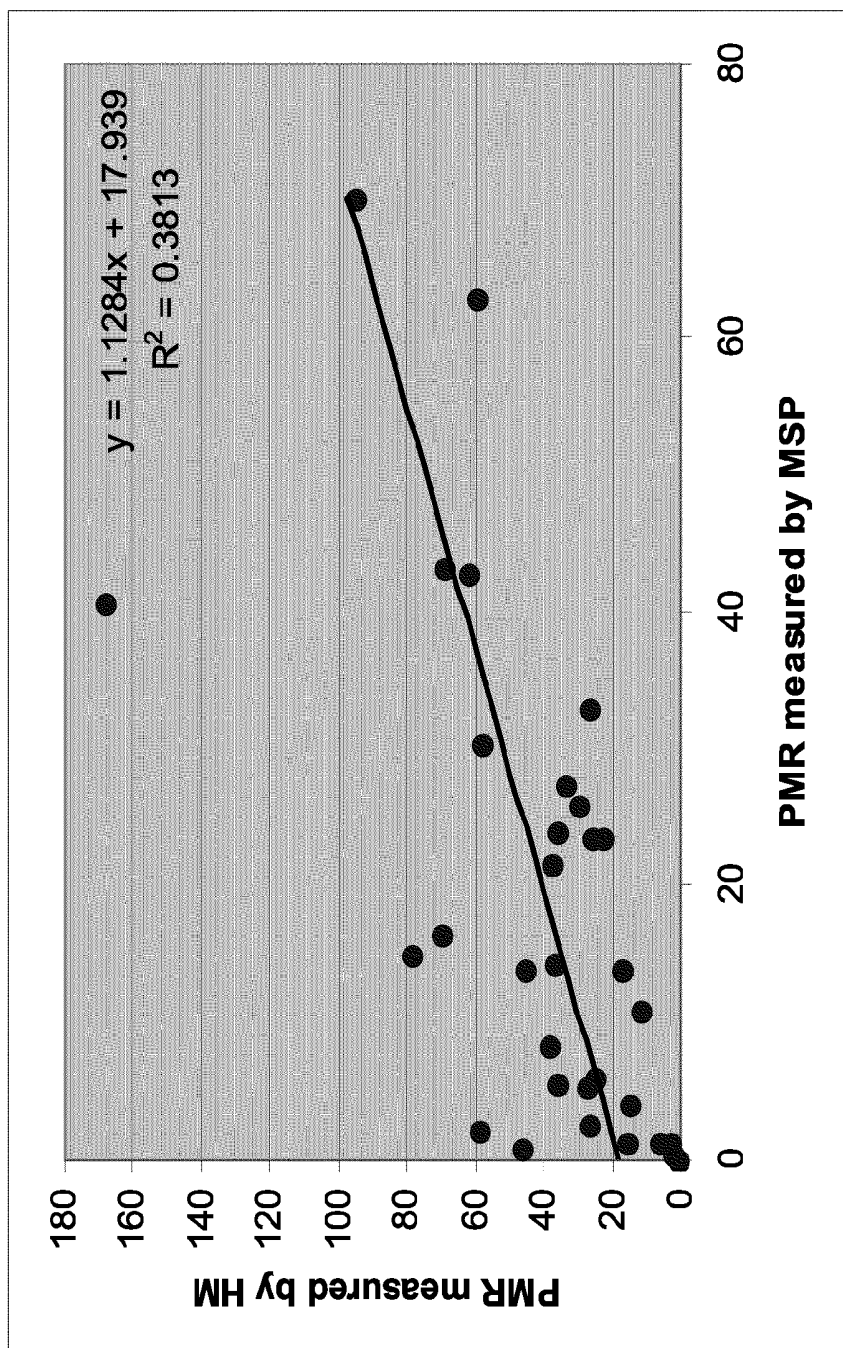
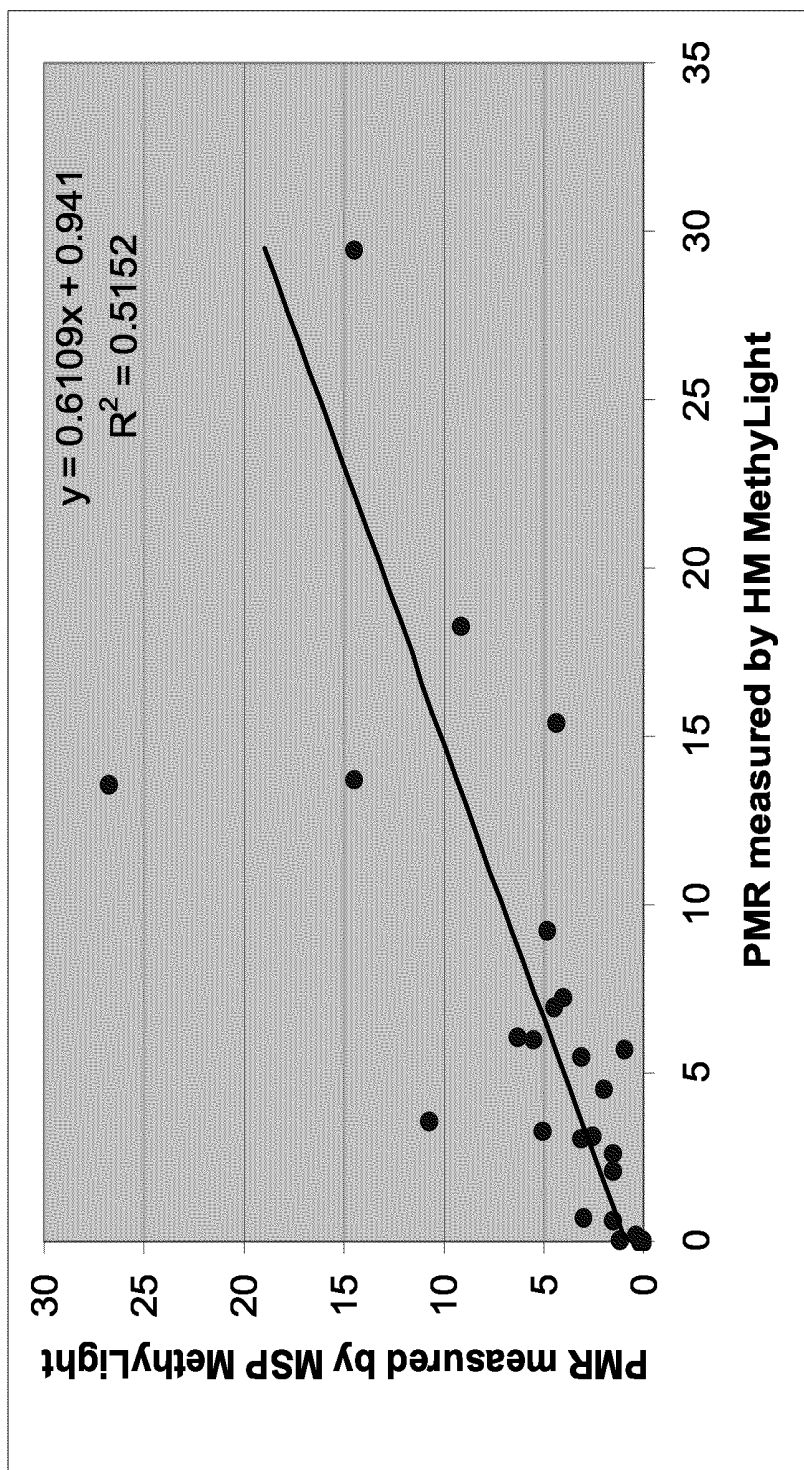
Figure 22

Figure 23

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

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation of U.S. patent application Ser. No. 10/506,089, which is the United States nationalization pursuant to 35 U.S.C. § 371 of International application number PCT/EP2003/002034, filed 27 Feb. 2003 and published 4 Sep. 2003 as WO 2003/072820, which claims priority to European Application EP 02004551.4, filed 27 Feb. 2002, both of which are incorporated by reference herein in their entireties.

FIELD OF THE INVENTION

Aspects of the invention relate generally to cellular proliferative disorders (e.g., cancer) and genomic DNA methylation, and more particularly to modified and genomic sequences, to oligonucleotides and/or PNA-oligomers for detecting the cytosine methylation state of genomic DNA, as well as to methods for ascertaining genetic and/or epigenetic parameters of genes for use in the differentiation, diagnosis, prognosis, treatment and/or monitoring of cellular proliferative disorders (e.g., cancer) including colon cell proliferative disorders (e.g., adenocarcinomas, polyps, squamous cell cancers, carcinoid tumours, sarcomas and lymphomas), or the predisposition to colon cell proliferative disorders.

BACKGROUND

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.

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.

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.

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.

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 micro satellite 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.

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.

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

The alpha-calcitonin gene encodes a small family of peptides: calcitonin, katacalcine, 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.

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:

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.

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.

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 co-methylation does not hold in the case of this gene.

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.

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.

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.

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.

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.

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.

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 (Gonzalzo 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).

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; Zeschnick 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.

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.

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.

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.

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 proportionally 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.

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.

DETAILED DESCRIPTION

The invention provides 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.

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.

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.

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.

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.

In one embodiment the method discloses the use of one or more genes selected from the group comprising Versican,

The sequences of SEQ ID NOS:6 to SEQ ID NO:25 provide modified versions of the nucleic acid according to SEQ ID NOS: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 NOS:1 to SEQ ID NO:5 as follows. (see also the following TABLE 1): SEQ ID NOS: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 NOS:6 to SEQ ID NO:15, converted SEQ ID NOS: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 NOS:16 to SEQ ID NO:25, converted SEQ ID NOS: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 NOS: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 NOS: 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 NOS: 6 to 15	Converted methylated strand	"C" to "T," but "CpG" remains "CpG" (all "C" residues of CpGs are methylated)
SEQ ID NOS: 16 to 25	Converted sense strand	"C" to "T" for all "C" residues (all "C" residues of CpGs are unmethylated)

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.

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 NOS:1 to SEQ ID NO:5 and sequences complementary thereto. SEQ ID NOS: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 NOS: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.

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 NOS:6 to SEQ ID NO:25 and sequences complementary thereto.

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

The described invention further disclose an oligonucleotide or oligomer for detecting the cytosine methylation state within pretreated DNA, according to SEQ ID NOS: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 NOS:6 to SEQ ID NO:25 and/or sequences complementary thereto.

Thus, the present invention includes nucleic acid molecules (e.g., oligonucleotides and peptide nucleic acid (PNA) molecules (PNA-oligomers)) that hybridise under moderately stringent and/or stringent hybridisation conditions to all or a portion of the sequences of SEQ ID NOS: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.

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 NOS:6 to SEQ ID NO:25, or to the complements thereof.

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 T_m , which is the temperature at which a probe dissociates from a target DNA. This melting temperature is used to define the required stringency conditions.

For target sequences that are related and substantially identical to the corresponding sequence of SEQ ID NOS: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 T_m , 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 T_m can be between 0.5° C. and 1.5° C. per 1% mismatch.

Examples of inventive oligonucleotides of length X (in nucleotides), as indicated by polynucleotide positions with reference to, e.g., SEQ ID NOS: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:

n to (n+(X-1));

where n=1, 2, 3, . . . (Y-(X-1));

where Y equals the length (nucleotides or base pairs) of SEQ ID NOS:1 to SEQ ID NO:5;

where X equals the common length (in nucleotides) of each oligonucleotide in the set (e.g., X=20 for a set of consecutively overlapping 20-mers); and

where the number (Z) of consecutively overlapping oligomers of length X for a given SEQ ID NO of length Y is equal to Y-(X-1). For example Z=2,785-19=2,766 for either sense or antisense sets of SEQ ID NOS:1 to SEQ ID NO:5, where X=20.

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

The present invention encompasses, for each of SEQ ID NOS: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.

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 NOS: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 NOS: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 NOS:11 to SEQ ID NO:15.

Particularly preferred oligonucleotides or oligomers according to the present invention are those in which the cytosine of the CpG dinucleotide (or the thymine of the TpG or the adenosine of the CpA dinucleotide) sequences is within the middle third of the oligonucleotide; that is, where

the oligonucleotide is, for example, 13 bases in length, the CpG, TpG or CpA dinucleotide is positioned within the fifth to ninth nucleotide from the 5'-end.

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

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.

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 NOS: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 NOS: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 NOS: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 NOS:6 to SEQ ID NO:25 and sequences complementary thereto) and genomic DNA (SEQ ID NOS: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 NOS:6 to SEQ ID NO:25, and sequences complementary thereto) and genomic DNA (SEQ ID NOS:1 to SEQ ID NO:5, and sequences complementary thereto).

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 NOS:6 to SEQ ID NO:25 and sequences complementary thereto, or segments thereof.

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.

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.

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.

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.

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.

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.

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.

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.

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.

In the third step of the method, fragments of the pretreated DNA are amplified. Wherein the source of the DNA is free DNA from serum, or DNA extracted from paraffin it is particularly preferred that the size of the amplificate fragment is between 100 and 200 base pairs in length, and wherein said DNA source is extracted from cellular sources (e.g. tissues, biopsies, cell lines) it is preferred that the amplificate is between 100 and 350 base pairs in length. It is particularly preferred that said 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.

In an alternate embodiment of the method, the methylation status of preselected CpG positions within the nucleic acid sequences comprising SEQ ID NOS: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 NOS: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 from the 3' end of the molecule. In addition to the CpG, TpG or CpA dinucleotides it is further preferred that said primers should further com-

prise several bisulfite converted bases (i.e. cytosine converted to thymine, or on the hybridizing strand, guanine converted to adenosine). In a further preferred embodiment said primers are designed so as to comprise no more than 2 cytosine or guanine bases.

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

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 amplicates 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).

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

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 subpopulation 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.

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 NOS: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 NOS:85 to SEQ ID NO:87, SEQ ID NOS:98, 103 and SEQ ID NO:108, and sequences complementary thereto.

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 derivatized at the 3' position with other than a "free" hydroxyl group. For example, 3'-O-acetyl oligonucleotides are representative of a preferred class of blocker molecule.

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

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

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 amplification product of said forward primer.

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.

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

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.

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.

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.

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 NOS:2 to SEQ ID NO:5; and the segment comprises at least one CpG, TpG or CpA dinucleotide.

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 NOS:1 to SEQ ID NO:5, and the equivalent positions within SEQ ID NOS: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, and SEQ ID NOS:88 and 89.

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).

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 NOS:78-84, 90, 99, 100, 104, 105, 109 and SEQ ID NO:110.

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 NOS:2 to SEQ ID NO:5.

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).

Additional embodiments of the invention provide a method for the analysis of the methylation status of genomic DNA according to the invention (SEQ ID NOS:1 to SEQ ID NO:5) without the need for pretreatment.

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.

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.

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.

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.

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.

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.

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 NOS: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.

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 NOS: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.

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 NOS:1 to SEQ ID NO:25 is used for manufacturing it, possibly together with suitable additives and ancillary agents.

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 NOS:1 to SEQ ID NO:5. Said kit may further comprise instructions for carrying out and evaluating the described method. In a further

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

Typical reagents (e.g., as might be found in a typical COBRA-based kit) for COBRA analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); restriction enzyme and appropriate buffer; gene-hybridisation oligo; control hybridisation oligo; kinase labelling kit for oligo probe; and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kits (e.g., precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

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

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.

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

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.

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."

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^MGG-3' (top strand): 3'-GGCC-5' (bottom strand)).

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

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

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

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

"Epigenetic parameters" are, in particular, cytosine 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.

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

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

The term "MS.AP-PCR" (Methylation-Sensitive Arbitrarily-Primed Polymerase Chain Reaction) refers to the art-recognized technology that allows for a global scan of the genome using CG-rich primers to focus on the regions most likely to contain CpG dinucleotides, and described by Gonzalgo et al., *Cancer Research* 57:594-599, 1997.

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

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

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

The term "MSP" (Methylation-specific PCR) refers to the art-recognized methylation assay described by Herman et al. *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996, and by U.S. Pat. No. 5,786,146.

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.

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.

"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.

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

BRIEF DESCRIPTION OF THE DRAWINGS

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.

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.

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.

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.

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.

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.

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.

FIG. 8 shows the level of methylation in different tumor and healthy tissues determined by a EYA4-HeavyMethyl MethylLight™ 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.

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

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

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.

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

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

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

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

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.

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.

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.

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.

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

FIG. 21 shows the ROC curve of a combined EYA4-Calcitonin-HeavyMethyl-MethylLight-Assay according to Example 6. The AUC is: 0.97.

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.

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

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

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.

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.

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: CGAAACCCTAAATATC-CCGAATAACGCCG (SEQ ID NO:81). The corresponding control assay was performed using the following primers and probes: Primer: TGGTGATGGAGGAGGTTTAG-TAAGT (SEQ ID NO:91); Primer: AACCAATAAAAC-CTACTCCTCCCTTAA (SEQ ID NO:92); and Probe: ACCACCACCAACACACAATAACAAACACA (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; 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).

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

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 cut-points of a diag-

nostic 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

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.

Methods.

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

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

(SEQ ID NO: 45)
Reverse Primer: CCCCTCAACCTAAAACTACAAC;

(SEQ ID NO: 87)
Forward Blocker: GTTATGGTTTGTGATTTGTGTGGG;

(SEQ ID NO: 86)
Reverse Blocker: AACTACAACCACTCAAATCAACCCA;
and

(SEQ ID NO: 84)
Probe: AAAATTACGACGACGCCACCCGAAA.

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

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);

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)).

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.

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

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.

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

The EYA4-HeavyMethyl-MethylLight-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

Twelve of the colon tissues analysed by real-time PCR also had paired serum taken before surgery. We extracted DNA from 1 ml of that serum using a Qiagen UltraSens DNA extraction kit, bisulfite treated the DNA sample, and ran the EYA4-HeavyMethyl-MethylLight-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

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: AGTTTATCGTCGTCGAGTGT; (SEQ ID NO: 34)

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

Probe: CGAATCTCTCGAACGATCGCATCCA. (SEQ ID NO: 78)

The corresponding control assay was performed as described above (Example 1).

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; 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)).

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

25

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

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).

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

(SEQ ID NO: 46)

Primer: GGATGTGAGAGTTGTTGAGGTTA;

(SEQ ID NO: 47)

Primer: ACACACCCAAACCCATTACTATCT;

(SEQ ID NO: 83)

Probe: ACCTCCGAATCTCTCGAACGATCGC;
and

(SEQ ID NO: 85)

Blocker: TGTGAGGTTATGTGTAATTGGGTGTGA.

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

Reaction Solution:

(300 nM primers; 450 nM probe; 3.5 mM magnesium chloride; 2 units of taq polymerase; 400 μ M dNTPs, 5 μ M blocker; and 7 μ l of DNA, in a final reaction volume of 20 μ l);

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)).

The corresponding control assay was performed as described above (Example 2).

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).

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.

26

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

Serum analysis with Calcitonin-HM MethylLight assay.

Twelve of the colon tissues analysed by real-time PCR also had paired serum taken before surgery. DNA was extracted from 1 ml of that serum using a Qiagen UltraSens® DNA extraction kit, the DNA sample was bisulfite treated, and the HeavyMethyl-MethylLight-assay was run 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 tumoursect methylation was detected in the serum of four of them. In contrast, no methylation was detected in any of the 11 serum samples taken from healthy donors.

Example 8

Identification of the methylation status of a CpG site within the Calcitonin Gene. A fragment of the upstream region of the calcitonin gene (SEQ ID NO:1) was amplified by PCR using the primers CCTTAGTCCCTACCTCTGCT (SEQ ID NO:94) and CTCATTTACACACACCCAAAC (SEQ ID NO:95). The resultant 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.

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).

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

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.

PCR products were detectable, with Nar I-hydrolyzed DNA isolated wherein the tissue in question contained unmethylated 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

Analysis of methylation within colon cancer using a Versican-MSP-MethylLight 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:

(SEQ ID NO: 36)
Forward Primer: TGGGATTAAGATTTTCGGTTAGTTTC;

(SEQ ID NO: 37)
Reverse Primer: CACTACAACGCTACGCGACTAAA;
and

(SEQ ID NO: 79)
Probe: TCGACGTTACCCAAACGAATCACATAAAAAAC

The corresponding control assay was performed as described above (Example 1). 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 units of taq polymerase; 200 µM dNTPs, 5 µM blocker; and 7 µl of DNA, in a final reaction volume of 20 µl);

Cycling Conditions:

95° C. for 10 minutes; (95° C. for 15 seconds, 60° C. for 1 minute) 50 cycles.

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

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.

This was further confirmed using a Versican-HeavyMethyl MethyLight assay, using dual LIGHTCYCLER™ probes.

Methods.

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

Forward Primer:	TGGATAGGAGTTGGGATTAAGATTTT	(SEQ ID NO: 96);
Reverse Primer:	CTTATTACAATTTAAAAAAAATTCCTACTACAA	(SEQ ID NO: 97);
Blocker:	AAATTCCTACTACAACACTACACAATAAATTCACATTAC	(SEQ ID NO: 98);
Probe:	TTTTTCGTATTTTTCGGGTTATTACGTTTT-Fluor	(SEQ ID NO: 99); and
Probe:	LC640-ATGTGATTGGTTGGGTAACTCGA-Phos	(SEQ ID NO: 100).

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

Reaction Conditions:

500 nM primers;

10 uM blocker; and

250 nM probes.

LightCycler FastStart Hybridization Probes Mix

4 mM Magnesium Chloride.

Cycling Profile:

95° C. denaturation for 10 minutes; and

50 cycles: 95° C. 10 seconds, 57° C. 30 seconds, 72° C. 20 seconds.

Example 10

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:

(SEQ ID NO: 38)
Forward Primer: TTTTTTTTCGGACGTCGTTG;

(SEQ ID NO: 39)
Reverse Primer: CCTCTACATACGCCGCGAAT;
and

(SEQ ID NO: 80)
Probe: AATTACCGAAAACATCGACCGA.

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 units of taq polymerase; 200 µM dNTPs, 5 µM blocker; and 7 µl of DNA, in a final reaction volume of 2 µl).

Cycling Conditions:

95° C. for 10 minutes; (95° C. for 15 seconds, 60° C. for 1 minute) 50 cycles.

The corresponding control assay was performed as described above (Example 1). The data was analysed using a PMR calculation previously described in the literature (Eads et al 2001).

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.

This was further confirmed using a TPEF-HeavyMethyl MethyLight assay (using dual labeled LIGHTCYCLER™ probes).

Methods.

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

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

(SEQ ID NO: 102)
Reverse Primer: TAAAAAATAAATAAATCTCTACATAC;

(SEQ ID NO: 103)
Blocker: AACTCTCTACATACACCACAAATAAAT;

(SEQ ID NO: 104)
Probe: CGAAACATCGACCGAACACG-Fluor;
and

-continued

(SEQ ID NO: 105)

Probe: LC640-GTCCGAAAAAAAAAACGAAGCTCC-Phos.

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

Reaction Conditions:

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

Cycling Conditions:

95° C. denaturation for 10 minutes; and
50 cycles: 95° C. for 10 seconds, 57° C. for 25 seconds, 72° C. for 10 seconds.

Example 11

Analysis of methylation within colon cancer using a TPEF-MSP-MethylLight 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.

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

(SEQ ID NO: 48)

Forward Primer: GGACGTTTTTTATCGAAGGCG;

(SEQ ID NO: 49)

Reverse Primer: GCCACCCAACCGCGA;
and

(SEQ ID NO: 90)

Probe: ACCCGAAATCACGCGCAAAAA.

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 units of taq polymerase; 200 μM dNTPs, 5 μM blocker; and 7 μl of DNA, in a final reaction volume of 20 μl).

Cycling Conditions:

95° C. for 10 minutes; (95° C. for 15 seconds, 60° C. for 1 minute) 50 cycles.

The corresponding control assay was performed as described above (Example 1).

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

Analysis of methylation within colon cancer using a H Cadherin-MSP-MethylLight Assay.

The colon cancer samples described in Example 1 were also analysed using a H Cadherin-MSP-MethylLight Assay.

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

(SEQ ID NO: 42)

Forward Primer: GACGGATTTTTTTTAACTTTTTC;

(SEQ ID NO: 43)

Reverse Primer: AAATAAAATACCACCTCCGCGA;
and

(SEQ ID NO: 82)

Probe: GCTCCTCGCGAAATACTCACCCCG

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 units of taq polymerase; 200 μM dNTPs, 5 μM blocker; and 7 μl of DNA, in a final reaction volume of 20 μl).

Cycling Conditions:

95° C. for 10 minutes; (95° C. for 15 seconds, 60° C. for 1 minute) 50 cycles.

The corresponding control assay was performed as described above (Example 1). The data was analysed using a PMR calculation previously described in the literature (Eads et al 2001).

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 ($p=0.0000000118$). The ROC curve of the assay is shown in FIG. 15. The AUC is 0.94.

This was further confirmed using a H Cadherin-HeavyMethyl MethylLight assay, using dual Lightcycler probes using Lightcycler style dual probe technology.

Methods.

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

(SEQ ID NO: 106)

Forward Primer: GTTAGTTAGTTAATTTTTAAATAGATTAGTAG;

(SEQ ID NO: 107)

Reverse Primer: CAAAAAACAATAAATACCACCTCC;

(SEQ ID NO: 108)

Blocker: CCTCCACAAAACCTCACTCCTCACAAAATAC;

(SEQ ID NO: 109)

Probe: red640 TTTCGTTTTGTATGGTAGATACGGGGTGA-

phosphate;
and

(SEQ ID NO: 110)

Probe: ATTAATGGTTTTATAAGACGGATTTTTTTTAACTG-

fluoresceine.

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

Reaction Conditions:

Forward primer: 600 nM;
Reverse primer: 300 nM;
Blocker: 10 uM;
Probes: 500 nM;
Taq polymerase: 0.1 U/ul;

31

dNTPs: 0.2 mM each;
 Magnesium Chloride: 4 mM;
 BSA: 0.25 mg/ml; and
 Roche buffer with no MgCl: 1×.
 Cycling Conditions:
 95° C. denaturation for 10 minutes; and
 50 cycles: 95° C. for 10 seconds, 57° C. for 25 seconds, 72°
 C. for 10 seconds.

Example 13

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.

Reaction Solution:

10 ng bisulfite treated DNA;
 3.5 mM MgCl₂;
 400 μM dNTPs;
 2 pmol each primer; and
 1 U Hot Star Taq (Qiagen).

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.

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.

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

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.

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

32

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.

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. Note that all rounds are performed without using any knowledge obtained in the previous runs. The number of correct classifications was averaged over all runs, which gives a good estimate of our test accuracy (percent of correct classified samples over all rounds).

Tables

TABLE 1

PCR primers and products		
Gene	Primers	Amplificate length
Versican (SEQ ID NO: 2)	GGATAGGAGTTGGGATTA AGAT AAATCTTTTCAACACCA AAAT	414
EYA4 (SEQ ID NO: 3)	GGAAGAGGTGATTAAATG GAT CCCAAAATCAACAACAA	226
H-Cadherin (SEQ ID NO: 4)	TTGTATTAGGTTGGAAGT GGT CCCAATAAATCAACAAC AACA	286
TPEF (SEQ ID NO: 5)	TTGTTGGGTTAATAAATG GA CTTCTCTCTTCTCCCCTCTC	295

TABLE 2

Hybridisation oligonucleotides	
Gene	Oligomer sequence
Versican (SEQ ID NO: 2)	AAGATTTTCGGTTAGTTT (SEQ ID NO: 88)
Versican (SEQ ID NO: 2)	AAGATTTTCGGTTAGTTT (SEQ ID NO: 89)

33

TABLE 2-continued

Hybridisation oligonucleotides	
Gene	Oligomer sequence
Versican (SEQ ID NO: 2)	ATGTGATTTCGTTGGGTA (SEQ ID NO: 50)
Versican (SEQ ID NO: 2)	ATGTGATTGTTTGGGTA (SEQ ID NO: 51)
Versican (SEQ ID NO: 2)	GGGTACGTCGAATTTAG (SEQ ID NO: 52)
Versican (SEQ ID NO: 2)	GGTAATGTTGAATTTAG (SEQ ID NO: 53)
Versican (SEQ ID NO: 2)	AAAAATTCGCGAGTTTAG (SEQ ID NO: 54)
Versican (SEQ ID NO: 2)	AAAAATTTGTGAGTTTAG (SEQ ID NO: 55)
EYA4 (SEQ ID NO: 3)	TATATATACGTGTGGGTA (SEQ ID NO: 56)
EYA4 (SEQ ID NO: 3)	TATATATATGTGTGGGTA (SEQ ID NO: 57)
EYA4 (SEQ ID NO: 3)	AGTGTATGCGTAGAAGGT (SEQ ID NO: 58)
EYA4 (SEQ ID NO: 3)	AGTGTATGTGTAGAAGGT (SEQ ID NO: 59)
EYA4 (SEQ ID NO: 3)	TTTAGATACGAAATGTTA (SEQ ID NO: 60)
EYA4 (SEQ ID NO: 3)	TTTAGATATGAAATGTTA (SEQ ID NO: 61)
EYA4 (SEQ ID NO: 3)	AAGTAAGTCGTTGTTGTT (SEQ ID NO: 62)
EYA4 (SEQ ID NO: 3)	AAGTAAGTTGTTGTTGTT (SEQ ID NO: 63)
H-Cadherin (SEQ ID NO: 4)	GAAGTGGTCGTTAGTTTT (SEQ ID NO: 64)
H-Cadherin (SEQ ID NO: 4)	GAAGTGGTTGTTAGTTTTT (SEQ ID NO: 65)
H-Cadherin (SEQ ID NO: 4)	TTGTTTAGCGTGATTGTT (SEQ ID NO: 66)
H-Cadherin (SEQ ID NO: 4)	TTGTTTAGTGTGATTGTT (SEQ ID NO: 67)
H-Cadherin (SEQ ID NO: 4)	AAGGAATTCGTTTGTAA (SEQ ID NO: 68)
H-Cadherin (SEQ ID NO: 4)	AAGGAATTTGTTTGTAA (SEQ ID NO: 69)
H-Cadherin (SEQ ID NO: 4)	AATGTTTTTCGTGATGTTG (SEQ ID NO: 70)
H-Cadherin (SEQ ID NO: 4)	AATGTTTTGTGATGTTG (SEQ ID NO: 71)
TPEF (SEQ ID NO: 5)	ATTTGTTTCGATTAATTT (SEQ ID NO: 72)
TPEF (SEQ ID NO: 5)	ATTTGTTTTCGATTAATTT (SEQ ID NO: 73)
TPEF (SEQ ID NO: 5)	ATAGGTTACGGTTGGAG (SEQ ID NO: 74)

34

TABLE 2-continued

Hybridisation oligonucleotides	
Gene	Oligomer sequence
TPEF (SEQ ID NO: 5)	ATAGGTTATGGTTGGAG (SEQ ID NO: 75)
TPEF (SEQ ID NO: 5)	AATTTGCGAACGTTTGGG
TPEF (SEQ ID NO: 5)	AATTTGTGAATGTTTGGG

TABLE 3

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

Gene	Oligo:
H-Cadherin (SEQ ID NO: 4)	AATGTTTTTCGTGATGTTG (SEQ ID NO: 70)
H-Cadherin (SEQ ID NO: 4)	AATGTTTTGTGATGTTG (SEQ ID NO: 71)
TPEF (SEQ ID NO: 5)	AATTTGCGAACGTTTGGG (SEQ ID NO: 76)
TPEF (SEQ ID NO: 5)	AATTTGTGAATGTTTGGG (SEQ ID NO: 77)
Versican (SEQ ID NO: 2)	GGGTACGTCGAATTTAG (SEQ ID NO: 52)
Versican (SEQ ID NO: 2)	GGTAATGTTGAATTTAG (SEQ ID NO: 53)
H-Cadherin (SEQ ID NO: 4)	AAGGAATTCGTTTGTAA (SEQ ID NO: 68)
H-Cadherin (SEQ ID NO: 4)	AAGGAATTTGTTTGTAA (SEQ ID NO: 69)
TPEF (SEQ ID NO: 5)	ATAGGTTACGGTTGGAG (SEQ ID NO: 74)
TPEF (SEQ ID NO: 5)	ATAGGTTATGGTTGGAG (SEQ ID NO: 75)
EYA4 (SEQ ID NO: 3)	AAGTAAGTCGTTGTTGTT (SEQ ID NO: 62)
EYA4 (SEQ ID NO: 3)	AAGTAAGTTGTTGTTGTT (SEQ ID NO: 63)
EYA4 (SEQ ID NO: 3)	AGTGTATGCGTAGAAGGT (SEQ ID NO: 58)
EYA4 (SEQ ID NO: 3)	AGTGTATGTGTAGAAGGT (SEQ ID NO: 59)
Versican (SEQ ID NO: 2)	AAAAATTCGCGAGTTTAG (SEQ ID NO: 54)
Versican (SEQ ID NO: 2)	AAAAATTTGTGAGTTTAG (SEQ ID NO: 55)
Versican (SEQ ID NO: 2)	AAGATTTTCGGTTAGTTT (SEQ ID NO: 88)
Versican (SEQ ID NO: 2)	AAGATTTTGGTTAGTTT (SEQ ID NO: 89)
TPEF (SEQ ID NO: 5)	ATTTGTTTCGATTAATTT (SEQ ID NO: 72)

35

TABLE 3-continued

Oligonucleotides used in differentiation between colon adenomas or carcinoma tissue and healthy colon tissue.	
Gene	Oligo:
TPEF (SEQ ID NO: 5)	ATTGTTTTGATTAATT (SEQ ID NO: 73)

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)	AATGTTTTGTGATGTTG (SEQ ID NO: 71)
TPEF (SEQ ID NO: 5)	AATTTGCGAACGTTTGGG (SEQ ID NO: 76)
TPEF (SEQ ID NO: 5)	AATTTGTGAATGTTTGGG (SEQ ID NO: 77)
H-Cadherin (SEQ ID NO: 4)	AAGGAATTCGTTTGTAA (SEQ ID NO: 68)
H-Cadherin (SEQ ID NO: 4)	AAGGAATTTGTTTGTAA (SEQ ID NO: 69)
Versican (SEQ ID NO: 2)	GGGTAACGTCGAATTTAG (SEQ ID NO: 52)
Versican (SEQ ID NO: 2)	GGGTAATGTTGAATTTAG (SEQ ID NO: 53)
EYA4 (SEQ ID NO: 3)	AGTGTATGCGTAGAAGGT (SEQ ID NO: 58)
EYA4 (SEQ ID NO: 3)	AGTGTATGTGTAGAAGGT (SEQ ID NO: 59)
EYA4 (SEQ ID NO: 3)	AAGTAAGTCGTTGTTGTT (SEQ ID NO: 62)
EYA4 (SEQ ID NO: 3)	AAGTAAGTTGTTGTTGTT (SEQ ID NO: 63)
TPEF (SEQ ID NO: 5)	ATAGGTTACGGTTGGAG (SEQ ID NO: 74)
TPEF (SEQ ID NO: 5)	ATAGGTTATGGTTGGAG (SEQ ID NO: 75)

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)	AATGTTTTGTGATGTTG (SEQ ID NO: 71)
TPEF (SEQ ID NO: 5)	AATTTGCGAACGTTTGGG (SEQ ID NO: 76)
TPEF (SEQ ID NO: 5)	AATTTGTGAATGTTTGGG (SEQ ID NO: 77)

36

TABLE 5-continued

Oligonucleotides used in differentiation between colon adenoma tissue and healthy colon tissue.	
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)	GGGTAACGTCGAATTTAG (SEQ ID NO: 52)
Versican (SEQ ID NO: 2)	GGGTAATGTTGAATTTAG (SEQ ID NO: 53)
H-Cadherin (SEQ ID NO: 4)	AAGGAATTCGTTTGTAA (SEQ ID NO: 68)
H-Cadherin (SEQ ID NO: 4)	AAGGAATTTGTTTGTAA (SEQ ID NO: 69)
EYA4 (SEQ ID NO: 3)	AAGTAAGTCGTTGTTGTT (SEQ ID NO: 62)
EYA4 (SEQ ID NO: 3)	AAGTAAGTTGTTGTTGTT (SEQ ID NO: 63)
EYA4 (SEQ ID NO: 3)	AGTGTATGCGTAGAAGGT (SEQ ID NO: 58)
EYA4 (SEQ ID NO: 3)	AGTGTATGTGTAGAAGGT (SEQ ID NO: 59)
Versican (SEQ ID NO: 2)	AAAAATTCGCGAGTTTAG (SEQ ID NO: 54)
Versican (SEQ ID NO: 2)	AAAAATTTGTGAGTTTAG (SEQ ID NO: 55)

TABLE 6

Genes analysed according to FIGS. 16 to 18	
Number in Figures	Gene name
Healthy vs Non-Healthy	
50-D	H-CADHERIN
20-C	CD44
54-C	TPEF (=TMEFF2; =HPP1)
21-C	VERSICAN
50-C	H-CADHERIN
25-B	GSTP1
43-C	TGFBR2
36-B	N33
49-A	CAV1
52-C	PTGS2
46-A	TP73
54-B	TPEF (=TMEFF2; =HPP1)
20-A	CD44
24-D	EYA4
24-B	EYA4
26-B	GTBP/MSH6
4-C	EGR4
15-E	CDH1
23-E	EGFR
30-B	LKB1
22-D	DAPK1
29-D	IGF2
10-A	HLA-F
29-C	IGF2
36-C	N33
21-D	VERSICAN
39-D	PTEN
32-B	MLH1
26-A	GTBP/MSH6

37

TABLE 6-continued

Genes analysed according to FIGS. 16 to 18	
Number in Figures	Gene name
14-C	CALCA
22-C	DAPK1
39-C	PTEN
9-D	WT1
23-A	EGFR
21-A	VERSICAN
30-A	LKB1
9-C	WT1
60-E	ESR1
12-A	APC
29-A	IGF2
8-D	MYOD1
36-A	N33
54-A	TPEF (=TMEFF2; =HPP1)
18-E	CDKN2a
15-D	CDH1
12-C	APC
Healthy vs Carcinoma	
50-D	H-CADHERIN
54-C	TPEF (=TMEFF2; =HPP1)
50-C	H-CADHERIN
21-C	VERSICAN
20-C	CD44
24-B	EYA4
12-A	APC
52-C	PTGS2
24-D	EYA4
39-B	PGR
25-B	GSTP1
49-A	CAV1
23-E	EGFR
36-B	N33
29-C	IGF2
10-D	HLA-F
54-B	TPEF (=TMEFF2; =HPP1)
46-A	TP73
Healthy vs Adenoma	
20-C	CD44
10-A	HLA-F
43-C	TGFBR2
26-A	GTBP/MSH6
26-B	GTBP/MSH6
30-B	LKB1
20-A	CD44
36-C	N33
50-D	H-CADHERIN
46-A	TP73
39-D	PTEN
36-B	N33
54-C	TPEF (=TMEFF2; =HPP1)

38

TABLE 6-continued

Genes analysed according to FIGS. 16 to 18	
Number in Figures	Gene name
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	GPIIb beta
14-C	CALCA
52-D	PTGS2
8-B	MYOD1
24-B	EYA4
21-D	VERSICAN
38-C	PGR
58-A	PCNA
34-D	MSH3
9-B	WT1
35-B	MYC
27-C	HIC-1
52-B	PTGS2
23-E	EGFR
30-A	LKB1
29-C	IGF2
39-C	PTEN
13-D	BCL2
5-B	AR
15-D	CDH1
Carcinoma vs Adenoma	
18-B	CDKN2a
7-E	GPIIb beta

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 110

<210> SEQ ID NO 1

<211> LENGTH: 965

<212> TYPE: DNA

<213> ORGANISM: Homo Sapiens

<400> SEQUENCE: 1

gatcaattaa gggcatctta gaagttagcg gttcccgctg cctccttga gcacggaggc 60

caccaacccc tagggggaag agatgtagcg cgaggcaggg gtgctgtgct aagaaatttc 120

gacgcttctg gggactgagg acaaaggtgc ggacacgacc ccggggtacc tggagtcccg 180

tgactgcgc cacggacggc acacctaggg gctaatttct gctctgcctc aaagaacctc 240

-continued

aagctagagt ccttgectcc gccacagcc ccgggatgcc gctgctgcgc tcaccgcaca	300
ggcagcgccc ggaccggctg cagcagatcg cgcgctgcgc gttccaccgg gagatgggtg	360
agacgctgaa aagcttcttt cttgccactc tggacgctgt gggcggaag cgccttagtc	420
cctacctctg ctgagctgaa cgctcaggca cagtggaaact gaaacccggg tccctacctc	480
tgctgagctg aacgctcagg cacagtggaa ctgaaacccg gttctgcggg atgtgagagc	540
tgttgaggtc acgcgtaatt ggggtgtgatg gagggcgctt gttcgtgatg tgtgcagggt	600
tgatgcaagc aggtcatcgt cgtgcgagtg tgtggatgcg accgcccag agactcggag	660
gcaggcttgg gacacgtttg agtgaacacc tcaggatact cttctggcca gtatctgttt	720
tttagtgtct gtgattcaga gtgggcacat gttgggagac agtaaatggg ttgggtgtgt	780
gtaaatgagt gtgaccggaa cgcagtgatg gcttgatcta ggcagggacc acacagcact	840
gtcacacctg cctgctcttt agtagaggac tgaagtgcgg ggggtggggg acggggccgg	900
aatagaatgt ctctgggaca tcttggcaaa cagcagccgg aagcaaaggg gcagctgtgc	960
aaacg	965

