



(22) Date de dépôt/Filing Date: 2006/09/15

(41) Mise à la disp. pub./Open to Public Insp.: 2007/03/15

(30) Priorité/Priority: 2005/09/15 (US60/717,790)

(51) Cl.Int./Int.Cl. *C07H 21/04* (2006.01),
C12Q 1/68 (2006.01), *G01N 33/574* (2006.01)

(71) Demandeur/Applicant:
VERIDEX, LLC, US

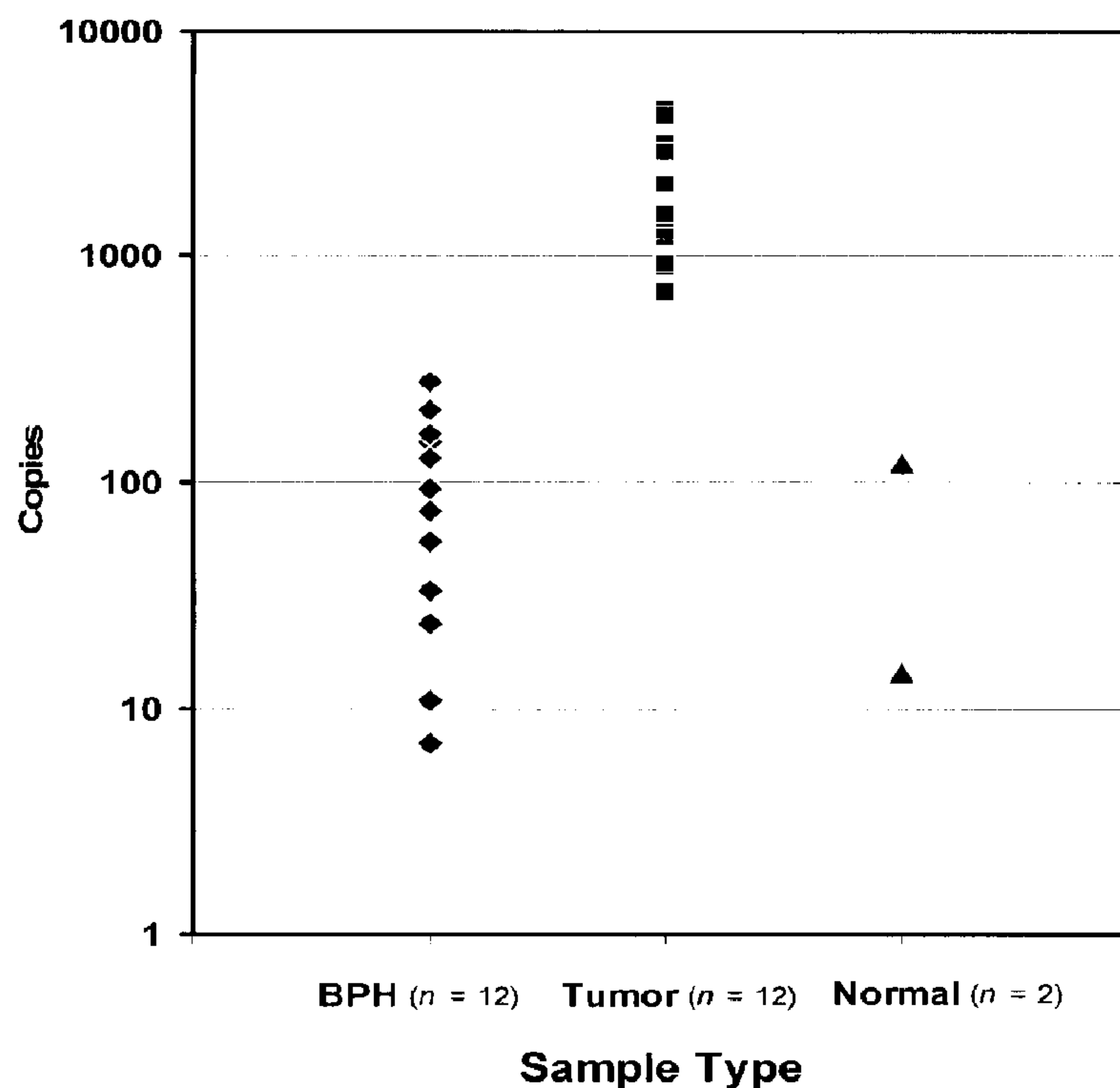
(72) Inventeurs/Inventors:
VENER, TATIANA, US;
MEHROTRA, JYOTI, US;
VARDE, SHOBHA, US;
MAZUMDER, ABHIJIT, US;
BADEN, JON, US;
BACKUS, JOHN, US

(74) Agent: OGILVY RENAULT LLP/S.E.N.C.R.L.,S.R.L.

(54) Titre : DETECTION DE LA METHYLATION DE GENES

(54) Title: DETECTING GENE METHYLATION

MSP Assay for GSTP1 According to the Invention



(57) Abrégé/Abstract:

Methods, kits, and compositions for detecting the methylation status of various genes are useful in various diagnostic applications involving suspected proliferative disorders such as prostate cancer.

Abstract of the Disclosure

DETECTING GENE METHYLATION

Methods, kits, and compositions for detecting the methylation status of various genes are useful in various diagnostic applications involving suspected proliferative disorders such as prostate cancer.

DETECTING GENE METHYLATION

BACKGROUND OF THE INVENTION

5 This invention relates to the interrogation of methylated genes in applications such as diagnostic and prognostic assays for cellular proliferative disorders such as cancer. The genes involved include the glutathione-S-transferase (GSTP1) gene or portions of it.

10 In higher order eukaryotes DNA is methylated only at cytosines located 5' to guanosine in the CpG dinucleotide. This modification has important regulatory effects on gene expression, especially when it involves CpG rich areas (CpG islands) located in gene promoter regions. Abberant methylation of normally unmethylated CpG islands is a frequent event in immortalized and transformed cells and has been
15 associated with transcriptional inactivation of certain tumor suppressor genes or genes otherwise associated with the amelioration of certain human cancers.

Glutathione S-transferases (GSTs) catalyze intracellular detoxification reactions, including the inactivation of electrophilic carcinogens, by conjugating chemically-
20 reactive electrophiles to glutathione (C. B. Pickett, et al., Annu. Rev. Biochem., 58:743, 1989; B. Coles, et al., CRC Crit. Rev. Biochem. Mol. Biol., 25:47, 1990; T. H. Rushmore, et al., J. Biol. Chem. 268:11475, 1993). Human GSTs, encoded by several different genes at different loci, have been classified into four families referred to as alpha, mu, pi, and theta (B. Mannervik, et al., Biochem. J., 282:305,
25 1992). Many human cancers such as prostate and hepatic cancers exhibit decreased GSTP1 expression relative to their tissues of origin.

In the case of prostate cancer, in particular, designing methylation specific PCR ("MSP") primers and probes that discriminate between benign prostatic hyperplasia
30 and prostate adenocarcinoma has proven to be complex. The best published primer/probe sets typically display up to about 75% sensitivity if one adjusts the assays to be 100% specific and less than 60% specific if one adjusts the assays to be

100% sensitive. A more robust and consistent assay might be obtained with new primer and probe combinations. The present invention fulfills such a need.

SUMMARY OF THE INVENTION

5

In one aspect of the invention, a method for detecting a cell proliferative disorder in prostatic tissue or other sample of a subject involves contacting a cellular component with a reagent useful in the amplification and detection of hypermethylated regions of certain genes.

10

In another aspect of the invention, at least one such hypermethylated gene is GSTP1 which can be assayed in combination with one or more other genes.

15

In another aspect of the invention, a nucleic acid sample suspected of having methylated target sequences is obtained from a biological sample, treating the sample with a reagent that can prime a portion of the target, amplifying the nucleic acid target, and comparing the degree of methylation of the amplified sample with that of a known normal sample. In yet another aspect of the invention, a sequence that is not likely to be methylated is amplified and detected for comparison with the amplified methylated sequence.

20

25

In yet another aspect, the invention provides a method for detecting a cell proliferative disorder. This is done by amplifying a gene whose methylation status is indicative of the cell proliferative disorder. The method of detecting may include contacting a nucleic acid-containing specimen or biological fluid with an agent that modifies unmethylated cytosine, amplifying the nucleic acid in the specimen with oligonucleotide primers or priming sequences that distinguish between modified methylated and nonmethylated nucleic acid, and detecting the methylated nucleic acid based on the presence or absence of amplification products produced during amplification.

30

The methods of the invention encompass determining a methylation ratio of a sample. The methylation ratio is a ratio between the level of methylation of a Marker (or a region of a Marker) that is hypermethylated in a diseased state relative to the level of methylation of a Marker that is not hypermethylated under the same condition (or
5 relative to a region of the same Marker that is not hypermethylated under the same condition). When the methylation ratio is determined via comparison of methylation of a Marker that is hypermethylated in a diseased state to another Marker that is not, the second Marker can be referred to as a reference Marker.

10 In another aspect of the invention, methylation ratios are determined via quantitative real time PCR.

In another aspect of the invention, reporter molecules for diagnostic and/or prognostic assays are provided.