<210> SEQ ID NO 2

<211> LENGTH: 16579

<212> TYPE: DNA

<213> ORGANISM: Homo Sapiens

<400> SEQUENCE: 2

tgtttttttaa aatgatgtct agctgaacca gccaatggag acaaaaatat ttatatataa	60
ttttcctcta agttcactag tcattttata gatcagttat aactttggct actccaatga	120
gttaatttct ttctagagac attataaaact ccttaaagat ggaaaataat tgttactttt	180
ttgaataatg cctactatct aaacaatgc acataaaagc tgctcaataa acttttggtg	240
aattgagaga aaatataaaa aattgaaagc agtctgtttg taataagtgt ttgctaagtc	300
ttacagatct tctctttttt ttggaattat tgcaacaaat atttgacctt tgattttag	360
gattaaaaatg tgatttttca tattttaata ggagcaagca ttagtcctgc aattcttttt	420
gtcaatagaa taaaaatgga atgtatgctt gctgcttcta catgccggtt gatacccatg	480
ctgaacactt ttcttacaac gcagagggtt ccttgtttct tgtcttgctg tctacctttg	540
ctcaccacag agacattaat gatgccatgt gtctttttgc accatcagtg cagcagcttt	600
cattatgctc cttgtctcgg tccactgta gtctaattaa ataaatggaa acgggtagtc	660
actgaaaata aacagcaact gcaaaagtgt tgctctttaa gaaaccttct gcattgtatt	720
ccgccctgcc tcattgatct taagatctct accgcttctt tgtgttcaga aaccagactc	780
ttcagtttgt cttgcttcag tcagacaaat gctttataaa aatgacctat atttttcaga	840
ctgtggggag agaagaaaaa gactttgaga aagtgactaa ggaggagta agagtgatac	900
aatctttgat tcaatttagt tggactgtgt tctttatgcc ttcactttt tctttataaa	960
aaagatagcg tttttgatgc tctgctccag gcagcctccc agggtagtg cagatgtgct	1020
gagccatttg gctgcttga ggaccggctc agagagcctc tctctcact aatgaagtca	1080
tctcctctcg gcagcccgcc ctctcccatc actccatgtt ggaacttggc tttggcttgc	1140
atagttctgt cgtgatgtag atggcttttc agttaagttc tagagaaatc tgctcaacca	1200
tgaacaggac atttacagca caggagaaaa aaaaatttat taaaacaaat attttactgt	1260
aacattttga tttgggggtt ggagagcaaa aactcaatat ctgatgtctg ggaaaaatct	1320
aatgttttat ggctttttaa ggaaatttgt ctgtgtttta tttatcagag caatttatat	1380

-continued

acttcggggtt aaagcacaat gagattgcc aacatcacttt gccttttgca agatttcctt	1440
tttaagaata aaaagaaggc aataaagacc ttatatcttct gccttttccta gaaagccaaa	1500
gagaattcaa agtgctggat gaaatgtctg tccctaattt gcagtgtttt atttaaaaaa	1560
aaaattatga cacaatctta gaggataggc ttcccttttct cagggtctttt gaaagttaaa	1620
agtacatttt taatctgtag cattactgta tcatccagtg aggataattt gcttagttgt	1680
gcctaggacc agtcagatca aggagttgaa actaagctct ttctgaattt gaatgggact	1740
ggaatcttgt catccctaga gcagcaatag atgatgccat gtctaagaac acttcacctg	1800
atcaataaaa aggagaagt ctgaaatgta atcctaaga acctcagaag gaggtttaga	1860
tgaagataat aaaattccct ggagacacat ctgccatcta cacctttcct ctccaagtca	1920
ttggcttgct gccctgagc tcagccaaat tcccaaagac aagatcaaag caaggaagca	1980
aggcacatag cttctggctt atggtttctt ttagacacaa gcattgtgaa tgaagatctt	2040
gctttattta tacctgggat cttggagttt tcacaaagaa tatcccagc acaaggatgt	2100
gatttaagg gggactcaca gaataatatt ttaattactg ttgtattcat acttgagtgt	2160
aatggacaaa taaatacaat tttagttata aaacctgaaa aatataatatt ttccattca	2220
aaaacttctg atagttttgt caggcattgg agaatacaga ataaatgctc tctcattttc	2280
tttctcttga tgaaggaaag tcattaaaaa cagttacata gaaataataa tacaactgta	2340
gactctattc taaatatagc tggtagttt tcccttttct ttttatcca ttctccattt	2400
tagacagcta gaggggaaaag tctgccctag agatagaaaag aagatatcca tggggataaa	2460
tttcatTTta gttatctgat agcctgaaga agtatttaat tttctgaata ataattgtac	2520
aagtcaaaga agagaaaaa caggcagttt aaagcttttc aacattgtgc atattcta	2580
tccttaagtc agccctgaga aagtgccttt gtccgttaga agtatttggc caagtgttgt	2640
gaagtctatt ggcattaaaa gaaactaaaa attttggagg attagagaat caaatttggc	2700
aacataaatg aaaacatctc agaattgtaa aagtatctaa cccatcactt tcaatcatgt	2760
tccattatat acagatggca ttatttatag tgaattgtta ctattaacat ttccctaaaa	2820
atgggcatta atctgctatt agttgcttgt ttctgatcta gttctccttt gttgaaaggt	2880
tttacttaac tgtatcataa agggacaaaa gggtaaaaga taataagctt ttcagactgg	2940
cagatctgtc acttacaaaa tggccttggg caggttactt gaactctttg aaactcaact	3000
agatttactc aaatttaaga atgagaatac aaatccacat cttgaagtgt ttcacagaaa	3060
ggctctatct aatgtctgga gtatatattt caatgaacat tcattttatt ttatttctct	3120
ccattctctg atcaagcaat cttgaatcta aagttgctat gattagcact gaaaagacca	3180
ctggactatt aattgtgtga ctttgggaca gtaactttct gcaccttagt ttgtttacat	3240
gttatacatg aaggttgaag tctgattctg ctctgtgact atcattctaa acatctgatg	3300
aaatcaaatt tcagtgtttg gaattgtagt acaataaatt tactaagaat aaataattca	3360
ctgcaaaaac acattgattt ccaaatgatg taactgacag ttatattact gcagagggt	3420
gataaataac aaaagaaatg aaagatgcac atggtgagaa ctgaaattat cctgacaagt	3480
cttctacctg tttatcactt aaaatcaatg accatgctga atgcctacaa attacaaaat	3540
ataaaagaaa tcttataaat gcgcattgtac agtagtctaa gttactaaaa gttttaaagc	3600
ataagtttaa accaaactaa tcaaagaagt tgagaggaaa aattggcttt catctttaat	3660
cactactggt ttgaggtcct atgtttaata taattttcta agtagaggct tcagagagaa	3720

-continued

gagttgtgag gatactttca tatttgtgta gaaggaaaag tttgccatcc attctagtat	3780
ccttagtggt atactgatgt gcaccttgga tttattttgt tctattgta taaactcata	3840
cttgacttca aagaaaagga aaatccaaag tccctctttt ctaaggggac agaaatcctt	3900
tgtgtcaact gtttgacctt tttctctgta aggtcctatt ggaaatcttt tgtaacacaa	3960
tgcaggggac tcttccatgt gttgatgctg tttacacagt ggggtgggac tgactgaaga	4020
aaaaaaatcg catatacgca tgaagatta tggctcttatt tccggaaagc atgaaagggtg	4080
attgatactt ccaagaagtc cctgttactc aggaaaatta tcaaatattc tactcagaga	4140
tacttggaag gactgaagga aggaagaac gaagaaagca gaatctagac ttatgtgggg	4200
agagatttgt ggcagaggaa aagtattctc tttgaatccg acaagggtt tgctggggg	4260
aatttctgt ccagcctttt attaccaggg tcttttgaag ccgggctccc cattgggcag	4320
ttccctggga gtgcagtggg gaattcttac actttccctc taggtccccg aaggatctcg	4380
ttttctcagt gtctctttca ggttggcagg agccttgagc ctgacacttc cctttgatgg	4440
gacaggcaag ctctgtgggc gcgtaaacac gctgtaacca agttctttgc tgattttaca	4500
gttttgtgtg ctcccagaaa gaagtgatcg tactcaattg tctattgctg gcctgcccc	4560
taagagcctg ggggctcctt tcccctaacc cagaactagc tgcacggggg gcggggaagt	4620
gggggtgggg aaggagtggg agggcagtgg tttccgcgag cagagcgatg ttactgagtg	4680
agtccctgaa tggggagcgc tgctgtcccc aagccgattg gtacttcttg tcaggaagaa	4740
acgccaagag gtgggagtgc ctggggaggg aggcaggcgg tccctaccgc aggcgcgggg	4800
agctgccttt ccgccccctc gcctgctttc caagcctgga ctcttaggag tggctgaagc	4860
tgcgagcgc ttttgagacc tgtgaatgaa ccctcctcct ctccctcctc cttcttctcg	4920
ctgagtctcc tctcggctc tgacggtaca gtgatataat gatgatgggt gtcacacccc	4980
gcatttgaac ttgcaggcga gctgccccga gcctttcttg ggaagaactc caggcgtgcg	5040
gacgcaacag ccgagaacat taggtgttgt ggacaggagc tgggaccaag atcttcggcc	5100
agccccgcat cctcccgcat cttccagcac cgtcccgcac cctccgcac cttccccggg	5160
ccaccacgct tctatgtga cccgctggg caacgccgaa ccagtcgcg cagcgtgca	5220
gtgaattttc ccccaaaact gcaataagcc gccttccaag gtaatcacgt ttcttttgtt	5280
cccccttaa aaaaacaaaa caaaaaactt atagaaaaaa acccgcgagc ttagaaaaaa	5340
gaagcaattg gtagaaggct ttaattaagg caaagagctg taaggcgaag ttaagaaaat	5400
gtaggcactt aaaaaatgca ggtaactttc ataaaggctt ttggggagag gcatacagag	5460
ggaccttggt gttgaaaaag attcagacaa aagaaacca ggggtgggtg ggggtaaaat	5520
gactaacgga attgggggaa gggagggaat aaattgtaaa gaaatcatag aaaagtgggtg	5580
ggttcttgag ctggagagaa gagagggacc tttggcaatt tgattttttt gttgtgtgtg	5640
ttcttaacac gctcaggcaa aaagtgtgaa tggggactac caagactgc cacagacaag	5700
tccccgaagc cgccttggtg caggccacct ggtttcccag cccctggtgt gtggtcagtg	5760
cctggtgttc ctggaaagcc actccgggc agtcctgac agtgcgaccc ggcgcccaag	5820
cagcctggga ccttgccgag acctgacccc ttcagaccgc aggcagtctg ggaggaggtc	5880
cggccggggg aggtgcagga tccccgccgt gtctctttga cgacttgggg actgtcacgg	5940
ttctctcccg gcgccccctg gttcttttgt cctgcacgag gtgcgaaggg gccagcaggg	6000
aaggagcaga ggatgggggg tggggtgtgt ggagccccgc ggaggctctg gaggccctg	6060
ggcgggaaaa gcctgttctg aatcggcagg gatgtgcaac aattttctc acctgaaga	6120

-continued

gtgaaatagg gtctgtcgct cccatctcaa caagcaaacc ggcacccaga ggcactgca	6180
gacaaagggtg gctcggggac ccgaatcagg ggctctggg tcagtctctc cgccagact	6240
acaggagtcc tccgtttcct tacatatcc ccacctctc ctagtctctc gctctagttg	6300
agcaacttta ctggcaggtc tagcggcagc ggcgcgcgcg tgtgccggga gccccggggg	6360
acgtcctcgg ctggagcgcc ccaccgtcct gcagaggcgt gggcgctgta gggcgacatc	6420
cctggtgcgt gcagacctag ggcacccggg ttgtctggc ccgcggtctg tgtcactggg	6480
cgtggagcgg tctgggtgtt aggcagagga gagcggggta gaaaaatagt gtcatagcct	6540
aaactatttg ctctccaatc caactgcgtc tcggcaagtc ctgcatgtcc ctgggagatg	6600
ctgcgggaag ggggagaaat ctacacggcg cctggagagt tgtcctgccg cgcgcacacc	6660
cgcggcagag tctctctggg tcggttctc catctttgtg tctccctctg tctccatctt	6720
ctctgcctcc ctctccctct cccagcctct gttctctctc tctagcttct gcgtccctc	6780
ccccacagct gacaaatgaa tggctagtgt gaaatccctg cctttccctg gctccaaggc	6840
agcaggaggagg gaggagcgag ggaggcggg cgctccttcg gatctgcccc cagttcagct	6900
cacaagactt gcagaaccta gatgtctagg aattgggagt tttgcggcgg gtgtgggcgc	6960
ccctgatgg agaagtcccg cacaggcgga gaaaaacaag cccccagac caagcgtgca	7020
tctttcacia ccttggttgc aaggatggaa ggccctagtt ttcccttaag tcattcattt	7080
tgcactccac aatctgcggc gtacatctag gagtttgag gacctgaaa aaaggtcttg	7140
gtctgtgcaa agtcagaga tgccttctcg cagtcggct ggaacgcacg cgcgccctg	7200
ggtccagttc gccgcttagt ggacgaacgc acatggccgg ggtggagcca ggtttcggag	7260
gtctagacgc gccacctggg gcgtctttaa gaaagataat acacataccg tgcctcaaa	7320
acccatgacc acttgaaccc aacggtaggg ccccgcccca ggatattgca aaagaagggc	7380
tcctaateca aaacttaaac ttctcaact tcaggcgggg cgtcggacgg gaaagtggag	7440
agaaggcggg cagtgggagg aaaaagaaa gggaagggaag ggaagtggaa ggaagagga	7500
ccgtggaggg gaggaaaagg agaaaggaa ggagatgggg cagatggcat tcaaacagcc	7560
aggaaacggc cggaagggtt cgtttcgggg ttgggtctg aatttcgaga ctcgttttag	7620
gatactcttt tccctttccc agcggcaaca aatgtgactg cagaggcgtg ccgggatgga	7680
aggggtggaa agaactagac aagcggaggt ggtccagccc cgcccaaagg ggtgggtct	7740
ggggcccgcc cccactcgct cacctcgcg caaggagga cccgcaacgg gcgctccccg	7800
cagcggctgg gttagccagc cgccacgtgg gttggcgctt cggctcccag ccgagacggg	7860
tgtgacttcc tgggctcaga gactgcagta cagccgcaca ggcggccggc gcagccgagg	7920
ccccaccca ccgaggccaa tggggggcggg gcggggaagg cagggccag cttaaccgag	7980
agagccgagt ggctcgccc ccgggattcg gggccggctg cttctggagc tgaccgagt	8040
tccatccaga cagcgtgaa cactcttatt aaaaggctat ttaagtttta tttatttta	8100
ttttattttt gcttgcttct tgggagtggg aggaggggca tcgtgctttt gttttgaggc	8160
aggggagagg ttaaaattta ctccaagcaa actttttgtc agaaaaatga gactttatac	8220
cgtattccat ttgttgactt tttaaataca ttcaggtcac tctaaagccc attcttagtc	8280
aattaacaga ttaatttcgt tccagactgt cataaagagt tcgcctggaa tggcaaatac	8340
tgccattcga caagactaat agaatttcac gggggcaaac atggaagggc tgcgagtgt	8400
actgaggttc ctgtgectca atgaggaacc ttagcgggga cctggcgggg actggaagct	8460

-continued

cagcacctat ttagggagtg actgagctgg aaacctggag aagcttggtg cagacgagac	8520
ggtgcttctt agagctgcat ttttaagtcct cccattttgc tgcattttgg aagaactgga	8580
ttttaaagat cttttccatc acctcagatt ctgtagctaa cagccacgtc tttttctttt	8640
tgtctgctga aagctgtgac ctaattgttt taagcacttc ttgacaattt cgccaaaccc	8700
atcctgaaaa tattcacata aagtcctaaa ataggatggt gtgagtgtaa tataaaaata	8760
ggaagggtacc aatggaagca acaattagtg ttctgagata gaagcgggtc ccagtttgta	8820
acttggaat gttaagagca gtacatcagg gagaaggctc ccaagagtct gacttcttgt	8880
tctgagctct gagtaaagaa agaaatatgt tagaattgta ttgtctgccc gcagtaatat	8940
tgtaaagggt tccaatggat tccaatgtga ttgagcaatt acttctcca gcaccaagtc	9000
cacagtcctt aatcttcttg tgagaagatt aggttaggaag cgggatgaga ttcaagacaa	9060
tcatttagga tttatgtcag aattcaaagt tttgattctt ttggtagata aaaacagagt	9120
tggttttaaa ttgtatttcc tgctttcaaa aaaacgaggt ttgaaaggaa tccccctaga	9180
tttacatagg tttacagatg taagtttctt ctttttcctt gtgtttgcaa actggaggga	9240
gatgcctttg atattccaga atacctgtat ctatacactt gtaaattctg ctaggaaagg	9300
acctaaacag ctaaaccaaa ggtgacctg attaatgtta aagatactac ccattgttct	9360
ctcttttggt gtattattag gcttaaatac tttggagtta atgattggtc acttaattga	9420
gtaacagagc ctttgagagc gcaaagatca actaactctg tctcccaaag tctggcaact	9480
aagccctcct ttctggtga gatttaaaaa tttttgttga cgtcattcat ggacaaacat	9540
cggtaaacac tgtatttaaa atgggatatt gccctgtgat ggagcagaac ttctcaagac	9600
acatggttct tttgaaagaa agaattaatg actgtagtaa ggtatttttt caccaccac	9660
gtagagaaac cagttaggac tgacaaacac tcattgttgg gagttacttg ctctggtctg	9720
agaaaatatg ctgattgtta cataggctg tattttacca gaaaacacaa agcatgtact	9780
ctccctgatt tggaatcaga aaagggggat tagatatgga tgtcttgccc aagctccatt	9840
tcaattagtg aatttttgc agtttctct tgtagcctag acttggtaaa ttacatatat	9900
ataaagtctg tgtgtgtgtg tgggggtcgt gtgtgtgtgt atgtatacag ggtttttaa	9960
cactgctaag ttcttaccat tgaatgtgtt agaggtggtt gtatttcaga ggtcagtaca	10020
aagaatcttt gtttaagtgt tcagttctgt ttctgaaac acacctctgg gtagatttac	10080
caactataaa aggatacaaa tgtatatact ttccacatct gtatacatcc tacctcatc	10140
ttctaaggat ttttaagggt gtttatgtct ataaaatgag attacatatt ttacaataaa	10200
catatttgat aaattataat ctttttatta ttaaattatt gatttgatat tacttgaggt	10260
atagtggaat gagcactaga cttggaatca gaaagtgtga gttggactca ctagctgaga	10320
aaaccaagca aagatattac ttttggaact ttagtgtcat aatctattaa atgaatatgc	10380
agttataaag atcaaaagag taagtacttt gaaaaactat acaaaaaagg gatgataata	10440
tattatgtct tgggattcat attgatctaa ttctataatt ttaataataa ctgcattacc	10500
caagccatgt agtttctact gaatctataa aataaatcag agaggggcat attgttttaa	10560
aaattgttta gtttcattaa tcagaagaca gattacagaa attgactcca atgaagttgg	10620
acattcagtg gaaaggaaga attattcaat aaatggttcc aggacaattg atgggtccat	10680
gcaaaaataag aatcagtcct tacctgattt ctcatcctaa aagaaatata agatgaatca	10740
aatatgaaat gtagggaata actttttaca aaatgataaa agaaaatcat gggattaaaa	10800
aaaataagtt aggaatagag aaacctgttc taagcaaac ataaaacca gtgccattct	10860

-continued

aggaaatact gataggtctg actaagtaaa aattacacaa aagcaataa tatatttgga	10920
aaacatctac gacatttgag agacaaaaag ctaataactc ttgatattcc atggactttt	10980
atgaataact ttgaaaagtc cacatctatc aaaatatgag caaaaattat gtacaggtag	11040
tggcagaaag agacatagag ttttaaaatt atagaaaaaa tcgctcagcc tcatctgtaa	11100
ttaaataaat agtaattaa acaccaatga tatgctacac ctttttttca ccaatcagag	11160
gagtaacaaa aaaaaataat ttttttcttt tttttttttt ttgagatgga gtctgttact	11220
caggttggag tgcagtggca cgatctcagc tccactgcaac ctccgccttc caggttcaag	11280
caattctcct gtgtcagcct ccctagtagc tgggattatg ggcacccacc accatgcttg	11340
gctaattttc gtgttttttag tagagtgggg tttcaccatg ttggccatgg cagttgatct	11400
tgaactcctg acctcaagtg attcaccac ctcagcctcc caaagtgctg ggattacag	11460
catgagccac tgtgcctggc aaaaatgaaa ttaaaaagtt tgatgcaagt ttgtgttggc	11520
caggttgtgg agaaactggg tctctcatat actgttgtat gaaagtaaat aggttcctct	11580
acttcaaaaa ataatttggc aatcactttt ttttgattct ttatcaaaag tttttaaag	11640
atgtatatac tctttgattc agcaattcca cttctcgaca ttattatatg gatggatatt	11700
ggtagcaata tcaccaaga caaatggaca aggatgttct ttgcagcatt gcttctaaga	11760
gcaaatatct ggaacagtg taacaggctc atcataggga accattttaa taaaagttgc	11820
tgtatccacc taatggaaaa ctatgcagct gtcaaaaaa acatggaaaa attgtatttg	11880
ctgatttga acattctcca aaatatattg tattctaagt gaaaaaattg atagataaaa	11940
ctgagtatat attatacat taattatgta aaatgggtgt atattcatc ttttaacaaa	12000
tatttactga gtgttctagg cattgtggat acaggagtga acaaacatc aaagactgct	12060
ctgttggaac ttttactatg gtaggtagag aaaggcaatt cccaatcaca tatgcaaat	12120
ggtgatatgt gctaggatga aaatagagca tgggaaaggg aatagagtac agtatgtctt	12180
aggtagagta atgagaaagg agttcaccag gcagagaaca gaggagtgtc aagggcgac	12240
aatgtgaaaa agcttagcca agtccagggt caggaattca catgatgcta ttacttgat	12300
gaagtaaatt catgaaattt cgcacctaaa tcacgtgatt tctctctgct gtacaatgat	12360
tctttctgtg agaaagaacc agtgagcaat atgatatgtg atgctcatta catccactt	12420
gtagctgaag taccctagaa agttggtgtg ctagggtacc taattagcag aatggtggac	12480
aaaaccaccg tgtctctgtc ctttaagtcac cccttgctcag ctgagcttcc ccaaatcaa	12540
gggagaatac taaatatcta ttttaatagaa tcttggtctt ctctcgagg ctgccatcc	12600
ctggtgtgtc ccttatctc cctgcagaca tttctcatgg ataggagtcc cattgacctt	12660
gtcttttatg tctcccaggc aattaaacag tctcagtacc ctgtacctgg attgttcaat	12720
gactttcttc ccttctcatt gacttcactg gagtagaagg ctaaatatag gcattcactc	12780
tcatatctct caggctcttc catgcagaca gagttgttat cactagtaga cggtgtttat	12840
aagagagaca tggaaaagtg gaagaatgga aatagggtatt tggagtgggtg ctaaaaacaa	12900
agaagaggc tcttgaaaag tttccattta atatgatcaa acatagaaga caaagatgt	12960
aacttaatga ccaaggatag ggacatagcc ttgagaaaat tacagaagtg gaatgaacgt	13020
cttgactcc aaagaagagg gaatggaaat gcagttctga ggcagctaaa gtcaataaag	13080
ccctctgata tatttactgt agaaataaat gatcatcttt taattcaagg tttgaaact	13140
taggtgacta cagggcgag gcattggagt gaaatgggta agagactaag tggggcgagt	13200

-continued

aatggccttg agcacacaga ctttaagggga gagactgctg cttgggttaat gctcatgac	13260
agccagctga aatatagact tagtggtgct agaccttctg attttttgaa agaagttaga	13320
gattttatat ttagtgtaaa cagtctaatt tttaaattga gtaacttatt caaacatatt	13380
tatgcaactt attcaaatgt atttagagca tgtctgatct cctccttgta ttcttaccta	13440
gtcttagggg taatcttaag gaaagtgatg cgtatacttg ttgtacctat gtagatggag	13500
gaccattctc ctctcccaa caaaagacac ttttataagg attgagtggt gcacaaagag	13560
ttcgattttt agtatttaaa ttgagtagga aactcaaata cagaagatcc tttcctgggt	13620
cacaccttg gtttttttaa atgagatttg actctctaga tattgacaaa catcatatta	13680
ttgggtcatg gtaaacttc aagtggactg attcaaaagt gttcaaaagt atttcagggtg	13740
ttaaggaatt gttggcaagg agtttcaaag gtgtttctgc ataagaaata gaggcgttga	13800
tcatgtattt gatctacaaa cattoctcaa ggatataccc tgaacctggc aatgaagaaa	13860
tattgtggga attattaagt tgaataaagt actagctctg ccttcagaa gctactgtt	13920
ctacagagga gattagacat gtttacagat aatttgaagt acatagcaat aaattatagg	13980
atgatttatt ccaagcataa agtgtatatt gctactatag tggtaaccta ggttataatt	14040
ttcaaagtga agagcccaa atactattct tttactctt agtgatatga aactatgtta	14100
gaacatctgc agaaagtctg aattagtaat ttaataata aatattgagt actgatatat	14160
tgtatatcac tgagatttat tttatataca ttattttatc aattcctatg atcacttcgg	14220
taagtactaa ttttcacaa taagtagtga cagaaaatga ggttagaga gatatagtat	14280
tttaccctaa gtcacagaat ataagtagag aatgcagaac ttgcattctg ggctatttga	14340
cagtggagcc cagactttta aaattaggct tataaagatt tctggaaaa aaagaaaaat	14400
aaaaataga gaacaaattt tctgaagaa aagtaggaaa ttattattta taagtataat	14460
taaagagaga tctttaattt cttttttttt ttttttttg agacggagtc tgcctctgtc	14520
gcccaggcgc aatctcagct cactgcaacc tccacctccc gggttcatgc cataaagaaa	14580
gtttctattt gattggtcg acctgattgt ctactagaac ataggcttcc cagtaggggt	14640
aacaagtcac aacagttcta gtgatataag ataaatacaa actaataata aggttaattt	14700
agaattatca acatatgttt ttcatgtga attatcagtt ggtctataga tggaaaatat	14760
aagatatata atgtattagc ttaaaaaatg ctcttcagag agtgtctaga attatagaat	14820
ttaaaaattt tgttttgaag aggtctaatt cttatttcaa aatatataaa gaaatcaa	14880
gcaaaaaaat taagagaaaa ggtcttaatt aattaaatgc tgtacaactc caagccctat	14940
ttatgttcaa gatcagtttt gtaatttatg actaatttca gcattagctt gttttcaatc	15000
taggtttgct aaataatata gaatttcttt tccaatgaag ttttccaggg tcccagctga	15060
aaaatatata ccagtttaga tgctttcaac tgtaaggaa caaaagtcac gattttattg	15120
ggtttaaaaca ataatgacat ttataatctc atataaaatg aagttcaaag gcattgtcca	15180
ctctagggtc aatccttaag gacttacgcc ttgcctttgc tgattcttcc tcagcttcat	15240
ttccagaggt atctcactta ctcatggatg caaaatccct acaacagttc cagcagaaa	15300
attccagcat gataatgtcc agggaagaaa actgaccacc tcttcctttg ttctcttctc	15360
aggatgaatc tatctttccc agggaccctt cactggagga agagatcccc tcacatctca	15420
ctggccaaaa ctatattata tactgttcca aatccaatca ttggaaaagg gaatggtacc	15480
tctataattg acttaggata tcaggatcat ccttggggt ggggtttgac ccaggattct	15540
ccaaagtatg tgactgaggg taggcactag aatggaacca gagttgtgtt agaaaggaga	15600

-continued

```

aacgcatggc tattatagaa tagttctaaa tgctaccgag gagggcacaa ctgtaaaaac 15660
caaataatct ttgtccatto ttttgaagtt ggcaactttga tttctagatg gttccccaac 15720
acagggttctt ctctctctca tcattatagc tgccctgaaa tttgagctgg aagggaacat 15780
tctgagaccc agattgttaa atgtcttttc caaagtcacg caataaatta aatggcaaag 15840
ccagggcagt ttcttgactc agtacagggt attttctttc attctttact cttgagactt 15900
tagaactggt ggtactgctt taaaattcat ggcaagaact ggtcactttt gtaattaaca 15960
cctccttata atacatttgt tttgtttgct tagccagcta gaaactacat ggagtctgtg 16020
ctttaaaaag cctgccgaag tccttattct ctgttttggg attatgtgca tgaaccacca 16080
attggttctt tctcacctta cacttgatga agatgccttt ctttcaacat ctttctctat 16140
tgctcccat cttctcttgc tctatttatg atcagctgtc tgtttctaaa tagactttgt 16200
ggtcacccat ttctttttgt gccagctcct atccactagt tacttgaact gtggtctcta 16260
ttgttcttca tacagcctac tcacttgtag ttgtcacaca catccacatt gttatatgtt 16320
cccaaactgt attctggaat gattttggta atggctgcat tggatgagat tcaaactaat 16380
aattaaagca ttgagatagc ttctgtctatt acaagtttac tcctgttttt atagtccaag 16440
aggagctact tcttctactc ttattactta atgcttaata ctacccttat tacatacaat 16500
gcacaaaaag catgtgattt atgaccact ttaaacctga atgcttgtga tttactttgt 16560
gtttttcttc atttctagg                                     16579

```

```

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

<400> SEQUENCE: 3

```

```

acattttaat tgcatttttg gcttgaata tattacatat ttaaaagaat tactattttt 60
ctgaatacca ttaatactaa taatagcatc acagtgacta agagtggact taaataaagg 120
ctggcttgaa actcaggtct gccatttatt agcaagcttc taaaaattct gagcccttag 180
ttttctcaca tgtgaaatgg agaaaaataa tctgcaaaat taattaaatc tgtccattca 240
tattttttct tccagtgtat attaaactggc attcctcggt aggccagaat gtgctctcaa 300
ccatgctcca aatccgcttt gtgccaaacc cactgccaga accctttcta ccttgagaac 360
cagaaaagga aacattatgc ctggcaatgc ctacaccctc caaaataaat ctgcaggaaa 420
gaacaccagc taagtgtatg gagcagcaac gactgccttt atcattttta atttacaaca 480
ccaccttttc tagagcctct taagcattgt agataattcc ccactcatta aaaaataaat 540
tgtaaccata agtattcagg gttgatactg cttttgaatt agacagtgtc catatcagtt 600
gcataagacc aacctaaagt agaggatgaa atcttttttc tgaacctttt tcagaacgta 660
acttagtgaa tatattaaaa ctaaactttc tttgaatggg agtaatttct acggattaat 720
ctgtaatctc ttagaccaca cctaaggtaa ttagagaggt gttgtatcat aggcctttgtg 780
attagagacc actggatttg tctttggaaa agtcactcat gattctttgg gctttgggtt 840
cctcatcttt tataccggct taataatgac caccgtagag ttgtcatgga agttaaatga 900
actcatgtac cataagtgc caatacaatg attgacactc agtgatttat caataaatta 960
aaacatttat tatcaatatg acagagaagg tgccgctaaa atagacaata ggtttttgga 1020
agaggtgatt aaatggatgc aaaatttat gattgtttat tccgtctacc tttgctgtgt 1080

```


-continued

```

cccttggttg tggcatacac acgtgtgggt ataaaatcgt aaatcctatg tagtcgcgta 1140
gtgcatgcgc agaaggctta gacacgaaat gtcatttcag caatgtgcct agagaagctc 1200
tgacgccgcc ttggaagtaa gtcgttgctg cctgaccttt gggcgtcttg gacggatgcc 1260
tatacctgca cccagcagca ctggaagggg cccaggccct tcgcagcaca gcctatcccc 1320
agaccgctta gtcttcata acatatactt ccacggaaaa gggatatttc tcccgtcaga 1380
aaaagcgccc cagtctggtc tgggttgggt ttattttcac gttgttgcaa gtaggcgaag 1440
tcccttctgt ctctccctt ggggtaagt gaaaggagtc cggcaggggg cccgcagtgg 1500
cctgcacagg ggaactgggt agcgagagag ttccaggcaa ttccgggggc tgccccacag 1560
aagcaggttg gcatcgacag tggctctccg gcccagggag gagagcgagg tcgcgggtcc 1620
ctccctcag cctggaggct gcagccgctc gagtccggcc gggtgggggc ggggtggggg 1680
cggcgcgagg ggcacggaga ttacggcggc gccacccggg acatccaggg ccccgaggcc 1740
ctgggcggtc cccacgcgag atcgcaaac atgacaatag gcagtcaccc gaggtcaa 1800
aaaaacggag tgggtccccc gcgcgcgcgc gccccccgct tccctggcgg cctcccccga 1860
ggcccccggc ggcttcacga gcccgagta gccggtggcg acgtcgcccc cgccccacct 1920
ccctgcgcaa gtgcgaggct gccggcagcg cggcgcacgc tccggccgtt cccggcttcc 1980
gcgcaaaact tccatcctgt ccacgtgaag ttgtcgctgc cttagagagg gggaaagagc 2040
tgcgggaaaa gccggggagt gacgactgcg gcggtgggc gcgtctctc attttctttt 2100
cttctctttt cccctctgtc gcagtcggga gttttggctc ctctctttc ctctccccc 2160
tcgggacggg cttctccctc cgccccgctt ctccccgct tgtgtacgt atttgttg 2220
gggtggccga aggggatgtc ctgttttcac cagaggcaca gcgcgaagg gaaacttcga 2280
cactggaagg aacgagaata aatacttaac tacggacgca ctgaaccgag gctgggacag 2340
acacttcggg aacccgaggc ggaccggggc acgaggtgag tgaccccttc ttccaacccc 2400
cgccccaggg ctcccggggg agcctgagtt gagagaaccc ccaaacttcc cgggaaagtg 2460
cgcgaggctc cgccggggac gccgagcgtt gggtaactgag gacgcgcagc tggacgggtc 2520
gtgggcgcct gcgtcccccg ggggcgcttg gaggcgggtt gccccacgcc tgagggcccg 2580
ggccgctcgg accgcagcgg tgcctctctc cctagaagac gtccccaagc cccaagggtc 2640
ccttcgagc ctgcctgtcc cttccggggt cggcgcgagg cctgcgcgta acggagttca 2700
tccagcagtc cagcgcgagg cttctacctg cccccgctt ccacctggca gaggcgcgag 2760
catcggggtc tccccacat cttctctatg acgtgtatta cttctgatg accccctaga 2820
tggtccaggc gcgaggatgc tgaccagag tccttcggag ggtcacaggc gcctgggctt 2880
tcccggtgcc gggtcgctgt gtactttaaa ggctcgcgtt ctaatctcca ggcactgatc 2940
gggcttttca actgcggcga tcccacttta atagttttta tgtggcgttg actgaatgct 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

```

-continued

<400> SEQUENCE: 4

```

agcttcagnc atcactnnca tctttatcaa tgnccctggg nccacaggnn gngncccttct    60
gnccaaacag cagccaatgn aggnatctta aaaggcacat cactttatat ttattcagtg    120
gttatgncat cacttaatcc tcaccacaac ctgtgngnt cagtacaatt tgtgntttca    180
ttttatctgt gaacaaatgg aggcagnnnt ttgngtaatt tacaagtaag tggcagatcc    240
gggggctgtg cctagacaat gtggctccag actccttgcc cctccctgcc ttctgctgca    300
tcttgattga gtctagccat ctatttcttt aacacacact aattgctgtc tccatgttcc    360
aggccctgtg ccagatgcta gaaatacatc cccatccact ggatcttggc atgggtccatt    420
gctcagtgtc ctgctatgtt acgtgttaca taatgcattt attccggtat ttagcccatg    480
gaaaaaatat ctgaaagaca ttgtatgcat gttttacaaa caaaactatg taacttatgt    540
ttcttttatt gctttggcta tcaggaagct cttatccaaa tcagagcaaa tacattagaa    600
tttgggcttg tcatttcagt ttgctgaact tttccttctg gccagatgtt tctatttttg    660
ttcataaatt ctattgcaca aatgtccttt attgtaaaac accttaaatt ctttctaagg    720
gaaggctgca tggaaatgat acagtaagggt ctccctgca ttttcttaga ttcctattag    780
ggaaggcaga cctagatgtc cctcttacc tcagtcccaa agccccccat ttataaaatc    840
ccttaagcag tgactactgc tgttctgagt acctggaggt agttcagagc ttctgaaagg    900
taaaatccat atacaaaaag aagtcccttc cgatccagat cccagctttg gtgccagatg    960
cacatttgag gagttaggtg ctatgcagac tctcacctga gcagttaata aaatctattg   1020
ccccctaatt gaatttttct tgcaagctcg aattgatctg tcatctttgt gattttgtgag   1080
atggcaggga agcaccaaac accatcatga cttgggccac agtgggtggga aaaaaggaaa   1140
aaagaaaaaa aaaatccact gccaaagcctt gccaggcgta naaagggtctg gaactgctgg   1200
ggccatttta tctgatttat tggaaataga gtggatctta ttaacatttt aataaagaga   1260
atctttttgc actaggctgg aagtggcgcg cagtcccccg tgcaattcca ttctctggaa   1320
aagtggaaat agctggcatt gccagcgtg atttgtgagg ctgagcccca acagtccaaa   1380
gaagcaaatg ggatgccacc tccgcggggc tcgctcctcg cgagggtgctc accccgtatc   1440
tgccatgcaa aacaggggag cgttaggaag gaatccgtct tgtaaagcca ttggtcctgg   1500
tcacagcct ctaccaaatg ctttctgat gctgctgctg atctatttgg gaagttggct   1560
ggctggcgag gcagagcctc tcctcaaagc ctggctccca cggaaaatat gctcagtgca   1620
gccgcgtgca tgaatgaaaa cgccgcgggg cgtttctagt cggacaaaat gcagccgaga   1680
actcgcgtcg ttctgtgcgt tctcctgtcc caggtaggga agaggggctg ccgggcgcgc   1740
tctgcgcccc gtttctgcat tcggatcgcc cggcacgggc aggggtgagg ggctttcggg   1800
gggtcggggc ctccggtcgc ggcggcgaag acagatcggg gctcggtagg gaggtcatte   1860
cgagcccaga gatectagcc acccccaca cacaggctcc cactctggcg tgcgtgcgtg   1920
tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtacgtt cgttaacggg   1980
agga

```

<210> SEQ ID NO 5

<211> LENGTH: 7833

<212> TYPE: DNA

<213> ORGANISM: Homo Sapiens

<400> SEQUENCE: 5

```

gtctttgggt agatatgtgt tttacaagtt ttaatggaga aaaatgtaag tattttacct    60

```

-continued

cctgaaactt ggctatttga gtaatgagaa aatagtcact ttccccagga cagtggttct	120
caatcatggc tatgtgttcc tccaggaaaa ctttaaaaat atatataac caatgcttct	180
gtgtcacttc tagggattcc aagtctttga atacgaactc tgcacagta ttctttaatt	240
atccaggtga ttgtgatgtg aaatcatgac tgagcccccac tgctctaaga tgaataaac	300
tttcctcagc actgaaatca caaacttaaa ctacccaaat taattaaggg catgggaatc	360
aataaggcat agggaagctt ttacattata aaattatttc tttaaatcac agctcattgt	420
ttatatgtta ttgtccattg tagaaaaggg tgaaaaata gcaaattaa ttactctcag	480
tttgaaaaat tatccagaaa tgaagatgac gactctgaaa cattgtcaat atcatttgac	540
ctataataa tgttctaata catttactac acactgatag atacttttcc atatgaatat	600
tatacattaa aactaaggca ataatgcatt tagaacattc tatctatata tatgtatctt	660
aagtaggcta gaaattaaga tatgagtat taagtatgag atgttaaggt gtggggttag	720
aaattatact gtacttcatt atcaataatc aacatatact tcaatatcac atacatttaa	780
ctttaatttg tacactctta actattttta attatgtgta taaatataag tacacacatc	840
tttatgtatt tatttattca tacctccatt cacttattta tataggggat cccccaaat	900
ccactaccat taaaccatac atttttattt taatctttag aacaagccca ggaggcaggt	960
attgttatta ctcacatttt acaaatgagg aaattgtcta cagtcacaaa gttactgtgt	1020
cagacatatt agaagcttaa tacatatttg gtgaacatat gcataaaaac agagagacag	1080
acatgtacaa cagctcatct ttacactgag taaaagcttt taacctgtct cagaaacctc	1140
tctgtgaaaa ctgagcaaaa atcgaggat cctttcattt gtcatatagg tataggtggt	1200
accttacttc tccaacaagg atgaatattg aaatgtggat cccaaggccc aactccagat	1260
tttctgaatc cctgatagtg ggacttggaa ttgtctatt gtttcaaagt ttctcaagga	1320
attcatatga tcaaccaggt tcagaaatca ctggatctta ttgccgaagt ttgagaatta	1380
aagtttgggc ctactgcgg ctccacagaa agggcaaatg aagtatcatg gacagaactg	1440
atcgttccc agttagtctt ccctctcaga agctaacagg cagcaatata gcagaaatta	1500
gtgacttatg tcttgtgctc tgaagtcagg cagaatttca cagagtccca gcagtgtcac	1560
tgacgagatt tgtttcttgg ggcaagttgc ctgatgcttt caaagccata ttctttttat	1620
ataaatgag ataattattc ttgtctcata ggggtgtttt aaagattaaa taaaaataac	1680
atgttctatc ctacatggca caatgcctga cacctaagaa gcaaaggata catcttacct	1740
ttattgaagc aatcagaaag tatgaaatca tgaaggagat aagagtcttg attggcagtg	1800
tatcttattt tcccaggctc atttatttat cttaacttat tctgttgga gaataactcc	1860
caagccccct acttaagctg tgagtaatct cacactttat aatgatgttc ttcccatgag	1920
aaaaaaaaat gttcttaagt ttcttgaga aaatatatct gcactatttc tactgaaaaa	1980
tctaacaact ggactctgct cctctgcac aattctagag tgtatatgcc acaataaag	2040
tgttctagct caagaagatt gaaagtaaat atggtatagt attttaaaat aagaattttg	2100
caaatacatg gtatgattgt gtcatattac tagcaatcat atgatacgca atgcaaagta	2160
cagttcatag acttaaatth aattctaata agtaactga ttttgcttg ctggggaaaa	2220
gttaaagcac taatccaatt gctaagcag tcttgtctac ttcttttgga cctagtgaca	2280
agtctaaata atgtatatat ttttatttac atattcagta atacaattct ctgctcaatg	2340
agtgatgttc ttctgccact tgggtgtgct tgccagtttc agaatttggt tcttggtggc	2400