15

In another embodiment, the invention provides a kit useful for the detection of a methylated nucleic acid. The kit includes one or more containers; a first container containing a reagent that modifies unmethylated cytosine and a second container containing a reagent that primes amplification of CpG-containing nucleic acid,
20 wherein the reagent distinguishes between modified methylated and nonmethylated nucleic acid.

In yet another aspect of the invention, the nucleic acid sequence that is detected in the methods of the invention or the kits for practicing them includes a promoter or a
25 portion of a promoter.

In yet another aspect of the invention, methods and kits used to detect methylated portions of certain genes also include steps and/or components for assaying the presence of another nucleic acid. The degree to which the purportedly methylated
30 gene is effected by the treatment (e.g., amplification) is compared to that of the other nucleic acid to determine the extent of methylation of the Marker.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a graph of the results of MSPCR assays using probes and primers of the prior art.

Fig. 2 is a graph of the results of MSPCR assays using hydrolysis probes and primers according to the invention.

DETAILED DESCRIPTION OF THE INVENTION

Hypermethylation of certain genes correlates with prostatic carcinogenesis. Nucleic acid sequences comprising such genes or portions of such genes are referred to as Markers in this specification. Markers include nucleic acids comprising the following genes or portions of such genes: GSTP1 (Seq. ID. No. 54), RARB2 (Seq. ID. No. 57), RASSF1A (Seq. ID. No. 64), TIMP3 (Seq. ID. No. 58), APC (Promoter= Seq. ID. No. 59, Gene= Seq. ID. No. 60), beta-Actin (Seq. ID. No. 55 and 56), PTGS2 (Promoter= Seq. ID. No. 61, Gene= Seq. ID. No. 62) and 14-3-3 sigma (Seq. ID. No. 63).

Assays for detecting such hypermethylation include such techniques as MSP and restriction endonuclease analysis. The promoter region is a particularly noteworthy target for detecting such hypermethylation analysis. Sequence analysis of the promoter region of GSTP1 shows that nearly 72% of the nucleotides are CG and about 10% are CpG dinucleotides.

The invention provides a method for determining the methylation status of certain regions of the Markers in a tissue or other biological sample of a subject in which the DNA associated with the proliferative disorder is amplified and detected. Since a decreased level of the protein encoded by the Marker (i.e., less transcription) is often the result of hypermethylation of a particular region such as the promoter, it is

desirable to determine whether such regions are hypermethylated. This is seen most demonstrably in the case of the GSTP1 gene. Hypermethylated regions are those that are methylated to a statistically significant greater degree in samples from diseased tissue as compared to normal tissue.

5

For purposes of the invention, a nucleic acid probe or reporter specific for certain Marker regions is used to detect the presence of methylated regions of the Marker gene in biological fluids or tissues. Oligonucleotide primers based on certain portions of the Marker sequence are particularly useful for amplifying DNA by techniques such as PCR. Any specimen containing a detectable amount of the relevant polynucleotide can be used. A preferred specimen of this invention is tissue of urogenital origin, specifically, prostate tissue. Preferably the sample contains epithelial cells.

15 Some of the primers/probes or reporter reagents of the invention are used to detect methylation of expression control sequences of the Marker genes. These are nucleic acid sequences that regulate the transcription and, in some cases, translation of the nucleic acid sequence. Thus, expression control sequences can include sequences involved with promoters, enhancers, transcription terminators, start codons (i.e., 20 ATG), splicing signals for introns, maintenance of the correct reading frame of that gene to permit proper translation of the mRNA, and stop codons.

The GSTP1 promoter is an expression control sequence exemplary of a useful Marker. It is a polynucleotide sequence that can direct transcription of the gene to 25 produce a glutathione-s-transferase protein. The promoter region is located upstream, or 5' to the structural gene. It may include elements which are sufficient to render promoter-dependent gene expression controllable for cell-type specific, tissue-specific, or inducible by external signals or agents; such elements may be located in the 5' or 3' regions of the of the polynucleotide sequence.

30

One method of the invention includes contacting a target cell containing a Marker

with a reagent that binds to the nucleic acid. The target cell component is a nucleic acid such as DNA or RNA. The reagents can include probes and primers such as PCR or MSP primers or other molecules configured to amplify and detect the target sequence. For example, the reagents can include priming sequences combined with or bonded to their own reporter segments such as those referred to as Scorpion reagents or Scorpion reporters and described in US Patents 6,326,145 and 6,270,967 to Whitcombe *et. al.* (incorporated herein by reference in their entirety). Though they are not the same, the terms “primers” and “priming sequences” may be used in this specification to refer to molecules or portions of molecules that prime the amplification of nucleic acid sequences.

One sensitive method of detecting methylation patterns involves combining the use of methylation-sensitive enzymes and the polymerase chain reaction (PCR). After digestion of DNA with the enzyme, PCR will amplify from primers flanking the restriction site only if DNA cleavage was prevented by methylation. Exemplary target regions to which PCR primers of the invention are designed include primers which flank the region that lies approximately between -71 and +59 bp according to genomic positioning number of M24485 (Genbank) from the transcription start site of GSTP1.

The method of the invention can also include contacting a nucleic acid-containing specimen with an agent that modifies unmethylated cytosine; amplifying the CpG-containing nucleic acid in the specimen by means of CpG-specific oligonucleotide primers; and detecting the methylated nucleic acid. The preferred modification is the conversion of unmethylated cytosines to another nucleotide that will distinguish the unmethylated from the methylated cytosine. Preferably, the agent modifies unmethylated cytosine to uracil and is sodium bisulfite, however, other agents that modify unmethylated cytosine, but not methylated cytosine can also be used. Sodium bisulfite (NaHSO₃) modification is most preferred and reacts readily with the 5,6-double bond of cytosine, but poorly with methylated cytosine. Cytosine reacts with the bisulfite ion to form a sulfonated cytosine reaction intermediate susceptible to

deamination, giving rise to a sulfonated uracil. The sulfonate group can be removed under alkaline conditions, resulting in the formation of uracil. Uracil is recognized as a thymine by Taq polymerase and therefore upon PCR, the resultant product contains cytosine only at the position where 5-methylcytosine occurs in the starting template.

5 Scorpion reporters and reagents and other detection systems similarly distinguish modified from unmodified species treated in this manner.

The primers used in the invention for amplification of a CpG-containing nucleic acid in the specimen, after modification (e.g., with bisulfite), specifically distinguish
10 between untreated DNA, methylated, and non-methylated DNA. In methylation specific PCR (MSPCR), primers or priming sequences for the non-methylated DNA preferably have a T in the 3' CG pair to distinguish it from the C retained in methylated DNA, and the complement is designed for the antisense primer. MSP
15 primers or priming sequences for non-methylated DNA usually contain relatively few Cs or Gs in the sequence since the Cs will be absent in the sense primer and the Gs absent in the antisense primer (C becomes modified to U (uracil) which is amplified as T (thymidine) in the amplification product).

The primers of the invention are oligonucleotides of sufficient length and appropriate
20 sequence so as to provide specific initiation of polymerization on a significant number of nucleic acids in the polymorphic locus. When exposed to the probes or reporters of the invention, the sequences that are amplified by the primers of the invention reveal methylation status and thus diagnostic information. They are more sensitive and specific than those of the prior art.

25 The inventive primers are most preferably eight or more deoxyribonucleotides or ribonucleotides capable of initiating synthesis of a primer extension product, which is substantially complementary to a polymorphic locus strand. Environmental conditions conducive to synthesis include the presence of nucleoside triphosphates and an agent for polymerization, such as DNA polymerase, and a suitable temperature
30 and pH. The priming segment of the primer or priming sequence is preferably single

stranded for maximum efficiency in amplification, but may be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent for polymerization. The exact length of primer will depend on factors such as temperature, buffer, and nucleotide composition. The oligonucleotide primers most preferably contain about 12-20 nucleotides although they may contain more or fewer nucleotides, preferably according to well known design guidelines or rules.

Primers of the invention are designed to be substantially complementary to each strand of the genomic locus to be amplified and include the appropriate G or C nucleotides as discussed above. This means that the primers must be sufficiently complementary to hybridize with their respective strands under conditions that allow the agent for polymerization to perform. In other words, the primers should have sufficient complementarity with the 5' and 3' flanking sequence(s) to hybridize and permit amplification of the genomic locus.

The primers of the invention are employed in the amplification process. That is, reactions (preferably, an enzymatic chain reaction) that produce greater quantities of target locus relative to the number of reaction steps involved. In a most preferred embodiment, the reaction produces exponentially greater quantities of the target locus. Reactions such as these include the PCR reaction. Typically, one primer is complementary to the negative (-) strand of the locus and the other is complementary to the positive (+) strand. Annealing the primers to denatured nucleic acid followed by extension with an enzyme, such as the large fragment of DNA Polymerase I (Klenow) and nucleotides, results in newly synthesized + and - strands containing the target locus sequence. The product of the chain reaction is a discrete nucleic acid duplex with termini corresponding to the ends of the specific primers employed.

The primers of the invention may be prepared using any suitable method, such as conventional phosphotriester and phosphodiester methods including automated

methods. In one such automated embodiment, diethylphosphoramidites are used as starting materials and may be synthesized as described by Beaucage, et al. (Tetrahedron Letters, 22:1859-1862, 1981). A method for synthesizing oligonucleotides on a modified solid support is described in U.S. Pat. No. 4,458,066.