-continued

actataacac taagtacaga gtaagtgcac caaaattgca gcattcccat tgaaaaggct	2460
ttgettcaaa ctgtttaata atttaaagga cctctgtgga agcaaccgca tttgttaacc	2520
agttacaacc agtaattaac tcctttggag ttttaactta cttttggcaa aacgtcttag	2580
gaagagcata tattattaga aagtatgcca aaaatttact tagcagaaaa ttcaaaaaca	2640
gttttctctc gctaagaggt tctctaaaat tctacttaca tagccaaact ctgaaatcct	2700
agcaggctcct gtttcattat cataattact gcataaacac ttttaaggac tttgccttta	2760
gtttcaagca tgacttattt tcataagcct gattagttag cacaccagcc ttgctatgga	2820
aatgacatg ttctcattct ctgctgtaga gttgttaaat cttgatctat atttatgttg	2880
ccttctctgc tgaagcctg tagcgaaaga aatttctaata tcttgtttt gcaatattag	2940
ttggcagctc tatctaagtg gtattctgtt tccttaaaga atttagctgc tctgtctaga	3000
agccgatttt ctgatgcctc caacgtctgg tctaattgat ctgttttaaat ggagtcttcg	3060
tcggtgagga gcgagatgcc accgactaga atgctgggat ctgctgctta attgccagga	3120
gtgagagaca ctgagattca gaaatctttg gaggtgggag gggagaggga cagtctcgga	3180
cggaggcgga gatgtaagat aaagggatgg atttcacaca ggaaaaaaaa aaagatttcg	3240
ttgaggcact gaggtgctgc acgatcacat ctctcaaagg agaagttaaa aagcaaggaa	3300
gtgggaggag gttggagggt aaagtactta aaaggattac tcgggtacaa tttgtttttc	3360
tgctggtgtc tgcaaaagat agatagtcct gttttcaaag tatatgaatg cctcttttaa	3420
gtgattggga atggacacta attgcctgtt aaatgttato aaatgctctc ctaaaattcag	3480
gggacacaga aagaggggca caaaaggaga atttaaatag aaaaaggag gatccggagg	3540
cttttgaaag cggggggaga agaaggagga gggataacag agaggaatag agaaggagag	3600
cggagagaag ataaacaaaa acaaaaacag gaatcactga ataatacac accaaaaaga	3660
aagctcttcc ctatggggca tccaaaacac tgagactgca atagtgaacc cggatcatgga	3720
agaaaagatg tcctctccac cctgtctccc gaaagctctt ggtcccgta ctggcgacta	3780
aaattccatt aggtcaaaaga gtgtgtctaa ctgcctgaag aatgcagcag acggaaggcg	3840
ggctccgcta tgccgtttgc ccttcccgct ggagagaatg aaagaaacgc gcagagccag	3900
agactcctgc cgagttagac cttctctcgt cgtcccgagt caccggccat ccggcaaaga	3960
cccagtaag gaacgcaggg tcaactgcctg ggccaacaaa tggagcccg tctcccttc	4020
cggagcccg ctgcccggcc gatgctcccc gcaaccacac cggcgctat gcagaggagc	4080
ctttctcttt ctctcagacc actgtctccg accaatctga ccttccaaac acatctgacc	4140
gcacctccca ggtggacaca ctaataggct acgggctgga gaggagcggg tgatgaggag	4200
agggattcaa acctgcgaac gcttgggctg ggtcggagct gcggggggcc tgggaggaga	4260
gaggggagaa gagagaagga aggagagcgc ctgccgggat ggctgagctg cctcggcgag	4320
cagccttggg gttgcacgct cttgtgggag atgctgctgt tgcttccagg tcggcaagag	4380
cggttctaac accatcgctc ctcacctctt ttctgtaaa tccctagaga aacgtccctg	4440
gcctctccgc cgcgacatto ccagcctgca tccccctaca gcctaggcgg cgcgctcccg	4500
cacgctggag cgcgggtcgc cagcaggacg cctctcccg cgcgactcg cccctctctg	4560
ccctgctgct gctgctctc tgacacctcc gccccacca tctccagctc ggagagacgc	4620
caccagccg cggcccgcac tcgcccgcg ggtcacgcg cggaagaggg gcgctagtcc	4680
ggaccccgcc ttcggtaggg ggcgtcctgg agcggagagt gagcggaatg gtatatgagt	4740
gtgcgggtag cccacctga agcccgagct tctcatttga gccatgcccc gcctagcccc	4800

-continued

actcgggcca	gcgcctggcg	agcgagccca	tctgtggctt	ccgcggccgc	ctcctccttg	4860
catccttgca	cctactcgto	gacccctccc	tcccgggacc	tgcacctgc	tccaccaatc	4920
agagcccgac	tgccctttcc	cacgtgaccc	cgggcgggct	gaggacctgc	tgcttcccaa	4980
acgccagagg	gatgcgggcg	gcagagctcg	agaggcggt	gccgggctgc	ggggcgccct	5040
gactctccct	ccaccctgcc	tctcggggt	ccactcgtct	gcccctggac	tcccgtctcc	5100
tctgtctctc	cggcttccca	gagctccctc	cttatggcag	cagcttcccg	cgtctccggc	5160
gcagcttctc	agcggacgac	cctctcgctc	cggggctgag	cccagtcctt	ggatgttgct	5220
gaaactctcg	agatcatgcy	cgggtttggc	tgctgcttcc	ccgcggggtg	ccactgccac	5280
cgcgcgcgcc	tctgtgcgcg	ccgtccgcgg	gatgtcagt	agcccgctgc	ccggcccccg	5340
cgatcctgtg	ttcctcggaa	gccgtttgct	gctgcagagt	tgacgaact	agtcagtgtg	5400
ctgtgggagt	cccccgcgca	gtgcagcagc	tgacactttt	gcgagggctt	ttgtcggtg	5460
ctgtctgtgc	ccgtcatgct	actcatcgta	gcccccccg	tgaagctcgc	tgctttccct	5520
acctccttaa	gtgactgcca	aacgcccacc	ggctggaatt	gctctggtaa	gtccagaacc	5580
cccgcccccg	accctttaac	tccgcagaag	aacacgcgta	tccagcacag	accagcctac	5640
cctagcgcgc	ctctcagccc	cctcacctcc	tactgcccta	gacccctaata	accacccacc	5700
tctatccaga	gaaacaaggg	gaactgttgc	agggccgggg	gtgaggggtg	gttctgggat	5760
gggcagaaag	tgacggtgta	gcaggaaaac	tttgcagtct	tgccgttaca	ttggagctgc	5820
gaggattttg	agaaatatta	aacgggatgg	ttttctgggt	tcaactgtttt	gaaagagcac	5880
caatcctagg	ggaaacactg	aaacagaagc	tttgtcatca	ttaaagaaaa	aagtcttact	5940
aggatgagga	agaaataact	ttatgagaaa	gaatgagcga	gaaagcaata	aatcaaatgg	6000
tgactgcagg	ggaatcgctg	attcctggca	aagggtccat	gaggtcgcac	tggtctcccg	6060
ttgaagacca	ggtcacacag	attctagagg	agctgggttt	caatagaatt	tctctctctc	6120
tctctctctc	tctctctctc	tctctctctc	tctctctctc	tatctatctc	tctctctctc	6180
tcattccctt	ctctcctagg	cggcaaaaag	cattgggttt	gcagtcacga	tatgccccct	6240
tctttgcttc	cctaagcttc	aaggtagtac	aggggagttg	agaaaaagaa	cactttgcgg	6300
gtctcccagg	ccggagtggg	catgactgag	gctggtcagg	ctccatgtag	gcgagccgag	6360
ggcggaaacc	acttcagtgg	gcgctgactc	ctccatttct	ggacaggctt	ctgtggagtg	6420
ggtcaggcac	tcttcttgct	cgctcgggtt	ccttcagatt	ctgacggcga	acgcttggca	6480
ggcttcgctc	tgctgaagct	tcctaattaa	atagggccag	aggatgggag	ttgctgcact	6540
cctagctggc	atagcattcg	gtttgacagc	ctgtagtata	gggtgtatgt	aatttttcat	6600
cttctgtgaa	tataattttg	ctgtagttaa	atctggctct	gaataaagtg	tctttcaaag	6660
atgtatataa	gctgaagtgt	atgtaacttt	agagaggagg	gaatgaccaa	ctgtaactca	6720
gggtgaaagc	ctgtatagtt	cctagtattt	actgatgtaa	atgccaaaag	gaaaattatt	6780
atgcatcatt	ctaattttat	ctttacaaag	acaagttgag	atatgcaacc	ctattagatt	6840
tgggtcaata	gattgttctc	ttttttggca	gtttctaaat	ttggcatttt	aataaaaactc	6900
aacatgtttc	tataactttc	tgattcatgc	gtacatgtgt	gttgtttttg	aaagaataag	6960
tttcaacttg	ctattgccta	atcaactttt	agatgcttta	ttatggtaat	aattatgagc	7020
ctgcaaaaac	aatttttggg	aatgttgatg	gctttgtagt	ccaacacaga	ctgggttgct	7080
tcattcctag	cccttcgatt	gttttaggaa	ataactaact	taaatgtgaa	gttgacattt	7140

-continued

```

gcaatcaaga aattacatat ttaccagata ttttaaaggg gactgcataa actaaagaga 7200
ataaactggg tttgcagata ggttgtcaag aacttggcac cccgcttcca cccctgttaa 7260
cttagagggt atcaatcttc atttgagcca aacagaccat cacagaaaac actgtgectg 7320
tttatcttta ttattgaggc tttgtttcct ctttgtctgg atacatttca aataaggggt 7380
tgtttcagtc gttgaagcaa aagaacaatt aaagatgggg aaatggtaaa aggggtattca 7440
gagatcatca ctagctcttt tccaaaatgt ggagttttgt ggtcataaat attgtccacc 7500
taatgagcaa aaaataaaaa taaaaaaaaa acaggaagca aatgttaagc ttctattcac 7560
cactgtcagt attaacgcaa gctttaaaaa atagcactat cagaaaagga tactaaagga 7620
gaattgacta gaaaagaatt gtggaaaatg gaaacgaata ttgatcactt aactagattt 7680
tgaggttatc agtagacagt gaccttgca gacagctata gttgttggat ttaaaattta 7740
ggacaagtat tttaaagctt caaagtagtg cttttttttg 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 gggatattta gaagttaggc gttttcgttg ttttttttga gtacggaggt 60
tattaatttt tagggggaag agatgtagcg cgaggtaggg gtgtcgtgtt aagaaatttc 120
gacgtttttg gggattgagg ataaagggtc ggatacgatt tgggggtatt tggagtttcg 180
tgattcgcgt tacggacggt atatttaggg gttatttttt gttttgtttt aaagaatttt 240
aagttagagt ttttgttttc gtttatagtt tgggatgtc gttgttgcgt ttatcgtata 300
ggtagcgttc ggatcggttg tagtagatcg cgcgttgcgc gttttatcgg gagatggtag 360
agacgttgaa aagttttttt tttgttattt tggacgttgt gggcggttaag cgttttagtt 420
tttatttttg ttgagttgaa cgtttaggta tagtgaatt gaaattcgg tttttatttt 480
tgttgagttg aacgttttag tatagtggaa ttgaaattcg gttttgcggg atgtgagagt 540
tgttgagggt acgcgtaatt ggggtgtgat gagggcggtt gttcgtgatg tgtgtagggt 600
tgatgtaagt aggttatcgt cgtgcgagtg tgtggatgag atcgttcgag agattcggag 660
gtaggtttgg gatacgtttg agtgaatatt ttaggatatt tttttggtta gtatttgttt 720
tttagtgttt gtgatttaga gtgggtatat gttgggagat agtaatgggt ttgggtgtgt 780
gtaaatgagt gtgatcggaa gcgagtgtga gtttgattta ggtagggatt atatagtatt 840
gttatatttg tttgtttttt agtagaggat tgaagtgcgg ggggtgggggt acggggtcgg 900
aatagaatgt ttttgggata ttttggtaaa tagtagtcgg aagtaaaggg gtagttgtgt 960
aaacg 965

```

```

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

```

```

<400> SEQUENCE: 7
cgtttgata gttgtttttt tgttttcggt tgttgtttgt taagatgttt tagagatt 60

```

-continued

```

ttatttcggt ttcgtatttt tattttcgta ttttagtttt ttattaaaga gtaggtaggt 120
gtgtagtgtgt tgtgtggttt ttgttttagat taagtttata ttcgttttcg gttatatatta 180
tttatatata tttaaattta ttattgtttt ttaatatgtg tttattttga attatagata 240
ttaaaaaata gatattgggt agaagagtat ttgaggtgt ttatttaaac gtgttttaag 300
tttgttttcg agtttttcgg gcggtcgtat ttatatattc gtacgacgat gatttgtttg 360
tattaaattt gtatatatta cgaataggcg ttttttatta tatttaatta cgcgtgattt 420
taatagtttt tatatttcgt agaatcgggt tttagtttta ttgtgtttga gcgttttagt 480
tagtagaggt aggggaatcgg gtttttagttt tattgtgttt gagcgttttag tttagtagag 540
gtagggatta aggcgtttgt cgtttatagc gtttagagtg gtaagaaaga agtttttttag 600
cgtttttatt attttttcgg ggaacgcgta gcgcgcgatt tgttgtagtc ggttcgggcg 660
ttgtttgtgc ggtgagcgta gtagcggat ttcggggttg tgggcggagg taaggatttt 720
agtttgaggt tttttgaggt agagtagaaa ttagttttta ggtgtgtcgt tcgtggcgcg 780
agttacggaa ttttaggtat ttcggggtcg tgttcgtatt tttgttttta gtttttagaa 840
gcgtcgaaat ttttagtac gatatttttg tttcgcgtta ttttttttt ttttaggggt 900
tgggtggttt cgtgtttaaa ggaggtagcg ggaacgttta atttttaaga tgtttttaat 960
tgatt 965

```

```

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

```

```

<400> SEQUENCE: 8

```

```

tgttttttaa aatgatgttt agttgaatta gttaatggag ataaaaatat ttatatataa 60
ttttttttta agtttattag ttattttata gattagttaa aattttggtt attttaatga 120
gttaattttt ttttagagat attataaatt ttttaaagat ggaaaataat tgttattttt 180
ttgaataatg tttattattt aaaataatgt atataaaagt tgtttaataa atttttgttg 240
aattgagaga aaatataaaa aattgaaagt agtttgtttg taataagtgt ttgttaagtt 300
ttatagattt tttttttttt ttggaattat tgtaataaat atttgatttt tgattttag 360
gattaaaatg tgatttttta tattttaata ggagtaagta ttagttttgt aatttttttt 420
gttaatagaa taaaaatgga atgtatgttt gttgttttta tatgtcgggt gatatttatg 480
ttgaatattt tttttataac gtagaggttt tttgtttttt tgttttcgtt tttatttttg 540
tttattatag agatattaat gatgttatgt gtttttttgt attattagtg tagtagtttt 600
tattatgttt tttgttttcg ttttattgta gtttaattaa ataaatggaa acgggtagtt 660
attgaaaata aatagtaatt gtaaaagtgt tgttttttaa gaaatttttt gtattgtatt 720
tcgttttggt ttattgattt taagattttt atcgtttttt tgtgtttaga aattagattt 780
tttagtttgt tttgttttag ttagataaat gttttataaa aatgatttat attttttaga 840
ttgtggggag agaagaaaaa gattttgaga aagtgattaa ggaggagta agagtgatat 900
aatttttgat ttaatttagt tggattgtgt tttttatgtt tttatttttt tttttataaa 960
aaagatagcg tttttgatgt tttgttttag gtagtttttt aggggtgagt tagatgtgtt 1020
gagttatttg gttgttttga ggatcgggtt agagagtttt tttttttatt aatgaagtta 1080

```

-continued

ttttttttcg gtagttcggg	tttttttatt	attttatggt	ggaatttggg	tttggtttgt	1140
atagttttgt cgtgatgtag	atggtttttt	agttaagttt	tagagaaatt	tgtttaatta	1200
tgaataggat atttatagta	taggagaaaa	aaaaatttat	taaaaataat	attttatgt	1260
aatattttga tttgggggtt	ggagagtaaa	aatttaatat	ttgatgtttg	ggaaaaattt	1320
aatgttttat ggttttttaa	ggaatttgt	ttgtgtttaa	tttattagag	taatttatat	1380
atttcgggtt aaagtataat	gagattgtta	ttattatttt	gttttttgta	agattttttt	1440
tttaagaata aaaagaaggt	aataaagatt	ttatattttt	gtttttttta	gaaagttaaa	1500
gagaatttaa agtggttgat	gaaatgtttg	tttttaattt	gtagtgtttt	atttaaaaaa	1560
aaaattatga tataatttta	gaggataggt	tttttttttt	tagggttttt	gaaagttaaa	1620
agtatatttt taatttgtag	tattattgta	ttatttagtg	aggataattt	gttttagttgt	1680
gtttaggatt agttagatta	aggagttaa	attaagtttt	ttttgaattt	gaatgggatt	1740
ggaattttgt tatttttaga	gtagtaatat	atgatgttat	gtttaagaat	attttatttg	1800
attaataaaa aaggagaagt	ttgaaatgta	attttaaa	attttagaag	gaggtttaga	1860
tgaagataat aaaatttttt	ggagatatat	ttgttattta	tatttttttt	ttttaagtta	1920
ttggtttggt gtttttgagt	ttagttaaat	ttttaagat	aagattaaag	taagggaagta	1980
aggtatatag tttttgggtt	atggtttttt	ttagatataa	gtattgtgaa	tgaagatttt	2040
gttttattta tatttgggat	tttggagt	ttataaagaa	tatttttagt	ataaggatgt	2100
gatttaaaag gggatttata	gaataatatt	ttaattattg	tttgatttat	atttgagtgt	2160
aatggataaa taaatataat	tttagttata	aaatttgaaa	aatatatatt	tttttattta	2220
aaaatttttg atagttttgt	taggtatttg	agaattagaa	ataaatgttt	ttttattttt	2280
ttttttttga tgaaggaaa	ttattaaaa	tagttatata	gaaataataa	tataattgta	2340
gattttattt taaatatagt	tggtatgttt	tttttttttt	tttttattta	ttttttattt	2400
tagatagtta gaggggaaa	tttgttttag	agatagaaa	aagatattta	tggggataaa	2460
ttttatttta gttatttgat	agtttgaaga	agtattta	tttttgaata	ataatgttat	2520
aagttaaaga agagaaaaga	taggttagtt	aaagtttttt	aatattgtgt	atattttaat	2580
tttttaagtt agttttgaga	aagtgttttt	gtttttaga	agtatttggg	taagtgttgt	2640
gaagtttatt ggtattaaaa	gaaattaaaa	attttggagg	attagagaat	taaatttggg	2700
aatataaatg aaaatatttt	agaatgtaaa	aagtatttta	tttattattt	ttaattatgt	2760
tttatttat atagatggta	ttatttatag	tgaatgttaa	ttattaatat	ttttttaaaa	2820
atgggtatta atttgttatt	agttgtttgt	ttttgattta	gttttttttt	gttgaaaggt	2880
tttattta atgtattata	agggataaaa	gggtaaaaga	taataagttt	tttagattgg	2940
tagatttggt atttataaaa	tggttttggg	taggttattt	gaattttttg	aaatttaatt	3000
agattttatt aaatttaaga	atgagaatat	aaatttatat	tttgaagtgt	tttatagaaa	3060
ggttttattt aatgtttgga	gtatatattt	taatgaatat	ttattttatt	ttattttttt	3120
ttatttttga attaagta	tttgaattta	aagttgttat	gattagtatt	gaaaagatta	3180
ttggattatt aattgtgtga	ttttgggata	gtaatttttt	gtattttagt	ttgtttatat	3240
gttatatatg aaggttgaag	tttgattttg	ttttgtgatt	attattttta	atatttgatg	3300
aaattaaatt ttaggttttg	gaatggtagt	ataataaatt	tattaagaat	aaataattta	3360
ttgtaaaa atattgattt	ttaaagtagt	taattgatag	ttatattatt	gtagagggtt	3420
gataaataat aaaagaaatg	aaagatgtat	atgggtgagaa	ttgaaattat	tttgataagt	3480

-continued

tttttatttg tttattattht aaaattaatg attatgttga atgtttataa attataaaat	3540
ataaaagaaa tttttataaat gcgtatgtat aggagtttaa gttattaaaa gtttttaaagt	3600
ataagtttta attaaattaa ttaaagaagt tgagaggaaa aattggtttt tatttttaaat	3660
tattattgtt ttgaggtttt atgttttaata taatttttta agtagaggtt ttagagagaa	3720
gagttgtgag gatattttta tatttgtgta gaaggaaaag tttgttattt atttttagtat	3780
tttttagtgt atattgatgt gtatttttga tttattttgt ttttattgta taaatttata	3840
tttgatttta aagaaaagga aaattttaag tttttttttt ttaaggggat agaaattttt	3900
tgtgttaatt gtttgatttt tttttttgta aggttttatt ggaaattttt tgtaataata	3960
tgtaggggat ttttttatgt gttgatgttg tttatatagt ggggtgggtt tgattgaaga	4020
aaaaaatcg tatatacgta tgaaagatta tggttttatt ttcggaaagt atgaaagggtg	4080
attgataatt ttaagaagtt tttgttattt aggaaaatta ttaaatattt tatttagaga	4140
tatttgaaa gattgaagga aaggaagaac gaagaaagta gaatttagat ttatgtgggg	4200
agagatttgt ggtagaggaa aagtattttt tttgaattcg ataagggtt tgtttggggg	4260
aattttttgt ttagtttttt attattagggt ttttttgaa tgggtttttt tattgggtag	4320
ttttttggga gtgtagtggg gaatttttat attttttttt taggttttcg aaggatttcg	4380
tttttttagt gtttttttta ggttggtagg agttttgagt ttgataattt tttttgatgg	4440
gataggtaa ttttgtggc gcgtaaatac gttgtaatta agttttttgt tgattttata	4500
gttttgtgtg ttttcgagaa gaagtgatcg tatttaattg tttattgttg gtttgttttt	4560
taagagtttg ggggtttttt ttttttaatt tagaattagt tgtacggggg gcggggaagt	4620
gggggtgggg aaggagtggg agggtagtgg ttttcgcgag tagagcgatg ttattgagtg	4680
agtttttgaa tggggagcgt gtgtgttttt aagtcgattg gtattttttg ttaggaagaa	4740
acgttaagag gtgggagtgt ttggggaggg aggtaggcgg tttttatcgt aggcgcgggg	4800
agttgttttt tctgtttttt gtttgttttt taagtttgga tttttaggag tggttgaagt	4860
tgcgagcgt ttttgagttt tgtgaatgaa tttttttttt tttttttttt ttttttttcg	4920
ttgagttttt ttttcggttt tgacggtata gtgatataat gatgatgggt gttataattc	4980
gtatttgaat ttgtaggcga gttgttttga gtttttttgg ggaagaattt taggcgtgcg	5040
gacgtaatag tgcagaatat taggtgttgt ggataggagt tgggattaag attttcgtt	5100
agtttcgtat ttttcgtat ttttttagtat cgtttcgtat ttttcgtatt ttttttcggg	5160
ttattacgtt ttttatgtga ttcgttttgg taacgtcgaa tttagtcgcg tagcgttgta	5220
gtgaattttt tttttaaatt gtaataagtc gttttttaag gtaattacgt tttttttgtt	5280
ttttttttta aaaaataaaa taaaaaattt atagaaaaaa attcgcgagt ttagaaaaaa	5340
gaagtaattg gtagaaggtt ttaattaagg taaagagttg taaggcgaag ttaagaaaat	5400
gtaggtattt aaaaaatga ggtaattttt ataaaggttt ttggggagag gtatatagag	5460
ggatttttgt gttgaaaaag atttagataa aagaaattta ggggtgggtg ggggtaaaat	5520
gattaacgga attgggggaa gggagggaat aaattgtaaa gaaattatag aaaagtgggtg	5580
ggtttttgag ttggagagaa gagagggtt tttggtattt tgattttttt gttgtgtgtg	5640
tttttaatac gttcaggata aaagtttgaa tggggattat taagatttgt tatagataag	5700
ttttcgaagt cgttttggtg taggttattt ggttttttag tttttggtgt gtggtttagtg	5760
tttggtgttt ttggaagtt attttcgggt agtttttgat agtgcgattc ggcgtttaag	5820

-continued

tagtttggga ttttgcgcgg atttgatttt tttagatcgt aggtagtttg ggaggagggt	5880
cggtcggggg aggtgtagga ttttcgctgt gtttttttga cgatttgggg attgttacgg	5940
ttttttttcg gcgttttttg gtttttttgt tttgtacgcg gtgcgaaggg gttagtaggg	6000
aaggagtaga ggatgggggg tggggttgtt ggagtttcgc ggaggtttg gaggtttttg	6060
ggcgggaaaa gtttgttttg aatcggtagg gatgtgtaat aatttttttt attttgaaga	6120
gtgaaatagg gtttgcgtt tttattttaa taagtaaatc ggtatttaga gcgtattgta	6180
gataaagggt gttcggggat tcgaattagg ggtttttggg ttagtttttt cgttttagatt	6240
ataggagttt ttcgtttttt tataatattt ttattttttt ttagtttttc gtttttagttg	6300
agtaatttta ttggtagggt tagcggtagc ggccgcgcgc tgtgtcggga gtttcggggg	6360
acgttttcgg ttggagcgtt ttatcgtttt gtagaggcgt gggcgttgta gggcgatatt	6420
tttggtcgt gtagatttag ggtattcggg ttgttttggg tcgcggtttg tgttattggg	6480
cgtggagcgg tttgggtgtt aggtagagga gagcggggta gaaaaatagt gttatagttt	6540
aaattatttg tttttaatt taattgcgtt tcggttaagt ttgtatgtt ttgggagatg	6600
ttgcgggaag ggggagaaat ttatacggcg ttggagagt tgtttgtcg cgcgtatatt	6660
cgcggtagag ttttttggg tcgcgttttt tatttttgtg ttttttttg tttttattt	6720
ttttgtttt ttttttttt tttagttttt gtttttttt tttagtttt gcgtttttt	6780
ttttatagtt gataaatgaa tggtttaggt gaaatttttg ttttttcgt gttttaaggt	6840
agtagggagg gaggagcag ggagggcgtt cgttttttcg gatttgttt tagtttagtt	6900
tataagattt gtagaattta gatgtttagg aattgggagt tttgcggcgg gtgtgggcgt	6960
tttttgatgg agaagtttcg tataggcga gaaaaataag ttttttagat taagcgtgta	7020
ttttttataa ttttgtgtt aaggatgga ggttttagtt ttttttaag ttatttattt	7080
tgtattttat aatttcgcgc gtatatttag gagtttgag gattttgaaa aaaggttttg	7140
gtttgtgtaa agttagaga tgttttttcg cgagtcggtt ggaacgtacg cgcgcttttg	7200
ggttttagtt ctcgttttag ggacgaacgt atatggtcgg ggtggagtta ggtttcggag	7260
gtttagacgc gttatttttg gcgtttttaa gaaagataat atatatatcg tgtttttaa	7320
atttatgatt atttgaattt aacggtaggg ttctgtttta ggatattgta aaagaagggt	7380
ttttaattta aaatttaaat ttttttaatt ttagggcggg cgtcggacgg gaaagtggag	7440
agaaggcggg tagtgggagg aaaaagaaa ggaaggaag ggaagtggaa ggaagagga	7500
tcgtggaggg gaggaagg agaaaggaac ggagatgggg tagatggtat ttaaattagtt	7560
aggaacggat cgaagggtt cgtttcgggg ttgggtttg aatttcgaga ttcgttttag	7620
gataattttt ttttttttt agcggtaata aatgtgattg tagaggcgtg tcgggatgga	7680
aggggtggaa agaatttagt aagcggaggt ggttttagtt cgtttaaagg ggttgggttt	7740
ggggttcgtt ttatttcgtt tatttcgcgt taaggagga ttcgtaacgg gcgtttttcg	7800
tagcggtttg gttagtttag cgttacgtgg gttggcgttt cggtttttag tcgagacggg	7860
tgtgattttt tgggtttaga gattgtagta tagtcgtata ggcggtcggc gtagtcgagg	7920
ttttatttta tcgaggttta tggggggcgg gcggggaagg tagggtttag ttttaatcgag	7980
agagtcgagt ggtttcgtt tcgggattcg ggttcggtt tttttggagt tgattcgagt	8040
tttatttaga tagcgttga tatttttatt aaaaggttat ttaagtttta ttttatttta	8100
ttttattttt gtttgtttt tgggagtggt aggggggta tcgtgtttt gttttgaggt	8160
aggggagagg ttaaaattta ttttaagtaa atttttgtt agaaaaatga gattttatat	8220

-continued

cgtattttat	ttgttgattt	tttaaatata	tttaggttat	tttaaagttt	attttttagtt	8280
aattaataga	ttaatttcgt	tttagattgt	tataaagagt	togtttgga	tggtaaatat	8340
tgttattcga	taagattaat	agaattttat	gggggtaaat	atggaagggt	tgcgagtgt	8400
attgaggttt	ttgtgtttta	atgaggaatt	ttagcgggga	tttggcgggg	attggaagtt	8460
tagtatttat	ttagggagtg	attgagttgg	aaatttgag	aagtttggtg	tagacgagac	8520
ggtgtttttt	agagttgtat	tttaagtttt	tttattttgt	tgtatattgg	aagaattgga	8580
ttttaaagat	tttttttatt	attttagatt	ttgtagtta	tagttacgtt	tttttttttt	8640
tgtttgtga	aagttgtgat	ttaattgttt	taagtatttt	ttgataattt	cgttaaattt	8700
attttgaaaa	tatttatata	aagtttttaa	ataggatgtt	gtgagtgtaa	tataaaaaata	8760
ggaaggatt	aatggaagta	ataattagtg	ttttgagata	gaagcggttt	ttagtttgta	8820
atttggaat	gttaagagta	gtatattagg	gagaagggtt	ttaagagttt	gattttttgt	8880
tttgagtttt	gagtaaagaa	agaaatatgt	tagaattgta	ttgtttgttc	gtagtaatat	8940
tgtaaagggt	tttaattgat	tttaattgta	ttgagtaatt	atttttttta	gtattaagtt	9000
tatagttttt	aatttttttg	tgagaagatt	aggtaggaag	cgggatgaga	tttaagataa	9060
ttatttagga	tttatgttag	aattttaagt	tttgattttt	ttggtagata	aaaatagagt	9120
tggttttaaa	ttgtattttt	tggttttaaa	aaaacgaggt	ttgaaaggaa	tttttttaga	9180
tttatatagg	tttatagatg	taagtttttt	tttttttttt	gtgtttgtta	attggaggga	9240
gatgtttttg	atatttttaga	atatttgtat	ttatatattt	gtaaattttg	ttaggaaagg	9300
atttaaatag	ttaaattaaa	ggtgattatg	attaatgtta	aagataattat	ttattgtttt	9360
ttttttgtt	gtattattag	gtttaaatat	tttgagttta	atgattgggt	atttaattga	9420
gtaatagagt	ttttgagagc	gtaaagatta	attaattttg	ttttttaaag	tttggttaatt	9480
aagttttttt	tttttggtga	gatttaataa	tttttggtga	cgttatttat	ggataaatat	9540
cggtaaatat	tgtatttaaa	atgggatatt	gttttggtat	ggagtagaat	tttttaagat	9600
atatggtttt	tttgaaagaa	agaattaatg	attgtagtaa	ggatattttt	tattattttac	9660
gtagagaaat	tagttaggat	tgataaatat	ttattgttgg	gagttatttg	ttttggtttg	9720
agaaaatatg	ttgattgtta	tataggtttg	tattttatta	gaaaatataa	agtatgtatt	9780
ttttttgatt	tggaattaga	aaagggggat	tagatatgga	tgttttggtt	aagttttatt	9840
tttaattagtg	aatttttgtt	agtttttttt	tgtagttag	atttggtaaa	ttatatatat	9900
ataaagtttg	tgtgtgtgtg	tggggtgcgt	gtgtgtgtgt	atgtatatag	ggtttttaaa	9960
tattgttaag	ttttttattat	tgaatgtgtt	agagggtggt	gtattttaga	ggtaggtata	10020
aagaattttt	gtttaatgtg	ttagttttgt	tttttgaaat	atatttttgg	gtagattttat	10080
taattataaa	aggatataaa	tgtatatatt	ttttatat	gtatatattt	tattttattt	10140
ttttaaggat	tttaaggagt	gtttatgttt	ataaaatgag	attatatatt	ttataataaa	10200
tatatattgat	aaattataat	ttttttatta	ttaaattatt	gatttgatat	tatttgaggt	10260
atagtgaat	gagtattaga	tttggaaata	gaaagtttga	gttggtttta	ttagttgaga	10320
aaattaagta	aagatattat	ttttggattt	ttagtggttat	aatttattaa	atgaatatgt	10380
agttataaag	attaaaagag	taagtatttt	gaaaaattat	ataaaaaagg	gatgataata	10440
tattatgttt	tgggatttat	attgatttaa	ttttataatt	ttataataa	ttgtattatt	10500
taagttatgt	agtttttatt	gaattttata	aataaattag	agaggggtat	attgttttaa	10560

-continued

aaattgttta gttttattaa ttagaagata gattatagaa attgatttta atgaagttgg	10620
atatttagtg gaaaggaaga attatttaaat aaatggtttt aggataattg atgggtttat	10680
gtaaaataag aattagtttt tatttgattt tttattttta aagaaatata agatgaatta	10740
aatatgaaat gtagggaata attttttata aaatgataaa agaaaattat gggattaaaa	10800
aaaaaagtt aggaatagag aaatttgttt taagtaaaat ataaaattta gtgttatttt	10860
aggaaatatt gataggtttg attaagtaaa aattatataa aagtaataa tatatttgga	10920
aaatatttac gatatttgag agataaaaag ttaataattt ttgatatttt atggattttt	10980
atgaataatt ttgaaaagt tatattttatt aaaatatgag taaaaattat gtataggtag	11040
tggtagaaag agatatagag ttttaaaatt atagaaaaa tcgtttagtt ttatttgtaa	11100
ttaataaat agtaattaa atattaatga tatgttatat tttttttta ttaattagag	11160
gagtaataaa aaaaaataat tttttttttt tttttttttt ttgagatgga gtttgttatt	11220
taggttgagg tgtagtggta cgattttagt ttattgtaat tttcgttttt taggtttaag	11280
taattttttt gtgttagttt ttttagtagt tgggattatg ggtatttatt attatgtttg	11340
gttaattttc gtgtttttag tagagtgggg ttttattatg ttggttatgg tagttgattt	11400
tgaatttttg attttaagt atttatttat tttagttttt taaagtgttg ggattatagg	11460
tatgagttat tgtgtttggt aaaaatgaaa ttaaaaagtt tgatgtaagt ttgtgttggt	11520
taggttgagg agaaattggg ttttttatat attgttgat gaaagtaaat aggttttttt	11580
attttaaaaa ataatttggt aattattttt ttttgatttt ttattaaaag tttttaaaag	11640
atgtatatat tttttgattt agtaatttta tttttcgata ttattatatg gatggatatt	11700
ggattaataa ttatttaaga taaatggata aggatgtttt ttgtagtatt gtttttaaga	11760
gtaaatattt ggaatagtg taataggttt attataggga attatttaa taaaagttgt	11820
tgtatttatt taatggaaaa ttatgtagtt gttaaaaa atagggaaaa attgtatttg	11880
ttgatttga atatttttta aaatatattg ttttttaagt gaaaaaattg atagataaaa	11940
ttgagtatat atttatatt taattatgta aaatgggtgt atatttttt ttttaataaa	12000
tatttattga gtgttttagg tattgtggat ataggagtga ataaaattt aaagattgtt	12060
ttgttgaat ttttattatg gtaggtagag aaaggtaatt ttaattata tatgtaaaat	12120
ggtgatatgt gtaggagta aaatagagta tgggaaaggg aatagagtat agtatgtttt	12180
aggtagagta atgagaaagg agttttattg gtagagaata gaggagtgtt aaggggcat	12240
aatgtgaaaa agtttagtta agtttaggtt taggaattta tatgatgta ttatttgat	12300
gaagtaaat tatgaaattt cgtattttaa ttacgtgatt tttttttggt gtataatgat	12360
tttttttggt agaaagaatt agtgagta atgatatgt atgtttatta tatttatttt	12420
gtagttgaag tatttttaga agttgggtgt ttaggtatt taattagtag aatggtggat	12480
aaaattatcg tgtttttgtt ttttaagttat tttttgtag ttgagttttt ttaaaattaa	12540
gggagaatat taaatattta tttaatagaa ttttggtttt ttttcggagg ttgtttattt	12600
ttggtgtggt ttttattttt tttgtagata ttttttatgg ataggagttt tattgatttt	12660
gttttttatg ttttttaggt aattaaatag ttttagtatt ttgtatttg attgtttaat	12720
gatttttttt tttttttatt gattttattg gagtagaagg ttaaatatag gtatttattt	12780
ttatattttt taggtttttt tatgtagata gaggttgtat tattagtaga cgggtgtttat	12840
aagagagata tggaaaagt gaagaatgga aatagggtt tggagtgggt ttaaaaaata	12900
agaaagaggt ttttgaaagt tttttattta atatgattaa atatagaaga taaaagatgt	12960

-continued

```

aatttaatga ttaaggatag ggatatagtt ttgagaaaat tatagaagtg gaatgaacgt 13020
tttggatttt aaagaagagg gaatggaaat gtagttttga ggtagttaaa gttaataaag 13080
ttttttgata tttttattgt agaaataaat gattatTTTT taatttaagg tttgtaaatt 13140
taggtgatta tagggcgag gtatggagtg gaaatgggta agagattaag tgggggtagt 13200
aatggttttg agtatataga ttaagggga gagattgttg tttggttaat gtttatgatt 13260
agttagttga aatatagatt tagtggtgtt agattttttg attttttgaa agaagttaga 13320
gattttatat ttagtgtaaa tagtttaatt tttaaattga gtaatttatt taaatatatt 13380
tatgtaattt atttaaatgt atttagagta tgtttgattt tttttttgta tttttattta 13440
gttttagggg taattttaag gaaagtgatg cgtatatttg ttgtatttat gtagatggag 13500
gattatTTTT ttttttttaa taaaagatat ttttataagg attgagtgtg gtataaagag 13560
ttcgattttt agtattttaa ttgagtagga aatttaaata tagaagattt tttttgggt 13620
tataTTTTtg gtttttttaa atgagatttg attttttaga tattgataaa tattatatta 13680
ttgggttatg gtaattttt aagtggattg atttaaaagt gtttaaaagt attttaggtg 13740
ttaaggaatt gtgtgtacgg agttttaaag gtgtttttgt ataagaaata gaggcgttga 13800
ttatgtattt gatttataaa ttttttttaa ggatatattt tgaatttggg aatgaagaaa 13860
tattgtggga attattaagt tgaataaagt attagttttg ttttttagaa gtttattgtt 13920
ttatagagga gattagatat gtttatagat aatttgaagt atatagtaat aaattatagg 13980
atgatttatt ttaagtataa agtgtatat gttattatag tggtaattta ggttataatt 14040
tttaaatgta agagttttta atattatttt ttttattttt agtgatatga aattatgtta 14100
gaatatttgt agaaagtttg aattagtaat ttaataata aatattgagt attgatatat 14160
tgtatattat tgagatttat tttatatata ttattttatt aatttttatg attatttcgg 14220
taagtattaa tttttataat taagtagtga tagaaaatga ggtttagaga gatatagtat 14280
tttattttaa gttatagaat ataagtagag aatgtagaat ttgtattttg ggttatttga 14340
tagtggagtt tagattttta aaattagggt tataaagatt ttttgaaaa aaagaaaaat 14400
aaaaataga gaataaattt ttttgaagaa aagtaggaaa ttattattta taagtataat 14460
taaagagaga tttttaattt tttttttttt tttttttttg agacggagtt tcgttttgtc 14520
gttttagcgt aattttagtt tattgtaatt tttatttttc gggtttatgt tataaagaaa 14580
gtttttattt gattggttcg atttgattgt ttattagaat ataggttttt tagtaggggt 14640
aataagttat aatagtttta gtgatataag ataaatataa attaataata aggttaattt 14700
agaattatta atatatgttt tttattgtga attattagtt ggtttataga tggaaaatat 14760
aagatatata atgtattagt ttaaaaaatg ttttttagag agtgttttaga attatagaat 14820
ttaaaaattt tgttttgaaa aggtttaatt tttattttta aatatataaa gaaattaaat 14880
gtaaaaaat taagagaaaa gggtttaatt aattaaatgt tgtataattt taagttttat 14940
ttatgtttta gattagtttt gtaatttatg attaatTTta gtattagttt gtttttaatt 15000
taggtttgtt aaataatata gaattttttt tttaatgaag ttttttaggg ttttagttga 15060
aaaaatatata ttagttagga tgtttttaat tgtaaggaaat aaaaagttac gattttattg 15120
ggtttaaata ataagatat ttataatttt atataaaatg aagtttaaag gtatggttta 15180
tttaggtttt aatttttaag gatttacgtt ttgtttttgt tgattttttt ttagttttat 15240
tttagaggtt attttattta tttatggatg taaaattttt ataatagttt tagtagaaag 15300