5

Any nucleic acid specimen, in purified or non-purified form, can be utilized as the starting nucleic acid or acids, provided it contains, or is suspected of containing, the specific nucleic acid sequence containing the target locus (e.g., CpG). Thus, the process may employ, for example, DNA or RNA, including messenger RNA. The DNA or RNA may be single stranded or double stranded. In the event that RNA is to be used as a template, enzymes, and/or conditions optimal for reverse transcribing the template to DNA would be utilized. In addition, a DNA-RNA hybrid containing one strand of each may be utilized. A mixture of nucleic acids may also be employed, or the nucleic acids produced in a previous amplification reaction herein, using the same or different primers may be so utilized. The specific nucleic acid sequence to be amplified, i.e., the target locus, may be a fraction of a larger molecule or can be present initially as a discrete molecule so that the specific sequence constitutes the entire nucleic acid.

20 The nucleic acid-containing specimen used for detection of methylated CpG may be from any source such as tissue (particularly, prostate tissue and lymphatic tissue), blood, lymph, urine, and ejaculate and may be extracted by a variety of techniques such as that described by Maniatis, et al. (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, N.Y., pp 280, 281, 1982).

25

If the extracted sample is impure, it may be treated before amplification with an amount of a reagent effective to open the cells, fluids, tissues, or animal cell membranes of the sample, and to expose and/or separate the strand(s) of the nucleic acid(s). This lysing and nucleic acid denaturing step to expose and separate the strands will allow amplification to occur much more readily.

30

Where the target nucleic acid sequence of the sample contains two strands, it is necessary to separate the strands of the nucleic acid before it can be used as the template. Strand separation can be effected either as a separate step or simultaneously with the synthesis of the primer extension products. This strand separation can be accomplished using various suitable denaturing conditions, including physical, chemical or enzymatic means. One physical method of separating nucleic acid strands involves heating the nucleic acid until it is denatured. Typical heat denaturation may involve temperatures ranging from about 80 to 105°C for up to 10 minutes. Strand separation may also be induced by an enzyme from the class of enzymes known as helicases or by the enzyme RecA, which has helicase activity, and in the presence of riboATP, is known to denature DNA. Reaction conditions that are suitable for strand separation of nucleic acids using helicases are described by Kuhn Hoffmann-Berling (CSH-Quantitative Biology, 43:63, 1978). Techniques for using RecA are reviewed in C. Radding (Ann. Rev. Genetics, 16:405-437, 1982). Refinements of these techniques are now also well known.

When complementary strands of nucleic acid or acids are separated, regardless of whether the nucleic acid was originally double or single stranded, the separated strands are ready to be used as a template for the synthesis of additional nucleic acid strands. This synthesis is performed under conditions allowing hybridization of primers to templates to occur. Generally synthesis occurs in a buffered aqueous solution, preferably at a pH of 7-9, most preferably about 8. A molar excess (for genomic nucleic acid, usually about 10^8 :1, primer:template) of the two oligonucleotide primers is preferably added to the buffer containing the separated template strands. The amount of complementary strand may not be known if the process of the invention is used for diagnostic applications, so the amount of primer relative to the amount of complementary strand cannot always be determined with certainty. As a practical matter, however, the amount of primer added will generally be in molar excess over the amount of complementary strand (template) when the sequence to be amplified is contained in a mixture of complicated long-chain nucleic acid strands. A large molar excess is preferred to improve the efficiency of the

process.

The deoxyribonucleoside triphosphates dATP, dCTP, dGTP, and dTTP are added to the synthesis mixture, either separately or together with the primers, in adequate amounts and the resulting solution is heated to about 90-100°C for up to 10 minutes, preferably from 1 to 4 minutes. After this heating period, the solution is allowed to cool to room temperature, which is preferable for the primer hybridization. To the cooled mixture is added an appropriate agent for effecting the primer extension reaction (the "agent for polymerization"), and the reaction is allowed to occur under conditions known in the art. The agent for polymerization may also be added together with the other reagents if it is heat stable. This synthesis (or amplification) reaction may occur at room temperature up to a temperature at which the agent for polymerization no longer functions.

The agent for polymerization may be any compound or system that will function to accomplish the synthesis of primer extension products, preferably enzymes. Suitable enzymes for this purpose include, for example, E. coli DNA polymerase 1, Klenow fragment of E. coli DNA polymerase I, T4 DNA polymerase, other available DNA polymerases, polymerase muteins, reverse transcriptase, and other enzymes, including heat-stable enzymes (e.g., those enzymes which perform primer extension after being subjected to temperatures sufficiently elevated to cause denaturing). A preferred agent is Taq polymerase. Suitable enzymes will facilitate combination of the nucleotides in the proper manner to form the primer extension products complementary to each locus nucleic acid strand. Generally, the synthesis will be initiated at the 3' end of each primer and proceed in the 5' direction along the template strand, until synthesis terminates, producing molecules of different lengths. There may be agents for polymerization, however, which initiate synthesis at the 5' end and proceed in the other direction, using the same process as described above.

Most preferably, the method of amplifying is by PCR. Alternative methods of amplification can also be employed as long as the methylated and non-methylated

loci amplified by PCR using the primers of the invention is similarly amplified by the alternative means.

The amplified products are preferably identified as methylated or non-methylated with a probe or reporter specific to the product as described in US Patent 4,683,195 to Mullis *et. al.*, incorporated herein by reference in its entirety. Advances in the field of probes and reporters for detecting polynucleotides are well known to those skilled in the art. Optionally, the methylation pattern of the nucleic acid can be confirmed by other techniques such as restriction enzyme digestion and Southern blot analysis.

Examples of methylation sensitive restriction endonucleases which can be used to detect 5'CpG methylation include SmaI, SacII, EagI, MspI, HpaII, BstUI and BssHII.

The most preferred method of the invention involves establishing a methylation ratio. This can be done by establishing a ratio between the amount of amplified methylated species of Marker attained and the amount of amplified reference Marker or non-methylated Marker region amplified. This is best done using quantitative realtime PCR. Ratios above an established or predetermined cutoff or threshold are considered hypermethylated and indicative of having a proliferative disorder such as cancer (prostate cancer in the case of GSTP1). Cutoffs are established according to known methods in which such methods are used for at least two sets of samples: those with known diseased conditions and those with known normal conditions. The reference Markers of the invention can also be used as internal controls. The reference Marker is preferably a gene that is constitutively expressed in the cells of the samples such as B Actin.

The inventive methods and kits can include steps and reagents for multiplexing. That is, more than one Marker can be assayed at a time.

Since a decreased level of transcription of the gene associated with the Marker is often the result of hypermethylation of the polynucleotide sequence and/or particular elements of the expression control sequences (e.g., the promoter sequence), primers

prepared to match those sequences were prepared. Accordingly, the invention provides methods of detecting or diagnosing a cell proliferative disorder by detecting methylation of particular areas within the expression control or promoter region of the Markers. Probes useful for detecting methylation of these areas are useful in such diagnostic or prognostic methods. Preferred molecules for the detection of Markers are shown below. The short name for the Marker gene is shown in parentheses along with the type of detection system employed. Antisense only refers to the orientation of the primer that is so designated in relationship to the priming sequence of the other member of the pair with which it is associated. It is not necessarily antisense with respect to genomic DNA.

SEQ ID NO. 1 (GSTP1 SCORPION):

CCCCGAACGTCGACCGCTCGGGG-BHQ-HEG-CGATTTCGGGGATTTTAGGGCGT

SEQ ID NO. 2 (GSTP1 SCORPION Antisense Primer):

AAAATCCCGCGAACTCCCGCC

SEQ ID NO. 3 (GSTP1 SCORPION):

CCCGAACGTCGACCGCTTTCGGG-BHQ-HEG-CGATTTCGGGGATTTTAGGGCGT

SEQ ID NO. 4 (GSTP1 SCORPION Antisense Primer):

AAAATCCCGCGAACTCCCGCC

SEQ ID NO. 5 (GSTP1 SCORPION):

CGGCGGGAGTTCGCGGGCGCCG-BHQ-HEG-ACTAAATCACGACGCCGACCGC

SEQ ID NO. 6 (GSTP1 SCORPION Antisense Primer):

CGGTTAGTTGCGCGGCGATTTC

SEQ ID NO. 7 (GSTP1 SCORPION):

CGGGAGTTCGCGGGTCCCG-BHQ-HEG-ACTAAATCACGACGCCGACCGC

SEQ ID NO. 8 (GSTP1 SCORPION Antisense Primer):

CGGTTAGTTGCGCGGCGATTTC

SEQ ID NO. 9 (GSTP1 SCORPION):

GTGGTTGATGTTTGGGGTATCAACCAC-BHQ-HEG_AATCCCACAAACTCCCACCAACC

SEQ ID NO. 10 (GSTP1 SCORPION Antisense Primer):

GTGGTGATTTTGGGGATTTTAGGGTGT

SEQ ID NO. 11 (GSTP1 SCORPION):

ACCCAGTGGTTGATGTTTGGGGT-BHQ-HEG-AATCCCACAAACTCCCACCAACC

SEQ ID NO. 12 (GSTP1 SCORPION Antisense Primer):

GTGGTGATTTTGGGGATTTTAGGGTGT

SEQ ID NO. 13 (GSTP1 SCORPION):

CCCCACAGGTTGGTGGGAGTTTGTGGGG-BHQ-HEG-CCCAATACTAAATCACAACACCAACCAC

SEQ ID NO. 14 (GSTP1 SCORPION Antisense Primer):

TGGTTAGTTGTGTGGTGATTTTGGGGA

SEQ ID NO. 15 (GSTP1 SCORPION):

CCCCGAACGTCGACCGCTCGGGG-BHQ-HEG- CGATTTCGGGGATTTTAGGGCGT
SEQ ID NO. 16 (GSTP1 SCORPION Antisense Primer):
 AAAATCCCGCGAACTCCCGCC

5 **SEQ ID NO. 17 (GSTP1 SCORPION):**
 CGCACGCCGAACGTCGACCGCAAACGTGCG-BHQ-HEG-
 CGATTTCGGGGATTTTAGGGCGT
SEQ ID NO. 18 (GSTP1 SCORPION Antisense Primer):
 AAAATCCCGCGAACTCCCGCC

10 **SEQ ID NO. 19 (GSTP1 SCORPION):**
 CGCACGGCGAACTCCCGCCGACGTGCG BHQ-HEG-TGTAGCGGTCGTCGGGGTTG
SEQ ID NO. 20 (GSTP1 SCORPION Antisense Primer):
 GCCCAATACTAAATCACGACG

15 **SEQ ID NO. 21 (GSTP1 SCORPION):**
 CCGACGCACAAAAAACACCCTAAAATCCGTCGG-BHQ-HEG-
 GGTTAGTTGTGTGGTGATTTT
SEQ ID NO. 22 (GSTP1 SCORPION Antisense Primer):
 20 CACAACACCAACCACTCTTC

SEQ ID NO. 23 (GSTP1 TAQMAN PRIMER):
 CGTGATTTAGTATTGGGGCGGAGCGGGGC

25 **SEQ ID NO. 24 (GSTP1 TAQMAN PRIMER):**
 ATCCCGAAAAACGAACCGCGCGTA

SEQ ID NO. 25 (GSTP1 TAQMAN PROBE):
 30 TCGGAGGTCGCGAGGTTTTCGTTGGA

SEQ ID NO. 26 (RARB2 SCORPION):
 GCCGCCGTCGAGAACGCGAGCGGGCGGC-BHQ-HEG-TACCCCGACGATACCCAAAC
SEQ ID NO. 27 (RARB2 SCORPION Antisense Primer):
 35 GGGATTAGAATTTTTTATGCGAGTTGT

SEQ ID NO. 28 (RARB2 SCORPION):
 CGGCAGGATTGGGATGTCGAGCTGCCG-BHQ-HEG-TACCCCGACGATACCCAAAC
SEQ ID NO. 29 (RARB2 SCORPION Antisense Primer):
 40 GGGATTAGAATTTTTTATGCGAGTTGT

SEQ ID NO. 30 (RASSF1A SCORPION):
 GCCGCGGTTTCGTTTCGGTTCGCGGC-BHQ-HEG-CCCGTACTTCGCTAACTTTAAACG
SEQ ID NO. 31 (RASSF1A SCORPION Antisense Primer):
 45 GCGTTGAAGTCGGGGTTC

SEQ ID NO. 32 (TIMP3 SCORPION):
 GCGGCGAGTTCGGGTTGTAGCGCCGC-BHQ-HEG-CGCCTCTCCAAAATTACCGTAC
SEQ ID NO. 33 (TIMP3 SCORPION Antisense Primer):
 50 GCGTCGGAGGTTAAGGTTGTT

SEQ ID NO. 34 (APC SCORPION):
 GCCGGCGGGTTTTTCGACGGGCCGCGC-BHQ-HEG-CGAACCAAAACGCTCCCCA
SEQ ID NO. 35 (APC SCORPION Antisense Primer):
 55 GTCGGTTACGTGCGTTTATATTAG

SEQ ID NO. 36 (ACTIN SCORPION):

GCGCCCAACCGCACAAAGGGCGC-BHQ-HEG-GGGTATATTTTCGAGGGGTACG
SEQ ID NO. 37 (ACTIN SCORPION Antisense Primer):
 CGACCCGCACTCCGCAAT

5 **SEQ ID NO. 38(ACTIN SCORPION):**
 CCGCGCATCACCACCCACACGCGCGG-BHQ-HEG-GGAGTATATAGGTTGGGGAAGTTTG
SEQ ID NO. 39 (ACTIN SCORPION Antisense Primer):
 AACACACAATAACAAACACAAATTCAC

10 **SEQ ID NO. 40 (ACTIN SCORPION):**
 CCCGGCTAAACCCACCATCCAGCCGGG-BHQ-HEG-GGGAGGGTAGTTTAGTTGTGGTT
SEQ ID NO. 41 (ACTIN SCORPION Antisense Primer):
 CAAAACAAAAAACTAAATCTACACAACC

15 **SEQ ID NO. 42 (ACTIN SCORPION):**
 CCGCGGAACATTCAACTCAACCGCGG-BHQ-HEG-GGAGGAGGAAGGTAGGTTTTT
SEQ ID NO. 43 (ACTIN SCORPION Antisense Primer):
 ACATACAACAATCAATAACATAAAACCAC

20 **SEQ ID NO. 44 (PTGS2/COX2 SCORPION):**
 CACGCCGCGGTATCTAGGCGTG-BHQ-HEG-GTTTGTTCGACGTGATTTTTTCGA
SEQ ID NO. 45 (PTGS2/COX2 Antisense Primer):
 GCAAAAAATCCCCTCTCCCGC

25 **SEQ ID NO. 46 (PTGS2/COX2 SCORPION):**
 GCCGCGCACAAATTTCCGCGGC-BHQ-HEG-GAATTGGTTTTTCGGAAGCGTTTCG
SEQ ID NO. 47 (PTGS2/COX2 Antisense Primer):
 CCCGAATTCCACCGCC

30 **SEQ ID NO. 48 (PTGS2/COX2 SCORPION):**
 GCGGGAACGCACAAATTTCCGCC-BHQ-HEG-GAATTGGTTTTTCGGAAGCGTTTCG
SEQ ID NO. 49 (PTGS2/COX2 Antisense Primer):
 CCCGAATTCCACCGCC

35 **SEQ ID NO. 50 (PTGS2/COX2 SCORPION):**
 TGCCGCCGCGGTATCTAATGGCGGCA-BHQ-HEG-GTTTGTTCGACGTGATTTTTTCGA
SEQ ID NO. 51 (PTGS2/COX2 Antisense Primer):
 GCAAAAAATCCCCTCTCCCGC

40 **SEQ ID NO. 52 (14-3-3 SCORPION):**
 CGGCCTTCGCTCTTCGAAAAGGCCG-BHQ-HEG-GGTAGTTTTTATGAAAGGCGTCGTG
SEQ ID NO. 53 (14-3-3 Antisense Primer):
 CTAACCGCCCACCACGTT

45 BHQ=Black Hole Quencher (BioSearch Technologies, San Fransisco, CA)

HEG= Hexaethylene glycol

The kits of the invention can be configured with a variety of components provided
 50 that they all contain at least one primer or probe of the invention or a detection
 molecule (e.g., Scorpion reporter). In one embodiment, the kit includes reagents for
 amplifying and detecting hypermethylated Marker segments. Optionally, the kit

includes sample preparation reagents and /or articles (e.g., tubes) to extract nucleic acids from samples.

In a preferred kit, reagents necessary for one-tube MSP are included such as, a corresponding PCR primer set, a thermostable DNA polymerase, such as Taq polymerase, and a suitable detection reagent(s) such as hydrolysis probe or molecular beacon. In optionally preferred kits, detection reagents are Scorpion reporters or reagents. A single dye primer or a fluorescent dye specific to double-stranded DNA such as ethidium bromide can also be used. The primers are preferably in quantities that yield high concentrations. Additional materials in the kit may include: suitable reaction tubes or vials, a barrier composition, typically a wax bead, optionally including magnesium; necessary buffers and reagents such as dNTPs; control nucleic acid (s) and/or any additional buffers, compounds, co-factors, ionic constituents, proteins and enzymes, polymers, and the like that may be used in MSP reactions. Optionally, the kits include nucleic acid extraction reagents and materials.

EXAMPLES

Example 1: Comparative

A hydrolysis probe assay was designed based upon the primers and probes in the literature and directed to the quantitative assay for detecting prostate cancer in patients with clinically localized disease. This assay includes a core CpG promoter region used to discriminate between neoplastic and non-neoplastic prostate tissue.