```

-continued

```

at t t t a g t a t g a t a a t g t t t a g g g a a g a a a a t t g a t t a t t t t t t t t t t g t t t t t t t t t 15360
a g g a t g a a t t t a t t t t t t t t a g g g a t t t t t t a t t g g a g g a a g a t t t t t t a t a t t a t t t a 15420
t t g g t t a a a a t t a t a t a t t a t a t t g t t t t a a a t t a a t t g g a a a a g g a a t g g t a t t 15480
t t t a t a a a t g a t t a g g a t a t a g g a t t a t t t t t t g g g t t g g g g t t t g a t t a g g a t t t t 15540
t t a a a g t a t g t g a t t g a g g g t a g g t a t t a g a a t g g a a t t a g a g t t g t g t t a g a a a g g a g a 15600
a a c g t a t g g t t a t t a t a g a a t a g t t t t t a a t g t t a t c g a g a g g g t a t a a t t g t a a a a a t 15660
t a a a t a a t t t t t g t t a t t t t t t t g a a g t t g g t a t t t t t g a t t t t t a g a t g g t t t t t a a t 15720
a t a g g t t t t t t t t t t t t t a t t a t a g t t g t t t t g a a a t t t g a g t t g g a a g g g a a t a t 15780
t t t g a g a t t t a g a t t g t t a a a t g t t t t t t t t a a g t t a t g t a a t a a a t t a a a t g g t a a a g 15840
t t a g g g t a g t t t t t g a t t t a g t a t a g g g t a t t t t t t t t t a t t t t t t a t t t t g a g a t t t 15900
t a g a a t t g t t g g t a t t g t t t t a a a a a t t a t g g t a a g a a t t g g t a t t t t t g t a a t t a a t a 15960
t t t t t t t a t a t a t t t t g t t t g t t t g t t a g t t a g t t a g a a t t a t a t g g a g t t t g t g 16020
t t t t a a a a g t t g t c g a a g t t t t t a t t t t t t t t g t t t t g g t a t t a t g t g t a t g a a t t a t t a 16080
a t t g g t t t t t t t t a t t t t a t t t a t t t g a t a g a t g t t t t t t t t a a t a t t t t t t t a t 16140
t g t t t t t t a t t t t t t t t g t t t a t t a t g a t t a g t t g t t g t t t t t a a a t a g a t t t t g t 16200
g g t t a t t t a t t t t t t t t g t t a g t t t t t a t t t a t t a g t t a t t t g a a t t g t g g t t t t t a 16260
t t g t t t t t t a t a t a g t t t a t t a t t t g t g g t t g t a t a t a t a t t a t t a t a t t g t t a t a t g t t 16320
t t t a a a t t g t a t t t t g g a a t g a t t t t g g t a t g g t t g t a t t g g a t g a g a t t a a a t t a a t 16380
a a t t a a g t a t t g a g a t a g t t t t t g t t a t t a t a a g t t t a t t t t g t t t t a t a g t t t a a g 16440
a g g a g t t a t t t t t t t a t t t a t t t t a t g t t t a a t a t t a t t t t a t t a t a t a a t 16500
g t a t a a a a g t a t g t g a t t t a t g a t t t a t t t a a a t t t g a a t g t t t g t g a t t a t t t t g t 16560
g t t t t t t t t t a t t t t a g g 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

```

t t t a g a a a t g a a g a a a a t a a a g t a a a t t a t a a g t a t t t a g g t t t a a a g t g g g t t a t a 60
a a t t a t a t g t t t t t g t g t a t t g t a t g t a a t a a g g g t a g t a t t a a g t a t t a a g t a a t a a g 120
a g t a g a a g a a g t a g t t t t t t t t g g a t t a t a a a a t a g g a g t a a a t t t g t a a t a g t a g a a g 180
t t a t t t t a a t g t t t t a a t t a t a g t t t g a a t t t t a t t t a a t g t a g t t a t t a t t a a a a t t a 240
t t t t a g a a t a t a g t t t t g g g a a t a t a t a t a a t a t g t g g a t g t g t g t a t a a t t a a g t g a g 300
t a g g t t g t a t g a a g a a t a a t a g a g a t t a t a g t t a a g t a a t a g t g g a t a g g a g t t g g t a 360
t a a a a g a a a t g g g t g a t t a t a a a g t t t a t t a g a a a t a g a t a g t t g a t t a t a a a t a g a g 420
t a a g a g a a g a t g g g g a g t a a t a g a g a a g a t g t t g a a a g a a g g t a t t t t t a t t a a g t g t 480
a a g g t g a g a a g g a a t t a a t t g g t g g t t t a t g t a t a t a a t t a a a a t a g a g a a t a g a a t a a g g a t 540
t t c g g t a g g t t t t t a a a g t a t a g a t t t t a t g t a g t t t t t a g t t g g t t a a g t a a a t a a a a 600
t a a a t g t a t t a t a a g g a g g t g t t a a t t a t a a a g t g a t t a g t t t t t g t t a t g a a t t t t a a 660
a g t a g t a t t a t a g t t t t t a a g t t t t a a g a g t a a a a t g a a g a a a t a t t t t g t a t t g 720
a g t t a a g a a a t g t t t t t g g t t t g t t a t t t a a t t t a t t g t a t g a t t t t t g g a a a g a t a t t 780

```

-continued

taataatttg ggttttagaa tgtttttttt tagttttaa	840
gaggaggaga agaatttgtg ttggggaatt atttagaa	900
aatggtaaaa gattatttgg tttttatagt tgtgttttt	960
tttataatag ttatgcgttt ttttttttta atataat	1020
ttttagtatt atattttgga gaattttggg ttaaatt	1080
attttaagtt aattatagag gtattatttt tttttt	1140
ataatatagt ttgtgttagt gagatgtgag gggatt	1200
ggaaagatag atttattttg agaagagaat aaagg	1260
gatattatta tgttgaatt tttttgttg aattgtt	1320
aagtgaata tttttgaaa tgaagttgag aaagaat	1380
ttaaggatta gatttagagt ggattatgtt ttgaat	1440
tgttattatt gtttaaattt aataaaatcg tgatt	1500
tttaattggt atatatattt tagttgggat ttgaaa	1560
atattattta gtaaatttag attgaaaata agtta	1620
aaattgattt tgaatataaa taggggttgg agttg	1680
ttttttttta ttttttgta ttgattttt ttatat	1740
tttaaaataa aatttttaaa ttttataatt ttaga	1800
ttaatatatt atatatatta tattttttat ttata	1860
aatatatgtt gataatttta aattaatttt attat	1920
agaattgtta tgatttgta attttattgg aaagt	1980
ggattaatta aatagaaatt ttttttatgg tatga	2040
gttgagattg cgtttggcg atagagcgag atttc	2100
aattaaagat ttttttttaa ttatatattt aaata	2160
aatttgtttt ttatttttta tttttttttt tttt	2220
aaaagtttgg gttttattgt taaatagttt aga	2280
ttttgtgatt ttgggtaaaa tattatatatt ttt	2340
ttgtgaaaat tagtatttat cgaagtgatt atagg	2400
taaatttttag tgatatataa tatattagta ttt	2460
agattttttg tagatgtttt aatatagttt tat	2520
tgggggtttt atatttgaaa attataattt aggt	2580
tatgttttga ataaattatt ttataattta ttgt	2640
tgtttaattt tttttgtaga atagtaggtt ttg	2700
tttaataatt ttataatat ttttttattg ttag	2760
ttgtagatta aatatatgat taacgttttt att	2820
cgtgttaata attttttaat atttgaaata ttt	2880
aagatttatt atgatttaat aatatgatgt ttgt	2940
taaaaaatt aaaggtgtga tttaggaaag gat	3000
ttaaatatta aaaatcgaat tttttgtgtt at	3060
tggggagagg agaatggttt ttattttata tag	3120

-continued

ttaagattat ttttaagatt aggtaagaat ataaggagga gattagatat gttttaaata	3180
tatttgaata agttgtataa atatgtttga ataagttatt taatttaaaa attagattgt	3240
ttatattaaa tataaaattt ttaatttttt ttaaaaaatt agaaggttta gtaattataa	3300
gtttatattt tagttggttg attatgagta ttaattaagt agtagttttt ttttttaagt	3360
ttgtgtgttt aagggtatta ttgtttttat ttagtttttt atttattttt attttatgtt	3420
tcgtttttgt agttatttaa gtttgtaaat ttgaattaa aagatgatta tttattttta	3480
tagtaaatat attagagggt tttattgatt ttagtgttt tagaattgta tttttatttt	3540
tttttttttg gagtttaaga cgtttatttt atttttgtaa tttttttaag gttatgtttt	3600
tatttttggg tattaagtta tattttttgt tttttatgtt tgattatatt aaatggaaag	3660
tttttagaag tttttttttt tgtttttagt attattttta atttttattt ttattttttt	3720
atttttttat gtttttttta taagtatcgt ttattagtga taataatttt gtttgtatgg	3780
aaagatttga gagatatgag agtgaatgtt tatatttagt tttttatttt agtgaagtta	3840
atgagaaggg aagaaagtta ttgaataatt taggtatagg gtattgagat tgtttaattg	3900
tttgggagat ataaaagata aggttaatgg aatttttatt tatgagaaat gttttagggg	3960
aggataaggg atatatagg gatgggtagt ttctgagaga aagttaagat tttattaaat	4020
agatatttag tatttttttt tgattttggg gaagttagt tgataagggg tgatttaagg	4080
atagagatac ggtggttttg tttattattt tgttaattag gtattttaga taattaattt	4140
tttaggatat tttagtata aagtggatgt aatgagtatt atatattata ttgtttattg	4200
gttttttttt atagaaagaa ttattgtata gtagagagaa attacgtgat ttaggtgcga	4260
aattttatga atttatttta tataagtaat agtattatgt gaatttttga atttgattt	4320
ggttaagttt ttttatattg tcgtttttta gtattttttt gttttttgtt tgggtgaattt	4380
tttttttatt attttattta agatatattg tattttattt ttttttttat gttttatttt	4440
tatttttagta tatattatta tttgtatat gtgattggga attgtttttt tttatttatt	4500
atagtaaaag ttttaataga gtagtttttg atgttttgtt tatttttgta tttataatgt	4560
ttagaatatt tagtaaatat ttgttaaaag aatgaatata tatttatttt atataattag	4620
atgtataata tatatttagt tttatttatt aattttttta tttagaatat aatatatttt	4680
ggagaatgtt ttaaatttagt aaatataatt tttttatgtt atttttgata gttgtatagt	4740
tttttattag gtggatatag taatttttat ttaaatgggt ttttatgatg agtttgttat	4800
attgttttta gatatttgtt tttagaagta atgttgtaaa gaatattttt gtttatttgt	4860
tttgggtgat attggtatta atatttattt atataataat gtcgagaagt ggaattgttg	4920
aattaaagag tatatatatt ttttaaaaat ttttgataaa gaattaaaaa aaagtgattg	4980
ttaaattatt ttttgaagta gaggaattta tttattttta tataatagta tatgagagat	5040
ttagtttttt tataatttgg ttaatataaa ttgtatttaa attttttaat tttatttttg	5100
ttaggatatag tggtttatgt ttgtaatttt agtatttttg gaggttgagg tgggtgaatt	5160
atttgagggt aggagtttaa gattaattgt tatggttaat atggtgaaat tttattttat	5220
taaaaatacg aaaattagtt aagtatgggt gtgggtgttt ataatttttag ttattaggga	5280
ggttgatata ggagaattgt ttgaatttgg aaggcggagg ttgtagttag ttgagatcgt	5340
gttattgtat ttttaattga gtaatagatt ttatttttaa aaaaaaaaaa aagaaaaaaaa	5400
ttattttttt ttgttatttt tttgattggg gaaaaaatgg tgtagtatat tattggtgtt	5460
ttaattatta tttatttaat tatagatgag gttgagcgt tttttttata attttaaaat	5520

-continued

tttatgtttt tttttgttat tatttgtata taatttttgt ttatatatttg atagatgtgg	5580
attttttaaa gttattttata aaagtttatg gaatattaag agttattagt tttttgtttt	5640
ttaaatgtcg tagatgtttt ttaaatatat tatttgtttt tgtgtaattt ttatttagtt	5700
agatttatta gtatttttta gaatggattt gggttttatg ttttgttttag aataggtttt	5760
tttattttta atttattttt ttttaatttta tgattttttt ttattatttt gtaaaaagtt	5820
attttttata ttttatattt gattttattt atattttttt taggatgaga aattaggtag	5880
agattgattt ttatttttga tggatttatt aattgttttg aaattattta ttgaataatt	5940
tttttttttt attgaatgtt taattttatt ggagttaatt ttgttaattt gttttttgat	6000
taatgaaatt aataattttt taaaataata tgtttttttt tgattttatt tatagattta	6060
gtagaaatta tatggtttgg gtaatgtagt tattattaaa attatagaat tagattaata	6120
tgaattttta gatataatat attattattt tttttttgta tagtttttta aagtatttat	6180
ttttttgatt tttataattg tatatttatt taatagatta tgatattaaa ggtttaaaag	6240
taattattttt gtttggtttt tttagttagt gagtttaatt taaatttttt gattttaagt	6300
ttagtgttta ttttattata ttttaagtaa tattaaatta ataatttaatt aataaaaaga	6360
ttataattta ttaaatatgt ttattgtaaa atatgtaatt ttattttata gatataaata	6420
ttttttaaaa tttttagaag aatgaggtag gatgtatata gatgtggaaa gtatatatat	6480
ttgtattttt ttatagttag taaatttatt tagaggtgtg ttttaggaaa tagaattgat	6540
atattaaata aagatttttt gtattgattt ttgaaatata attattttta atatatttaa	6600
tggtagaat tttagttagt ttaaaaattt tgtatatata tatatatata cgtattttat	6660
atatatatat agattttata tatatgtaatt ttattaagtt taggttataa gaggaattg	6720
gtaaaaattt attaatgaa atggagtttg gttaagatat ttatatttta ttttttttt	6780
ttgattttta attagggaga gtatatgttt tgtgtttttt ggtaaaatat agatttatgt	6840
aataattagt atattttttt agatttagagt aagtaatttt taataatgag tgtttgttag	6900
ttttaattgg tttttttacg tgggtggtga aaaaattttt tattatagtt attaattttt	6960
ttttttaaaa gaattatgtg ttttgagaag ttttgtttta ttatagggtta atattttatt	7020
ttaaatatag tgtttatcga tgtttgttta tgaatgacgt taataaaaat gtttaaaatt	7080
tattaggaaa ggagggttta gttgttagat tttgggagat agagtttagt gatttttgcg	7140
tttttaagg ttttgttatt taattaagtg attaattatt aattttaaag tatttaagtt	7200
taataatata ataaaagaga gaataatggg tagtattttt aatattaatt atgggtattt	7260
ttggtttagt tgtttaggtt tttttttagt agaatttata agtgtataga tataggattt	7320
ttggaatatt aaaggtattt ttttttagtt tgtaaatata gggaaaaaga agaaatttat	7380
atttgtaaat ttatgtaaat ttagggggat tttttttaaa tttcgttttt ttgaaagtag	7440
gaaatataat ttaaaattta ttttgttttt atttattaaa agaattaaaa ttttgaattt	7500
tgatataaat tttaaatgat tgttttgaat tttatttcgt tttttattta atttttttat	7560
aagaagatta aggattgtgg atttgggtgt ggaggaagta attgtttaat tatattggaa	7620
tttattggaa ttttttataa tattattgag ggtagataat ataattttta tatatttttt	7680
tttttattta gagtttagaa taagaagtta gattttttgga gatttttttt ttgatgtatt	7740
gtttttaata tttttaagtt ataaattggg aatcgttttt attttagaat attaatgtt	7800
gtttttattg gtattttttt atttttatat tatatttata atattttatt ttaggatttt	7860

-continued

atgtgaatat ttttaggatg gggttgccga aattgttaag aagtgttta aataattagg	7920
ttatagtttt tagtagataa aaagaaaaag acgtggttgt tagttataga atttgaggtg	7980
atggaaaaga tttttaaaat ttagtttttt taatatgtag taaaaaggga ggatttaaaa	8040
tgtagtttta agaagtatcg ttctgttgtt attaagtttt tttaggtttt tagtttagtt	8100
attttttaaa taggtgttga gtttttagtt ttctgttaggt ttctgttaag gttttttatt	8160
gaggtatagg aatttttagtt atattcgtag tttttttatg ttgttttta tgaaatttta	8220
ttagtttgtt cgaatggtag tatttgttat ttaggcgaa tttttatga tagtttgaa	8280
cgaattaat ttgttaattg attaagaatg gggttttagag tgatttgaat gtatttaaaa	8340
agttaataaa tggaatacgg tataaagttt ttttttttg ataaaaagtt tgtttggagt	8400
aaattttaat ttttttttg ttttaaaata aaagtacgat gttttttta ttatttttaa	8460
gaagtaagta aaaataaaat aaaataaaat aaaatttaa tagttttta ataagagtgt	8520
ttagcgttgt ttggatggaa ttccgggttag ttttagaagt agtcggttc gaatttcggg	8580
ggcgaggtta ttccgttttt tcggttaagt tgggttttgt ttttttcgtt tcgtttttat	8640
tggtttcggt ggggtggggt ttccggttgc tcggtcgtt gtgcggttgt attgtagttt	8700
ttgagtttag gaagtatat tcgttcggtt tgggagtcga agcgttaatt tacgtggcgg	8760
ttggttaatt tagtcgttgc ggggagcgtt cgttcggtt ttttttttg cgcgaggtga	8820
gcgagtgagg gcgggtttta gatttagttt ttttgggcgg ggttggaata ttttcgtttg	8880
tttagttttt ttttattttt ttatttcggt acgtttttgt agttatattt gttgtcgttg	8940
ggaaagggaa aagagtattt taaaacgagt ttcgaaattt agatttaaat ttcgaaacga	9000
attttttcgg ttctgttttt gttgtttgaa tgttatttgt tttatttcg tttttttttt	9060
tttttttttt tttttacggt tttttttttt tttatttttt tttttttttt tttttttttt	9120
tttttattgt tcgttttttt tttatttttt cgttcgacgt tcgttttgaa gttgaggaag	9180
tttaagtttt ggattaggag tttttttttt gtaatatatt ggggcggggt tttatcgttg	9240
ggtttaagtg gttatgggtt ttgagagtac ggtatgtgta ttattttttt taaagacgtt	9300
tagggtggcg cgtttagatt ttcgaaattt ggttttattt cggttatgtg cgttcgttta	9360
ttaagcggcg aattggattt agggcgtcgc gtgcgtttta gtcgattcgc gagaaggat	9420
ttttgtattt tgtatagatt aagatttttt tttagggttt tttaaatttt tagatgtacg	9480
tcgtagattg tggagtgtaa aatgaatgat ttaagggaat attagggttt tttatttttg	9540
aatataaggt tgtgaagat gtacgttttg tttggggggt ttgttttttt tcgtttgtgc	9600
gggatttttt tattaggggg cgtttatatt cgtcgtaaaa tttttaattt ttagatattt	9660
aggttttgta agttttgtga gttgaattgg gggtagattc gaaggagcgt atcgtttttt	9720
tcgttttttt ttttttgttg ttttgagta cgggaaaggt agggatttta tattagttat	9780
ttatttgta gttgtggggg aggggacgta gaagttagag agagagaata gaggttggga	9840
gaggggaagg gaggtagaga agatggagat agaggagat ataaagatga ggaacgcgat	9900
ttagagagat ttgtcgcgg gtgtgcgcgc ggtaggataa ttttttaggc gtcgtgtaga	9960
tttttttttt ttttcgtagt attttttagg gatatgtagg atttgcgag acgtagtgtg	10020
attggagagt aaatagttta ggttatgata ttattttttt atttcgtttt tttttgttta	10080
atatttagat cgttttacgt ttagtgatat agatcgcggg ttagaataat tcggatgttt	10140
taggtttgta cgtattaggg atgtcgtttt atagcgttta cgtttttgta ggacggtggg	10200
gcgttttagt caggagcgtt ttccggggtt ttccgtatac gcgcgcgtcg ttgtcgttag	10260

-continued

atttgtagt aaagttgttt aattagagcg gagaattagg agagggtggg gaatatgtaa 10320
 ggaaacggag gatttttgta gtttgggcga gaggattgat ttagaggttt ttgattcggg 10380
 ttttcgagtt atttttgttt gtagtgcggt ttgggtgctg gtttgtttgt tgagatggga 10440
 gcgatatgatt ttattttatt ttttaggggt agaaaaattg ttgtatat ttgtcgattt 10500
 agaatagggt ttttcggtt aggggttttt tagattttcg cggggtttta ataattttat 10560
 tttttatttt ttgttttttt tttgttggtt ttttcgtatc gcgtgtagga taaaagaatt 10620
 taggggcgtc gggagagaat cgtgatagtt ttaagtcgt taaagagata cggcggggat 10680
 tttgtatttt tttcggtcgg attttttttt agattgtttg cggtttgaag gggtagggtt 10740
 cgcgtaagggt ttaggttgt ttgggcgtcg ggtcgtattg ttaggagttg ttcgggagtg 10800
 gttttttagg aatattagggt attgattata tattaggggt tgggaaatta ggtggtttgt 10860
 attaagcggg tttcggggat ttgtttgtgg taagttttgg tagtttttat ttaaattttt 10920
 gtttcgagcg tgttaagaat aataataata aaaaaattaa agtgtaaag gttttttttt 10980
 ttttttagt ttaagaattt attatttttt tatgattttt ttataattta tttttttttt 11040
 ttttttaatt tcgttagtta ttttattttt attttatttt gggttttttt tgtttgaatt 11100
 ttttttaata ttaaggtttt tttgtatggt tttttttaaa agtttttatg aaagtatttt 11160
 gtatttttta agtgtttata tttttttaat ttcgttttat agtttttgt ttaatttaa 11220
 gttttttatt aattgttttt tttttttaag ttcgcgggtt tttttttata agtttttgt 11280
 ttttgttttt taagggggga ataaaagaaa cgtgattatt ttggaaggcg gtttattgta 11340
 gtttgggggg aaaatttatt gtagcgttc gcgattgggt tcgpcgttgt ttaggcgggt 11400
 tatataggaa gcgtggtggt tcggggaagg atcgcgaggg tgcgggacgg tgttgaaga 11460
 tgcgggagga tgcgggggtg gtccaagatt ttggttttag tttttgttta taatatttaa 11520
 tgttttcggt tgttcggttc gtacgtttgg agtttttttt tagaaagggt cgggtagtt 11580
 cgtttgtaag tttaaatgcg ggttggtgata tttattatta ttatattatt gtatcgttag 11640
 agtcgaggag gagatttagc gagaagaagg aggagggaga ggaggagggt ttatttatag 11700
 gttttaaaag cgttcgttag ttttagttat ttttaagagt ttaggttttg aaagtaggcg 11760
 gagggcgga aaggtagttt ttcgcgtttg cggtagggat cgtttgtttt ttttttagg 11820
 tatttttatt ttttggcgtt tttttttgat aagaagtatt aatcggtttg gggatagtag 11880
 cgttttttat ttagggattt atttagtaat atcgttttgt tcgcggaaat tattgttttt 11940
 ttattttttt tttattttta ttttttcgtt tttcgtgtag ttagtttttg gttaggggaa 12000
 aggagttttt aggttttttag gggtagggtt agtaatagat aattgagtac gattattttt 12060
 tttcgggagt atataaaatt gtaaaattag taaagaattt ggttatagcg tgtttacgcg 12120
 tttatagagt ttgtttgttt tattaaaggg aagtgttagg ttttaaggtt ttgttaattt 12180
 gaaagagata ttgagaaaa gagatttttc ggggatttag agggaaagtg taagaatttt 12240
 ttattgtatt tttagggaat tgtttaatgg ggagttcgggt tttaaaagat tttggttaata 12300
 aaaggttga taggaaattt ttttaggtaa attttttgtc ggatttaaag agaataattt 12360
 tttttgtta taaatttttt tttatataag tttagatttt gtttttttcg tttttttttt 12420
 ttttagttt ttttaagtat ttttagtag aatatttgat aatttttttg agtaataggg 12480
 attttttgga agtattaatt attttttatg tttttcggaa ataagattat aattttttat 12540
 gcgtatatgc gatttttttt ttttagttag gtttatttta ttgtgtaaat agtattaata 12600

-continued

tatggaagag	ttttttgtat	tgtgttataa	aagattttta	ataggatttt	atagagaaaa	12660
gggttaaata	gttgatataa	aggatttttg	tttttttaga	aaagagggat	tttgattttt	12720
tttttttttt	gaagttaagt	atgagtttat	ataataggaa	taaaaataat	ttaagggtga	12780
tattagtata	atattaggga	tattagaatg	gatggtaaat	tttttttttt	atataaatat	12840
gaaagtattt	ttataatttt	tttttttgaa	gtttttattt	agaaaattat	attaaatata	12900
ggatttttaa	atagtagtga	ttaaagatga	aagttaattt	ttttttttaa	tttttttgat	12960
tagtttggtt	taaatttatg	ttttaaaatt	tttagtaatt	tagatttttg	tatatgcgta	13020
tttataagat	tttttttata	ttttgtaatt	tgtaggtatt	tagtatgggt	attgatttta	13080
agtataaat	aggtagaaga	tttgtagga	taattttagt	ttttattatg	tgtatttttt	13140
attttttttg	ttatttatta	gttttttgta	gtaataaat	tgttagttat	attatttgga	13200
aattaatgtg	ttttttagt	gaattattta	tttttagtaa	atttattgta	ttattatttt	13260
aaatattgaa	atttgatttt	attagatggt	tagaatgata	gttatagagt	agaattagat	13320
tttaattttt	atgtataata	tgtaaaaaa	ttaagggtga	gaaagttatt	gttttaaagt	13380
tatataatta	atagtttagt	ggttttttta	gtgttaatta	tagtaatttt	agatttaaga	13440
ttgtttgatt	taggaatgga	gagaaaaaa	ataaatgaa	tgtttattga	aatatatatt	13500
ttagatatta	agatagattt	ttttgtgaaa	tattttaaga	tgtggatttg	tatttttatt	13560
tttaaatttg	agtaaattta	gttgagtttt	aaagagttta	agtaatttgt	ttaaggttat	13620
tttgtaagtg	atagatttgt	tagtttgaaa	agtttattat	tttttatttt	tttgtttttt	13680
tatgatatag	ttaagtaaaa	ttttttaata	aaggagaatt	agattagaaa	taagtaatta	13740
atagtagatt	aatgtttatt	tttagggaaa	tgttaatagt	taatatttat	tataaataat	13800
gttatttgta	tataatggaa	tatgattgaa	agtgatgggt	tagataattt	ttatattttg	13860
agatgttttt	atttatgttg	ttaaatttga	ttttttaatt	ttttaaaatt	tttagttttt	13920
tttaatgtta	atagatttta	taatatttga	ttaaatattt	ttataggata	aagggtattt	13980
tttaggggtg	atttaaggaa	ttagaatatg	tataatgttg	aaaagtttta	aattgtttgt	14040
tttttttttt	ttttgatttg	taatattatt	atttagaaaa	ttaaatattt	ttttagggtta	14100
ttagataatt	aaaatgaaat	tatattttat	ggatattttt	tttttatttt	tagggtagat	14160
tttttttttt	agttgtttta	aatggagaat	ggataaaaaa	gaaaaaggaa	aatatattag	14220
ttatatattg	aatagagttt	atagttgtat	tattattttt	atgtaattgt	ttttaatgat	14280
ttttttttat	taagagaaa	aaaatgagag	agtattttat	tttgattttt	taatgtttga	14340
taaaattatt	agaagttttt	gaatgggaaa	atatatattt	tttaggtttt	ataattaaaa	14400
ttgtatttat	ttgtttatta	tatttaagta	tgaattaaat	agtaattaaa	atattatttt	14460
gtgagttttt	ttttaaatga	tatttttgta	ttggggatat	tttttgtgaa	aattttaaga	14520
ttttagggtat	aaataaagta	agatttttat	ttataatggt	tgtgttttaa	agaaattata	14580
agttagaagt	tatgtgtttt	gtttttttgt	tttgattttg	tttttgggaa	tttggttgag	14640
tttaggggta	gtaagttaat	gatttgagga	ggaaagggtg	agatggtaga	tgtgttttta	14700
gggaatttta	ttatttttat	ttaaattttt	ttttgaggtt	ttttaggatt	atatttttaga	14760
tttttttttt	tttattgatt	agggtgaagt	tttttagata	tgggtattatt	tattgttggt	14820
ttagggatga	taagatttta	gttttattta	aatttagaaa	gagtttagtt	ttaatttttt	14880
gatttgattg	gttttaggtg	taattaagta	aattattttt	attggatgat	atagtaatgt	14940
tatagattaa	aaatgtattt	ttaattttta	aagatttttg	aaaaaggaag	tttatttttt	15000

-continued

```

aagattgtgt tataatTTTT tttttaaata aaatattgta aattagggat agatatttta 15060
tttagtattt tgaatTTTT ttggtTTTT aggaaaggta gaaatataag gtttttattg 15120
tttttttttt atttttaaaa aagaaatTTT gtaaaaggta aagtgatggg ggtaatttta 15180
ttgtgtttta attcgaagta tataaattgt ttgataaat taaatataga taaatTTTT 15240
tagaaagtta taaaatatta gatTTTTTT agatattaga tattgagttt ttgtttttta 15300
atTTTtaaT taaaatgTta tagtaaaata ttgtttttaa taaatTTTT ttttttttgt 15360
gttgtaaatg ttttgTttat ggttgagtag attTTTTtag aatttaattg aaagattatt 15420
tatattacga tagaattatg taagttaaag ttaagtTTta atatggagtg atgggagagg 15480
tcgggtgtgc gagaggagat gatTTTtatta gtgagaagag aggtTTTTtg agtcggtttt 15540
taaggtagtt aaatggTTta gtatatTTgt atttattTTg ggaggtTgtt tggagtagag 15600
tattaaaaac gttattTTTT ttataaagaa aaagatgaag gtataaagaa tatagtTTaa 15660
ttaaattgaa ttaaagattg tattattTTTT attTTTTttt tagttatttt tttaaagttt 15720
tttttttttt tttttatagt ttgaaaaata taggttattt ttataaagta ttgtttgat 15780
tgaagtaaga taaattgaag agtttggttt ttgaatataa ggaagcggta gagattttag 15840
aattaatgag gtagggcgga atataatgta gaaggTTTT taagaggtaa aatttttgta 15900
gttgTgtttt atTTTtagtg attattcgtt tttatttatt taattagatt atagtgggat 15960
cgagataagg agtataatga aagttgtTgt attgatggTg taaaaagata tatggtatta 16020
ttaatgtttt tgtggtgagt aaaggtagaa cgtaagatag aaaataagga aatttttgcg 16080
ttgtaagaaa agtGtttagt atgggtatta atcggtatgt agaagtagta agtatatatt 16140
ttatttttat tttattgata aaaagaattg taggattaat gtttGttttt attaaaaat 16200
gagaaattat atTTtaattt tataaaTTaa aggttaaata tttgttgtaa taattTTaaa 16260
aaaaagagaa gatTtgtaag atttagtaaa tatttattat aaatagattg tttttaattt 16320
tttatatttt tttttaattt aataaaaagt tattgagtag ttttatgtg tattgtttta 16380
gatagtaggt attattTTaaa aaagtaataa ttattTTtta ttttaagga gtttataatg 16440
tttttagaaa gaaattaatt tattggagta gttaaagtta taattgattt ataaaatgat 16500
tagtgaattt agaggaaaat tatatataaa tatttttgTt tttattggTt ggtttagtta 16560
gatattattt taaaaaata
16579

```

```

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

<400> SEQUENCE: 10

```

```

atattTaaat tgtatatttg gtttGtaata tattatatat tTaaaagaat tattattttt 60
ttgaatatta ttaatatTaa taatagtatt atagtgatta agagtggatt taaataaagg 120
ttggtttgaa atttaggttt gttatttatt agtaagtttt taaaaatttt gagtttttag 180
tttttttata tgtgaaatgg agaaaataaa tttgtaaaat taattaaatt tgtttattta 240
tatttttttt tttagtgtat attaattggT atttttcgtt aggttagaat gtgtttttta 300
ttatgtttta aattcgtttt gtgttaattt tattgttaga atttttttta ttttgagaat 360
tagaaaagga aatattatgt ttggtaatgt ttatattttt taaaaataat ttgtaggaaa 420

```

-continued

gaatatattag taagtgatga gagtagtaac gattgttttt attattttta atttataata	480
ttattttttt tagagttttt taagtattgt agataatttt ttattttatta aaaaataaat	540
tgtaattata agtatttagg gttgatattg tttttgaatt agatagtgtt tatattagtt	600
gtataagatt aatttaaagt agaggatgaa attttttttt tgaatttttt ttagaacgta	660
atttagtgaa tatattaaaa ttaaattttt tttgaatggg agtaattttt acggattaat	720
ttgtaatttt ttagattata ttttaaggtaa tgtagagggt gttgtattat aggttttgtg	780
attagagatt attggatttg tttttggaaa agttatttat gatttttttg gttttggttt	840
ttttattttt tatatcgggt taataatgat tatcgtagag ttgttatgga agttaaata	900
atttatgtat tataagtgat taatataatg attgatattt agtgatttat taataaatta	960
aaatatttat tattaatatg atagagaagg tgcgttaaa atagataata ggtttttgga	1020
agaggtgatt aaatggatgt aaaatttatg gattgtttat ttcgtttatt tttgttgtgt	1080
tttttggttg tggatatat acgtgtgggt ataaaatcgt aaattttatg tagtcgcgta	1140
gtgtatcgt agaaggttta gatacgaat gttatttttag taatgtgttt agagaagttt	1200
tgacgtcgt ttggaagtaa gtcgttgttg ttgatattt gggcgtttg gacggatgtt	1260
tatatttgta ttttagtaga ttggaagggg tttaggtttt tcgtagtata gtttattttt	1320
agatcgttta gttttttata atatatattt ttacggaaaa gggatttttt ttcgttaga	1380
aaaagcgttt tagtttggtt tgggttggtt tttattttac gttgttgtaa gtaggcgaag	1440
ttttttttgt tttttttttt ggggtaagt gaaaggagt cggtaggggg ttcgtagtgg	1500
tttgatatagg ggaattgggt agcgagagag ttttaggtta tttcgggggt tgttttata	1560
aagtaggtgg ggcgcgatag tggtttttcg gtttagggag gagagcgcgg tcgcgggtt	1620
tttttttttag tttggaggtt gtagtcgttc gagtcggttc gggcgggggc ggggtggggg	1680
cggcgcggag ggtacggaga ttacggcggc gttattcggg atatttaggg ttcgaggtt	1740
ttgggcggtt tttacgcgag atcgtaaatt atgataatag gtagttatc gaggttaa	1800
aaaaacggag tgggtttttc gcgcgtcgtc gtttttcggt tttttggcgg tttttttcga	1860
ggttttcggc ggttttacga gttcgtagta gtcggtagcg acgtcgtttt cgttttattt	1920
ttttgcgtaa gtgcgaggtt gtcggtagcg cggcgtacgt ttcggtcgtt ttcggttttc	1980
gcgtaaaaatt tttattttgt ttacgtgaag ttgtcgttgt tttagagagg gggaaagagt	2040
tgccggaaaa gtcggggagt gacgattgcg gcggttgggc gcgttttttt attttttttt	2100
tttttttttt tttttttgtc gtagttcgga gttttgggtt tttttttttt tttttttttt	2160
tcggagtcgg tttttttttt cgtttcgttt ttttttcgtt tgtgtacgtt atttgttgt	2220
gggtggtcga aggggatgtt ttgtttttat tagaggtata gcgcgaagg gaaatttcga	2280
tattggaagg aacgagaata aatatttaac tacggacgta ttgaatcgcg gttgggatag	2340
atatttcggg aattcgaggc ggatcggggc acgaggtgag tgattttttt ttttaatttt	2400
cgttttaggg ttttcggggg agtttgagtt gagagaattt ttaaattttt cgggaaagt	2460
gcgcgaggtt cgtcggggac gtcgagcgtt gggatttag gacgcgtagt tggacgggtc	2520
gtgggcgttt gcgttttcgg ggggcgtttg gaggtcgggt gttttacgtt tgaggggtcg	2580
ggtcgttcgg atcgtagcgg tgttttttgt tttagaagac gtttttaagt ttttaagggt	2640
tttttcgagt ttgtttgttt ttttcggggc cggcgcggag tttgcgcgta acggagttta	2700
tttagtagtt tagcgcgcgg tttttatttg tatttcgttt ttatttggtta gaggcgcgag	2760
tatcgggggt ttttttatat tttttttatg acgtgtatta ttttttgatg attttttaga	2820

-continued

```

tggtttaggc gcgaggatgt tgatttagag tttttcggag gggtatagc gtttgggttt 2880
tttcgggtgc gggcgctgt gtattttaaa ggctcgcgtt ttaattttta ggtattgac 2940
gggtttttta attgcggcga ttttatttta atagttttta tgtggcgtgg attgaatgtt 3000

```