The primer and dual-labeled hydrolysis probe sequences tested for this design were as follows:

Forward primer GSTP1;(-192)MU17 AGTTGCGCGGCGATTTC (Seq. ID. No. 65)

Reverse primer GSTP1;(-74)ML22 GCCCCAATACTAAATCACGACG (Seq. ID. No. 20)

Probe (5'FAM/3'TAMRA)GSTP1;(-152)MP23 CGGTCGACGTTTCGGGGTGTAGCG
(Seq. ID. No. 66)

GSTP1 methylation levels were analyzed in 12 samples from benign prostatic
hyperplasia (BPH), 12 samples from clinically localized prostate adenocarcinoma
(Tumor), and 2 normal samples. Data is presented in total copies of GSTP1 detected
and is shown in Fig. 1. The assay was found to display 75% sensitivity for detecting
prostate adenocarcinoma when a cutoff was set to the highest copy number exhibited
by a benign sample. The moderate distribution of GSTP1 methylation levels in
prostate tissues displaying benign prostatic hyperplasia and clinically localized
prostate adenocarcinoma suggests an assay that would not be clinically relevant
without being complemented by another marker.

Example 2 (Hydrolysis Probe Assay According to the Invention)

MSP primers and probe were designed to further improve sensitivity and specificity
for the GSTP1 assay and were tested on the same sample set as in Example 1. This
design was located further downstream of the CpG island and encompassed part of
the core promoter and exon 1. The primer and dual-labeled hydrolysis probe
sequences tested for this example are as follows:

Forward primer: (-71)MU29 CGTGATTAGTATTGGGGCGGAGCGGGGC (Seq.
ID No. 23)

Reverse primer: (+59)ML25 ATCCCCGAAAAACGAACCGCGCGTA (Seq. ID No.
24)

Probe (5'FAM/3'TAMRA)GSTP1;(-23)MP26
TCGGAGGTCGCGAGGTTTTCGTTGGA (Seq. ID No. 25)

Data is presented in total copies of GSTP1 detected and is shown in Fig. 2. The assay was found to display 100% sensitivity and 100% specificity for detecting prostate adenocarcinoma when a cutoff is set to the highest copy number exhibited by a benign sample.

5

The 1.5kb CpG island of GSTP1 spans through the promoter into exon 3 and has been shown to be extensively methylated throughout. Two transcription binding sites AP1 and Sp1 are located just upstream of exon 1 and were the focus of the design of the probes and primers used in this example due to the high content of CpG's within SP1. This proved to be a good region for assay design due to competition between transcription factors and methyl-CpG binding proteins binding to the AP1 and Sp1 sites in order to induce chromatin condensation and gene silencing.

10

Example 3: Scorpion MSP

15

Scorpion reagents were designed for the following Markers and reference Markers:

Seq. ID No. 19, 20 (GSTP1)

Seq. ID No. 38, 39 (Actin)

Seq. ID No. 28, 29 (RAR2)

20

Seq. ID No. 30, 31 (RASSF1A)

Seq. ID No. 32, 33 (TIMP3)

Seq. ID No. 34, 35 (APC)

Seq. ID No. 50, 51 (PTGS2)

Seq. ID No. 52, 53 (14-3-3 sigma)

25

The designs were tested for secondary structures and inappropriate dimerizations using Oligo6 primer design software (Molecular Biology Insights, Cascade, CO, USA) and Visual OMP (DNA Software, Inc., 120½ West Washington Street Ann Arbor, MI, USA). Scorpion folding and amplicon probing was modeled using the DNA mfold program (<http://bioinfo.math.rpi.edu/~mfold/dna/form1.cgi>) and Visual OMP (DNA Software, Inc., 120½ West Washington Street Ann Arbor,

30

MI, USA). Scorpions and primers were synthesized in an automated DNA synthesizer using standard phosphoramidite chemistry (Scandinavian Gene Synthesis AB, Koping, Sweden) and Biosearch Technologies (Novato, CA, USA).

5 **Example 3 Scorpion Reagent Testing:**

The reagents described in Example 2 were tested on various human DNA samples. These samples included commercially available methylated (all cytosines preceding guanine are methylated at position 5) and unmethylated human genomic DNA
10 (Chemicon, Temecula, CA, USA) as positive and negative template respectively. Genomic DNA from human breast cancer cell line, MCF-7 and colorectal cancer cell line, HCT116 as positive and negative controls for GSTP1 methylation were also used (American Type Culture Collection, Manassas, VA, USA).

15 Genomic DNA was modified using a commercially available sodium bisulfite conversion reagent kit (Zymo Research, Orange, CA, USA). This treatment converted all Cytosines in unmethylated DNA into Uracil, whereas in methylated DNA only cytosines not preceding guanine were converted into Uracil. All cytosines preceding guanine (in a CpG dinucleotide) remained as cytosine.

20 Sodium bisulfite modified genomic DNA (100 – 150 ng) was amplified in a 25 µl reaction containing the following components: 67mM Tris pH 8.8, 16.6mM (NH₄)₂SO₄, 6.7mM MgCl₂, 10mM beta mercaptoethanol; 1.25mM each dATP, dCTP, dGTP, dTTP, 1 U Hot start Taq DNA Polymerase, 250 nM Scorpion probe, 250 nM
25 reverse or forward primer (depending on scorpion design), 625 nM of passive reference dye, ROX (for ABI7900 runs only).

The samples were then tested in a quantitative real-time PCR assay on an ABI 7900HT sequence detection system and /or Cepheid SmartCycler[®] PCR instrument.

30

The PCR conditions used were:

For ABI 7900 runs:

95⁰C for 5 mins.; then 40 cycles of 95⁰C for 30sec, 59⁰C for 30 sec and 72⁰C for 30 sec; and a final extension at 72⁰C for 5 mins. Optical data was collected at 59⁰C for every cycle.

For Cepheid smart cycler runs:

95⁰C for 60 sec; then 40 cycles of 95⁰C for 30sec, 59⁰C for 30 sec, and a final extension at 72⁰C for 5 min. Optical data was collected at 59⁰C for every cycle.

The Scorpion reagents showed a PCR efficiency of 91.2% in the case of GSTP1 reagents and 95.2% in the case of Actin reagents. The average Ct values were:

GSTP1	28
B Actin	30
APC	29
TIMP3	30
RASSF1A	29
RARB2	29
PTGS2	33
14-3-3 S	32

Example 4: Comparison with Molecular Beacons.

Molecular beacons were obtained for detecting GSTP1 hypermethylation along the portion of the gene described in Example 1. The beacons were applied to PCR reactions involving samples as described in Example 2. In each case, B Actin was co-amplified so that methylation ratios could be obtained. Both GSTP1 and B-actin Scorpion probe-primer sets displayed better performance in detection of methylated GSTP1 DNA and unmethylated B Actin DNA. The Ct threshold values for the designed Scorpion sets were more than 1.5 Ct better than for the molecular beacon probes (average Ct of 28.82 for beacons v/s 27.30 for Scorpion reporters). Results are as follows:

Universal Methylated DNA, ng **500 250 100 50 10 5**

Scorpion (GSTP1/Bactin copy# ratio) 272 310 285 296 339 288
Mol Beacon (GSTP1/Bactin copy# ratio) 759 735 758 767 797 814

5 **Example 5 Methylation Ratio Cutoffs**

Prostate tissue samples from patients with known clinical conditions were then subjected to PCR as described in Example 2. Methylation ratios were calculated and are shown below. Data was analyzed by using the following procedure. Prior to
10 employing an algorithm, the following pre-requisites were followed.

For Actin: CT values < 40 for both replicates (as defined by real-time PCR equipment)

For “methylated” GSTP1 result: CT values < 40 for both replicates (as defined by real-time PCR equipment)

15 The slope of both standard curves above -4 (PCR efficiency > 80%)

R2: 0.99, 4 relevant data points on the standard curve

Positive assay control has to generate a positive result

Routinely included no-template controls (standard curves) have to be negative

If the pre-requisites were met, copy numbers were calculated for each of the genes
20 (GSTP1 and Actin) based on the standard curves and a ratio was calculated for each sample (including the positive control) by using the following formula: (GSTP1 Copy Number/ β -Actin Copy Number) * 1000. A cutoff was then established by using a ratio which generated high specificity and high sensitivity. The cutoff was 10 for GSTP1 to provide a sensitivity of 87% and a specificity of 100% for 142 formalin
25 fixed paraffin embedded prostate tissues. Ratios generated were used to determine methylation status of a given sample by employing the following criteria.

Results	copy# (mGSTP1)	copy# (B- Actin)	Ratio x 1000
Methylated GSTP1	>2	>2	>10
UnMethylated GSTP1	any	>200	<10
Undetermined	<2	<200	

One skilled in the art will recognize that the cutoff can be adjusted, preferably to set the sensitivity and specificity values to those suitable for the purpose for which they will be used (e.g., screening, therapy monitoring, reflex testing of ambiguous diagnoses).

5

Example 6 Combination Markers (Prophetic)

Scorpion reagents for the following combinations are prepared as described in

10

Example 1: GSTP1 Marker and RAR β 2, GSTP1 and APC, and GSTP1 and PTGS2.

The Markers are run in combination as multiplexed PCR according to Example 2 using repetitive runs of 50 prostate tissue samples with known conditions (i.e., normal, benign hyperplasia, prostate cancer). Expected results are shown below:

15

Marker/s	Seq. ID No.	Sensitivity	Specificity
GSTP1 + RAR β 2		90%	100%
GSTP1 + APC		96.2%	92.9
GSTP1 + PTGS2		96.2	100

20

25

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Veridex, LLC
- (ii) TITLE OF INVENTION: Detecting Gene Methylation
- (iii) NUMBER OF SEQUENCES: 66
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Ogilvy Renault LLP
 - (B) STREET: 200 Bay St. Suite 3800
 - (C) CITY: Toronto
 - (D) STATE: Ontario
 - (E) COUNTRY: Canada
 - (F) ZIP: M5J2Z4
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: not assigned yet
 - (B) FILING DATE: 15-SEP-2006
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 60/717790
 - (B) FILING DATE: 15-SEP-2005
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 416-216-4000
 - (B) TELEFAX: 416-216-3930

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 46 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: artificial
- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 23..23
 - (D) OTHER INFORMATION: /standard_name= "Black Hole Quencher (BHQ)-Hexaethylglycol (HEG)"
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CCCCGAACGT CGACCGCTCG GGGCGATTTC GGGGATTTTA GGGCGT

46

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

AAAATCCCGC GAACTCCCGC C

21

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 46 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: artificial

- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 23..23
 - (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CCCGAACGTC GACCGCTTTC GGGCGATTTC GGGGATTTTA GGGCGT

46

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AAAATCCCGC GAACTCCCGC C

21

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: artificial

- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 22..22
 - (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CGGCGGGAGT TCGCGGGCGC CGACTAAATC ACGACGCCGA CCGC

44

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CGGTTAGTTG CGCGGCGATT TC

22

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 41 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: artificial

- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 19..19
 - (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature

- (B) LOCATION: 25..25
- (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CGGGAGTTTCG CGGGTCCCGA CTAAATCACG ACGCCGACCG C

41

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CGGTTAGTTG CGCGGCGATT TC

22

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 50 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: artificial

- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 27..27
 - (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GTGGTTGATG TTTGGGGTAT CAACCACAAT CCCACAAACT CCCACCAACC

50

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GTGGTGATTT TGGGGATTTT AGGGTGT

27

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: artificial

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 24..24
- (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ACCCAGTGG TTGATGTTTG GGGTAATCCC ACAAACTCCC ACCAACC

47

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GTGGTGATTT TGGGGATTTT AGGGTGT

27

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 56 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: artificial

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 28..28
- (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CCCCACAGGT TGGTGGGAGT TTGTGGGGCC CAATACTAAA TCACAACACC AACCAC

56

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TGGTTAGTTG TGTGGTGATT TTGGGGA

27

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 46 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: artificial

- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 23..23
 - (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CCCCGAACGT CGACCGCTCG GGGCGATTTC GGGGATTTTA GGGCGT

46

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (vi) ORIGINAL SOURCE:

(A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

AAAATCCCGC GAACTCCCGC C

21

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 53 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: artificial

(ix) FEATURE:

(A) NAME/KEY: misc_feature

(B) LOCATION: 30..30

(D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CGCACGCCGA ACGTCGACCG CAAACGTGCG CGATTTCGGG GATTTTAGGG CGT

53

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

AAAATCCCGC GAACTCCCGC C

21

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: artificial

(ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION: 27..27
 (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CGCACGGCGA ACTCCCGCCG ACGTGCGTGT AGCGGTCGTC GGGGTTG

47

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GCCCCAATAC TAAATCACGA CG

22

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 55 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: artificial

(ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION: 34..34
 (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CCGACGCACA AAAAAACACC CTAAAATCCG TCGGGGTIAG TTGTGTGGTG ATTTT

55

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CACAACACCA ACCACTCTTC

20

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CGTGATTTAG TATTGGGGCG GAGCGGGGC

29

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

ATCCCCGAAA AACGAACCGC GCGTA

25

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TCGGAGGTCG CGAGGTTTTTC GTTGGA

26

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: artificial

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 28..28
- (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GCCGCCGTCG AGAACGCGAG CGGGCGGCTA CCCCACGAT ACCCAAAC

48

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GGGATTAGAA TTTTATATGC GAGTTGT

27

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: artificial

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 27..27
- (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CGGCAGGATT GGGATGTCGA GCTGCCGTAC CCCGACGATA CCCAAAC

47

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GGGATTAGAA TTTTATATGC GAGTTGT

27

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 49 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: artificial

(ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION: 25..25
 (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GCCGCGGTTT CGTTCGGTTC GCGGCCCCGT ACTTCGCTAA CTTTAAACG

49

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 18 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GCGTTGAAGT CGGGGTTC

18

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 48 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: artificial

- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 26..26
 - (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GCGGCGAGTT CGGGTTGTAG CGCCGCCGCC TCTCCAAAAT TACCGTAC

48

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GCGTCGGAGG TTAAGGTTGT T

21

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: artificial

- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 25..25
 - (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GCCGGCGGGT TTTCGACGGG CCGGCCGAAC CAAAACGCTC CCCA

44

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 25 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GTCGGTTACG TCGGTTTATA TTTAG

25

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 44 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: artificial

(ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION: 22..22
 (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GCGCCCAACC GCACAAGGGC GCGGGTATAT TTTCGAGGGG TACG

44

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 18 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CGACCCGCAC TCCGCAAT

18

(2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 52 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: artificial

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 27..27
- (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CCGCGCATCA CCACCCACACA CGCGCGGGGA GTATATAGGT TGGGGAAGTT TG

52

(2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

AACACACAAT AACAAACACA AATTCAC

27

(2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 50 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: artificial

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 27..27
- (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

CCCGGCTAAA CCCACCATCC AGCCGGGGGG AGGGTAGTTT AGTTGTGGTT

50

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 29 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

CAAAACAAAA AACTAAATC TACACAACC

29

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 47 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: artificial

(ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION: 26..26
 (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

CCGCGGAACA TTCAACTCAA CCGCGGGGAG GAGGAAGGTA GGTTTTTT

47

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 29 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

ACATACAACA ATCAATAACA TAAACCAC

29

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 47 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: artificial

(ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION: 22..22
 (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

CACGCCGCCG TATCTAGGCG TGGTTTGTTT CGACGTGATT TTTTCGA

47

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GCAAAAAATC CCCTCTCCCG C

21

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 45 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: artificial

(ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION: 22..22

(D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

GCCGCGCACA AATTTCGCG GCGAATTGGT TTTCGGAAGC GTTCG

45

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

CCCGAATTCC ACCGCC

16

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: artificial

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 23..23
- (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GGCGGAACGC ACAAATTTC GCGAATTGG TTTTCGGAAG CGTTCG

46

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: artificial
- (B) STRAIN: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CCCGAATTCC ACCGCC

16

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: artificial

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 26..26
- (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

TGCCGCCGCC GTATCTAATG GCGGCAGTTT GTTTCGACGT GATTTTTTCG A

51

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

GCAAAAAATC CCCTCTCCCG C

21

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: artificial

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 26..26
- (D) OTHER INFORMATION: /standard_name= "BHQ-HEG"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

CGGCCTTCGC TCTTCGCAAA AGGCCGGGTA GTTTTATGA AAGGCGTCGT G 51

(2) INFORMATION FOR SEQ ID NO:53:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

CTAACCGCCC ACCACGTT 18

(2) INFORMATION FOR SEQ ID NO:54:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 4260 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens

- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 1..4260
 - (D) OTHER INFORMATION: /standard_name= "HUMGSTP1G Homo Sapiens

(clone pHGST-pi)
glutathione S-transferase pi (GSTP1) gene, complete cds"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

AACAAGAGAT CAATATCTAG AATAAATGGA GATCTGCAAA TCAACAGAAA GTAGGCAGCA	60
AAGCCAAAGA AAATAGCCTA AGGCACAGCC ACTAAAAGGA ACGTGATCAT GTCCTTTGCA	120
GGGACATGGG TGGAGCTGGA AGCCGTTAGC CTCAGCAAAC TCACACAGGA ACAGAAAACC	180
AGCGAGACCG CATGGTCTCA CTTATAAGTG GGAGCTGAAC AATGAGAACA CATGGTCACA	240
TGGCGGCGAT CAACACACAC TGGTGCCTGT TGAGCGGGGT GCTGGGGAGG GAGAGTACCA	300
GGAAGAATAG CTAAGGGATA CTGGGCTTAA TACCTGGGTG ATGGGATGAT CTGTACAGCA	360