```

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

```

```

<400> SEQUENCE: 11

```

```

gatatattagt ttacgttata taaaaattat taaagtggga tcgtcgtagt tgaaaagttc 60
gattagtgtt tggagattag aacgcgagtt tttaaagtat atacgtattc ggtatcggga 120
aagtttaggc gtttgtgatt ttctgaagga ttttggtta gtattttcgc gtttgatta 180
tttagggggt tattagaaag taatatacgt tataagaaag atgtggggga gatttcgatg 240
ttcgcgtttt tgttaggttg aggcgggggt taggtagaag tcgcgcgttg gattgttgga 300
tgaatttcgt tacgcgtagg ttctgcgtcg atttcggaag ggataggtag gttcggaaag 360
gatttttggg gtttggggac gtttttagg gtagagagta tcgttcggtt tcgagcgggt 420
cgggttttta ggcgtggggt attcggtttt taagcgtttt tcggggacgt aggcgtttac 480
gtatcgttta gttgcgcgtt ttagtatatt agcgttcggt gttttcggcg gagtttcgcg 540
tattttttcg gaaagtttgg ggggtttttt aatttaggtt ttttcgggag ttttggggcg 600
ggggttggaa gaaggggtta tttatttcgt cgttcgggtt gtttcgggtt ttcgaagtgt 660
ttgttttagt cgcggtttag tgcgttcgta attaagtatt tattttcgtt ttttttagtg 720
tcgaagtttt tttttcgcgt tgtgtttttg gtgaaaatag gatatttttt tcggttattt 780
tataataaat agcgtatata agcgggggag aagcggggcg gagggagaag tcggtttcga 840
gggggaggag gaaaggagag gagttaaaat ttcggtattc gatagggggg aaaggagaag 900
aaaagaaaat gagagagcgc gtttagtcgt cgtagtcgtt atttttcggt ttttttcgta 960
gttttttttt tttttttaag gtagcgataa ttttacgttg ataggatgga agtttttcgc 1020
ggaagtcggg aacggtcggg gcgtgcgtcg cgttgcgttg agtttcgtat ttgcgtaggg 1080
aggtggggcg ggggcgacgt cgttatcgtt tattgcgggt tcgtgaggtc gtcgggggtt 1140
tcgggggagg tcgttaggga cgcggggggc gccggcgccg gggggattta tttcgttttt 1200
atttgatttc gggtgattgt ttattgttat ggtttgcgat ttcgcgtggg gatcgtttag 1260
ggtttcgggg ttttgatgtt ttccgggtgg gtcgtcgtaa ttttcgtgtt tttcgcgtcg 1320
ttttattttc gtttttattc gggtcgattc gagcgttgtt agtttttagg ttgaggggag 1380
ggattcgcga tcgcgttttt ttttttgggt cggagagtta ttgtcgattt ttatttgttt 1440
ttgtggggtg gttttcggaa ttgtttggaa ttttttcgtt atttagtttt ttgtgtagg 1500
ttattgcggg ttttttgcg gatttttttt tattttttt aagggaggag atagaaggga 1560
tttcgtttat ttgtaataac gtgaaataaa aattaattta gattagattg gggcgttttt 1620
tttgacggga gaaatattt tttttcgttg agatatatgt tatgaaggat taagcgggtt 1680
gggtagaggt tgtgttcgca agggtttggg ttttttttag tgtgttggg tgtaggata 1740
ggtattcgtt ttgacgttt aaaggttagg tagtaacgat ttatttttaa gccggcggtt 1800
gagttttttt aggtatattg ttgaaatgat atttcgtgtt taagtttttt gcgtatgtat 1860

```

-continued

tacgcgatta tataggattt acgattttat atttatacgt gtgtatgtta taattagggg	1920
atatagtaaa ggtagacgga ataaataatt tataaatttt gtattttatt aattattttt	1980
tttaaaaatt tattgtttat ttttagcggta ttttttttgt tatattgata ataaatgttt	2040
taattttattg ataaattatt gagtggttaatt tattgtattg gttattttatg gtatatgagt	2100
ttattttaatt tttatgataa ttttacgggtg gttattatta agtcgggtata aaagatgagg	2160
aaattaaagt ttaaagaatt atgagtgatt tttttaaaga taaatttagt gggttttaatt	2220
tataaagttt atgatataat aatttttata ttatttttagg tgtgggttaa gagattatag	2280
attaattcgt agaaattatt tttattttaa gaaagtttag ttttaataata tttattaagt	2340
tacgttttga aaaagggtta gaaaaaagat tttatttttt attttagggtt gggttttatgt	2400
aattgatatg agtattgttt aatttaaaag tagtattaat tttgaatatt tatgggtata	2460
attttatttt taatgagtg ggaattattt ataattgtta agagggttta gaaaagggtg	2520
tgttgtaaat ttaaaatgat aaaggtagtc gttgtgtttt ttattattta ttgggtgttt	2580
ttttttgtag atttttttg gaggggttag gtattgttag gtataatgtt ttttttttg	2640
gtttttaagg tagaaagggt tttggtagt ggggttggtat aaagcggatt tggagtatgg	2700
ttgagagtat attttgggtt aacgaggaat gttagttaat atatattgga gagaaaaata	2760
tgaatggata gatttaatta attttgtaga tttatttttt ttattttata tgtgagaaaa	2820
ttaagggttt agaattttta gaagttgtt aataaatggt agatttgagt ttttaagttag	2880
ttttatttta agttttattt tagttattgt gatgttatta ttagtattaa tggattttag	2940
aaaaatagta atttttttaa atatgtaata tattataagt taaatatgta atttaaatgt	3000

<210> SEQ ID NO 12

<211> LENGTH: 1984

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

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

<220> FEATURE:

<221> NAME/KEY: unsure

<222> LOCATION: (9, 17, 18, 33, 41, 49, 50, 52, 54, 62, 80, 84, 127, 157, 159)

<223> OTHER INFORMATION: unknown base

<220> FEATURE:

<221> NAME/KEY: unsure

<222> LOCATION: (175, 207..209, 214, 1181)

<223> OTHER INFORMATION: unknown base

<400> SEQUENCE: 12

agtttttagnt attattnnnta tttttattaa tgnntttggg nttatagggn gngntttttt	60
gnntaaatag tagttaatgn aggnatttta aaaggatat tattttatat ttatttagtg	120
gttatgntat tatttaattt ttattataat tttgtgngnt tagtataatt tgtgntttta	180
ttttatttgt gaataaatgg aggtagnnnt ttgngtaatt tataagtaag tggtagattc	240
gggggttggt ttagataaat gtggttttag attttttgtt tttttttgtt tttgttgta	300
ttttgattga gtttagttat ttattttttt aatatatatt aattgttggt tttatgtttt	360
aggttttggt ttagatgtta gaaatatatt tttatttatt ggattttggt atgggtttatt	420
gtttagtgtt ttgttatgtt acgtgttata taatgtattt atttcggtat ttagtttatg	480
gaaaaaatg ttgaaagata ttgtatgtat gtttttataa taaaattatg taatttatgt	540
tttttttatt gttttggta ttaggaaggt tttattttaa ttagagtaaa tatattagaa	600
ttggggttg ttatttttag ttgtgaatt ttttttttg gtttagattt tttattttgg	660

-continued

```

tttataaatt ttattgtata aatgtttttt attgtaaaat attttaaatt ttttttaagg 720
gaaggttgta tggaaatgat atagtaaggt tttttttgta tttttttaga tttttattag 780
ggaaggtaga tttagatggt ttttttattt ttagttttta agttttttat ttataaaatt 840
ttttaagtag tgattattgt tgttttgagt atttgagagt agtttagagt ttttgaaagg 900
taaaatttat atataaaaag aagttttttt cgatttagat tttagttttg gtgttagatg 960
tatatttgag gagtaggtgg ttagttagat ttttatttga gtagttaata aaatttattg 1020
ttttttaatg gaattttttt tgtaagttcg aattgatttg ttatttttgt gatttgtgag 1080
atggtaggga agtattaaat attattatga tttgggttat agtgggtgga aaaaaggaaa 1140
aaagaaaaaa aaaatttatt gttaagtttt gttaggcgta naaagggttg gaattgttgg 1200
ggttatttta tttgatttat tggaaataga gtggatttta ttaatathtt aataaagaga 1260
atttttttgt attaggttgg aagtggtcgt tagtttttcg tgtaatttta ttttttgga 1320
aagtgaatt agttggtatt gtttagcgtg atttgtgagg ttgagtttta atagtttaaa 1380
gaagtaaatg ggaagtattt ttcgcggggg tcgtttttcg cgagggtttt atttcgtatt 1440
tgttatgtaa aacgagggag cgtaggaag gaattcgttt tgtaaagtta ttggttttgg 1500
ttattagttt ttatttaatg ttttcgtgat gttgtgttg atttatttg gaagtgtggt 1560
ggttggcgag gttagtttt tttttaagt ttggttttta cggaaaatat gtttagtgta 1620
gtcgcgtgta tgaatgaaaa cgtcgtcggg cgttttttag cgataaaaat gtagtcgaga 1680
atttcgttcg ttttgtcgt tttttgttt taggtaggga agagggttg tcgggcgcgt 1740
tttgcgttc gtttttgtat tcggatcgtt cggtagcggg agggtagagg ggttttcggg 1800
gggtcggggg tttcgtcgc gccggcgaag atagatcggg gttcggtagg gaggttattt 1860
cgagtttaga gattttaggt attttttata tataggtttt tattttggcg tgcgtgcgtg 1920
tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtacgtt cgttaacggg 1980
agga 1984

```

```

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

<400> SEQUENCE: 13

```

```

tttttcggt aacgaacgta tatatatata tatatatata tatatatata tatatatata 60
tatatacgta cgtacgttag agtgggaggt tgtgtgtggg ggggtgttag gatttttggg 120
ttcggaatga tttttttatc gagtttcgat ttgttttcgt cgtcgcgac ggaggtttcg 180
atttttcgaa agttttttta ttttgttcgt gtcgggcgat tcgaatgtag aaacggggcg 240
tagagcgcgt tcggtagttt ttttttttta ttgggtag gagaacgtat agaacgagcg 300
gagttttcgg ttgtattttg ttcgattaga agcgttcggc ggcgttttta tttatgtacg 360
cggttgtatt gagtatattt ttcgtgggag ttaggttttg aggagagggt ttgtttcgtt 420

```

-continued

```

agttagttaa ttttttaaat agattagtag tagtattacg aaagtattgg gtagagggtg 480
atgattagga ttaatgggtt tataagacgg attttttttt aacgtttttt cgttttgtat 540
ggtagatacg gggtaggat ttcgcgagga gcgagtttcg cggagggtgg attttatttg 600
tttttttga ttgtgggggt ttagttttat aaattacgtt gggtaatgtt agttgatttt 660
atttttttag agaatggaat tgtacggggg attggcggtt atttttagtt tagtgtaaaa 720
agattttttt tattaataatg ttaataagat ttattttatt ttaataaat tagataaaat 780
ggtttttagta gtttttagtt tttntacgtt tggtaaggtt tggtagtgga tttttttttt 840
tttttttttt ttttttttat tattgtgggt taagttatga tgggtgttgg tgtttttttg 900
ttattttata aattataaag atgatatgatt aattcgagtt ttagaaaaa attttattaa 960
ggggaatag attttattaa ttgtttaggt gagagtttga ttagttattt attttttaaa 1020
tgtgtatttg gtattaaagt tgggatttgg atcggaagga attttttttt gtatatggat 1080
ttattttttt agaagttttg aattattttt aggtattttg aatagtagta gttattgttt 1140
aagggttttt ataaatgggg gggttttggga ttgagggtaa gagggatatt taggtttgtt 1200
ttttttaata ggaatttaag aaaatgtagg gaagatttta ttgtattatt tttatgtagt 1260
tttttttttag aaagaattta aggtgtttta taataaagga tatttgtgta atagaattta 1320
tgaattaaaa tagaaaattt ggggtagaag gaaaagtta gtaaattgaa atgataagtt 1380
taaattttta tgtatttggt ttgatttggg taagagtttt ttgatagta aagtaataaa 1440
agaaatataa gttatatagt ttgtttgtag aaatatgtat ataagttttt ttagtattat 1500
tttttatggg ttaaatatcg gaataaatgt attatgtaac acgtaataa gtaggatatt 1560
gagtaatgga ttatgttaag atttagtgga tggggatgta ttttttagtat ttggtatagg 1620
gtttggaata tggagatagt aattagtgtg tgttaaagaa atagatgggt agatttaatt 1680
aagatgtagt agaaggtagg gaggggtaag gagtttgag ttatattgtt taggtatagt 1740
tttcggattt gttattttat tgtaaatat ntaannntt gttttttttt gtttatagat 1800
aaaatgaaan tataaattgt attgantnta taaggttgtg gtgaggatta agtgatgnta 1860
taattattga ataaatataa agtgatgtgt tttttaagat nttntattg gttgtgtttt 1920
ggntagaagg ntntnntttg tggntttaag gntattgata aagatgnnag tgatgnttga 1980
agtt 1984

```

<210> SEQ ID NO 14

<211> LENGTH: 7833

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

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

<400> SEQUENCE: 14

```

gtttttgggt agatatgtgt ttataagtt ttaatggaga aaaatgtaag tttttattt 60
tttgaaattt ggttatttga gtaatgagaa aatagttatt ttttttagga tagtggttt 120
taattatggt tatgtgtttt tttaggaaaa ttttaaaaat atatatatat taatgtttt 180
gtgttatttt tagggatttt aagtttttga atacgaattt tgtattagta ttttttaatt 240
atttagtgta ttgtgatgtg aaattatgat tgagttttat tgttttaaga tgaataaat 300
tttttttagt attgaaatta taaatttaaa ttattaaaat taattaaggg tatgggaatt 360
aataaggat agggaagttt ttatattata aaattatttt tttaaattat agttttattgt 420

```

-continued

ttatatgtta tttgttattg tagaaaaggg tgaaaaaata gtaaatTTaa ttatttttag	480
tttgaaaaat tatttagaaa tgaagatgac gattttgaaa tattgttaat attatttgat	540
ttataaataa tgttttaata tatttattat atattgatag atattttttt atatgaatat	600
tatatattaa aattaaggtata ataagtatt tagaatattt tatttatatt tatgtatttt	660
aagtaggtta gaaattaaga tatgagttat taagtatgag atgttaaggt gtggggtag	720
aaattatatt gtattttatt attaataatt aatatatatt ttaattattat atatatttaa	780
ttttaatttg tatattttta attattttta attatgtgta taaatataag tatatatatt	840
tttatgtatt tattttattt tatttttatt tattttattt tataggggat ttttttaa	900
ttattattat taaattatat atttttattt taatttttag aataagttta ggaggtagg	960
attgttatta tttatatatt ataatgagg aaattgttta tagttataaa gttattgtgt	1020
tagatatatt agaagtttaa tatatatatt gtgaatatat gtataaaaat agagagatag	1080
atagtataa tagtttattt ttatatagg taaaagtttt taatttgttt tagaaatttt	1140
tttgtgaaaa ttgagtaaaa atcgaggtat ttttttattt gttatatagg tatagggtg	1200
attttatttt tttaataagg atgaatattg aaatgtggat tttaaggttt aatttttag	1260
tttttgaatt ttgatagtg ggatttgga tttgtttatt gttttaaagt tttttaagga	1320
atttatatga ttaattaggt tttagaaatta ttggatttta ttgtcgaagt ttgagaatta	1380
aagtttgggt tttattgagg ttttatagaa agggtaaatg aagtattatg gatagaattg	1440
atagcttttt agttagtttt tttttttaga agttaatagg tagtaataa gttagaaatta	1500
gtgatttatg ttttgtttt tgaagttagg tagaatttta tagagtttta gtagtgttat	1560
tgacgagatt tgtttttttg ggtaagttgt ttgatgtttt taaagttata ttttttttat	1620
ataaaatgag ataatttttt ttgttttata ggggtgtttt aaagattaaa taaaaataat	1680
atgttttatt ttatatggtta taatgtttga tttttaagaa gtaaggata ttttttatt	1740
ttattgaagt aattagaaag tatgaaatta tgaaggagat aagagttttg attggtagtg	1800
tattttattt ttttaggttt atttatttat tttaaattat tttgttgga gaataatttt	1860
taagtttttt atttaagttg tgagtaattt tatattttat aatgatgttt tttttatgag	1920
aaaaaaaaat gtttttaagt tttttggaga aaatatattt gtattatttt tattgaaaaa	1980
tttaataatt ggattttgtt tttttgtatt aatttttagag tgtatatgtt ataaataaag	2040
tgtttttagt taagaagatt gaaagtaaat atggatatagt attttaaaat aagaattttg	2100
taaatatatg gtatgattgt gttatattat tagtaattat atgatacgta atgtaaagta	2160
tagtttatag atttaaat ttttttaata agtaaatgta tttgttttg ttggggaaaa	2220
gttaaagtat taatttaatt gttaatgtag tttgttttat ttttttgga tttagtata	2280
agtttaata atgtatatat ttttatttat atatttagta atataatttt ttgtttaatg	2340
agtgatgttt ttttgttatt tgggtgtgtt tgttagtttt agaattgttt ttttggtggt	2400
attataatat taagtataga gtaagtgtta taaaattgta gtatttttat tgaaaagggt	2460
ttgttttaaa ttgtttaata atttaagga tttttgtgga agtaatcgta tttgttaatt	2520
agttataatt agtaattaat ttttttgag ttttaattta tttttggtta aacgttttag	2580
gaagagtata tattattaga aagtatgtta aaaatttatt tagtagaaaa tttaaaaata	2640
gttttttttt gttaagaggt tttttaaaat ttattttata tagttaaatt ttgaaatttt	2700
agtaggtttt gttttattat tataattatt gtataaatat ttttaaggat tttgttttta	2760
gttttaagta tgattttatt ttataagttt gattagttat tatattagtt ttgttatgga	2820

-continued

aatgatatg tttttatgtt ttgtgttaga gttgttaaat tttgatttat atttatgttg	2880
ttttttttgt tgaaagtgtg tagcgaaaga aatttttaaat tttttgtttt gtaatattag	2940
ttggtagttt tatttaaatgg gtattttgtt tttttaaaga atttagttgt tttgtttaga	3000
agtcgatttt ttgatgtttt taacgttttg tttaattgat ttgttttaaat ggagttttcg	3060
tcggtgagga gcgagatgtt atcgattaga atgttgggat ttgttggtta attgtagga	3120
gtgagagata ttgagattta gaaatttttg gaggtgggag gggagaggga tagtttcgga	3180
cggaggcgga gatgtaagat aaagggatgg attttatata ggaaaaaaaa aaagatttcg	3240
ttgaggtatt gaggtgtgtt acgattatat tttttaaagg agaagttaa aagtaaggaa	3300
gtgggaggag gttggaggtt aaagtattta aaaggattat tcgggtataa tttgtttttt	3360
tggtggtgtt tgtaaggat agatagtttc gtttttaag tatatgaatg ttttttttaa	3420
gtgattgga atggaatta attgtttgtt aaatgttatt aaatgttttt ttaaatttag	3480
gggatataga aagaggggta taaaggaga atttaaatag aaaaaggag gattcggagg	3540
tttttgaaag cggggggaga agaaggagga ggataatag agaggaatag agaaggagag	3600
cggagagaag ataaataaaa ataaaaatag gaattattga ataattatat attaaaaaga	3660
aagttttttt ttatggggta tttaaaatat tgagattgta atagtattt cggttatgga	3720
agaaagatgt ttttttttat tttgttttc gaaagttttt gggttcgtta ttggcgatta	3780
aaattttatt aggttaaaga gtgtgtttaa ttgtttgaag aatgtagtag acggaaggcg	3840
gggttcgtta tgctgtttgt ttttttcgtt ggagagaatg aaagaaacgc gtagagttag	3900
agatttttgt cgagtttagt ttttttcgt cgttttaggt ttcggttaa	3960
ttcgagtaag gaacgtaggg ttattgtttg ggttaataaa tggagttcgt tttttttttt	4020
tcggacgtcg ttgttcggtc gatgttttcg gtaatttatt cgcggcgat gtagaggagt	4080
tttttttttt ttttagatt attgttttcg attaatgtga ttttttaaat atatttgatc	4140
gtatttttta ggtggatata ttaatagggt acgggttgga gaggagcggg tgatgaggag	4200
agggatttaa atttgcaac gtttggttg ggtcggagt gcggggggtt tgggaggaga	4260
gaggggagaa gagagaagga aggagagcgt ttgtcggat ggttgagttg tttcggcgag	4320
tagttttggg gttgtacgtt ttttggggag atgttgttgt tgtttttagg tcggtaaagag	4380
cggttttaat attatcgttt tttatttttt tttttgtaa ttttagaga aacgtttttg	4440
gttttttcgt cgcgatattt ttagtttgta tttttttata gtttagcgg cgcgttttcg	4500
tacgttgag cgtcggtcgt tagtaggacg ttttttttcg cgtcgattcg tttttttttg	4560
ttttgttgtt gttgtttttt tgatattttc gtttttatta ttttagttc ggagagacgt	4620
tatttagtcg cggttcgtat tcgcggttcg ggttacgcg cggagaggg gcgttagttc	4680
ggatttcgtt ttcggtaggg ggcgttttg agcggagagt gaggcgaatg gtatatgagt	4740
gtgcgggtag tttattttga agtcgaggt ttttatttga gttatgttc gtttagtttt	4800
attcgggtta gcgtttggc agcaggttta tttgtggttt tcgcggtcgt tttttttttg	4860
tatttttgta tttattcgtc gatttttttt tttcgggatt tgattttgt tttattaatt	4920
agagttcgat tgtttttttt tacgtgattt cgggcgggtt gaggatttgt tgttttttaa	4980
acgttagagg gatgcggcg gtagagttcg agaggcgtt gtcgggttc ggggcgtttt	5040
gatttttttt ttattttgtt ttttcgggtt ttattcgttt gtttttggt tttcgttttt	5100
ttttgttttt cggtttttta gagttttttt tttatggtag tagtttttcg cgttttcggc	5160

-continued

gtagtttttt agcggacgat tttttcgttt cggggttgag tttagttttt ggatgttgtt	5220
gaaattttcg agattatgcg cgggttttgt tgtgtttttt tcgtcgggtg ttattgttat	5280
cgtcgtcgtt tttgttgtcg tcgttcgcgg gatgtttagt agtcgttgtt tcggtttttcg	5340
cgattttgtg tttttcggaa gtcgttttgt gttgtagagt tgtaagaatt agttatgggtg	5400
ttgtgggagt tttcgcggta gtgtagtagt tggatatattt gcgaggggtt ttgttggttg	5460
ttgttgttgt tcgttatgtt atttatcgta gtctgttcgg tgaagttcgt tgtttttttt	5520
atttttttaa gtgattgtta aacgtttatc ggttgaatt gttttggtta gtttagaatt	5580
ttcgttttcg attttttaatt ttcgtagaag aatacgcgta tttagtatag attagtttat	5640
tttagcgcgt ttttttagtt ttttattttt tattgtttta gatttttaatt attatttatt	5700
tttatttaga gaaataaggg gaattgttgt aggttcgggg gtgaggggtg gttttgggat	5760
gggtagaaaag tgtaggtgta gtaggaaatt tttgtatgtt tgcgtttata ttggagtgtc	5820
gaggattttg agaaatatta aacgggatgg ttttttgggt ttattgtttt gaaagagtat	5880
taattttagg ggaatatattg aaatagaagt tttgttatta ttaaagaaaa aagttttatt	5940
aggatgagga agaaataatt ttatgagaaa gaatgagcga gaaagtaata aattaaatgg	6000
tgattgtagg ggaatcgttg atttttggta aaggtgttat gaggtcgtat tggtttttcg	6060
ttgaagatta ggttatatag attttagagg agttgggttt taatagaatt tttttttttt	6120
tttttttttt tttttttttt tttttttttt tttttttatt tattattttt tttttttttt	6180
ttattttttt ttttttttagg cggtaaaaga tattgttttt gtagtttaga tatgtttttt	6240
tttttgtttt ttttaagtttt aaggtagtat aggggagttg agaaaaagaa tattttgcgg	6300
gtttttttagg tcggagtggg tatgattgag gttggttagg ttttatgtag gcgagtcgag	6360
ggcggaaatcg attttagtgg gcgttgattt ttttattttt ggatagggtt ttgtggagtg	6420
ggttagggtat ttttttttgt cgttcgggtt ttttttagatt ttgacggcga acgttttggtta	6480
ggtttcgttt tgttgaagtt ttttaattaa atagggttag aggatgggag ttgttgtatt	6540
tttagttggt atagtattcg gtttgatagt ttgtagtata ggggtgatgt aattttttat	6600
tttttgtgaa tataattttt ttgtagttaa atttggtttt gaataaagtg ttttttaaag	6660
atgtatataa gttgaagtgt atgtaatttt agagaggagg gaatgattaa ttgtaattta	6720
gggtgaaagt ttgtatagtt tttagtattt attgatgtaa atgttaaaag gaaaattatt	6780
atgtattatt ttaattttatt ttttataaag ataagttgag atatgtaatt ttattagatt	6840
tgggttaata gattgttttt ttttttggtta gtttttaaat ttggtatttt aataaaattt	6900
aatatgtttt tataattttt tgatttatgc gtatatgtgt gttgtttttg aaagaataag	6960
ttttattttg ttattgttta attatttttt agatgtttta ttatggtaat aattatgagt	7020
ttgtaaaaat aatttttgga aatgttgatg gttttgtagt ttaatataga ttgggttgtt	7080
ttatttttag tttttgtatt gttttaggaa ataattaatt taaatgtgaa gttgatattt	7140
gtaattaaga aattatataa ttattagata ttttaagggt gattgtataa attaaagaga	7200
ataaattggt tttgtagata ggttggttaag aatttgggtat ttcgttttta tttttgttaa	7260
tttagagggt attaatTTTT atttgagtta aatagattat tatagaaaat attgtgtttg	7320
tttattttta ttattgaggt tttgtttttt tttgttttgg atatatttta aataaggggt	7380
tgttttagtc gttgaagtaa aagaataatt aaagatgggg aaatggtaaa aggggtattta	7440
gagattatta ttagtttttt tttaaaatgt ggagttttgt ggttataaat attgtttatt	7500
taatgagtaa aaaataaaaa taaaaaaa ataggaagta aatgttaagt ttttatttat	7560

-continued

tattgttagt attaacgtaa gttttaaaaa atagtattat tagaaaagga tattaagga	7620
gaattgatta gaaaagaatt gtggaaaatg gaaacgaata ttgattatTT aattagattt	7680
tgaggttatt agtagatagt gattttgtag tatagttata gttgttggtat ttaaaattta	7740
ggataagtat tttaaagttt taaagtagtg tttttttttg ttaaaaattt gtaagatgtt	7800
ttaatgattg gagtgttttt tttgaatttg agg	7833

<210> SEQ ID NO 15

<211> LENGTH: 7833

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

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

<400> SEQUENCE: 15

ttttaattt aaagagaata ttttagttat taaaatttt tatagatttt taataaaaaa	60
aagtattatt ttgaagtttt aaaatatttg ttttaattt taaatttaatt aattatagtt	120
gtattgtaag gttattgttt attgataatt ttaaaattta gtttaagtgt taatattcgt	180
ttttattttt tataattttt ttttagttaa ttttttttta gtattttttt ttgatagtgt	240
tattttttta agtttgcgtt aatattgata gtggtgaatg aaagttaatt atttgttttt	300
tgtttttttt ttatttttat tttttgttta ttaggtggat aatatttatg attataaaat	360
tttatatttt ggaaaagagt tagtgatgat ttttgaatat tttttatta tttttttatt	420
tttaattgtt tttttgtttt aacgattgaa ataatttttt atttgaatg tatttagata	480
aagaggaaat aaagttttaa taataaagat aaatagggtat agtggttttt gtgatggttt	540
gtttggttta aatgaagatt gattattttt aagttaatag ggggtgaagc ggggtgttaa	600
gtttttgata atttatttgt aaaattagtt tatttttttt agtttatgta gtttttttta	660
aaatatttgg taaatatgta attttttgat tgtaaatgtt aattttatat ttaagttagt	720
tattttttta aataatgtaa gggtaggaa tgaagtaaat tagtttgtgt tggattataa	780
agttattaat atttttaaaa attgtttttg taggtttata attattatta taataaagta	840
tttaaaaagt gattaggtta tagtaagtg aaatttattt ttttaaaaat aatatatatg	900
tacgtatgaa ttaagaagtt atagaaatat gttgagtttt attaaaatgt taaatttaga	960
aattgttaaa aaagagaata atttattgat ttaaatttaa tagggttgta tattttaatt	1020
tgtttttgta aaggataaat tagaatgatg tataataatt ttttttttgg tatttatatt	1080
agtaataatt aggaattata taggttttta ttttgagtta tagttgggta tttttttttt	1140
tttaaaagtta tatatatttt agtttatata tttttttgaa agatatttta ttttagagttta	1200
gatttaatta tagtaaaatt atatttatag aagatgaaaa attatatata ttttatatta	1260
taggttggtta aatcgaatgt tatgttagtt aggagtgtag taatttttat tttttggttt	1320
tatttaatta ggaagtttta gtagagcgaa gtttggttaag cgttcgtcgt tagaatttga	1380
aggaattcga gcgagtaaga agagtgtttg atttatttta tagaagtttg ttagaaaatg	1440
gaggagttag cgtttattga agtcggtttc gttttcggtt cgtttatatg gagtttgatt	1500
agtttttagtt atgtttattt cggtttggga gattcgtaaa gtgttttttt ttttaatttt	1560
tttgatttat tttgaagttt agggaaagtaa agagaggggt atatttggat tgtaaaatta	1620
atgttttttg tcgttttaga gagaagggaa tgagagagag agagagatag atagatagag	1680
agagagagag agagagagag agagagagag agagagagag agaaatttta ttgaaattta	1740

-continued

gttttttttag aatttgtgtg atttggtttt taacgggaga ttagtgcatg tttatggat	1800
ttttgttagg aattagcatg tttttgttag ttattatttg atttattggt ttttcgttta	1860
ttttttttta taaagtattt ttttttttat tttagtaaga tttttttttt taatgatgat	1920
aaagtttttg ttttagtggt ttttttagga ttggtgtttt tttaaaatag tgaatttaga	1980
aaattatttc gtttaatat ttttaaaatt ttcgtagttt taatgtaagc gtaagtatgt	2040
aaagggtttt tgttatattt gtattttttt ttatttttag aattattttt tattttcggg	2100
tttgtaatag ttttttttgt ttttttggt agagggtggg ggtattaggg gtttagggta	2160
gtaggagggt aggggttagg gaggcgcgtt agggtagggt ggtttgtgtt ggatacgcgt	2220
gtttttttgc ggagttaaag ggtcggggac gggggttttg gatttattag agtaatttta	2280
gtcggtaggc gtttgtagt tatttaagga gtagggaaa gtagcagatt ttatcgggcg	2340
ggttacgatg agtagtatga cgggtagtag tagtagttag taaaagtttt cgtaaagtgt	2400
ttagttgttg tattgtcgcg gggattttta tagtattatg attagtctgt gtaattttgt	2460
agtagtaaac ggttttcgag gaatatagga tcgcgggggt cgggtagcgg gttattgagt	2520
atttcgcgga cggcggtagt agaggcggcg gcggtggtag tggatttcgg cggggaagta	2580
gtagttaaat tcgcgtatga tttcgagagt tttagtaata tttagggtt ggttttagtt	2640
tcggagcgag agggtcgttc gttgagaagt tcggtcggag acgcgggaag ttgttggtat	2700
aaggaggag ttttgggaag tcggaggata ggaggagacg ggagtttagg ggtagacgag	2760
tggagtcca ggaggtaggg tggagggaga gttaggcgt ttcgtagtgc ggtagtcgtt	2820
tttcgagttt tgcgttcgt atttttttg cgtttggga gtagtaggt tttagttcgt	2880
tcggggttac gtgggaagag gtagtcgggt tttgattggt ggagtaggat gtaggtttcg	2940
ggaggaggag gtcgacgagt aggtgtaagg atgtaaggag gaggcggtcg cggaagtat	3000
agatgggttc gttcgttagg cgttggttcg agtggggta ggcgggtat ggtttaaatg	3060
agaagtccg gttttagggt gggttattcg tatattata tattattcgt tttatttttc	3120
gttttaggac gttttttatc gaaggcggg ttcggattag cgttttttt tcgcgcgtga	3180
tttcgggtcg cgagtgcggg tcgcggttg gtggtgttt ttcgagttg agatggtggg	3240
ggcggagggt ttagaggagt agtagtagta gggtagagag gggcgagtcg gcgcgggaga	3300
gggcgttttg ttggcgatcg cgtttttagc gtgcgggagc gcgtcgttta ggttgtaggg	3360
ggatgtagg tgggaatgtc gcggcgga ggttagggac gtttttttag ggatttatag	3420
gaaagagggt gagaggcat ggtgttagaa tcgtttttgt cgatttgga gtaatagtag	3480
tattttttat aagagcgtgt aattttaagg ttgttcgtcg aggtagtta gttatttcgg	3540
taggcgtttt tttttttttt tttttttttt ttttttttt ttaggtttt cgtagtttcg	3600
atttagttta agcgttcgta ggtttgaatt ttttttttta ttattcgttt ttttttagtt	3660
cgtagtttat tagtggtttt atttgggagg tcggttaga tgtgtttgga aggttagatt	3720
ggtcgggata agtggtttga gagaagaga aagggttttt tgtatacgtc gcgggtgggt	3780
tgtcgggagt atcggtcggg tagcggcgtt cgggaagggg agagcgggtt ttattgtgtg	3840
gtttaggtag tgattttcgt ttttttatc gggtttttgt cggatggtcg gtgatttggg	3900
gcgacgagag aaggtttaat tcggtaggag tttttgggtt tgcgcgtttt tttattttt	3960
tttagcggga agggtaaacg gtatagcggg attcgttttt cgtttgttgt atttttagg	4020
tagttagata tatttttttag tttaatggaa ttttagtcgt tagtaacggg attaagagtt	4080
ttcggggata aggggtggaga ggaatatatt ttttttatga tcgggggttat tattgtagtt	4140

-continued

ttagtgtttt	ggatgtttta	tagggaagag	ttttttttt	ggtgtgtgat	tatttagtga	4200
tttttgtttt	tgtttttgtt	tatttttttt	tcgttttttt	ttttattttt	tttttgttat	4260
tttttttttt	tttttttttt	tcgtttttta	aagtttttcg	attttttttt	tttttattta	4320
aatttttttt	ttgtgttttt	tttttgtgt	tttttgaatt	taggagagta	tttgataata	4380
tttaataagg	aattagtgtt	tatttttaat	tatttaaaag	aggatattat	atattttgaa	4440
aacgggatta	tttatttttt	gtagatatta	gtagaaaaat	aaattgtatt	cgagtaattt	4500
ttttaagtat	tttaattttt	aatttttttt	tatttttttg	ttttttaatt	ttttttttga	4560
gagatgtgat	cgtgtagtat	tttagtgttt	taacgaaatt	tttttttttt	ttttgtgtga	4620
aatttatttt	tttattttat	attttcgttt	tcgttcgaga	ttgttttttt	ttttttttat	4680
ttttaagat	ttttgaattt	tagtggtttt	tatttttggt	aattaagtag	tagatttttag	4740
tatttttagtc	ggtggtattt	cgttttttat	cgacgaagat	tttattaaaa	tagattaatt	4800
agattagacg	ttggagggat	tagaaaatcg	gttttttagat	agagtagtta	aattttttta	4860
ggaaaatagaa	tattttattag	atagagtgtt	taattaatat	tgtaaaaata	ggaattagaa	4920
atttttttcg	ttatagggtt	ttagtagaga	aggtaatata	aatatagatt	aagatttaat	4980
aattttatag	tagagaatga	gaatatgtta	ttttttatag	taaggttggt	gtggtaatta	5040
attaggttta	tgaataaag	ttatgtttga	aattaaagg	aaagttttta	aaagtgttta	5100
tgtagtaatt	atgataatga	aataggattt	gttaggattt	tagagtttgg	ttatgtaagt	5160
agaatttttag	agaatttttt	agttagaggaa	aattgttttt	gaattttttg	tttaagtaa	5220
ttttggtata	ttttttaata	atatatgttt	tttttaagac	gttttgttaa	aagtaagtta	5280
aaattttaaa	ggagttaatt	attggttgta	attggttaat	aatgcgggt	gtttttatag	5340
aggtttttta	aattattaaa	tagtttgaag	taaagttttt	ttaatgggaa	tgttgtaatt	5400
ttgttgattt	tattttgtat	ttagtgttat	agtgttatta	agaaataaat	ttgaaattg	5460
gtaagtatta	tttaagtgtg	gaagaatatt	atttattgag	tagagaattg	tattattgaa	5520
tatgtaata	aaaatatata	tattatttag	atttggtatt	aggattataa	gaagtagata	5580
agattgtatt	agtaattgga	ttagtgtttt	aatttttttt	tagtaaggta	aaattagttt	5640
atttattaga	attaaattta	agtttatgaa	ttgtattttg	tattgcgtat	tatatgattg	5700
ttagtaatat	gatataatta	tattatgtat	ttgtaaaatt	tttattttta	aatattatat	5760
tatatattatt	tttaattttt	ttgagttaga	atattttatt	tgtgggtatat	atattttaga	5820
attgatgtag	aggagtagag	tttagttggt	agattttttt	gtagaaatag	tgtagatata	5880
tttttttttag	aaaatttaag	aatatttttt	tttttttatgg	aaagaatatt	attataaagt	5940
gtgagattat	ttatagttta	agtagggggt	ttgggagtta	ttttttaata	agaatagttt	6000
aagataaata	aatgaatttg	ggaaaataag	atatattggt	aattagaatt	tttatttttt	6060
ttatgatttt	atattttttg	attgttttaa	taaaggttaag	atgtattttt	tgttttttag	6120
gtgttaggta	ttgtgttatg	taggatagaa	tatgttattt	ttatttaatt	tttaaaatat	6180
ttttatgaga	taaagaatat	tattttattt	tatatataag	gaatatggtt	ttgaaagtat	6240
taggtaattt	gttttaagaa	ataaatttcg	ttagtgatat	tgttgggatt	ttgtgaaatt	6300
ttgtttgatt	ttagagtata	agatataagt	tattaatttt	tgttgatttg	ttgtttgtta	6360
gtttttgaga	ggggaaatta	attgggaacg	tattagtttt	gtttatgata	ttttatttgt	6420
tttttttggt	gagtcgtagt	aaggttttaa	ttttaatttt	taaatttcgg	taataagatt	6480

-continued

tagtgatttt tgaatttggt tgattatatg aattttttga gaaattttga aataatagat	6540
aaattttaag ttttattatt agggatttag aaaatttgga gttgggtttt gggatttata	6600
ttttaatatt tatttttggt ggagaagtaa ggtattattt atatttatat gataaatgaa	6660
aggatatttc gatttttggt tagtttttat agagagggtt ttgagatagg ttaaaagttt	6720
ttatttagtg taaagatgag ttgttgata tgtttgtttt tttgttttta tgtatatgtt	6780
tattaaatat gtattaagtt tttaatatgt ttgatatagt aattttgtga ttgtagataa	6840
tttttttatt tgtaaatgt gagtaataat aatatttgtt ttttgggttt gttttaaaga	6900
ttaaaataaa aatgtatggt ttaatggtag tggatttggg gggatttttt atataaataa	6960
gtgaatggag gtatgaataa ataaatatat aaagatgtgt gtatttatat ttatatatat	7020
aattaaat agttaagat gtataaatta aagttaaatg tatgtgatat tgaagtatat	7080
gttgattatt gataatgaag tatagtataa tttttaattt tatattttaa tattttatat	7140
ttaataattt atattttaat ttttagttta tttagatat atagatatag atagaatgtt	7200
ttaaatgtat tattgtttta gttttaatgt ataataatta tatgaaaaag tatttattag	7260
tgtgtagtaa atgtattaga atattattta taggttaaat gatattgata atgttttaga	7320
gtcgttattt ttatttttgg ataatttttt aaattgagag taattaaatt tgttattttt	7380
ttattttttt ttataatggt aaataatata taaataatga gttgtgattt aaagaaataa	7440
ttttataatg taaaagtttt tttatgtttt attgattttt atgtttttaa ttaattttgg	7500
tagtttaagt ttgtgatttt agtggtgagg aaagtttatt ttattttaga gtagtggggt	7560
ttagttatga ttttatatta taattatttg gataattaaa gaattattgat gtagagttag	7620
tatttaaga tttggaattt ttagaagtgat tatagaagta ttggtatata tatattttta	7680
aagttttttt ggagaaatat atagttatga ttgagaatta ttgttttggg gaaagtgatt	7740
atttttttat tatttaataa gttaagtttt agggagtaaa atatttatat ttttttttat	7800
taaaatttgt aaaatatata ttttattaaa gat	7833

<210> SEQ ID NO 16

<211> LENGTH: 965

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

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

<400> SEQUENCE: 16

gattaattaa gggattttta gaagttaggt gtttttggtt tttttttga gtatggaggt	60
tattaatttt tagggggaag agatgtagtg tgaggtaggg gtgtgtgtt aagaaatttt	120
gatgtttttg gggattgagg ataaaggtgt ggatatgatt ttggggattt tggagttttg	180
tgatttggtt tatggaatgt atatttaggg gttaattttt gttttgtttt aaagaatttt	240
aagtttagagt ttttgttttt ttttatagtt ttgggatgtt gttgttgtgt ttattgtata	300
ggtagtggtt ggattgggtt tagtagattg tgtgttgtgt gttttatttg gagatgggtg	360
agatgttgaa aagttttttt tttgttattt tggatgttgt gggtggttaag tgttttagtt	420
tttatttttg ttgagttgaa tgttttaggt tagtggaatt gaaatttggg tttttatttt	480
tgttgagttg aatgtttagg tatagtggaa ttgaaatttg gttttgtggg atgtgagagt	540
tgttgagggt atgtgtaatt ggggtgtgat gaggggtgtt gtttgtgatg tgtgtaggtt	600
tgatgtaagt aggttattgt tgtgtgagtg tgtggatgtg attgtttgag agatttgag	660
gtaggttttg gatatttttg agtgaatttt ttaggatatt tttttggtta gtatttggtt	720

-continued

```

tttagtggtt gtgatttaga gtgggtatat gttgggagat agtaatgggt ttgggtgtgt    780
gtaaatgagt gtgattggaa gtgagtgtga gtttgattta ggtagggatt ataatagtatt    840
gttatatttg tttgtttttt agtagaggat tgaagtgtgg ggggtgggggt atggggtttg    900
aatagaatgt ttttgggata ttttggtaaa tagtagttgg aagtaaaggg gtagttgtgt    960
aatg                                           965

```

```

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

```

```

<400> SEQUENCE: 17
tgtttgtata gttgtttttt tgtttttggt tgttgtttgt taagatgttt tagagatatt    60
ttattttggt tttgtatttt tatttttgta ttttagtttt ttattaaaga gtaggtaggt    120
gtgatatgtg tgtgtgtgtt ttgttttagat taagtttata tttgtttttg gttatattta    180
tttatatata tttaaattta ttattgtttt ttaatatgtg tttattttga attatagata    240
ttaaaaaata gatattggtt agaagagtat ttgaggtgtt ttatttaaatt gtgttttaag    300
ttgttttttg agtttttttg gtggttgtat ttatatattt gtatgatgat gatttgtttg    360
tattaaattt gtatatatta tgaatagggt ttttttatta tatttaatta tgtgtgattt    420
taatagtttt tatattttgt agaattgggt tttagtttta ttgtgtttga gtgtttagtt    480
tagtagaggt aggggaattg gttttagttt tattgtgttt gagtgtttag tttagtagag    540
gtagggatta aggtgtttgt tgtttatagt gtttagagtg gtaagaaaga agttttttag    600
tgtttttatt attttttggt ggaatgtgta gtgtgtgatt tgtgtagatt ggtttgggtg    660
ttgtttgtgt ggtgagtgtg gtagtggtat tttggggttg tgggtggagg taaggatttt    720
agtttgaggt tttttgaggt agagtagaaa ttagttttta ggtgtgttgt ttgtgtgtgt    780
agttatggaa ttttaggtat tttggggttg tgtttgtatt tttgttttta gtttttagaa    840
gtgttgaaat ttttagtat gatatttttg ttttgtgtta tatttttttt ttttaggggt    900
tgggtggttt tgtgtttaaa ggaggtagtg ggaatgttta atttttaaga tgtttttaat    960
tgatt                                           965

```

```

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

```

```

<400> SEQUENCE: 18
tgttttttaa aatgatgttt agttgaatta gttaatggag ataaaaatat ttatatataa    60
tttttttta agtttattag ttattttata gattagttaa aattttggtt attttaatga    120
gttaattttt ttttagagat attataaatt ttttaaagat ggaaaataat tgttattttt    180
ttgaataatg tttattattt aaaataatgt atataaaagt tgtttaataa atttttgttg    240
aattgagaga aaatataaaa aattgaaagt agtttggttg taataagtgt ttgttaagtt    300
ttatagattt tttttttttt ttggaattat tgtaataaat atttgatttt tgattttag    360
gattaaaaat tgatttttta tattttaata ggagtaagta ttagttttgt aatttttttt    420

```

-continued

gttaatagaa taaaaatgga atgtatgttt gttgttttta tatgttggtt gatatttatg	480
ttgaatatatt tttttataat gtagagggttt ttttgttttt tgttttgtgt tttatttttg	540
tttattatag agatattaat gatgttatgt gtttttttgt attatttagt tagtagtttt	600
tattatgttt tttgttttggt ttttattgta gtttaattaa ataaatggaa atgggtagtt	660
attgaaaata aatagtaatt gtaaaagtgt tgttttttaa gaaatttttt gtattgtatt	720
ttgttttggt ttattgattt taagattttt attgtttttt tgtgtttaga aattagattt	780
tttagtttgt tttgttttag ttagataaat gttttataaa aatgatttat attttttaga	840
ttgtggggag agaagaaaaa gattttgaga aagtgattaa ggaggagta agagtgatat	900
aatttttgat ttaatttagt tggatttgtt tttttatgtt tttatttttt tttttataaa	960
aaagatagtg tttttgatgt tttgttttag gtagtttttt aggggtgagtg tagatgtgtt	1020
gagttatttg gttgtttga ggattgggtt agagagtttt tttttttatt aatgaagtta	1080
tttttttttg gtagtttggt tttttttatt attttatgtt ggaatttggt tttggtttgt	1140
atagtttgtt tgtgatgtag atggtttttt agttaagttt tagagaaatt tgtttaatta	1200
tgaataggat atttatagta taggagaaaa aaaaatttat taaaaaaat attttattgt	1260
aatattttga tttgggggtt ggagagtaaa aatttaatat ttgatgtttg gaaaaaattt	1320
aatgttttat ggttttttaa ggaaatttgt ttgtgtttta tttattagag taatttatat	1380
attttgggtt aaagtataat gagattgtta ttattatttt gttttttgta agattttttt	1440
tttaagaata aaaagaaggt aataaagatt ttatattttt gtttttttta gaaagttaaa	1500
gagaatttaa agtgttggat gaaatgtttg tttttaattt gtagtgtttt atttaaaaaa	1560
aaaattatga tataatttta gaggataggt tttttttttt taggggtttt gaaagttaa	1620
agtatatttt taattttagt tattattgta ttatttagtg aggataattt gtttagttgt	1680
gtttaggatt agttagatta aggagtga aaagttttt ttttgaattt gaatgggatt	1740
ggaattttgt tatttttaga gtagtaatag atgatgttat gtttaagaat attttatttg	1800
attaataaaa aaggagaagt ttgaaatgta attttaaaaga attttagaag gaggtttaga	1860
tgaagataat aaaatttttt ggagatatat ttgttattta tatttttttt ttttaagtta	1920
ttggtttgtt gtttttgagt ttagttaaat ttttaaatag aagattaaag taaggagta	1980
aggatatag tttttggttt atgggttttt ttagatataa gtattgtgaa tgaagatttt	2040
gttttattta tatttgggat tttggagttt ttataaagaa tatttttagt ataaggatgt	2100
gatttaaagg gggatttata gaataatatt ttaattattg tttgatttat atttgagtgt	2160
aatggataaa taaatataat ttagttata aaatttgaaa aatatatatt tttttattta	2220
aaaatttttg atagttttgt taggtatttg agaattagaa ataaatgttt ttttattttt	2280
tttttttga tgaaggaaaag ttattaaaaa tagttatata gaaataataa tataattgta	2340
gattttattt taaatatagt tggtagtttt tttttttttt tttttattta tttttattt	2400
tagatagtta gaggggaaag tttgttttag agatagaaag aagatattta tggggataaa	2460
ttttatttta gttatttgat agtttgaaga agtatttaaat ttttgaata ataattgtat	2520
aagttaaaga agagaaaaga taggtagttt aaagtttttt aatatttgtt atattttaat	2580
tttttaagtt agttttgaga aagtgttttt gttttgtaga agtatttggt taagtgttgt	2640
gaagtttatt ggtattaaaa gaaattaaaa attttggagg attagagaat taaatttggt	2700
aatataaatg aaaatatttt agaagttaa aagtatttaa tttattttt ttaattatgt	2760
tttatttat atagatggta ttatttatag tgaatgttaa ttattaatat ttttttaaaa	2820

-continued

atgggtatta atttgttatt agttgtttgt ttttgattta gttttttttt gttgaaaggt	2880
tttatttaaat tgtattataa agggataaaa gggtaaaaga taataagttt ttttagattgg	2940
tagatttggt atttataaaa tggttttggg taggttattt gaattttttg aaatttaatt	3000
agatttattt aaatttaaga atgagaatat aaatttatat tttgaagtgt tttatagaaa	3060
ggtttatttt aatgtttgga gtatatattt taatgaatat ttattttatt ttattttttt	3120
ttatttttga attaagtaat tttgaattta aagttgttat gattagtatt gaaaagatta	3180
ttggattatt aattgtgtga ttttgggata gtaatttttt gtattttagt ttgtttatat	3240
gttatatatg aagggtgaag tttgattttg ttttgtgatt attattttaa atatttgatg	3300
aaattaaatt ttagtgtttg gaatggtagt ataataaatt tattaagaat aaataattta	3360
ttgtaaaaat atattgattt ttaaatgatg taattgatag ttatattatt gtagagggtt	3420
gataaataat aaaagaaatg aaagatgtat atgggtagaa ttgaaattat tttgataagt	3480
tttttatttg tttattattt aaaattaatg attatgttga atgtttataa attataaaat	3540
ataaaagaaa ttttataaat gtgtatgtat aggagttaa gttattaaaa gttttaaagt	3600
ataagtttaa attaaattaa ttaagaagt tgagaggaaa aattggtttt tatttttaatt	3660
tattattgtt ttgaggtttt atgtttaata taatttttta agtagagggt ttagagagaa	3720
gagttgtgag gatattttta tatttgtgta gaaggaaaag tttgttattt atttttagtat	3780
ttttagtgtt atattgatgt gtattttgga tttattttgt ttttattgta taaatttata	3840
tttgatttta aagaaaagga aaattttaag tttttttttt ttaaggggat agaaattttt	3900
tgtgttaatt gtttgatttt tttttttgta aggttttatt ggaaattttt tgtaataata	3960
tgtaggggat ttttttatgt gttgatgttg tttatatagt ggggtgggtt tgattgaaga	4020
aaaaaaattg tatatatgta tgaaagatta tggttttatt tttggaaagt atgaaaggtg	4080
attgataattt ttaagaagtt tttgttattt aggaaaatta ttaaatattt tatttagaga	4140
tatttggaag gattgaagga aaggaagaat gaagaaagta gaatttagat ttatgtgggg	4200
agagatttgt ggtagaggaa aagtattttt tttgaatttg ataagggtt tgtttggggg	4260
aattttttgt ttagtttttt attattaggg ttttttgaa gtgggttttt tattgggtag	4320
ttttttggga gtgtagtgga gaatttttat attttttttt taggtttttg aaggattttg	4380
tttttttagt gtttttttta ggttggtagg agttttgagt ttgatatttt tttttgatgg	4440
gataggtaag ttttgtgggt gtgtaaatat gttgtaatta agttttttgt tgattttata	4500
gttttgtgtg tttttgagaa gaagtgattg tatttaattg tttattgttg gtttgttttt	4560
taagagtttg ggggtttttt ttttttaatt tagaattagt tgtatggggg gtggggaagt	4620
gggggtgggg aaggagtggg agggtagtgg tttttgtgag tagagtgatg ttattgagtg	4680
agtttttgaa tggggagtgt tgttgttttt aagttgattg gtattttttg ttaggaagaa	4740
atgttaagag gtgggagtgt ttggggaggg aggtaggtgg tttttattgt aggtgtgggg	4800
agttgttttt ttgttttttt gtttgttttt taagtttgga tttttaggag tggttgaagt	4860
tgtggagtgt ttttgagatt tgtgaatgaa tttttttttt tttttttttt tttttttttg	4920
ttgagttttt tttttggttt tgatggtata gtgatataat gatgatgggt gttataattt	4980
gtatttgaat ttgtaggtga gttgttttga gtttttttgg ggaagaattt taggtgtgtg	5040
gatgtaatag ttgagaatat taggtgttgt ggataggagt tgggattaag atttttggtt	5100
agttttgtat ttttttgtat ttttttagtat tgttttgtat tttttgtatt tttttttggg	5160

-continued

ttattatggt ttttatgtga tttgtttggg taatgttgaa tttagtgtg tagtgttgta	5220
gtgaattttt tttttaaat gtaataagtt gttttttaag gtaattatgt tttttttgtt	5280
ttttttttaa aaaataaaaa taaaaaatat atagaaaaaa atttgtgagt ttagaaaaaa	5340
gaagtaattg gtagaagggt ttaattaagg taaagagttg taagggtgaag ttaagaaaat	5400
gtaggtattt aaaaaatgta ggtaattttt ataaagggtt ttggggagag gtatatagag	5460
ggatttttgtt gttgaaaaag atttagataa aagaaattta ggggtgggtg ggggtaaaat	5520
gattaatgga attgggggaa gggagggaat aaattgtaaa gaaattatag aaaagtgggtg	5580
ggtttttgag ttggagagaa gagagggtt tttggtattt tgattttttt gttgtgtgtg	5640
tttttaatat gtttgaggta aaagtttgaa tggggattat taagatttgt tatagataag	5700
tttttgaagt tgttttggtg taggttattt ggttttttag tttttggtgt gtggttagtg	5760
tttgggtgtt ttggaaagt atttttgggt agtttttgat agtgtgattt ggtgtttaag	5820
tagtttggga ttttgtgtgg atttgatttt tttagattgt aggtagtttg ggaggagggt	5880
tgggtggggg aggtgtagga tttttgtgtt gtttttttga tgatttgggg attgttatgg	5940
tttttttttg gtgttttttg gtttttttgt tttgtatgtg gtgtgaagg gttagtaggg	6000
aaggagtaga gtagggggg tgggggtgtt ggagttttgt ggaggtttg gaggtttttg	6060
ggtgggaaaa gtttgttttg aattggtagg gatgtgtaat aatttttttt attttgaaga	6120
gtgaaatagg gtttgtgtt tttattttta taagtaaatt ggtatttaga gtgtattgta	6180
gataaagggt gtttggggat ttgaattagg ggtttttggg ttagtttttt tgtttagatt	6240
ataggagttt tttgtttttt tatatatatt ttattttttt ttagtttttt gttttagttg	6300
agtaatttta ttggtaggtt tagtggtagt ggtgtgtgtg tgtgttggga gttttggggg	6360
atgttttttg ttggagtgtt ttattgtttt gtagagggtt ggggtgtgta ggggtgatatt	6420
tttgggtgtt gtagatttag ggtatttggg ttgttttgtt ttgtggtttg tgttattggg	6480
tgtggagtgg tttgggtgtt aggttagagga gagtggggt gaaaaatagt gttatagttt	6540
aaattatttg ttttttaatt taattgtgtt ttggtaagtt ttgtatgttt ttgggagatg	6600
ttgtgggaag ggggagaaat ttatatgggt tttggagagt tgttttgttg tgtgtatatt	6660
tgtggtagag tttttttggg ttgtgttttt tatttttgtg ttttttttg tttttatttt	6720
ttttgttttt tttttttttt tttagttttt gttttttttt tttagttttt gtgttttttt	6780
ttttatagtt gataaatgaa tggtttaggt gaaatttttg ttttttttgt gttttaaggt	6840
agtagggagg gagtagtgag ggagggtgtg tgtttttttg gatttgtttt tagtttagtt	6900
tataagattt gtagaatgta gatgtttagg aattgggagt tttgtggtgg gtgtgggtgt	6960
tttttgatgg agaagttttg tatagggtga gaaaaataag ttttttagat taagtgtgta	7020
ttttttataa ttttgtgttt aaggatggaa ggttttagtt tttttttaag ttatttattt	7080
tgtattttat aatttgttgt gtatattag gagtttgag gattttgaaa aaaggttttg	7140
gtttgtgtaa agtgtagaga tgtttttttg tgagtgtgtt ggaatgtatg tgggtgtttg	7200
ggtttagttt gttgttttag ggatgaatgt atatggttg ggtggagtta ggttttgag	7260
gtttagatgt gttatttttg gtgtttttta gaaagataat atatatatg tgtttttaaa	7320
atttatgatt atttgaattt aatggtaggg ttttgtttta ggatattgta aaagaagggt	7380
ttttaattta aaatttaaat ttttttaatt ttagggtggg tgttggtagg gaaagtggag	7440
agaagggtgg tagtgggagg aaaaaagaaa ggaagggaag ggaagtggaa ggaagaggga	7500
ttgtggaggg gaggaaaagg agaaaggaat ggagatgggg tagatggtat ttaaatagtt	7560

-continued

aggaatggat	tggaaggggt	tgttttgggg	tttgggtttg	aattttgaga	tttgttttag	7620
gataatTTTT	TTTTTTTT	agtggtaata	aatgtgattg	tagagggtgtg	ttgggatgga	7680
aggggtgggaa	agaattagat	aagtggaggt	ggttttagttt	tgtttaaagg	ggttgggttt	7740
gggtttgtt	tttatttgtt	tatttttgtt	taagggagga	tttgaatgg	gtgttttttg	7800
tagtggttgg	gttagttagt	tgttatgtgg	gttgggtttt	tggtttttag	ttgagatggg	7860
tgtgattttt	tgggtttaga	gattgtagta	tagttgtata	ggtgggttgg	gtagttgagg	7920
ttttatttta	ttgaggttaa	tgggggtggg	gtggggaagg	tagggtttag	tttaattgag	7980
agagttgagt	ggttttgttt	ttgggatttg	gggttggttg	tttttgaggt	tgatttgagt	8040
tttatttaga	tagtgttgaa	tatttttatt	aaaaggttat	ttaagtttta	ttttatttta	8100
ttttattttt	gtttgttttt	tgggagtggt	aggaggggta	ttgtgttttt	gttttgaggt	8160
aggggagagg	ttaaaattta	ttttaagtaa	atTTTTgtt	agaaaaatga	gattttatat	8220
tgtattttat	ttgttgattt	tttaaata	tttaggttat	tttaaagttt	atTTTTtagtt	8280
aattaataga	ttaattttgt	tttagattgt	tataaagagt	ttgtttggaa	tggtaaatat	8340
tgttatttga	taagattaat	agaattttat	gggggtaaat	atggaaggggt	tgtgagtgtg	8400
attgaggttt	ttgtgtttta	atgaggaatt	ttagtgggga	tttgggtggg	attggaagtt	8460
tagtattttat	ttagggagtg	attgagttgg	aaatttgagg	aagtttggtg	tagatgagat	8520
ggtgtttttt	agagttgtat	tttaagtttt	tttattttgt	tgtatattgg	aagaattgga	8580
ttttaagat	tttttttatt	atTTtagatt	ttgtagttaa	tagttatggt	tttttttttt	8640
tgtttgttga	aagtgtgtat	ttaattgttt	taagtatttt	ttgataattt	tgtaaatttt	8700
atTTtgaaaa	tatttatata	aagttttaaa	ataggatgtt	gtgagtgtaa	tataaaaaata	8760
ggaaggtatt	aatggaagta	ataattagtg	ttttgagata	gaagtgggtt	ttagtttgta	8820
atTTtgaaat	gttaagagta	gtatattagg	gagaagggtt	ttaagagttt	gattttttgt	8880
tttgagtttt	gagtaaagaa	agaaatatgt	tagaattgta	ttgtttgttt	gtagtaatat	8940
tgtaaaggggt	tttaattgat	tttaattgta	ttgagtaatt	atTTTTttta	gtattaagtt	9000
tatagttttt	aatttttttg	tgagaagatt	aggtaggaag	tgggatgaga	tttaagataa	9060
ttatttagga	tttatgttag	aatttaaagt	tttgattttt	ttggtagata	aaaatagagt	9120
tggtttttaa	ttgtattttt	tgtttttaaa	aaaatgaggt	ttgaaaggaa	tttttttaga	9180
tttatatagg	tttatagatg	taagtttttt	tttttttttt	gtgtttgtta	attggaggga	9240
gatgtttttg	atatttttaga	atatttgtat	ttatatattt	gtaaattttg	ttaggaaagg	9300
atTTaaatag	ttaaattaaa	ggtgattatg	attaatgtta	aagataattat	ttattgtttt	9360
ttttttgtt	gtattattag	gtttaaatat	tttggagtta	atgattgggt	atTTaattga	9420
gtaatagagt	ttttgagagt	gtaaagatta	attaattttg	ttttttaaag	tttggtaatt	9480
aagttttttt	tttttgggtga	gatttaaata	tttttgttga	tgttatttat	ggataaatat	9540
tggtaaatat	tgtattttaa	atgggatatt	gttttgtgat	ggagtagaat	tttttaagat	9600
atatggtttt	tttgaaagaa	agaattaatg	attgtagtaa	ggtatttttt	tattattttat	9660
gtagagaaat	tagttaggat	tgataaatat	ttattgttgg	gagttatttg	ttttggtttg	9720
agaaaaatag	ttgattgtta	tataggtttg	tattttatta	gaaaaataa	agtatgtatt	9780
ttttttgatt	tggaattaga	aaagggggat	tagatatgga	tgttttgggt	aagttttatt	9840
ttaattagtg	aatttttgtt	agtttttttt	tgtagttag	atttggtaaa	ttatatatat	9900

-continued

ataaagtttg	tgtgtgtgtg	tgggggtgtg	gtgtgtgtgt	atgtatatag	ggttttttaa	9960
tattgttaag	tttttattat	tgaatgtgtt	agagggtggt	gtattttaga	ggtagtata	10020
aagaattttt	gtttaatgtg	ttagttttgt	tttttgaaat	atatttttgg	gtagatttat	10080
taattataaa	aggatataaa	tgtatatatt	ttttatattt	gtatatattt	tattttattt	10140
ttttaaggat	tttaagaggt	gtttatgttt	ataaaatgag	attatatatt	ttataataaa	10200
tatatgtgat	aaattataat	ttttttatta	ttaaattatt	gatttgatat	tatttgaggt	10260
atagtgaat	gagtattaga	tttgaatta	gaaagtttga	gttggaattt	ttagttgaga	10320
aaattaagta	aagatattat	ttttggattt	ttagtgttat	aatttattaa	atgaatatgt	10380
agttataaag	attaaaagag	taagtatttt	gaaaaattat	ataaaaaagg	gatgataata	10440
tattatgttt	tgggatttat	attgatttaa	ttttataatt	ttaataataa	ttgtattatt	10500
taagttatgt	agtttttatt	gaattttata	aataaattag	agaggggtat	attgttttaa	10560
aaattgttta	gttttattaa	ttagaagata	gattatagaa	attgatttta	atgaagttgg	10620
atatttagtg	gaaaggaaga	attattttaa	aaatggtttt	aggataattg	atgggtttat	10680
gtaaaataag	aattagtttt	tatttgattt	tttattttta	aagaaatata	agatgaatta	10740
aatatgaaat	gtaggggaata	tttttttata	aaatgataaa	agaaaattat	gggattaaaa	10800
aaaataagtt	aggaatagag	aaatttgttt	taagtaaaat	ataaaattta	gtgttatttt	10860
aggaaatatt	gataggtttg	attaagttaa	aattatataa	aagtaataaa	tatatgtgga	10920
aaatatttat	gatatttgag	agataaaaa	ttataaattt	ttgatatttt	atggattttt	10980
atgaataaatt	ttgaaaagtt	tatatattat	aaaatatgag	taaaaattat	gtataggtag	11040
tggtagaaag	agatatagag	ttttaaaatt	atagaaaaaa	ttgtttagtt	ttatttgtaa	11100
ttaaataaat	agtaattaaa	atattaatga	tatgttatat	tattttttta	tttaattagag	11160
gagtaataaa	aaaaataaat	tttttttttt	tttttttttt	ttgagatgga	gtttgttatt	11220
taggttggag	tgtagtggta	tgatttttagt	ttattgtaat	ttttgttttt	taggtttaag	11280
taattttttt	gtgttagttt	tttttagtagt	tgggattatg	ggtatttatt	attatgtttg	11340
gttaattttt	gtgttttttag	tagagtgggg	ttttattatg	ttggttatgg	tagttgattt	11400
tgaatttttg	attttaagtg	atttatttat	tttagttttt	taaagtgttg	ggattatagg	11460
tatgagttat	tgtgtttggt	aaaaatgaaa	ttaaaaagtt	tgatgtaagt	ttgtgttggt	11520
taggttgtgg	agaaattggg	tttttttatat	attgttgat	gaaagtaaat	aggttttttt	11580
attttaaaaa	ataatttggt	aattattttt	ttttgatttt	ttattaaaa	tttttaaaag	11640
atgtatatat	tttttgattt	agtaatttta	ttttttgata	ttattatatg	gatggatatt	11700
ggtattaata	ttatttaaga	taaatggata	aggatgtttt	ttgtagtatt	gtttttaaga	11760
gtaaatattt	ggaaatagtg	taataggttt	attataggga	attattttaa	taaaagttgt	11820
tgtatttatt	taatggaaaa	ttatgtagtt	gttaaaaaata	atatggaaaa	attgtatttg	11880
ttgatttgga	atatttttta	aaatatattg	tatttttaagt	gaaaaaattg	atagataaaa	11940
ttgagtatat	attatatatt	taattatgta	aaatgggtgt	atattttatt	ttttaataaa	12000
tatttattga	gtgttttagg	tattgtggat	ataggagtga	ataaaatatt	aaagattggt	12060
ttgttgaat	ttttattatg	gtaggtagag	aaaggtaatt	tttaattata	tatgtaaaat	12120
ggtgatatgt	gttaggatga	aaatagagta	tgggaaaggg	aatagagtat	agtatgtttt	12180
aggtagagta	atgagaaagg	agttttattag	gtagagaata	gaggagtgtt	aaggggtgat	12240
aatgtgaaaa	agtttagtta	agtttaggtt	taggaattta	tatgatgtta	ttatttgat	12300

-continued

gaagtaaatt tatgaaattt tgtattttaa ttatgtgatt tttttttgtt gtataatgat	12360
tttttttgtg agaaagaatt agtgagtaat atgatatgtg atgtttatta tttttatttt	12420
gtagttgaag ttttttagaa agttggttgt ttaggttatt taattagtag aatggtggat	12480
aaaattattg tgtttttgtt ttttaagttat tttttgttag ttgagttttt ttaaaattaa	12540
gggagaatat taaatattta ttttaatagaa ttttggtttt tttttggagg ttgtttattt	12600
ttggtgtgtt tttttatttt tttgtagata ttttttatgg ataggagttt tattgatattt	12660
gttttttatg ttttttaggt aattaaatag ttttagtatt ttgtatttgg attgtttaat	12720
gatttttttt tttttttatt gattttattg gagtagaagg ttaaatatag gtattttattt	12780
ttatattttt taggtttttt tatgtagata gagttgttat tattagtaga tgggttttat	12840
aagagagata tggaaaagtg gaagaatgga aatagggatt tggagtgggtg ttaaaaaata	12900
agaagaggt ttttgaaagt tttttattta atatgattaa atatagaaga taaagatgt	12960
aatttaatga ttaaggatag ggatatagtt ttgagaaaat tatagaagtg gaatgaatgt	13020
tttggtttt aaagaagagg gaatggaaat gtagttttga ggtagttaa gtttaataaag	13080
ttttttgata tattttattg agaaataaat gattttttt taatttaagg tttgtaaatt	13140
taggtgatta taggggtgag gtatggagtg gaaatgggtg agagattaag tgggggtagt	13200
aatggttttg agtatataga ttttaagggga gagattgttg tttggttaat gtttatgatt	13260
agttagtga aatatagatt tagtggtgtt agattttttg attttttgaa agaagtaga	13320
gattttatat ttagtgtaa tagtttaatt tttaaattga gtaatttatt taaatatatt	13380
tatgtaattt atttaaatgt atttagagta tgtttgattt tttttttgta tttttattta	13440
gttttagggg taattttaag gaaagtgatg tgtatatttg ttgtatttat gtagatggag	13500
gattattttt ttttttttaa taaagatat ttttataagg attgagtgtg gtataaagag	13560
tttgattttt agtattttaa ttgagtagga aatttaaata tagaagattt tttttgggt	13620
tatatttttg gtttttttaa atgagatttg attttttaga tattgataaa tattatatta	13680
ttgggttatg gtaattttt aagtggattg atttaaaagt gtttaaaagt attttagggtg	13740
ttaaggaatt gttggtatgg agttttaaag gtgtttttgt ataagaaata gaggtgttga	13800
ttatgtattt gatttataaa tattttttta ggatatattt tgaatttggg aatgaagaaa	13860
tattgtggga attattaagt tgaataaagt attagttttg ttttttagaa gtttattgtt	13920
ttatagagga gattagatat gtttatagat aatttgaagt atatagtaat aaattatagg	13980
atgatttatt ttaagtataa agtgtatat gttattatag tggtaattta ggttataatt	14040
tttaaatgta agagttttaa atattatttt tttatttttt agtgatatga aattatgtta	14100
gaatatttgt agaaagtttg aattagtaat ttttaataata aatattgagt attgatatat	14160
tgtatattat tgagatttat tttatatata ttattttatt aatttttatg attatttttg	14220
taagtattaa tttttataat taagtagtga tagaaaatga ggttagaga gatatagtat	14280
ttattttaa gttatagaat ataagtagag aatgtagaat ttgtattttg ggttatttga	14340
tagtggagtt tagattttta aaattaggtt tataaagatt ttttgaaaa aaagaaaaat	14400
aaaaataga gaataaattt ttttgaagaa agtaggaaa ttattattta taagtataat	14460
taaagagaga tttttaattt tttttttttt tttttttttg agatggagtt ttgttttgtt	14520
gtttagggtg aattttagtt tattgtaatt tttatttttt gggtttatgt tataaagaaa	14580
gtttttattt gattggtttg atttgattgt ttattagaat ataggtttt tagtagggtt	14640

-continued

```

aataagttat aatagtttta gtgatataag ataaatataa attaataata aggttaattt 14700
agaattatta atatatgttt tttattgtga attattagtt ggtttataga tggaaaatat 14760
aagatatata atgtattagt ttaaaaaatg ttttttagag agtgtttaga attatagaat 14820
ttaaaaaatt tgttttgaaa aggtttaatt tttattttaa aatatataaa gaaattaaat 14880
gtaaaaaaat taagagaaaa gggtttaatt aattaaatgt tgtataattt taagttttat 14940
ttatgtttta gattagtttt gtaatttatg attaatttta gtattagttt gtttttaatt 15000
taggtttgtt aaataatata gaattttttt ttaatagaag ttttttaggg ttttagttga 15060
aaaatatata ttagttagga tgtttttaat tgtaaggaat aaaaagtat gattttattg 15120
ggtttaaata ataatgatat ttataatttt atataaaatg aagtttaaag gtatggttta 15180
ttttagggtt aatttttaag gatttatggt ttgtttttgt tgattttttt ttagttttat 15240
ttttagaggt attttattta tttatggatg taaaattttt ataatagttt tagtagaaaag 15300
attttagtat gataatgttt aggaagaaa attgattatt tttttttttg tttttttttt 15360
aggatgaatt tatttttttt agggattttt tattggagga agagattttt ttatatttta 15420
ttggttaaaa ttatattata tattgtttta aatttaatta ttggaagg gaatggattt 15480
tttataattg atttaggata ttaggattat tttttgggtt ggggtttgat ttaggatttt 15540
ttaaagtatg tgattgaggg taggtattag aatggaatta gagttgtgtt agaaaggaga 15600
aatgtatggt tattatagaa tagttttaaa tgttattgag gaggtataa ttgtaaaaat 15660
taaataattt ttgtttattt ttttgaagtt ggtattttga tttttagatg gttttttaat 15720
ataggttttt ttttttttta ttattatagt tgttttgaaa tttaggttg aaggaatat 15780
tttgagattt agattgttaa atgttttttt taaagttatg taataaatta aatggtaaag 15840
ttagggtagt tttttgattt agtatagggt attttttttt attttttatt tttgagattt 15900
tagaattggt ggattgtttt taaaatttat ggtaagaatt ggttattttt gtaattaata 15960
tttttttata atatatttgt tttgtttgtt tagttagtta gaaattatat ggagtttgtg 16020
ttttaaaaag ttgttggaag tttttatttt ttgttttgtt attatgtgta tgaattatta 16080
attgggtttt ttttttttta tatttgatga agatgttttt ttttaatat ttttttttat 16140
tgttttttat tttttttgt tttatttatg attagtgtt ttgttttaaa tagattttgt 16200
ggttatttat tttttttgt gttagttttt atttattagt tatttgaatt gtgggtttta 16260
ttgtttttta tatagtttat ttatttgttg ttgttatata tatttatatt gttatatgtt 16320
tttaaattgt attttggaat gattttggta atggttgtat tggatgagat ttaaattaat 16380
aattaaagta ttgagatagt ttttgttatt ataagtttat tttgttttt atagtttaag 16440
aggagtattt tttttatttt ttattattta atgtttaata ttatttttat tatatataat 16500
gtataaaaag tatgtgattt atgatttatt ttaaatttga atgtttgtga tttattttgt 16560
gttttttttt atttttagg                                     16579

```

<210> SEQ ID NO 19

<211> LENGTH: 16579

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

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

<400> SEQUENCE: 19