AACCATCATG GCGCACACAC CTATGTAACA AACCTGCACA TCCTGCACAT GTACCCCAGA	420
ACTTCAAATA AAAGTTGGAC GGCCAGGCGT GGTGGCTCAC GCCTGTAATC CCAGCACTTT	480
GGGAAGCCGA GGCGTGCAGA TCACCTAAGG TCAGGAGTTC GAGACCAGCC CGGCCAACAT	540
GGTGAAACCC CGTCTCTACT AAAAATACAA AAATCAGCCA GATGTGGCAC GCACCTATAA	600
TTCCACCTAC TCGGGAGGCT GAAGCAGAAT TGCTTGAACC CGAGAGGCGG AGGTTGCAGT	660
GAGCCGCCGA GATCGCGCCA CTGCACTCCA GCCTGGGCCA CAGCGTGAGA CTACGTCATA	720
AAATAAAATA AAATAACACA AAATAAAATA AAATAAAATA AAATAAAATA AAATAATAAA	780
ATAAAATAAA ATAAAATAAA ATAAAATAAA ATAAAGCAAT TTCCTTTCCT CTAAGCGGCC	840
TCCACCCCTC TCCCCTGCCC TGTGAAGCGG GTGTGCAAGC TCCGGGATCG CAGCGGTCTT	900
AGGGAATTTT CCCCCGCGAT GTCCCGGCGC GCCAGTTCGC TGCGCACACT TCGCTGCGGT	960
CCTCTTCCTG CTGTCTGTTT ACTCCCTAGG CCCCCTGCGG GACCTGGGAA AGAGGGAAAG	1020
GCTTCCCCGG CCAGCTGCGC GCGGACTCCG GGGACTCCAG GGCGCCCCTC TCGGGCCGAC	1080
GCCCGGGGTG CAGCGGCCGC CGGGGCTGGG GCCGGCGGGA GTCCGCGGGA CCCTCCAGAA	1140
GAGCGGCCGG CGCCGTGACT CAGCACTGGG GCGGAGCGGG GCGGGACCAC CCTTATAAGG	1200
CTCGGAGGCC GCGAGGCCTT CGCTGGAGTT TCGCCGCCGC AGTCTTCGCC ACCAGTGAGT	1260
ACGCGCGGCC CGCTCCCCGG GGATGGGGCT CAGAGCTCCC AGCATGGGGC CAACCCGCAG	1320
CATCAGGCCC GGGCTCCCCG CAGGGCTCCT CGCCACCTC GAGACCCGGG ACGGGGGCCT	1380
AGGGGACCCA GGACGTCCCC AGTGCCGTTA GCGGCTTTCA GGGGGCCCCG AGCGCCTCGG	1440
GGAGGGATGG GACCCCGGGG GCGGGGAGGG GGGGCAGGCT GCGCTCACCG CGCCTTGGCA	1500
TCCTCCCCCG GGCTCCAGCA AACTTTTCTT TGTTGCTGC AGTGCCGCC TACACCGTGG	1560
TCTATTTCCC AGTTCGAGGT AGGAGCATGT GTCTGGCAGG GAAGGGAGGC AGGGGCTGGG	1620
GCTGCAGCCC ACAGCCCCTC GCCCACC CGG AGAGATCCGA ACCCCCTTAT CCCTCCGTCC	1680
TGTGGCTTTT ACCCCGGGCC TCCTTCCTGT TCCCCGCCTC TCCCGCCATG CCTGCTCCCC	1740
GCCCCAGTGT TGTGTGAAAT CTTGCGAGGA ACCTGTTTAC CTGTTCCCTC CCTGCACTCC	1800
TGACCCCTCC CCGGGTTGCT GCGAGGCGGA GTCGGCCCGG TCCCCACATC TCGTACTTCT	1860
CCCTCCCCGC AGGCCGCTGC GCGGCCCTGC GCATGCTGCT GGCAGATCAG GGCCAGAGCT	1920
GGAAGGAGGA GGTGGTGACC GTGGAGACGT GGCAGGAGGG CTCACTCAAA GCCTCCTGCG	1980
TAAGTGACCA TGCCCGGGCA AGGGGAGGGG GTGCTGGGCC TTAGGGGGCT GTGACTAGGA	2040
TCGGGGGACG CCCAAGCTCA GTGCCCCCTC CTGAGCCATG CCTCCCCCAA CAGCTATACG	2100
GGCAGCTCCC CAAGTTCCAG GACGGAGACC TCACCCTGTA CCAGTCCAAT ACCATCCTGC	2160

GTCACCTGGG	CCGCACCCTT	GGTGAGTCTT	GAACCTCCAA	GTCCAGGGCA	GGCATGGGCA	2220
AGCCTCTGCC	CCCGGAGCCC	TTTTGTTTAA	ATCAGCTGCC	CCGCAGCCCT	CTGGAGTGGA	2280
GGAAACTGAG	ACCCACTGAG	GTTACGTAGT	TTGCCCAAGG	TCAAGCCTGG	GTGCCTGCAA	2340
TCCTTGCCCT	GTGCCAGGCT	GCCTCCCAGG	TGTCAGGTGA	GCTCTGAGCA	CCTGCTGTGT	2400
GGCAGTCTCT	CATCCTTCCA	CGCACATCCT	CTTCCCCTCC	TCCCAGGCTG	GGGCTCACAG	2460
ACAGCCCCCT	GGTTGGCCCA	TCCCCAGTGA	CTGTGTGTTG	ATCAGGCGCC	CAGTCACGCG	2520
GCCTGCTCCC	CTCCACCCAA	CCCCAGGGCT	CTATGGGAAG	GACCAGCAGG	AGGCAGCCCT	2580
GGTGGACATG	GTGAATGACG	GCGTGGAGGA	CCTCCGCTGC	AAATACATCT	CCCTCATCTA	2640
CACCAACTAT	GTGAGCATCT	GCACCAGGGT	TGGGCACTGG	GGGCTGAACA	AAGAAAGGGG	2700
CTTCTTGTGC	CCTCACCCCC	CTTACCCCTC	AGGTGGCTTG	GGCTGACCCC	TTCTTGGGTC	2760
AGGGTGCAGG	GGCTGGGTCA	GCTCTGGGCC	AGGGGCCCAG	GGGCCTGGGA	CAAGACACAA	2820
CCTGCACCCT	TATTGCCTGG	GACATCAACC	AGCCAAGTAA	CGGGTCATGG	GGGCGAGTGC	2880
AAGGACAGAG	ACCTCCAGCA	ACTGGTGGTT	TCTGATCTCC	TGGGGTGGCG	AGGGCTTCCT	2940
GGAGTAGCCA	GAGGTGGAGG	AGGATTTGTC	GCCAGTTTCT	GGATGGAGGT	GCTGGCACTT	3000
TTAGCTGAGG	AAAATATGCA	GACACAGAGC	ACATTTGGGG	ACCTGGGACC	AGTTCAGCAG	3060
AGGCAGCGTG	TGTGCGCGTG	CGTGTGCGTG	TGTGTGCGTG	TGTGTGTGTA	CGCTTGCAATT	3120
TGTGTCGGGT	GGGTAAGGAG	ATAGAGATGG	GCGGGCAGTA	GGCCCAGGTC	CCGAAGGCCT	3180
TGAACCCACT	GGTTTGGAGT	CTCCTAAGGG	CAATGGGGGC	CATTGAGAAG	TCTGAACAGG	3240
GCTGTGTCTG	AATGTGAGGT	CTAGAAGGAT	CCTCCAGAGA	AGCCAGCTCT	AAAGCTTTTG	3300
CAATCATCTG	GTGAGAGAAC	CCAGCAAGGA	TGGACAGGCA	GAATGGAATA	GAGATGAGTT	3360
GGCAGCTGAA	GTGGACAGGA	TTTGGTACTA	GCCTGGTTGT	GGGGAGCAAG	CAGAGGAGAA	3420
TCTGGGACTC	TGGTGTCTGG	CCTGGGGCAG	ACGGGGGTGT	CTCAGGGGCT	GGGAGGGATG	3480
AGAGTAGGAT	GATACATGGT	GGTGTCTGGC	AGGAGGCGGG	CAAGGATGAC	TATGTGAAGG	3540
CACTGCCCCG	GCAACTGAAG	CCTTTTGAAG	CCCTGCTGTC	CCAGAACCAG	GGAGGCAAGA	3600
CCTTCATTGT	GGGAGACCAG	GTGAGCATCT	GGCCCCATGC	TGTTCCCTTC	TCGCCACCCT	3660
CTGCTTCCAG	ATGGACACAG	GTGTGAGCCA	TTTGTTTAGC	AAAGCAGAGC	AGACCTAGGG	3720
GATGGGCTTA	GGCCCTCTGC	CCCCAATTCC	TCCAGCCTGC	TCCCGCTGGC	TGAGTCCCTA	3780
GGCCCCCTGC	CCTGCAGATC	TCCTTCGCTG	ACTACAACCT	GCTGGACTTG	CTGCTGATCC	3840
ATGAGGTCCT	AGCCCCTGGC	TGCCTGGATG	CGTTCCCCCT	GCTCTCAGCA	TATGTGGGGC	3900
GCCTCAGTGC	CCGGCCCAAG	CTCAAGGCCT	TCCTGGCCTC	CCCTGAGTAC	GTGAACCTCC	3960
CCATCAATGG	CAACGGGAAA	CAGTGAGGGT	TGGGGGGACT	CTGAGCGGGA	GGCAGAGTTT	4020

GCCTTCCTTT CTCCAGGACC AATAAAATTT CTAAGAGAGC TACTATGAGC ACTGTGTTTC	4080
CTGGGACGGG GCTTAGGGGT TCTCAGCCTC GAGGTCG3TG GGAGGGCAGA GCAGAGGACT	4140
AGAAAACAGC TCCTCCAGCA CAGTCAGTGG CTTCTG3AG CCCTCAGCCT GGCTGTGTTT	4200
ACTGAACCTC ACAAACCTAGA AGAGGAAGAA AAAAAAAGAG AGAGAGAAAC AAAGAGAAAT	4260

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2011 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 1..2011
- (D) OTHER INFORMATION: /standard_name= "HSACTBPR Human beta-actin gene 5' flanking region, CpG Island 1656 to 1955"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