```

tttagaaatg aagaaaaata taaagtaaat tataagtatt taggtttaaa gtgggttata      60

```

```

aattatatgt tttttgtgta ttgtatgtaa taagggtagt attaagtatt aagtaataag      120

```

-continued

agtagaagaa gtagtttttt ttggattata aaaataggag taaatttgta atagtagaag	180
ttatttttaaat gttttaatta ttagtttgaa ttttatttaa ttagtttatt attaaaatta	240
ttttagaata tagtttggga atatataata atgtggatgt gtgtgataat tataagttag	300
taggttgat gaagaataat agagattata gtttaagtaa ttagtgata ggagttggta	360
taaaaagaaa tgggtgatta taaagtttat ttagaaatag atagttgatt ataaatagag	420
taagagaaga tggggagtaa tagagaaaga tgttgaaaga aaggtatttt tattaagtgt	480
aaggtgagaa ggaattaatt ggtggtttat gtatataata ttaaaataga gaataaggat	540
tttggtaggt tttttaagt atagatttta ttagtttttt agttgggttaa gtaataaaaa	600
taaatgtatt ataaggaggt gtttaattata aaagtgatta gtttttgta tgaattttta	660
agtagtatta atagttttta agttttaaga gtaagaatg aaagaaaata ttttgtattg	720
agttaagaaa ttttttggt tttgttattt aatttattgt atgatttttg aaaagattt	780
taataatttg ggttttagaa tgttttttt tagttttaa ttaaggttag ttataatgat	840
gaggaggaga agaatttttg tggggaatt atttagaaat taaagtgtta attttaaaag	900
aatggtaaaa gattattttg tttttatagt tgtgttttt ttggtagtat ttagaattat	960
tttataatag ttatgtgttt tttttttta atataatttt ggttttattt tagtgtttat	1020
tttagttat atattttgga gaattttggg ttaattttta gtttaaggga tgattttgat	1080
attttaagtt aattatagag gtatttttt tttttttaat gattggattt ggaatagtat	1140
ataatatagt tttggttag gagatgtgag gggatttttt tttttagtga aggggttttg	1200
ggaaagatag atttattttg agaagagaat aaaggaagag gtggttagtt ttttttttg	1260
gatattatta tgttggaatt tttttgttg aattgttgta gggattttgt atttatgagt	1320
aagttagata tttttgaaa tgaagttgag aaagaattag taaaggtag gtgtaagttt	1380
ttaaggatta gatttagagt ggattatgtt tttgaatttt attttatatg agattataaa	1440
tgttattatt gtttaaat aataaaattg tgattttttg ttttttatag ttgaaagtat	1500
tttaattggt atatatattt tagttgggat tttgaaaaat tttattgga aagaaatttt	1560
atattattta gtaattttag attgaaaata agttaatgtt gaaattagtt ataaattata	1620
aaattgattt tgaatataaa tagggtttg agttgtatag tatttaatta attagattt	1680
tttttttaa ttttttgta tttgatttt ttatatattt tgaataaga attagattt	1740
tttaataaa aatttttaa ttttataatt ttagatattt tttgaaggt attttttaag	1800
ttaatatatt atatatattt tttttttat ttatagatta attgataatt tataatgaaa	1860
aatatatgtt gataatttta aattaatttt attattagtt tgtatttatt ttatattatt	1920
agaattgtta tgatttgta attttattg aaagtttatg ttttagtaga taattaggtt	1980
ggattaatta aatagaaatt ttttttatg tatgaatttg ggaggtggag gttgtagtga	2040
gttgagattg tgtttgggtg atagagttag attttgtttt aaaaaaaaa aaaaaaaga	2100
aattaaagat ttttttttaa ttatatattt aaataaat tttttatttt ttttaggaa	2160
aatttgtttt ttatttttta ttttttttt ttttaggaa atttttataa gtttaatttt	2220
aaaagtttg gttttattgt taaatagttt agaagtgaag tttgtattt ttattttata	2280
ttttgtgatt ttgggtaaaa tattatatatt ttttaattt tattttttgt tattatttag	2340
ttgtgaaaat tagtatttat tgaagtgatt ataggaattg ataaataat gtatataaaa	2400
taaatttttag tgatatataa tatatttaga ttttaatttt attttaaaaa ttattaattt	2460

-continued

agattttttg tagatgtttt aatatagttt tatattatta agagtaaaaa gaatagttatt	2520
tgggggttttt atatttgaaa attataattt aggttattat tatagtagta atatatattt	2580
tatgttttga ataaattatt ttataattta ttgttatgta ttttaaatta tttgtaaata	2640
tgtttaattt tttttgtaga atagtaggtt tttggaaggt agagtttagta ttttatttaa	2700
tttaataatt ttataaatat ttttttattg ttaggttttag ggtatatattt tgaggaaatgt	2760
ttgtagatta aatatatgat taatgttttt attttttatg tagaaatatt tttgaaattt	2820
tgtgttaata attttttaatt atttgaaata tttttgaata tttttgaatt agttttattg	2880
aagatttatt atgatttaatt aatatgatgt ttgttaatat tttagagagtt aaattttatt	2940
taaaaaatt aaaggtgtga tttaggaaag gattttttgt atttgagttt tttatttaat	3000
ttaaatatta aaaattgaat tttttgtgtt atattttaatt tttataaaag tgttttttgt	3060
tggggagagg agaattggttt tttatttata taggtataat aagtatatgt attatttttt	3120
ttaagattat ttttaagatt aggtaagaat ataaggagga gattagatat gttttaaata	3180
tatttgaata agttgtataa atagtgttga ataagttatt taatttaaaa attagattgt	3240
ttatattaaa tataaaattt ttaatttttt ttaaaaaatt agaaggttta gtaatattaa	3300
gtttatattt tagttgggtg attatgagta ttaattaagt agtagttttt ttttttaagt	3360
ttgtgtgttt aaggttatta ttgtttttat tttagttttt atttattttt attttatgtt	3420
ttgtttttgt agttatttaa gtttgtaaat ttgtaattaa aagatgatta tttattttta	3480
tagtaaatat attagagggt tttattgatt ttagtgtttt tagaattgta tttttatttt	3540
tttttttttg gagtttaaga tgtttatttt atttttgtaa tttttttaag gttatgtttt	3600
tatttttggt tattaagtta tattttttgt tttttatgtt tgattatatt aaatggaaag	3660
tttttagaag tttttttttt tgtttttagt attattttta atttttattt ttattttttt	3720
atttttttat gtttttttta taagtattgt ttattagtga taataatttt gtttgtatgg	3780
aaagatttga gagatatgag agtgaatgtt tatatttagt tttttatttt agtgaagtta	3840
atgagaaggg aagaaagtta ttgaataatt taggtatagg gtattgagat tgtttaattg	3900
tttgggagat ataaaaata aggttaatgg aatttttatt tatgagaaat gttttagggg	3960
aggataaggg atatatagg gatgggtagt ttttgagaga aagttaagat tttattaaat	4020
agatatttag tatttttttt tgattttggg gaagtttagt tgataagggg tgatttaagg	4080
atagagatat ggtgggtttg tttattattt tgttaattag gtattttaga taattaattt	4140
tttaggatat tttagttata aagtggatgt aatgagtatt atatatata ttgtttattg	4200
gttttttttt atagaaagaa ttattgtata gtagagagaa attatgtgat ttaggtgtga	4260
aattttatga atttatttta tataagtaat agtattatgt gaatttttga atttggaatt	4320
ggttaagttt ttttatattg ttgtttttta gtattttttt gttttttgtt tgggtgaattt	4380
tttttttatt attttattta agatatattg tattttattt ttttttttat gttttatttt	4440
tatttttagta tatattatta ttttgatat gtgattggga attgtttttt tttatttatt	4500
atagtaaaag ttttaataga gtagtttttg atgttttgtt tatttttgta tttataatgt	4560
ttagaatatt tagtaaatat ttgttaaaag aatgaatata tatttatttt atataattag	4620
atgtataata tatatttagt tttatttatt aattttttta tttagaatat aatatatttt	4680
ggagaatgtt ttaaaattag aaatataatt tttttatgtt atttttgata gttgtatagt	4740
tttttatttag gtggatatag taatttttat ttaaatgggt ttttatgatg agtttgttat	4800
attgttttta gatatttgtt tttagaagta atgttgtaaa gaatattttt gtttatttgt	4860

-continued

tttgggtgat attggtatta atatttattt atataataat gttgagaagt ggaattgttg	4920
aattaaagag tatatatatt ttttaaaaat ttttgataaa gaattaaaaa aaagtgattg	4980
ttaaaattatt ttttgaagta gaggaattta tttattttta tataatagta tatgagagat	5040
ttagtttttt tataatttgg ttaatatataa tttgtattaa attttttaat tttatttttg	5100
ttaggtatag tggtttatgt ttgtaatttt agtatttttg gaggttgagg tgggtgaatt	5160
atttgagggt aggagtttaa gattaattgt tatggttaat atggtgaaat tttattttat	5220
taaaaatag aaaattagtt aagtatgggt gtgggtgttt ataattttag ttattaggga	5280
ggttgatata ggagaattgt ttgaatttgg aagggtggagg ttgtagtgag ttgagattgt	5340
gttattgtat ttaatttga gtaatagatt ttattttaaa aaaaaaaaaa aagaaaaaaa	5400
ttattttttt ttgttatttt tttgattggt gaaaaaatgg tgtagtatat tattggtgtt	5460
ttaattatta tttattttaa tatagatgag gttgagtgat tttttttata attttaaaat	5520
tttatgtttt tttttgttat tatttgata taatttttgt ttatatttg atagatgtgg	5580
attttttaaa gttatttata aaagtttatg gaatattaag agttattagtt tttttgtttt	5640
ttaaatgttg tagatgtttt ttaaatatat tttttgtttt tgtgtaattt ttatttagtt	5700
agatttatta gtatttttta gaatggattt gggttttatg ttttgtttag aataggtttt	5760
ttatttttta atttattttt ttaattttta tgattttttt ttattatttt gtaaaaagtt	5820
attttttata ttttatattt gatttatttt atattttttt taggatgaga aattaggtag	5880
agattgattt ttatttttga tggatttatt aattgttttg aaattattta ttgaataatt	5940
tttttttttt attgaatgtt taattttatt ggagttaatt tttgtaattt gttttttgat	6000
taatgaaatt aaataatttt taaaataata tgtttttttt tgatttattt tatagattta	6060
gtagaaatta tatggttttg gtaatgtagt tattattaaa attatagaat tagattaata	6120
tgaattttta gatataatat attattattt tttttttgta tagtttttta aagtatttat	6180
ttttttgatt ttataaattg tatatttatt taatagatta tgatattaaa ggtttaaaag	6240
taatattttt gtttggtttt tttagttagt gagtttaatt taaatttttt gattttaagt	6300
ttagtgttta ttttattata ttttaagtaa tattaaatta ataatttaatt aataaaaaga	6360
ttataattta ttaaatatgt ttattgtaaa atatgtaatt ttattttata gatataaata	6420
ttttttaaaa tttttagaag aatgaggtag gatgtatata gatgtggaaa gtatatatat	6480
ttgtattttt ttatagtttg taaatttatt tagaggtgtg ttttaggaaa tagaattgat	6540
atattaaata aagatttttt gtattgattt ttgaaatata attattttta atatatttaa	6600
tggtagaagt ttagtagtgt ttaaaaattt tgtatatata tatatatata tgtattttat	6660
atatatatat agattttata tatatgtaatt ttattaagtt taggttataa gaggaattg	6720
gtaaaaattt attaatgaa atggagtttg gttaagatat ttatatttaa tttttttttt	6780
ttgattttta attagggaga gtatagtgtt tgtgtttttt ggtaaaatat agatttatgt	6840
aataattagt atattttttt agattagagt aagtaatttt taataatgag tgtttgttag	6900
ttttaattgg tttttttatg tgggtggtga aaaaatattt tattatagtt attaattttt	6960
ttttttaaaa gaattatgtg ttttgagaag ttttgtttta ttatagggtat atattttatt	7020
ttaaatatag tgtttattga tgtttgttta tgaatgatgt taataaaaat gtttaaat	7080
tattaggaaa ggagggttta gttgttagat tttgggagat agagtttagtt gatttttgtg	7140
tttttaagg ttttgttatt taattaagtg attaatattt aattttaaag tatttaagtt	7200

-continued

taataatata ataaaagaga gaataatggg tagtatTTTT aatattaatt atgggtattt	7260
ttggttttagt tgttttaggtt ttttttttagt agaatttata agtgtataga tataggtatt	7320
ttggaatatt aaaggtattt ttttttagtt tgtaaatata gggaaaaaga agaaatttat	7380
atttgtaaatt ttatgtaaat ttagggggat tttttttaa ttttgttttt ttgaaagtag	7440
gaaatataat ttaaaattaa ttttgttttt atttattaaa agaattaaaa ttttgaattt	7500
tgatataaat tttaaatgat tgttttgaat tttattttgt tttttattta atttttttat	7560
aagaagatta aggattgttg atttggtgtt ggaggaagta attgtttaat tatattggaa	7620
tttattggaa ttttttataa tattattgtg ggtagataat ataattttaa tatatttttt	7680
ttttattta gagtttagaa taagaagtta gatttttgga gatttttttt ttgatgtatt	7740
gtttttaata tttttaagtt ataaattggg aattgttttt attttagaat attaattgtt	7800
gtttttattg gtattttttt atttttatat tatatttata atattttatt ttaggatttt	7860
atgtgaatat ttttaggatg ggtttggtga aattgttaag aagtgttaa aataattagg	7920
ttatagtttt tagtagataa aaagaaaaag atgtggttgt tagttataga atttgaggtg	7980
atggaaaaga tttttaaaat ttagtttttt taatatgtag taaaatggga ggatttaaaa	8040
tgtagtttta agaagtattg ttttgtttgt attaagtttt tttaggtttt tagtttagtt	8100
attttttaaa taggtgttga gtttttagtt tttgttaggt tttgttaag gttttttatt	8160
gaggatatagg aattttagtt atattttagt tttttttatg tttgttttta tgaaatttta	8220
ttagttttgt tgaatggtag tatttgttat tttaggtgaa ttttttatga tagtttgga	8280
tgaaattaat ttgttaattg attaagaatg ggttttagag tgatttgaat gtatttaaaa	8340
agttaataaa tggaatatgg tataaagttt tatttttttg ataaaaagtt tgtttggagt	8400
aaattttaat tttttttttg ttttaaaata aaagtatgat gtttttttta ttatttttaa	8460
gaagtaagta aaaaataaat aaaaataaat aaaatttaaa tagtttttta ataagagtgt	8520
ttagtgttgt ttggatggaa tttgggttag ttttagaagt agttggtttt gaattttggg	8580
ggtgaggtta tttggttttt ttggttaagt tgggttttgt tttttttgtt ttgtttttat	8640
tggttttggg ggggtggggg tttggttgtg ttggttgttt gtgtggttgt attgtagttt	8700
ttgagtttag gaagttatat ttgttttggt tgggagtga agtgtaatt tatgtggtgg	8760
ttggttaatt tagttgttgt ggggagtgtt tgttgtgggt ttttttttgg tgtgagggtga	8820
gtgagtggg gtgggtttta gatttagttt ttttgggtgg ggttggtta tttttgtttg	8880
tttagttttt ttttattttt ttattttggt atgtttttgt agttatattt gttgttgttg	8940
ggaaaggga aagagtattt taaaatgagt ttgaaattt agatttaaat ttgaaatga	9000
atttttttgg tttgtttttg gttgtttgaa tgttatttgt tttatttttg tttttttttt	9060
tttttttttt tttttatggt tttttttttt tttatttttt tttttttttt tttttttttt	9120
tttttattgt ttgttttttt tttatttttt tgtttgatgt ttgttttgaa gttgagggaag	9180
tttaagtttt ggattaggag tttttttttt gtaatatattt ggggtggggg tttattgttg	9240
ggtttaagtg gttatgggtt ttgagagtat ggtatgtgta ttattttttt taaagatgtt	9300
taggtgggtg tgtttagatt tttgaaattt ggttttattt tggttatgtg tgtttgttta	9360
ttaagtggtg aattggattt aggggtgtgt gtgtgtttta gttgatttgt gagaagggtat	9420
ttttgtattt tgtatagatt aagatttttt tttagggttt tttaaatttt tagatgtatg	9480
ttgtagattg tggagtgtaa aatgaatgat ttaagggaaa attagggttt tttatttttg	9540
aatataaggt tgtgaaagat gtatgtttgg tttggggggg ttgttttttt ttgtttgtgt	9600

-continued

gggatttttt	tattaggggg	tgtttatatt	tggtgtaaaa	tttttaattt	ttagatattt	9660
aggttttgta	agttttgtga	gttgaattgg	gggtagattt	gaaggagtgt	attgtttttt	9720
ttgttttttt	ttttttgttg	ttttggagta	tgggaaaggt	agggatttta	tattagtatt	9780
ttattttgta	gttgtggggg	aggggatgta	gaagttagag	agagagaata	gagggtggga	9840
gaggggaagg	gaggtagaga	agatggagat	agagggagat	ataaagatga	ggaatgtgat	9900
ttagagagat	tttgttgttg	gtgtgtgtgt	ggtaggataa	ttttttaggt	gttgtgtaga	9960
tttttttttt	ttttttagt	attttttagg	gatagttagg	atttgttgag	atgtagttag	10020
attggagagt	aaatagttta	ggttatgata	ttattttttt	attttgtttt	tttttgttta	10080
atatttagat	tgttttatgt	ttagtgatat	agattgtggg	ttagaataat	ttggatgttt	10140
taggtttgta	tgtattaggg	atgttgtttt	atagtgttta	tgtttttgta	ggatggtggg	10200
gtgttttagt	tgaggatgtt	ttttgggggt	tttgggtata	gtgtgtgttg	ttgttgttag	10260
atttgttagt	aaagtgtttt	aattagagtg	gagaattagg	agagggtggg	gaatatgtaa	10320
ggaaatggag	gattttttga	gtttgggtga	gaggattgat	ttagaggttt	ttgatttggg	10380
tttttgagtt	attttttgtt	gtagtgtgtt	ttgggtgttg	gtttgtttgt	tgagatggga	10440
gtgatagatt	ttattttatt	ttttaggggt	agaaaaattg	ttgtatat	ttgtttgatt	10500
agaatagggt	ttttttgttt	aggggttttt	tagatttttg	tggggtttta	ataattttat	10560
tttttatttt	ttgttttttt	tttgttgggt	tttttgtatt	gtgtgtagga	taaaagaatt	10620
taggggtgtt	gggagagaat	tgtgatagtt	tttaagttgt	taaagagata	tgggtgggat	10680
tttgattttt	ttttggttgg	attttttttt	agattgtttg	tggtttgaag	gggttagggt	10740
tggtgaaggt	tttaggttgt	ttgggtgttg	ggttgtattg	ttaggagttg	tttgggagtg	10800
gttttttagg	aatattaggt	attgattata	tattaggggt	tgggaaatta	ggtggtttgt	10860
attaaggtgg	ttttggggat	ttgtttgttg	taagtttttg	tagtttttat	ttaaattttt	10920
gttttgagtg	tgtaagaat	aataataata	aaaaaattaa	agtgttaaag	gttttttttt	10980
tttttttagt	ttagaatttt	attatttttt	tatgattttt	ttataattta	tttttttttt	11040
ttttttaatt	ttgttagtta	ttttattttt	attttatttt	gggttttttt	tgtttgaatt	11100
ttttttaata	ttagggtttt	tttgtatgtt	tttttttaaa	agtttttatg	aaagttattt	11160
gtatttttta	agtgtttata	tttttttaat	tttgttttat	agttttttgt	tttaattaaa	11220
gttttttatt	aattgttttt	tttttttaag	tttgtgggtt	tttttttata	agttttttgt	11280
ttttgttttt	taagggggga	ataaaagaaa	tgtgattatt	ttggaagggt	gtttattgta	11340
gtttgggggg	aaaattttat	gtagtgttgt	gtgattgggt	ttggtgttgt	ttaggtgggt	11400
tatataggaa	gtgtggttgt	ttggggaagg	atgtggaggg	tgtgggatgg	tggtggaaga	11460
tgtgggagga	tgtgggggtg	gttgaagatt	ttgggttttag	tttttgttta	taatatttaa	11520
tgtttttgtt	tggtgtgttt	gtatgttttg	agtttttttt	tagaaagggt	tggggtagtt	11580
tgtttgtaag	tttaaatgtg	ggttgtgata	tttattatta	ttatattatt	gtattgttag	11640
agttgaggag	gagatttagt	gagaagaagg	aggagggaga	ggaggagggt	ttatttatag	11700
gttttaaaag	tgttttgtag	tttttagttat	ttttaagagt	ttaggttttg	aaagtaggtg	11760
gaggggtgga	aaggtagttt	tttgtgtttg	tggtagggat	tgtttgtttt	tttttttagg	11820
tatttttatt	ttttggtgtt	tttttttgat	aagaagtatt	aattggtttg	gggatagtag	11880
tgttttttat	ttagggattt	atttagtaat	attgttttgt	ttgtggaaat	tattgttttt	11940

-continued

ttattttttt	ttatttttta	tttttttgtt	ttttgtgtag	ttagtttttg	gttaggggaa	12000
aggagttttt	aggtttttag	ggggtaggtt	agtaatagat	aattgagtat	gattattttt	12060
ttttgggagt	atataaaatt	gtaaaattag	taaagaattt	ggttatagt	tgtttatgtg	12120
tttatagagt	ttgtttgttt	tattaaaggg	aagtgttagg	tttaaggttt	ttgttaattt	12180
gaaagagata	ttgagaaaa	gagatttttt	ggggattttg	agggaaagt	taagaatttt	12240
ttattgtatt	tttagggaat	tgtttaattg	ggagtttggt	tttaaaagat	tttggtaata	12300
aaaggttgga	taggaaattt	ttttaggtaa	attttttgtt	ggattttaa	agaatatttt	12360
ttttttgtta	taaatttttt	tttatataag	tttagatttt	gttttttttg	tttttttttt	12420
tttttagttt	ttttaagtat	ttttgagtag	aatatttgat	aatttttttg	agtaaatagg	12480
attttttgga	agtattaatt	attttttatg	ttttttggaa	ataagattat	aatttttttat	12540
gtgtatatgt	gatttttttt	ttttagttag	gtttatttta	ttgtgtaaat	agtattaata	12600
tatggaagag	ttttttgtat	tgtgttataa	aagattttta	ataggatttt	atagagaaaa	12660
gggttaaata	gttgatataa	aggatttttg	tttttttaga	aaagagggat	tttggatttt	12720
tttttttttt	gaagttaa	atgagtttat	ataataggaa	taaaataaat	ttaagggtga	12780
tattagtata	atattaggga	tattagaatg	gatggtaaat	tttttttttt	atataaatat	12840
gaaagtattt	ttataatttt	tttttttgaa	gtttttattt	agaaaattat	attaaatata	12900
ggatttttaa	atagtagtga	ttaaagatga	aagttaattt	ttttttttta	tttttttgat	12960
tagtttggtt	taaatttatg	ttttaaaatt	tttagtaatt	tagatttttg	tatatgtgta	13020
tttataagat	tttttttata	ttttgtaatt	tgtaggtatt	tagtatgggt	attgatttta	13080
agtataaat	aggtagaaga	tttgtagga	taatttttagt	ttttattatg	tgtatttttt	13140
attttttttg	ttatttatta	gttttttgta	gtaataata	tgtagttat	attatttgga	13200
aattaatgtg	tttttgtagt	gaattatttta	tttttagtaa	atttattgta	ttattatttt	13260
aaatattgaa	atttgatttt	attagatggt	tagaatgata	gttatagagt	agaattagat	13320
tttaattttt	atgtataata	tgtaataaaa	ttaagggtga	gaaagttatt	gttttaaagt	13380
tataataatta	atagtttagt	ggttttttta	gtgttaatta	tagtaatttt	agatttaaga	13440
ttgtttgatt	taggaatgga	gagaaataaa	ataaatgaa	tgtttattga	aatatatatt	13500
ttagatatta	agatagattt	ttttgtgaaa	tatttttaaga	tgtggatttg	tattttttatt	13560
tttaaatgtg	agtaaattta	gttgagtttt	aaagagttta	agtaatttgt	ttaagggttat	13620
tttgtaagt	atagatttgt	tagtttgaaa	agtttattat	tttttatttt	tttgtttttt	13680
tatgatata	taaagtaaaa	ttttttaata	aaggagaatt	agattagaaa	taagtaatta	13740
atagtagatt	aatgtttatt	tttagggaaa	tgtaaatagt	taatatttat	tataaataat	13800
gttatttgta	tataatggaa	tatgattgaa	agtgatgggt	tagatatttt	ttatattttg	13860
agatgttttt	atttatgttg	ttaaaattga	ttttttaatt	ttttaaaatt	tttagttttt	13920
tttaattgta	atagatttta	taataattga	ttaaatattt	ttataggata	aagggtatttt	13980
tttaggggtg	atttaaggaa	ttagaatatg	tataatgttg	aaaagtttta	aattgtttgt	14040
tttttttttt	ttttgatttg	taataattatt	atttagaaaa	ttaaatattt	ttttagggtta	14100
ttagataaatt	aaaatgaaat	tataatttat	ggatattttt	tttttatttt	tagggtagat	14160
tttttttttt	agttgtttta	aatggagaat	ggataaaaaa	gaaaaaggaa	aatatatttag	14220
ttatatttag	aatagagttt	atagttgtat	tattattttt	atgtaattgt	ttttaatgat	14280
tttttttttat	taagagaaag	aaaatgagag	agtattttatt	tttgattttt	taatgtttga	14340

-continued

taaaattatt	agaagttttt	gaatgggaaa	atatatattt	tttaggtttt	ataattaaaa	14400
ttgtatttat	ttgtttatta	tatttaagta	tgaattaaat	agtaattaaa	atattatttt	14460
gtgagttttt	ttttaaatta	tatttttgta	ttggggatat	tttttgtaa	aattttaaga	14520
ttttaggtat	aaataaagta	agatttttat	ttataatggt	tgtgtttaaa	agaaattata	14580
agttagaagt	tatgtgtttt	gtttttttgt	tttgattttg	tttttgggaa	tttggttgag	14640
tttaggggta	gtaagttaat	gatttgagga	ggaaagggtg	agatggtaga	tgtgttttta	14700
gggaatttta	ttatttttat	ttaaattttt	ttttgagggt	ttttaggatt	atatttttaga	14760
tttttttttt	tttattgatt	agggtgaagt	tttttagata	tgggtattatt	tattgttggt	14820
ttagggatga	taagatttta	gttttattta	aatttagaaa	gagtttaggt	ttaatttttt	14880
gatttgattg	gttttaggta	taattaagta	aattattttt	attggatgat	atagtaatgt	14940
tatagattaa	aaatgtattt	ttaattttta	aagatttttg	aaaaaggaag	tttatttttt	15000
aagattgtgt	tataattttt	tttttaata	aaatattgta	aattagggat	agatatttta	15060
tttagtattt	tgaatttttt	ttggtttttt	aggaaaggta	gaaatataag	gtttttattg	15120
tttttttttt	atttttaaaa	aagaaatttt	gtaaaaggta	aagtgtgggt	ggtaatttta	15180
ttgtgtttta	atttgaagta	tataaattgt	tttgataaat	taaatataga	taaatttttt	15240
tagaaagtta	taaaatatta	gatttttttt	agatatttaga	tattgagttt	ttgtttttta	15300
atttttaaat	taaaatgtta	tagtaaaata	tttgttttta	taaatttttt	ttttttttgt	15360
gttgtaaatg	ttttgtttat	ggttgagtag	attttttttg	aatttaattg	aaagattatt	15420
tatattatga	tagaattatg	taagttaaag	ttaaagttta	atatggagtg	atgggagagg	15480
ttgggttggt	gagaggagat	gattttatta	gtgagaagag	aggttttttg	agttggtttt	15540
taaggtagtt	aaatggttta	gtatatttgt	atttattttg	ggagggtggt	tggagtagag	15600
tattaaaaat	gttatttttt	ttataaagaa	aaagatgaag	gtataaagaa	tatagtttta	15660
ttaaattgaa	ttaaagattg	tattattttt	attttttttt	tagttatttt	tttaaagttt	15720
tttttttttt	tttttatagt	ttgaaaaata	taggttattt	ttataaagta	tttgtttgat	15780
tgaagtaaga	taaattgaag	agtttggttt	ttgaatataa	ggaagtggta	gagatttttag	15840
aattaatgag	gtaggggtga	atataatgta	gaaggttttt	taagaggtta	aatttttgta	15900
gttggttggt	atttttagtg	attatttggt	tttatttatt	taattagatt	atagtgggat	15960
tgagataagg	agtataatga	aagttgttgt	attgatgggt	taaaaagata	tatggtatta	16020
ttaatgtttt	tgtggtgagt	aaaggtagaa	tgtaagatag	aaaataagga	aatttttggt	16080
ttgtaagaaa	agtgtttagt	atgggtatta	attggtatgt	agaagtagta	agtatatatt	16140
ttatttttat	tttattgata	aaaagaattg	taggattaat	gtttgttttt	attaaaaatat	16200
gagaaattat	attttaattt	tataaattaa	aggttaaata	tttggtgtaa	taatttttaa	16260
aaaaagagaa	gatttgtaag	atttagtaaa	tatttattat	aaatagattg	tttttaattt	16320
tttatatttt	tttttaattt	aataaaagtt	tattgagtag	tttttatgtg	tattgtttta	16380
gatagtaggt	attattttaa	aaagtaataa	ttatttttta	tttttaagga	gtttataatg	16440
tttttagaaa	gaaattaatt	tattggagta	gttaaagtta	taattgattt	ataaaatgat	16500
tagtgaattt	agaggaaaat	tatatataaa	tatttttggt	tttattgggt	ggttttagtta	16560
gatattattt	taaaaaata					16579

-continued

```

<211> LENGTH: 3000
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 20
atattttaa tgtatat tttgttaata tattatata ttaaaagaat tattattttt    60
ttgaatatta ttaattatta taatagtatt atagtgatta agagtggatt taaataaagg    120
ttggtttgaa atttaggttt gttatttatt agtaagtttt taaaaatttt gagtttttag    180
tttttttata tgtgaaatgg agaaaataaa tttgtaaaat taattaaatt tgtttattta    240
tatttttttt tttagtgtat attaattggt attttttggt aggttagaat gtgttttta    300
ttatgtttta aatttgtttt gtgttaattt tattgttaga atttttttta ttttgagaat    360
tagaaaagga aatattatgt ttggtaatgt ttatatTTTT taaaataaat ttgtaggaaa    420
gaatatttag taagtatga gagtagtaat gattgttttt attattttaa atttataata    480
ttattttttt tagagttttt taagtattgt agataatttt ttatttatta aaaaataaat    540
tgtaattata agtatttagg gttgatattg tttttgaatt agatagtgtt tatattagtt    600
gtataagatt aatttaaagt agaggatgaa attttttttt tgaatttttt ttagaatgta    660
atttagtgaa tatattaaaa ttaaattttt tttgaatggg agtaattttt atggattaat    720
ttgtaatttt ttagattata ttttaaggtaa tgttagagggt gttgtattat aggtttttgtg    780
attagagatt attggatttg tttttggaaa agttatttat gatttttttg gttttggttt    840
ttttattttt tatattggtt taataatgat tattgtagag ttgttatgga agttaaatga    900
atttatgtat tataagtgat taatataatg attgatattt agtgatttat taataaatta    960
aaatatttat tattaatatg atagagaagg tgttggttaa atagataata ggttttttga    1020
agaggtgatt aaatggatgt aaaatttatg gattgtttat tttgtttatt tttgttgtgt    1080
tttttggttg tggatatata atgtgtgggt ataaaattgt aaattttatg tagttgtgta    1140
gtgtatgtgt agaaggttta gatatgaaat gttatttttag taatgtgttt agagaagttt    1200
tgatgttgtt ttggaagtaa gttgttgttg tttgattttt ggggtgttgg gatggatgtt    1260
tatatttgta tttagtagta ttggaagggg tttaggtttt ttgtagtata gtttattttt    1320
agattgttta gttttttata atatataatt ttatggaaaa gggatttttt ttttgtaga    1380
aaaagtgttt tagtttggtt tgggttggtt tttattttat gttgttgtaa gtaggtgaag    1440
ttttttttgt tttttttttt ggggtaagtg gaaaggagtt tggtaggggg tttgtagtgg    1500
tttgatatagg ggaattgggt agtgagagag ttttaggttaa ttttgggggt tgttttatag    1560
aagtaggtgg ggattgatag tggttttttg gtttagggag gagagtgtgg ttgtgggttt    1620
tttttttttag tttggaggtt gtagttgttt gagttgggtt ggggtgggggt ggggtggggg    1680
tgggtgtggg ggtatggaga ttatgggtgt gttatttggg atatttaggg ttttgaggtt    1740
ttgggtgggt tttatgtgag attgtaaatt atgataatag gtagttattt gaggttaaatt    1800
aaaaatggag tgggtttttt gtgtgttgtt gttttttgtg ttttgggtgg ttttttttga    1860
ggtttttggg ggttttatga gttttagtag gttggtgggt atgttgtttt tgttttatatt    1920
ttttgtgtaa gtgtgagggt gttggtagtg tgggttatgt tttgggtgtt tttgggtttt    1980
gtgtaaaatt tttattttgt ttatgtgaag ttgttgttgt ttttagaggg gggaaagagt    2040
tgtgggaaaa gttggggagt gatgattgtg gtgggtgggt gtgttttttt attttttttt    2100
tttttttttt tttttttgtt gtagtttggg gttttgggtt tttttttttt tttttttttt    2160

```

-continued

```

ttggagttgg tttttttttt tgttttggtt tttttttggt tgtgtatggt atttgttggt 2220
gggtggttga aggggatggt ttgtttttat tagaggata gtgtgaagg gaaattttga 2280
tattggaagg aatgagaata aatatttaaat tatggatgta ttgaattgtg gttgggatag 2340
atattttggg aatttgaggt ggattgggtg atgaggtgag tgattttttt ttttaatttt 2400
tgttttaggg tttttggggg agtttgaggt gagagaattt ttaaattttt tgggaaagtg 2460
tgtgaggttt tgttggggat gttgagtgtt gggatttgag gatgtgtagt tggatgggtg 2520
gtgggtgttt gtgtttttgg ggggtgtttg gaggttgggt gttttatggt tgaggggtttg 2580
ggttgtttgg attgtagtgg tgttttttgt tttagaagat gtttttaagt ttttaagggtt 2640
ttttttgagt ttgtttgttt tttttgggtt tgggtgggag tttgtgtgta atggagtta 2700
tttagtagtt tagtgtgtgg tttttatttg tattttgttt ttatttggtg gaggtgtgag 2760
tattggggtt tttttatat tttttttatg atgtgtatta ttttttgatg attttttaga 2820
tggtttaggt gtgaggatgt tgatttagag ttttttgag gggtataggt gtttgggttt 2880
ttttggtgtt ggggtgtgtg gtattttaaa ggtttgtgtt ttaattttta ggtattgatt 2940
gggtttttta attgtggtga ttttatttta atagttttta tgtgggtgtg attgaatgtt 3000

```

<210> SEQ ID NO 21

<211> LENGTH: 3000

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

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

<400> SEQUENCE: 21

```

gatatttagt ttatgttata taaaaattat taaagtggga ttgtttagt tgaaggttt 60
gattagtgtt tggagattag aatgtgaggt tttaaagtat atatgtattt ggtattggga 120
aagtttaggt gtttgtgatt tttgaagga ttttgggtta gtatttttgt gtttgatta 180
tttagggggt tattagaaag taatatatgt tataagaaag atgtggggga gattttgatg 240
tttgtgtttt tgttaggtgg aggtgggggt taggtagaag ttgtgtgttg gattgttgga 300
tgaattttgt tatgttagg ttttgtgttg attttggaag gtagaggtag gtttgaagg 360
gatttttggg gtttggggat gtttttagg gtagagagta ttgttgtggt ttgagtgggt 420
tgggttttta ggtgtgggggt atttggtttt taagtgtttt ttggggatgt aggtgtttat 480
gtattgttta gttgtgtgtt tttagtattt agtgtttggt gtttttggtg gagttttgtg 540
tatttttttg gaaagtttgg ggggtttttt aatttaggtt tttttgggag ttttgggtg 600
ggggttgga gaaggggtta tttattttgt tgtttgggtt gttttgggtt tttgaagtgt 660
ttgttttagt tgtggtttg tgtgtttgta attaagtatt tatttttgtt ttttttagtg 720
ttgaagtttt ttttttgtgt tgtgttttg gtgaaaatag gatatttttt ttggttattt 780
tataataaat agtgtatata agtgggggag aagtgggggt gagggagaag ttggttttga 840
gggggaggag gaaaggagag gagttaaaat tttgatttgt gatagggggg aaaggagaag 900
aaaagaaat gagagagtgt gtttagttgt ttagttgtt attttttggt tttttttgta 960
gttttttttt tttttttaag gtagtgataa ttttatgtgg ataggatgga agttttgtgt 1020
ggaagtggg aatggttgga gtgtgtgttg tgtgttggt agttttgtat ttgttaggg 1080
aggtgggggt ggggtgatgt tgttattggt tattgtgggt ttgtgaggt gttgggggtt 1140
ttgggggagg ttgttaggga tgtggggggt ggtggtgtgt gggggattta tttgtttttt 1200

```

-continued

```

atttgatttt gggtgattgt ttattgttat ggtttgtgat tttgtgtggg gattgttttag 1260
ggttttgggg ttttggaagt tttgggtggg gttgttgtaa tttttgtgtt ttttgtgttg 1320
tttttatttt gtttttattt ggggtgattt gagtggttgt agtttttagg ttgaggggag 1380
ggatttgtga ttgtgttttt ttttttgggt tggagagtta ttgttgattt ttatttgttt 1440
ttgtggggta gtttttggaa ttgtttggaa tttttttgtt atttagtttt tttgtgtagg 1500
ttattgtggg ttttttgttg gatttttttt tatttatttt aagggaggag atagaaggga 1560
ttttgtttat ttgtaataat gtgaaataaa aattaattta gattagattg ggggtgtttt 1620
tttgatggga ggaatatatt ttttttgttg agatatatgt tatgaaggat taagtgggtt 1680
ggggataggt tgtgttgga aggggttggg ttttttttag tgtgttggg tgtaggtata 1740
ggtatttgtt ttagatgttt aaaggtagg tagtaatgat ttatttttaa ggtggtgtta 1800
gagttttttt aggtatattg ttgaaatgat attttgtgtt taagtttttt gtgtatgtat 1860
tatgtgatta tataggattt atgattttat atttatatgt gtgtatgtta taattagggg 1920
atatagtaaa ggtagatgga ataaataatt tataaatatt gtatttattt aattattttt 1980
tttaaaaatt tattgtttat ttagtggtta tttttttgt tatattgata ataaatgttt 2040
taatttattg ataaattatt gagtgtaaat tattgtattg gttatttatg gtatatgagt 2100
ttatttaatt ttatgataa ttttatggtg gttattatta agttgggata aaagatgagg 2160
aaattaaagt ttaaagaatt atgagtgatt tttttaaaga taaatttagt ggtttttaat 2220
tataaagttt atgatataat aatttttata ttattttagg tgtgggttaa gagattatag 2280
attaatttgt agaaattatt tttattttaa gaaagttag ttttaataa tttattaagt 2340
tatgttttga aaaagggtta gaaaaaagat tttatttttt attttagggt ggttttatgt 2400
aattgatatg agtattgttt aatttaaaag tagtattaat tttgaatatt tatggttata 2460
atttattttt taatgagtgg ggaattattt ataattgtta agagggttta gaaaagggtg 2520
tgttgtaaat ttaaaatgat aaaggtaggt gttgtgtttt ttattattta ttgggtgttt 2580
ttttttgtag atttattttg gaggggtgtag gtattgttag gtataatgtt ttttttttg 2640
gtttttaagg tagaaagggt tttggtagtg ggggttggtat aaagtggatt tggagtatgg 2700
ttgagagtat attttgggtt aatgaggaat gttagttaat atatattgga gagaaaaata 2760
tgaatggata gatttaatta attttgtaga tttatttttt ttattttata tgtgagaaaa 2820
ttaagggttt agaattttta gaagtgtgtt aataaatggt agatttgagt ttttaagttag 2880
tttttattta agtttatttt tagttattgt gatgttatta ttagtattaa tggatttttag 2940
aaaaatagta atttttttta atatgtaata tattataagt taaatatgta atttaaatgt 3000

```

```

<210> SEQ ID NO 22
<211> LENGTH: 1984
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (9, 17, 18, 33, 41, 49, 50, 52, 54, 62, 80, 84, 127,
157, 159)
<223> OTHER INFORMATION: unknown base
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (175, 207..209, 214, 1181)
<223> OTHER INFORMATION: unknown base
<400> SEQUENCE: 22

```

-continued

```

agtttttagnt attattnnnta tttttattaa tgnnttttggg nttataggnn gngntttttt 60
gnttaaatag tagttaatgn aggnatttta aaaggatat tttttatat ttatttagtg 120
gttatgntat tatttaattt ttattataat tttgtgngnt tagtataatt tgtgntttta 180
ttttatttgt gaataaatgg aggtagnnnt ttgngtaatt tataagtaag tggtagattt 240
gggggttggtg tttagataat gtggttttag attttttggt tttttttggt ttttggtgta 300
tttgattga gtttagttat ttattttttt aatataatatt aattggtggt tttatgtttt 360
aggttttgtg tttagatgta gaaatatatt tttatttatt ggattttgggt atgggtttatt 420
gtttagtgtt ttgttatgtt atgtgttata taatgtattt attttggtat ttagtttatg 480
gaaaaaatg ttgaaagata ttgtatgtat gtttttataa taaaattatg taatttatgt 540
tttttttatt gttttgggta ttaggaaggt tttattttaa ttagagtaaa tatattagaa 600
tttgggtttg ttatttttagt ttgttgaatt ttttttttg gtttagattt tttattttgg 660
tttataaatt ttattgtata aatgtttttt attgtaaaat attttaaatt ttttttaagg 720
gaagggtgta tggaaatgat atagtaagggt tttttttgta tttttttaga tttttattag 780
ggaaggtaga tttagatgtt ttttttattt ttagttttta agttttttat ttataaaatt 840
ttttaagtag tgattattgt tgttttgagt atttgagggt agtttagagt tttgaaagg 900
taaaatttat atataaaaag aagttttttt tgatttagat tttagttttg gtgttagatg 960
tatatttgag gagtaggtgg ttagttagat ttttatttga gtagttaata aaatttattg 1020
ttttttaatg gaattttttt tgtaagtttg aattgatttg ttatttttgt gatttgtgag 1080
atggtaggga agtattaaat attattatga tttgggttat agtgggtggga aaaaaggaaa 1140
aaagaaaaaa aaaatttatt gttaagtttt gttaggtgta naaagggttg gaattgttgg 1200
ggttatttta tttagtttat tggaaataga gtggatttta ttaatathtt aataaagaga 1260
atttttttgt attaggttgg aagtgggtgt tagttttttg tgtaatttta ttttttggaa 1320
aagtggaaatt agttggtatt gtttagtgtg atttgtgagg ttgagtttta atagtttaaa 1380
gaagtaaatg ggatgttatt tttgtggggt ttgttttttg tgagggtgtt attttgtatt 1440
tgttatgtaa aatgagggag tgttaggaag gaatttgttt tgtaaagtta ttggttttgg 1500
ttattagttt ttatttaatg tttttgtgat gttgtgttg atttatttgg gaagttgggt 1560
ggttggtgag gtagagtttt tttttaaagt ttgggttttta tggaaaatat gtttagtgta 1620
gttgtgtgta tgaatgaaaa tgttgttggg tgtttttagt tggataaaat gtagttgaga 1680
attttgtttt ttttgttgt ttttttgttt taggtaggga agaggggttg ttgggtgtgt 1740
tttgtgtttt gtttttgtat ttggattgtt tggatgggt aggggtgagg ggtttttggg 1800
gggttggggt ttttggttgt ggtggtgaag atagattggg gtttggtagg gaggttattt 1860
tgagtttaga gatttttagt attttttata tataggtttt tattttggtg tgtgtgtgtg 1920
tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtatggt tgttaatggg 1980
agga 1984

```

```

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

```

-continued

```

<223> OTHER INFORMATION: unknown base
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (1923, 1931, 1933, 1935, 1936, 1944, 1952, 1967, 1968,
1976)
<223> OTHER INFORMATION: unknown base

<400> SEQUENCE: 23

tttttttgtt aatgaatgta tatatatata tatatatata tatatatata tatatatata      60
tatatatgta tgtatgttag agtgggagtt tgtgtgtggg ggggtgttag gatttttggg      120
tttggaatga tttttttatt gagttttgat ttgtttttgt tgggtgtgatt ggaggttttg      180
attttttgaa agttttttta ttttgtttgt gttgggtgat ttgaatgtag aaatgggggtg      240
tagagtgtgt ttggtagttt ttttttttta ttgggtagat gagaatgtat agaatgagtg      300
gagtttttgg ttgtattttg tttgattaga agtgtttggg ggtgttttta tttatgtatg      360
tggttgtatt gagtatattt tttgtgggag ttaggttttg aggagagggt ttgttttgtt      420
agttagttaa ttttttaaat agattagtag tagtattatg aaagtattgg gtagagggtg      480
atgattagga ttaaatggtt tataaatgg attttttttt aatgtttttt tgttttgtat      540
ggtagatatg gggtagatgt tttgtgagga gtgagttttg tggagggtgt attttatttg      600
tttttttga ttgttggggg ttagttttat aaattatgtt gggtaatgtt agttgatatt      660
atttttttag agaattggaat tgtatggggg attggtggtt attttttagtt tagtgtaaaa      720
agattttttt tattaataatg ttaataagat ttattttatt ttaataaat tagataaaat      780
ggtttttagta gtttttagtt tttntatggt tggtaagggt tggtagtgga tttttttttt      840
tttttttttt ttttttttat tattgtgggt taagtattga tgggtgttgg tgtttttttg      900
ttattttata aattataaag atgatatagtt aatttgagtt tgtagaaaaa attttattaa      960
ggggaatag attttattaa ttgtttaggt gagagtttga ttagttattt attttttaaa     1020
tgtgtatttg gtattaaagt tgggatttgg attggaagga attttttttt gtatatggat     1080
tttatttttt agaagttttg aattattttt aggtattttg aatagtagta gttattgttt     1140
aagggttttt ataaatgggg gggttttgga ttgagggtaa gagggatatt taggtttgtt     1200
ttttttaata ggaatttaag aaaatgtagg gaagatttta ttgtattatt tttatgtagt     1260
ttttttttag aaagaattta aggtgtttta taataaagga tatttgtgta atagaattta     1320
tgaattaaaa tagaaaattt gggttagaag gaaaagttta gtaaattgaa atgataagtt     1380
taaattttta tgtatttgtt ttgatttgga taagagtttt ttgatagtta aagtaataaa     1440
agaaatataa gttatatagt tttgttgtag aaatatgtat ataagtgttt ttagtattat     1500
tttttatggg ttaaatattg gaataaatgt attatgtaat atgtaataata gtaggatatt     1560
gagtaatgga ttatgttaag atttagtgga tggggatgta tttttagtat ttggtatagg     1620
gtttggaata tggagatagt aattagtgtg tgtaaagaa atagatgggt agatttaatt     1680
aagatgtagt agaaggtagg gaggggtaag gagtttgagg ttatattgtt taggtatagt     1740
ttttggattt gttatttatt tgtaaattat ntaaanntt gtttttattt gtttatagat     1800
aaaatgaaan tataaattgt attgantnta taaggttgtg gtgaggatta agtgatgnta     1860
taattattga ataaatataa agtgatgtgt tttttaagat nttntattg gttgttgttt     1920
ggntagaagg ntntnntttg tggntttaag gntattgata aagatgnnag tgatgnntga     1980
agtt                                                                 1984

```

<210> SEQ ID NO 24

-continued

```

<211> LENGTH: 7833
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<400> SEQUENCE: 24
gttttttggtg agatatgtgt tttataagtt ttaatggaga aaaatgtaag tattttattt    60
tttgaaattt ggttatttga gtaatgagaa aatagttatt ttttttagga tagtggtttt    120
taattatggt tatgtgtttt tttaggaaaa ttttaaaaat atatatatat taatgttttt    180
gtgttatttt tagggatttt aagtttttga atatgaattt tgtattagta ttttttaatt    240
atthaggtga ttgtgatgtg aaattatgat tgagttttat tgttttaaga tgaaataaat    300
tttttttagt attgaaatta taaattttaa ttattaaaat taattaaggg tatgggaatt    360
aataaggtat aggggaagtt ttatattata aaattatttt tttaaattat agttttattgt    420
ttatatgtta tttgttattg tagaaaaggg tgaaaaaata gtaaatttaa ttatttttag    480
tttgaaaaat tatttagaaa tgaagatgat gattttgaaa tattgttaat attatttgat    540
ttataaataa tgttttaata tatttattat atattgtag atattttttt atatgaatat    600
tatatatata aattaaggtg ataatgtatt tagaatattt tatttatatt tatgtatttt    660
aagtaggtta gaaattaaga tatgagttat taagtatgag atgttaaggt gtggggtag    720
aaattatatt gtattttatt attaataatt aatatatatt ttaattatt atatttttaa    780
ttttaatttg tatattttta attattttta attatgtgta taaatataag tatatatatt    840
tttatgtatt tattttattt tatttttatt tattttattt tataggggat ttttttaaat    900
ttattattat taaattatat atttttattt taatttttag aataagttta ggaggtaggt    960
attgttatta tttatatttt ataatgagg aaattgttta tagttataaa gttattgtgt    1020
tagatatatt agaagtttaa tatatatttg gtgaatatat gtataaaaat agagagatag    1080
atatgtataa tagtttattt ttatattgag taaaagtttt taatttgttt tagaaatttt    1140
tttgtagaaa ttgagtaaaa attgaggtat ttttttattt gttatatagg tataggtggt    1200
attttatttt ttttaataagg atgaatattg aaatgtggat ttttaagttt aatttttagat    1260
tttttgaatt tttgatagtg ggatttgga tttgtttatt gttttaaagt tttttaagga    1320
atttataatga ttaattaggt ttagaaatta ttggatttta ttgttgaagt ttgagaatta    1380
aagtttggtt tttattgttg ttttatagaa agggtaaatg aagtattatg gatagaattg    1440
atatgttttt agttagtttt tttttttaga agttaatagg tagtaatata gtagaaatta    1500
gtgatttatg ttttgtgttt tgaagttagg tagaatttta tagagtttta gtagtgttat    1560
tgatgagatt tgttttttgg ggtaagttgt ttgatgtttt taaagttata ttttttttat    1620
ataaaatgag ataatttttt ttgttttata ggggtgtttt aaagattaaa taaaaataat    1680
atgttttatt ttatatggta taatgtttga ttttaagaa gtaaaggata tattttattt    1740
ttattgaagt aattagaaag tatgaaatta tgaaggagat aagagttttg attggtagtg    1800
tattttattt ttttaggttt atttatttat tttaaattat tttgttggga gaataatttt    1860
taagtttttt atttaagttg tgagtaattt tatattttat aatgatgttt tttttatgag    1920
aaaaaaaaat gtttttaagt tttttggaga aaatatattt gtattatttt tattgaaaaa    1980
tttaataaatt ggattttggt tttttgtatt aatttttagag tgtatatggt ataaataaag    2040
tgtttttagt taagaagatt gaaagtaaat atggtatagt attttaaaat aagaattttg    2100
taaatatatg gtatgattgt gttatattat tagtaattat atgatatgta atgtaaagta    2160

```

-continued

tagtttatag attttaaattt aattttaata agtaaattga ttttgttttg ttggggaaaa	2220
gttaaagtat taatttaatt gttaatgtag ttttgtttat ttttttggtta ttttagtgata	2280
agtttaaata atgtatatat ttttatttat atatttagta atataatttt ttgtttaatg	2340
agtgatgttt ttttgttatt tgggtggtgtt tgtagttttt agaatttgtt ttttggtggt	2400
attataatat taagtataga gtaagtgtaa taaaattgta gtatttttat tgaaaagggt	2460
ttgttttaaa ttgtttaata atttaaagga tttttgtgga agtaattgta ttgtttaatt	2520
agttataatt agtaattaat ttttttgag ttttaattta tttttggtaa aatgttttag	2580
gaagagtata tattattaga aagtatgtta aaaatttatt tagtagaaaa tttaaaaata	2640
gttttttttt gttaagaggt tttttaaaat tttatttata tagttaaatt ttgaaatttt	2700
agtaggtttt gttttattat tataattatt gtataaatat ttttaaggat tttgttttta	2760
gttttaagta tgattttatt ttataagttt gattagttat tatattagtt ttgttatgga	2820
aatgatatg tttttatttt ttgttgtaga gttgttaaat tttgatttat atttatgttg	2880
ttttttttgt tgaaagtttg tagtgaaaga aatttttaat tttttgtttt gtaatttag	2940
ttggtagttt tatttaatgg gtattttgtt tttttaaaga atttagttgt ttgttttaga	3000
agttgatttt ttgatgtttt taatgtttgg ttttaattgat ttgttttaat ggagtttttg	3060
ttggtgagga gtgagatgtt attgattaga atgttgggat ttgttgttta attgttagga	3120
gtgagagata ttgagattta gaaatttttg gaggtgggag gggagaggga tagttttgga	3180
tggagggtga gatgtaagat aaagggatgg attttatata ggaaaaaaa aaagattttg	3240
ttgaggtatt gaggtgttgt atgattatat tttttaaagg agaagttaaa aagtaaggaa	3300
gtgggaggag gttggagggt aaagtattta aaaggattat ttgggtataa tttgtttttt	3360
tgttggtgtt tgtaaggat agatagtttt gtttttaaag tatatgaatg ttttttttaa	3420
gtgattggga atggatatta attgtttgtt aaatgttatt aaatgttttt ttaaatttag	3480
gggatataga aagagggtga taaaaggaga atttaaatag aaaaggag gatttggagg	3540
tttttgaaag tggggggaga agaaggagga gggataatag agaggaatag agaaggagag	3600
tggagagaag ataaataaaa ataaaaatag gaattattga ataattatat attaaaaaga	3660
aagttttttt ttatgggtga tttaaaatat tgagattgta atagtattt tggttatgga	3720
agaaagatgt ttttttttat ttttgttttt gaaagttttt ggttttggtta ttggtgatta	3780
aaattttatt aggttaaaga gtgtgtttaa ttgtttgaag aatgtagtag atggaagggtg	3840
ggtttttggtta tggtgtttgt tttttttgtt ggagagaatg aaagaaatgt gtagagttag	3900
agatttttgt tgagtttagt ttttttttgt tgttttaggt tattggttat ttggtaaaaga	3960
tttgagtaag gaatgtaggg ttattgtttg ggttaataaa tggagtttgt tttttttttt	4020
ttggatgttg ttgtttggtt gatgtttttg gtaatttatt tgtggtgtat gtagaggagt	4080
tttttttttt tttttagatt atttgtttt attaatttga ttttttaaat atatttgatt	4140
gtatttttta ggtggatata ttaatagggt atgggttga gaggagtggg tgatgaggag	4200
agggatttaa atttgtgaat gtttgggtt ggttgaggt gtggggggtt tgggaggaga	4260
gaggggagaa gagagaagga aggagagtgt ttgttgggat ggttgagttg ttttggtgag	4320
tagttttggg gttgtatgtt ttttggggag atgttgttgt tgtttttagg ttggtaaagag	4380
tggttttaat attattgttt tttatttttt tttttgtaaa ttttttagaga aatgtttttg	4440
gtttttttgt tgtgatattt ttagtttgta tttttttata gtttaggtgg tgtgtttttg	4500

-continued

tatgttgagg	tgttggtgt	tagtaggatg	tttttttttg	tgttgatttg	tttttttttg	4560
ttttgttgtt	gttggttttt	tgatattttt	gtttttatta	tttttagttt	ggagagatgt	4620
tatttagttg	tggtttgtat	ttgtggtttg	gggttatgtg	tggagagagg	gtgttagttt	4680
ggattttgtt	tttggtaggg	ggtgttttgg	agtggagagt	gaggtgaatg	gtatatgagt	4740
gtgtgggtag	tttattttga	agtttgagtt	ttttatttga	gttatgtttt	gttttagttt	4800
atttgggtta	gtgtttgttg	agttagttta	tttgtggttt	ttgtggttgt	tttttttttg	4860
tatttttgta	tttatttgtt	gatttttttt	ttttgggatt	tgtattttgt	tttattaatt	4920
agagtttgat	tgtttttttt	tatgtgattt	tgggtgggtt	gaggatttgt	tgttttttaa	4980
atgttagagg	gatgtgggtg	gtagagtttg	agagtggtt	gttgggttgt	ggggtgtttt	5040
gatttttttt	ttatttttgt	tttttgggtt	ttatttggtt	gtttttggat	ttttgttttt	5100
ttttgttttt	tgggttttta	gagttttttt	tttatggtag	tagttttttg	tgtttttggt	5160
gtagtttttt	agtggatgat	ttttttgttt	tggggttgag	tttagttttt	ggatgttgtt	5220
gaaatttttg	agattatgtg	tgggttttgt	tgttgttttt	ttgttgggtg	ttattgttat	5280
tgttgttgtt	tttgttgttg	ttgtttgttg	gatgtttagt	agtttgttgt	ttggtttttg	5340
tgattttgtg	ttttttggaa	gttgttttgt	gtttagagtg	tgtatgaatt	agttatgggtg	5400
ttgtgggagt	ttttgtggta	gttagtagtg	tggatatttt	gtgaggggtt	ttgttgggtg	5460
ttgttgttgt	ttgttatgtt	atttatttga	gtttgttttg	tgaagtgtgt	tgtttttttt	5520
atttttttaa	gtgattgtta	aatgtttatt	ggttgaatt	gttttggtta	gtttagaatt	5580
tttgtttttg	attttttaat	ttttagaag	aatatgtgta	tttagtatag	attagtttat	5640
tttagtgtgt	tttttttagt	ttttattttt	tattgtttta	gatttttaat	attatttatt	5700
tttatttaga	gaaataaggg	gaattgttgt	aggtttgggg	gtgaggggtg	gttttgggat	5760
gggtagaaag	tgtagggtga	gtaggaaatt	tttgtatgtt	tgtgtttata	ttggagtgtg	5820
gaggattttg	agaaatatta	aatgggatgg	ttttttgggt	ttattgtttt	gaaagagtat	5880
taatttttagg	ggaaatattg	aaatagaagt	ttgtttatta	ttaaagaaaa	aagttttatt	5940
aggatgagga	agaaataatt	ttatgagaaa	gaatgagtga	gaaagtaata	aattaaatgg	6000
tgattgtagg	ggaattgttg	atttttggtg	aaggtgttat	gaggttgtat	tggttttttg	6060
ttgaagatta	ggttatatag	atttttagagg	agttgggttt	taatagaatt	tttttttttt	6120
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tatttatttt	tttttttttt	6180
ttattttttt	tttttttagg	tggtaaaaga	tattggtttt	gtagttaga	tatgtttttt	6240
tttttgtttt	tttaagtttt	aaggtagtat	aggggagttg	agaaaaagaa	tattttgtgg	6300
gttttttagg	ttggagtggg	tatgattgag	gttggtagg	ttttatgtag	gtgagttgag	6360
ggtggaattg	attttagtgg	gtgttgattt	ttttattttt	ggatagggtt	ttgtggagtg	6420
ggttagggtat	tttttttgtt	tgtttgggtt	tttttagatt	ttgatgggtga	atgtttggta	6480
ggtttttgtt	tgttgaagtt	ttttaattaa	atagggttag	aggatgggag	ttgttgtatt	6540
tttagttggt	atagtatttg	gtttgatagt	ttgtagtata	gggtgtatgt	aattttttat	6600
tttttgtgaa	tataattttg	ttgtagttaa	atttggtttt	gaataaagtg	ttttttaaag	6660
atgtatataa	gttgaagtgt	atgtaatttt	agagaggagg	gaatgattaa	ttgtaattta	6720
gggtgaaagt	ttgtatagtt	tttagttatt	attgatgtaa	atgttaaaag	gaaaattatt	6780
atgtattatt	ttaatatttt	ttttataaag	ataagttgag	atatgtaatt	ttattagatt	6840
tgggttaata	gattgttttt	ttttttggta	gtttttaaat	ttggtatttt	aataaaattt	6900

-continued

```

aatatgtttt tataattttt tgatttatgt gtatatgtgt gttgtttttg aaagaataag 6960
ttttattttg ttattgttta attatttttt agatgtttta ttatggtaat aattatgagt 7020
ttgtaaaat aatttttgga aatgttgatg gttttgtagt ttaatataga ttggtttggt 7080
ttatttttag tttttgtatt gttttaggaa ataattaatt taaatgtgaa gttgatattt 7140
gtaattaaga aattatatat ttattagata ttttaaaggg gattgtataa attaaagaga 7200
ataaattggt ttgtagata gggtgttaag aatttggtat ttgttttta tttttgttaa 7260
tttagagggt attaatTTTT atttgagtta aatagattat tatagaaaat attgtgtttg 7320
tttattttta ttattgaggt ttgttttttt tttgttttgg atatatttta aataaggggt 7380
tgtttttagt gttgaagtaa aagaataatt aaagatgggg aaatggtaaa aggggtattta 7440
gagattatta ttatgttttt tttaaaatgt ggagttttgt gggtataaat attgtttatt 7500
taatgagtaa aaaataaaaa taaaaaaaa ataggaagta aatgttaagt tttattttat 7560
tattgttagt attaatgtaa gttttaaaaa atagtattat tagaaaagga tattaagga 7620
gaattgatta gaaaagaatt gtggaaaatg gaaatgaata ttgattttt aattagattt 7680
tgaggttatt agtagatagt gatTTTtag tatagttata gttgttggt ttaaaattta 7740
ggataagtat tttaaagttt taaagtagtg tttttttttg ttaaaaattt gtaagatgtt 7800
ttaatgattg gagtgTTTT tttgaatttg agg 7833

```

```

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

<400> SEQUENCE: 25

```

```

ttttaaattt aaagagaata ttttagttat taaaatattt tatagatttt taataaaaaa 60
aagtattatt ttgaagtttt aaaatatttg ttttaaattt taaatttaat aattatagtt 120
gtattgtaag gttattgttt attgataatt ttaaaattta gttaagtgat taatatttgt 180
ttttattttt tataattttt ttttagttaa ttttttttta gtattttttt ttgatagtgt 240
tattttttta agtttgtgtt aatattgata gtggtgaatg aaagttaaat attgtttttt 300
tgtttttttt ttatttttat tttttgttta ttagggtgat aatatttatg attataaaat 360
tttatatttt ggaaaagagt tagtgatgat ttttgaatat ttttttatta tttttttatt 420
tttaattggt tttttgtttt aatgattgaa ataatttttt atttgaaatg tatttagata 480
aagaggaaat aaagttttaa taataaagat aaatagggtat agtggttttt gtgatgggtt 540
gtttggttta aatgaagatt gattattttt aagttaatag gggtggaagt ggggtgttaa 600
gtttttgata atttatttgt aaaattagtt tatttttttt agtttatgta gtttttttta 660
aaatatttgg taaatatgta attttttgat tgtaaatgtt aattttatat ttaagttagt 720
tattttttta aataatgtaa gggtaggaa tgaagtaaat tagtttgtgt tggattataa 780
agttattaat atttttaaaa attgtttttg taggtttata attattatta taataaagta 840
tttaaaaagt gattaggtta tagtaaagtg aaatttattt ttttaaaaat aatatatatg 900
tatgtatgaa ttaagaagtt atagaaatat gttgagtttt attaaatgt taaatttaga 960
aattgttaaa aaagagaata atttattgat ttaaatttaa tagggttgta tattttaatt 1020
tgtttttgta aaggataaat tagaatgatg tataataatt ttttttttgg tatttatatt 1080

```

-continued

agtaataatt aggaattata taggttttta ttttgagtta tagttggtta tttttttttt	1140
tttaaagtta tatatatattt agtttatata tttttttgaa agatatttta ttttagagtta	1200
gatttaatta tagtaaaatt atatttatag aagatgaaaa attatatata ttttatatta	1260
taggttgtta aattgaatgt tatgttagtt aggagtgtag taatttttat tttttggttt	1320
tatttaatta ggaagtttta gtagagttaa gtttgtaaag tgtttgttgt tagaatttga	1380
aggaatttga gtgagtaaga agagtgtttg atttatttta tagaagtttg tttagaaatg	1440
gaggagttag tgtttattga agttggtttt gtttttggtt tgtttatatg gagtttgatt	1500
agtttttagtt atgtttattt tggtttggga gatttgtaaa gtgttttttt ttttaatttt	1560
tttgatttat tttgaagttt agggagtaa agagaggggt atatttggtat tgtaaaatta	1620
atgttttttg ttgtttagga gagaagggaa tgagagagag agagagatag atagatagag	1680
agagagagag agagagagag agagagagag agagagagag agaaatttta ttgaaattta	1740
gttttttttag aatttgtgtg atttggtttt taatgggaga ttagtgtgat tttatggtat	1800
ttttgttagg aattagtgat tttttgttag ttattatttg atttattggt tttttgttta	1860
ttttttttta taaagttatt ttttttttat tttagttaaga tttttttttt taatgatgat	1920
aaagtttttg ttttagtgtt ttttttagga ttggtgtttt tttaaaatag tgaatttaga	1980
aaattatttt gttaaatttt ttttaaaatt ttttagttt taatgtaagt gtaagtatgt	2040
aaaggttttt tgttatattt gtattttttg tttatttttag aattattttt tatttttggg	2100
tttgtaatag ttttttttgt ttttttggt agagggtgggt ggtattagggt gtttagggta	2160
gtaggaggtg aggggttgag gaggtgtgtt agggtaggtt ggtttgtgtt ggatatgtgt	2220
gtttttttgt ggagttaaag ggttggggat gggggttttg gatttattag agtaatttta	2280
gttggtgggt gtttggtagt tatttaagga ggtaggaaa gtagtgagtt ttattgggtg	2340
ggttatgatg agtagtatga tgggtagtag tagtagttag taaaagttt tgtaaagtgt	2400
ttagttgttg tattgttgtg gggattttta tagtattatg attagtgtgt gtaattttgt	2460
agtagtaaat ggtttttgag gaatatagga ttgtgggggt tgggtagtgg gttattgagt	2520
attttgtgga tgggtgtagt agaggtgggt gtggtgtagt tggattttg tggggaagta	2580
gtagttaaat ttgtgtatga ttttgagagt tttagttaata tttagggatt ggttttagtt	2640
ttggagttag aggggtgttt gttgagaagt tgtgttgtag atgtgggaag ttgttgttat	2700
aaggagggag ttttgggaag ttggaggata ggaggagatg ggagtttagg ggtagatgag	2760
tggagttaga ggaggtagggt tggagggaga gtttaagggtt tttgtagttt ggtagtgtt	2820
ttttgagttt tgttgtttgt atttttttg tgtttgggaa gtagtaggtt tttagtgtt	2880
ttggggttat gtgggaagag gtagttgggt tttgattggt ggagtaggat gtaggtttt	2940
ggagggagggt gttgatgagt aggtgtaagg atgtaaggag gaggtggtt tggaaagtat	3000
agatgggttt tttgttaggt tgttggtttg agtgggggta ggtgggttat ggtttaaatg	3060
agaagtttgg gttttagggt gggttatttg tatatttata tattatttgt tttatttttt	3120
gttttaggat gttttttatt gaagggtggg tttggattag tgtttttttt ttgtgtgtga	3180
ttttgggttg tgagtgtggg ttgtggttgg gtggtgtttt tttgagttgg agatggtggg	3240
ggtggagggt ttagaggagt agtagtagta gggtagagag gggtaggttg gtgtgggaga	3300
gggtgttttg ttggtgattg gtgttttagt gtgtgggagt gtgtgtttta ggtttagggg	3360
ggatgtagggt tgggaatggt gtggtggaga ggttagggat gtttttttag ggatttatag	3420
gaaagaggggt gagaggtgat ggtgttagaa ttgttttttg tgatttggaa gtaatagtag	3480

-continued

tattttttat aagagtgtgt aattttaagg ttgtttgttg aggtagttaa gttatttttg	3540
taggtgtttt tttttttttt tttttttttt tttttttttt ttaggttttt ttagtatttg	3600
atttagttta agtggttgta gggttggaatt ttttttttta ttatttggtt ttttttagtt	3660
tgtagtttat tagtggtttt atttgggagg tgtggttaga tgtgtttgga aggttagatt	3720
ggttgggata agtggtttga gagaaagaga aagggttttt tgtatatgtt gtgggtgggt	3780
tgttgggagt attggttggg tagtggtgtt tgggaagggg agagtgggtt ttatttggtg	3840
gtttaggtag tgattttgtg ttttttattt gggtttttgt tggatggttg gtgatttggg	3900
gtgatgagag aagggttaatt ttggtaggag tttttgggtt tgtgtgtttt tttttttttt	3960
tttagtggga agggtaaatg gtatagtggg atttggtttt tgtttgttgt attttttagg	4020
tagttagata tatttttttag tttaatggaa ttttagttgt tagtaatggg attaagagtt	4080
tttggggata aggtgggaga ggaatatttt ttttttatga ttggggttat tattgtagtt	4140
ttagtggttt ggatgtttta tagggaagag tttttttttt ggtgtgtgat tatttagtga	4200
tttttggttt tgtttttgtt tatttttttt ttgttttttt tttttatttt tttttgttat	4260
tttttttttt tttttttttt ttgtttttaa aagtttttgg attttttttt tttttattta	4320
aatttttttt ttgtgttttt ttttttgtgt tttttgaatt taggagagta ttgataata	4380
tttaataggt aattagtgtt tatttttaatt tttttaaaag aggtatttat atattttgaa	4440
aatgggatta tttatttttt gtagatatta gtagaaaaat aaattgtatt tgagtaattt	4500
ttttaagtat ttttaatttt aatttttttt tatttttttg ttttttaatt ttttttttga	4560
gagatgtgat tgtgtagtat tttagtgttt taatgaaatt tttttttttt ttttgtgtga	4620
aatttatttt tttattttat atttttggtt ttgtttgaga ttgttttttt ttttttttat	4680
ttttaagat ttttgaattt tagtggtttt tatttttggt aattaagtag tagattttag	4740
tattttagtt ggtggtattt tgttttttat tgatgaagat tttattaaaa tagattaatt	4800
agattagatg ttggaggat tagaaaaatt gtttttagat agagttagta aattttttta	4860
ggaaaatagaa tattttattag atagagttgt taattaatat tgtaaaataa ggaattagaa	4920
attttttttg ttatagggtt ttagtagaga aggtaatata aatatagatt aagatttaatt	4980
aattttatag tagagaatga gaatatgtta ttttttatag taagggtggt gtggttaatta	5040
attaggttta tgaaaaaag ttatgtttga aattaaaggt aaagttttta aaagtgttta	5100
tgtagtaatt atgataatga aataggattt gttaggattt tagagtgttg ttatgtaagt	5160
agaatttttag agaatttttt agtagaggaa aattgttttt gaattttttg ttaagtaaat	5220
ttttggtata ttttttaata atatatgttt tttttaagat gttttgttaa aagtaagtta	5280
aaattttaaa ggagttaatt attggttgta attggttaatt aaatgtggtt gtttttatag	5340
aggtttttta aattattaaa tagtttgaag taaagttttt ttaatgggaa tgttgaatt	5400
ttgtgtgatt tattttgtat ttagtggtat agtggtatta agaaataaat ttgaaattg	5460
gtaagtatta ttaagtggta gaagaatatt atttattgag tagagaattg tattattgaa	5520
tatgtaaata aaaatatata tattatttag atttggtatt aggtattaaa gaagtagata	5580
agattgtatt agtaattgga ttagtggttt aatttttttt tagtaaggta aaattagttt	5640
atttattaga attaaattta agtttatgaa ttgtattttg tattgtgtat tatatgattg	5700
ttagtaatat gatataatta tattatgtat ttgtaaaatt tttattttta aatattatat	5760
tatatattatt ttttaatttt ttgagttaga atattttatt tgtggtatat atattttaga	5820

-continued

attgatgtag aggagtagag tttagtgtgt agatttttta gtagaaatag tgtagatata	5880
tttttttttag aaaatttaag aatatttttt ttttttatgg aaagaatatt attataaagt	5940
gtgagattat ttatagttaa agtaggggggt ttgggagtta ttttttaata agaatagttt	6000
aagataaata aatgaatttg ggaaaaaag atatatgttt aattagaatt tttatttttt	6060
ttatgatttt atattttttg attgttttaa taaaggttaag atgtattttt tgttttttag	6120
gtgttaggta ttgtgttatg taggatagaa tatgttattt ttatttaatt tttaaaatat	6180
ttttatgaga taaagaatat tattttattt tatataaaag gaatatgggt ttgaaagtat	6240
taggtaattt gttttaagaa ataaattttg ttagtgatat tgttgggatt ttgtgaaatt	6300
ttgtttgatt ttagagtata agatataagt tattaatttt tgttgattg ttgtttgtta	6360
gtttttgaga ggggaaatta attgggaatg tattagtattt gtttatgata tttattttgt	6420
ttttttgtg gagttgtagt aagggtttaa ttttaatttt taaattttgg taataagatt	6480
tagtgatttt tgaatttggt tgattatatg aattttttga gaaattttga aataatagat	6540
aaattttaag ttttattatt agggatttag aaaatttgga gttgggtttt gggattttata	6600
ttttaatatt tatttttggt ggagaagtaa ggtattattt atatttatat gataaatgaa	6660
aggatatttt gatttttggt tagtttttat agagagggtt ttgagatagg ttaaaagttt	6720
ttatttagtg taaagatgag ttgttgata tgtttgtttt tttgttttta tgtatatgtt	6780
tattaaatat gtattaagtt tttaatatgt ttgatatagt aattttgtga ttgtagataa	6840
tttttttatt tgtaaaatgt gagtaataat aatatttggt ttttgggttt gttttaaaga	6900
ttaaaataaa aatgtatggt ttaatggtag tggatttggg gggatttttt atataaataa	6960
gtgaatggag gtatgaataa ataaatatat aaagatgtgt gtatttatat ttatatatat	7020
aattaaaaat agttaaagat gtataaatta aagttaaatg tatgtgatat tgaagtatat	7080
gttgattatt gataatgaag tatagtataa tttttaattt tatattttta tattttatat	7140
ttaataattt atattttaat ttttagttta tttagatat atagatatag atagaatgtt	7200
ttaaatgtat tattgtttta gttttaatgt ataattttta tatgaaaaag tatttattag	7260
tgtgtagtaa atgtattaga atattattta taggttaaat gatattgata atgttttaga	7320
gttgttattt ttatttttgg ataatttttt aaattgagag taattaaatt tgttattttt	7380
ttattttttt ttataatggt aaataatata taaataatga gttgtgattt aaagaaataa	7440
ttttataatg taaaagtttt tttatgtttt attgattttt atgtttttta ttaatttttg	7500
tagtttaagt ttgtgatttt agtggtgagg aaagtttatt ttattttaga gtagtggggt	7560
ttagttatga ttttatatta taattatttg gataattaaa gaattattgat gtagagttag	7620
tatttaaaga tttggaattt ttagaagtga tatagaagta ttggtatata tatattttta	7680
aagttttttt ggagaaatat atagtatatga ttgagaatta ttgttttggg gaaagtgatt	7740
atttttttat tatttaaata gtttaagttt aggaggtaaa atatttatat ttttttttat	7800
taaaatttgt aaaatatata ttttattaaa gat	7833

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

<400> SEQUENCE: 26

ggataggagt tgggattaag at

22

-continued

<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

aaatcttttt caacacaaaa 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

cccaaaaaatc aaacaacaa 19

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

<400> SEQUENCE: 30

tttgtattag gttggaagtg gt 22

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

<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
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: 0002393:1381020

<400> SEQUENCE: 33

cttctctctt ctccccctctc 20

<210> SEQ ID NO 34

<211> LENGTH: 21

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: MSP PRIMER CALCITONIN

<400> SEQUENCE: 34

aggttatcgt cgtgcgagtg t 21

<210> SEQ ID NO 35

<211> LENGTH: 24

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: MSP PRIMER CALCITONIN

<400> SEQUENCE: 35

tcactcaaac gtatcccaaa ccta 24

<210> SEQ ID NO 36

<211> LENGTH: 26

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: MSP PRIMER VERSICAN

<400> SEQUENCE: 36

tgggattaag attttcggtt agtttc 26

<210> SEQ ID NO 37

<211> LENGTH: 23

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: MSP PRIMER VERSICAN

<400> SEQUENCE: 37

cactacaacg ctacgcgact aaa 23

<210> SEQ ID NO 38

<211> LENGTH: 21

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: MSP PRIMER TPEF

<400> SEQUENCE: 38

tttttttttc ggacgtcggtt g 21

<210> SEQ ID NO 39

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: MSP PRIMER TPEF

<400> SEQUENCE: 39

cctctacata cgccgcgaat 20

-continued

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

<400> SEQUENCE: 40

cggagggtac ggagattacg 20

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

<400> SEQUENCE: 41

cgacgacgcg cgaaa 15

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

<400> SEQUENCE: 42

gacggatttt tttttaacgt tttttc 26

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

<400> SEQUENCE: 43

aaataaaata ccacctccgc ga 22

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

<400> SEQUENCE: 44

ggtgattggt tattgttatg gtttg 25

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

<400> SEQUENCE: 45

cccctcaacc taaaaactac aac 23

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

-continued

<400> SEQUENCE: 46

ggatgtgaga gttgttgagg tta

23

<210> SEQ ID NO 47

<211> LENGTH: 24

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: HM PRIMER CALCITONIN

<400> SEQUENCE: 47

acacacccaa acccattact atct

24

<210> SEQ ID NO 48

<211> LENGTH: 21

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: TPEF FORWARD PRIMER

<400> SEQUENCE: 48

ggacgttttt tatcgaaggc g

21

<210> SEQ ID NO 49

<211> LENGTH: 15

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: TPEF BACKWARD PRIMER

<400> SEQUENCE: 49

gccacccaac cgcga

15

<210> SEQ ID NO 50

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: 0002044:2183A188

<400> SEQUENCE: 50

atgtgattcg ttgggta

18

<210> SEQ ID NO 51

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: 0002044:2183B188

<400> SEQUENCE: 51

atgtgatttg ttgggta

18

<210> SEQ ID NO 52

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: 0002044:2194A186

<400> SEQUENCE: 52

gggtaacgtc gaatttag

18

<210> SEQ ID NO 53

<211> LENGTH: 18

-continued

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002044:2194B186

<400> SEQUENCE: 53

gggtaatggt gaatttag 18

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

<400> SEQUENCE: 54

aaaaattcgc gagtttag 18

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

<400> SEQUENCE: 55

aaaaatttgt gagtttag 18

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

<400> SEQUENCE: 56

tatatatatcg tgtgggta 18

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

<400> SEQUENCE: 57

tatatatatg tgtgggta 18

<210> SEQ ID NO 58
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 0002064:1147A188

<400> SEQUENCE: 58

agtgtatgtag tagaaggt 18

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

<400> SEQUENCE: 59

-continued

agtgtatgtg tagaaggt 18

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

<400> SEQUENCE: 60

ttagatgacg 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

ttagatgacg 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 ttgttggt 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

aagtaagtcg ttgttggt 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

gaagtggctg ttagtttt 18

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

<400> SEQUENCE: 65

gaagtggctg ttagttttt 19

<210> SEQ ID NO 66
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

-continued

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

<400> SEQUENCE: 71

aatgtttttg tgatgttg 18

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

<400> SEQUENCE: 72

attgttttcg attaat 18

-continued

<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

atttgtttg attaat

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 ggttgag

18

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

<400> SEQUENCE: 75

ataggttatg ggttgag

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:
<223> OTHER INFORMATION: 0002393:1332B186

<400> SEQUENCE: 77

aatttgtgaa tgtttggg

18

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

<400> SEQUENCE: 78

cgaatctctc gaacgatcgc atcca

25

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

-continued

<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 aatatcccgataacgccg 29

<210> SEQ ID NO 82

<211> LENGTH: 24

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: HCadherin ML probe

<400> SEQUENCE: 82

gctcctcgcg aaatactcac cccg 24

<210> SEQ ID NO 83

<211> LENGTH: 25

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Calcitonin HM probe

<400> SEQUENCE: 83

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

tgttgagggt atgtgtaatt ggggtgtga 28

<210> SEQ ID NO 86

-continued

<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: EYA4 HM blocker

<400> SEQUENCE: 86

aaactacaac cactcaaadc 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

gttatgggtt gtgattttgt gtggg 25

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

<400> SEQUENCE: 88

aagattttcg gttagttt 18

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

<400> SEQUENCE: 89

aagatttttg gttagttt 18

<210> SEQ ID NO 90
<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

tggatgatga ggaggttttag 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

-continued

aaccaataaa acctactcct cccttaa	27
 <210> SEQ ID NO 93 <211> LENGTH: 30 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: control probe <400> SEQUENCE: 93	
accaccaccc aacacacaat aacaaacaca	30
 <210> SEQ ID NO 94 <211> LENGTH: 20 <212> TYPE: DNA <213> ORGANISM: Homo Sapiens <400> SEQUENCE: 94	
ccttagtccc tacctctgct	20
 <210> SEQ ID NO 95 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Homo Sapiens <400> SEQUENCE: 95	
ctcatttaca cacacccaaa c	21
 <210> SEQ ID NO 96 <211> LENGTH: 26 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: primer <400> SEQUENCE: 96	
tggaaggag ttgggattaa gatatt	26
 <210> SEQ ID NO 97 <211> LENGTH: 33 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: primer <400> SEQUENCE: 97	
cttattacaa tttaaaaaaa aaattcacta caa	33
 <210> SEQ ID NO 98 <211> LENGTH: 39 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: blocker <400> SEQUENCE: 98	
aaattcacta caacactaca caactaaatt caacattac	39
 <210> SEQ ID NO 99 <211> LENGTH: 32 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Probe	

-continued

<400> SEQUENCE: 99

ttttcgtatt ttttttcggg ttattacgtt tt 32

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

<400> SEQUENCE: 100

atgtgattcg tttgggtaac gtcga 25

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

<400> SEQUENCE: 101

gtagggttat tgtttgggtt aataaat 27

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

<400> SEQUENCE: 102

taaaaaaaaa aaaaaaactc ctctacatac 30

<210> SEQ ID NO 103
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: blocker

<400> SEQUENCE: 103

aactcctcta catcacccac aaataaatt 29

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

<400> SEQUENCE: 104

cgaaaacatc gaccgaacaa cg 22

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

<400> SEQUENCE: 105

gtccgaaaaa aaaaaaacga actcc 25

<210> SEQ ID NO 106
<211> LENGTH: 33

-continued

<212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Forward Primer

<400> SEQUENCE: 106

gttagttagt taatttttta aatagattag tag

33

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

<400> SEQUENCE: 107

caaaaaaaca aataaaatac cacctcc

27

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

<400> SEQUENCE: 108

cctccacaaa actcactcct cacaaaatac

30

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

attaatgggtt ttataagacg gatttttttt taacgt

36

The invention claimed is:

1. A method for detecting cytosine methylation of one or more CpG dinucleotides in an EYA4 gene, or a fragment thereof, in genomic DNA in a sample from a human subject, the method comprising:

- a) isolating genomic DNA comprising an EYA4 gene from the biological sample from the human subject;
- b) chemically modifying the genomic DNA comprising the EYA4 gene, or a fragment thereof, to convert non-methylated cytosine bases to uracil by treating the genomic DNA with bisulfite, thereby producing modified genomic DNA comprising a modified EYA4 gene, or a fragment thereof;
- c) amplifying the modified genomic DNA comprising the modified EYA4 gene, or the fragment thereof, with an amplification enzyme and at least a forward primer and a reverse primer, each primer comprising a contiguous

sequence at least 16 nucleotides in length that is complementary to a contiguous sequence of at least 16 nucleotides of a sequence selected from the group consisting of SEQ ID NOs: 10, 11, 20 and 21, or complements thereof, to produce one or more amplificates, wherein the one or more amplificates comprise the one or more CpG dinucleotides; and

- d) detecting the cytosine methylation of the one or more CpG dinucleotides in the EYA4 gene or the fragment thereof by determining the presence of the one or more amplificates.

2. The method of claim 1, wherein in step a) the biological sample is selected from the group consisting of histological slides, biopsies, paraffin-embedded tissue, bodily fluid, stool, blood, serum, plasma, urine, sputum, and combinations thereof.

3. The method of claim 1, wherein treating in step b) is subsequent to embedding the genomic DNA comprising the EYA4 gene, or the fragment thereof, in agarose.

4. The method of claim 1, wherein step b) comprises treating the genomic DNA comprising the EYA4 gene, or the fragment thereof, with bisulfite in the presence of at least one of a DNA denaturing agent or a radical scavenger.

5. The method of claim 1, wherein 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 an amplicate nucleic acid molecule carrying a detectable label; and combinations thereof.

6. The method of claim 5, wherein the detectable label is selected from the 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.

7. The method of claim 6, wherein detecting in step d) comprises use of mass spectrometry for detecting the one or more amplicates or detecting detachable amplicate fragment mass labels carried by the one or more amplicates.

8. The method of claim 1, wherein amplifying in step c) comprises amplifying two or more different portions of the modified EYA4 gene or the fragment thereof.

9. The method of claim 1, wherein one or more of the at least two primers comprise sequences selected from the group consisting of SEQ ID NOs: 40, 41, 44, and 45.

10. The method of claim 1, wherein one or more of the at least two primers comprise one or more CpG, TpG or CpA dinucleotides.

11. The method of claim 10, wherein one or more of the at least two primers comprise between 2 to 5 CpG, TpG or CpA dinucleotides.

12. The method of claim 10, wherein the one or more CpG, TpG or CpA dinucleotides are located within the half of the primer that terminates at the 3' end of the primer.

13. The method of claim 1, wherein the one or more amplicates obtained in step c) comprise at least one 20 base pair sequence that comprises 3 or more CpG, TpG or CpA dinucleotides.

14. The method of claim 1, further comprising in step c) hybridizing at least one nucleic acid molecule or peptide nucleic acid molecule at least 18 base pairs in length to the one or more amplicates, the at least one nucleic acid molecule or peptide nucleic acid molecule comprising one or more CpG, TpG or CpA dinucleotides, and wherein the sequence of the at least one nucleic acid molecule or peptide nucleic acid molecule is complementary or identical to at least 18 base pairs of a sequence selected from the group consisting of SEQ ID NOs: 10, 11, 20 and 21, and complements thereof, and wherein said nucleic acid molecule or peptide nucleic acid molecule suppresses amplification of the one or more amplicates to which it is hybridized.

15. The method according to claim 14, wherein the sequence of the at least one nucleic acid molecule or peptide nucleic acid molecule is selected from the group consisting of SEQ ID NOs: 85, 86, 87, 98, 103, and 108, and complements thereof.

16. The method of claim 14, wherein amplification of DNA that was unmethylated prior to treatment of step b) is suppressed.

17. The method of claim 14, wherein the at least one 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.

18. The method of claim 14, wherein the at least one nucleic acid molecule or peptide nucleic acid molecule in each case lacks a 3' hydroxyl group.

19. The method of claim 14, wherein the amplification enzyme is a polymerase lacking 5'-3' exonuclease activity.

20. The method of claim 14, wherein the binding site of the at least one nucleic acid molecule or peptide nucleic acid molecule is identical to, or overlaps with, one of the at least two primers and thereby hinders hybridization of the primer to its binding site.

21. The method of claim 14, wherein the at least one nucleic acid molecule or peptide nucleic acid molecule is at least two nucleic acid molecules or peptide nucleic acid molecules, and wherein the binding sites of the at least two nucleic acid molecules or peptide nucleic acid molecules are identical to, or overlap with those of the at least two primers, and thereby hinder hybridization of the at least two primers to their binding site.

22. The method of claim 21, wherein hybridization of at least one of the nucleic acid molecules or peptide nucleic acid molecules hinders hybridization of a forward primer, and the hybridization of at least one of the nucleic acid molecules or peptide nucleic acid molecules hinders the hybridization of a reverse primer that binds to the elongation product of said forward primer.

23. The method of claim 14, wherein the at least one nucleic acid molecule or peptide nucleic acid molecule hybridizes between the binding sites of the forward and reverse primers.

24. The method of claim 1, wherein step c) comprises hybridization of at least one nucleic acid molecule or peptide nucleic acid molecule in each case comprising a contiguous sequence at least 9 nucleotides in length comprising one or more CpG, TpG or CpA dinucleotides and wherein the sequence of the molecule is complementary or identical to at least 9 nucleotides of a sequence selected from the group consisting of SEQ ID NOs: 10, 11, 20 and 21.

25. The method of claim 24, wherein the at least one nucleic acid molecule or peptide nucleic acid molecule is bound to a solid phase.

26. The method of claim 25, wherein a plurality of the 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 group consisting of linear arrays, hexagonal arrays, rectangular arrays, and combinations thereof.

27. The method of claim 24, wherein the sequence of the at least one nucleic acid molecule or peptide nucleic acid molecule is selected from the group consisting of SEQ ID NOs: 84, 86, 87, and complements thereof.

28. The method according to claim 24, wherein the at least one nucleic acid molecule or peptide nucleic acid molecule is fluorescently labelled, and wherein detection thereof is by either an increase or a decrease in fluorescence or fluorescence polarization.

29. The method according to claim 24, wherein the hybridization of the at least one nucleic acid molecule or peptide nucleic acid molecule is detectable by fluorescence resonance energy transfer, and wherein the detection is by either an increase or a decrease in fluorescence.

30. The method of claim 1, wherein determining the presence of the one or more amplicates comprises sequencing the one or more amplicates.

31. The method of claim 1, wherein step c) further comprises: hybridizing at least one nucleic acid molecule comprising a contiguous sequence at least 9 nucleotides in length that is complementary to a sequence selected from the group consisting of SEQ ID NOs: 10, 11, 20 and 21, and complements thereof; and extending the at least one hybridized nucleic acid molecule by at least one nucleotide. 5

32. The method of claim 1, wherein the biological sample has a background DNA concentration between 100 to 1000 fold excess of the concentration of the genomic DNA comprising the EYA4 gene or the fragment thereof. 10

33. The method of claim 1, wherein the at least two primers used in step c) are methylation-specific primers and the methylation status of the EYA4 gene or portion thereof is determined based on detecting the presence or absence of the one or more amplicates generated by the methylation-specific primers. 15

34. The method of claim 1, wherein the subject has colorectal cancer.

35. The method of claim 1, wherein the biological sample is blood, plasma or serum. 20

* * * * *

专利名称(译)	用于分析结肠细胞增殖性疾病的方法和核酸		
公开(公告)号	US9988683	公开(公告)日	2018-06-05
申请号	US12/794449	申请日	2010-06-04
[标]申请(专利权)人(译)	埃皮吉諾米克斯股份公司		
申请(专利权)人(译)	AG EPIGENOMICS		
当前申请(专利权)人(译)	AG EPIGENOMICS		
[标]发明人	ADORJAN PETER BURGER MATTHIAS MAIER SABINE LESCHE RALF COTTRELL SUSAN MOONEY SUZANNE		
发明人	ADORJAN, PETER BURGER, MATTHIAS MAIER, SABINE LESCHE, RALF COTTRELL, SUSAN MOONEY, SUZANNE		
IPC分类号	C12Q1/68 G01N27/62 A61K48/00 A61P35/00 A61P43/00 C12M1/00 C12N15/09 C12Q1/02 G01N21/78 G01N33/53 G01N33/543 G01N33/574 G01N33/58 G01N37/00		
CPC分类号	C12Q1/6886 Y10T436/143333 C12Q2600/154 A61P35/00 A61P43/00		
优先权	2002004551 2002-02-27 EP PCT/EP2003/002034 2003-02-27 WO		
其他公开文献	US20110046005A1		
外部链接	Espacenet USPTO		

摘要(译)

提供了通过分析Versican，TPEF，H-钙黏着蛋白，降钙素和EYA4中的一种或多种基因来检测，区分或区分结肠细胞增殖性病症的方法和核酸。进一步提供了用于所述基因的细胞增殖性疾病特异性分析以及其方法，测定和试剂盒的新型核酸序列。