GAGCTCTGTC TCTTGGCCAG CTGAATGGAG GCCCAGCGGC AACACAGGTC CTGCCTGGGG	60
ATCAGGTCTG CTCTGCACCC CACCTTGCTG CCTGGAGCCG CCCACCTGAC AACCTCTCAT	120
CCCTGCTCTG TAGATCCGGT CCCATCCCCA CTGCCCACCC CACCCCCCCA GCACTCCACC	180
CAGTTCAACG TTCCACGAAC CCCCAGAACC AGCCCTCATC AACAGGCAGC AAGAAGGGCC	240
CCCCGCCCAT CGCCCCACAA CGCCAGCCGG GTGAACTGTA GCGTTGGCAG GTCCTGAGGC	300
AGCTGAAAGA TACAAGGCCA GGGACAGGAC AGTCCCATCC CCAGGAGGCA GGGAGTATAC	360
AGGCTGGGGA AGTTTGCCCT TGCGTGGGGT GGTGATGGAG GAGGCTCAGC AAGTCTTCTG	420
GACTGTGAAC CTGTGTCTGC CACTGTGTGC TGGGTGGTGG TCATCTTTCC CACCAGGCTG	480
TGGCCTCTGC AACCTTCAAG GGAGGAGCAG GTCCCATTGG CTGAGCACAG CCTTGTACGT	540
GAAGTGAACA AGCAGCCTCC TTCCTGGCCA CAGGTTCCAT GTCCTTATAT GGACTCATCT	600
TTGCCTATTG CGACACACAC TCAATGAACA CCTACTACGC GCTGCAAAGA GCCCCGCAGG	660
CCTGAGGTGC CCCCACCTCA CCACTCTTCC TATTTTGTG TAAAAATCCA GCTTCTTGTC	720
ACCACCTCCA AGGAGGGGGA GGAGGAGGAA GGCAGGTTCC TCTAGGCTGA GCCGAATGCC	780
CCTCTGTGGT CCCACGCCAC TGATCGCTGC ATGCCACCA CCTGGGTACA CACAGTCTGT	840
GATTCCCGGA GCAGAACGGA CCCTGCCCAC CCGGTCITGT GTGCTACTCA GTGGACAGAC	900

```

CCAAGGCAAG AAAGGGTGAC AAGGACAGGG TCTTCCCAGG CTGGCTTTGA GTTCCTAGCA      960
CCGCCCCGCC CCAATCCTC TGTGGCACAT GGAGTCTTGG TCCCCAGAGT CCCCAGCGG      1020
CCTCCAGATG GTCTGGGAGG GCAGTTCAGC TGTGGCTGCG CATAGCAGAC ATACAACGGA      1080
CGGTGGGCCC AGACCCAGGC TGTGTAGACC CAGCCCCCCC GCCCCGCAGT GCCTAGGTCA      1140
CCCACTAACG CCCCAGGCCT GGTCTTGGCT GGGCGTGACT GTTACCCTCA AAAGCAGGCA      1200
GCTCCAGGGT AAAAGGTGCC CTGCCCTGTA GAGCCCCTT CTTTCCCAGG GCTGCGGCTG      1260
GGTAGGTTTG TAGCCTTCAT CACGGGCCAC CTCCAGCCAC TGGACCGCTG GCCCCTGCCC      1320
TGTCTGGGG AGTGTGGTCC TGC GACTCTA ATGGCCGCAA GCCACCTGAC TCCCCAACA      1380
CCACACTCTA CCTCTCAAGC CCAGGTCTCT CCCTAGTGAC CCACCCAGCA CATTTAGCTA      1440
GCTGAGCCCC ACAGCCAGAG GTCCTCAGGC CCTGCTTTCA GGGCAGTTGC TCTGAAGTCG      1500
GCAAGGGGGA GTGACTGCCT GGCCACTCCA TGCCCTCCAA GAGCTCCTTC TGCAGGAGCG      1560
TACAGAACCC AGGGCCCTGG CACCCGTGCA GACCCTGGCC CACCCACCT GGGCGCTCAG      1620
TGCCCAAGAG ATGTCCACAC CTAGGATGTC CCGCGGTGGG TGGGGGGCCC GAGAGACGGG      1680
CAGGCCGGGG GCAGGCCTGG CCATGCGGGG CCGAACCGGG CACTGCCCAG CGTGGGGCGC      1740
GGGGGCCACG GCGCGCGCCC CCAGCCCCCG GGCCCAGCAC CCAAGGCGG CCAACGCCAA      1800
AACTCTCCCT CCTCCTCTTC CTCAATCTCG CTCTCGCTCT TTTTTTTTTT CGCAAAGGA      1860
GGGGAGAGGG GTAAAAAAA TGCTGCACTG TCGGCGAAGC CGGTGAGTGA GCGGCGCGGG      1920
GCCAATCGCG TGCGCCGTTT CGAAAGTTGC CTTTATGGC TCGAGCGGCC GCGGCGGCGC      1980
CCTATAAAAC CCAGCGGCGC GACGCGCCAC C                                2011

```

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1792 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 1..1792
- (D) OTHER INFORMATION: /standard_name= "Homo sapiens actin, beta (ACTB), mRNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

CGCGTCCGCC CCGCGAGCAC AGAGCCTCGC CTTTGCCGAT CCGCCGCCCCG TCCACACCCG	60
CCGCCAGCTC ACCATGGATG ATGATATCGC CGCGCTCGTC GTCGACAACG GCTCCGGCAT	120
GTGCAAGGCC GGCTTCGCGG GCGACGATGC CCCCCGGGCC GTCTTCCCCT CCATCGTGGG	180
GCGCCCCAGG CACCAGGGCG TGATGGTGGG CATGGGTCAG AAGGATTCCT ATGTGGGCGA	240
CGAGGCCCAG AGCAAGAGAG GCATCCTCAC CCTGAAGTAC CCCATCGAGC ACGGCATCGT	300
CACCAACTGG GACGACATGG AGAAAATCTG GCACCACACC TTCTACAATG AGCTGCGTGT	360
GGCTCCCGAG GAGCACCCCG TGCTGCTGAC CGAGGCCCCC CTGAACCCCA AGGCCAACCG	420
CGAGAAGATG ACCCAGATCA TGTTTGAGAC CTTCAACACC CCAGCCATGT ACGTTGCTAT	480
CCAGGCTGTG CTATCCCTGT ACGCCTCTGG CCGTACCACT GGCATCGTGA TGGACTCCGG	540
TGACGGGGTC ACCCACACTG TGCCCATCTA CGAGGGGTAT GCCCTCCCCC ATGCCATCCT	600
GCGTCTGGAC CTGGCTGGCC GGGACCTGAC TGA CTACCTC ATGAAGATCC TCACCGAGCG	660
CGGCTACAGC TTCACCACCA CGGCCGAGCG GGAAATCGTG CGTGACATTA AGGAGAAGCT	720
GTGCTACGTC GCCCTGGACT TCGAGCAAGA GATGGCCACG GCTGCTTCCA GCTCCTCCCT	780
GGAGAAGAGC TACGAGCTGC CTGACGGCCA GGTCATCACC ATTGGCAATG AGCGGTTCCG	840
CTGCCCTGAG GCACTCTTCC AGCCTTCCTT CCTGGGCATG GAGTCCTGTG GCATCCACGA	900
AACTACCTTC AACTCCATCA TGAAGTGTGA CGTGGACATC CGCAAAGACC TGTACGCCAA	960
CACAGTGCTG TCTGGCGGCA CCACCATGTA CCCTGGCATT GCCGACAGGA TGCAGAAGGA	1020
GATCACTGCC CTGGCACCCA GCACAATGAA GATCAAGATC ATTGCTCCTC CTGAGCGCAA	1080
GTACTCCGTG TGGATCGGCG GCTCCATCCT GGCCTCGCTG TCCACCTTCC AGCAGATGTG	1140
GATCAGCAAG CAGGAGTATG ACGAGTCCGG CCCCTCCATC GTCCACCGCA AATGCTTCTA	1200
GGCGGACTAT GACTTAGTTG CGTTACACCC TTTCTTGACA AAACCTAACT TGCGCAGAAA	1260
ACAAGATGAG ATTGGCATGG CTTTATTTGT TTTTTTTGTT TTGTTTTGGT TTTTTTTTTT	1320
TTTTTGGCTT GACTCAGGAT TTAAAACTG GAACGGTGAA GGTGACAGCA GTCGGTTGGA	1380
GCGAGCATCC CCCAAAGTTC ACAATGTGGC CGAGGACTTT GATTGCACAT TGTTGTTTTT	1440
TTAATAGTCA TTCAAATAT GAGATGCATT GTTACAGGAA GTCCCTTGCC ATCCTAAAAG	1500
CCACCCCACT TCTCTCTAAG GAGAATGGCC CAGTCCTCTC CCAAGTCCAC ACAGGGGAGG	1560
TGATAGCATT GCTTTCGTGT AAATTATGTA ATGCAAAATT TTTTAAATCT TCGCCTTAAT	1620
ACTTTTTTAT TTTGTTTTAT TTTGAATGAT GAGCCTTCGT GCCCCCCTT CCCCCTTTTT	1680
GTCCCCCAAC TTGAGATGTA TGAAGGCTTT TGGTCTCCCT GGGAGTGGGT GGAGGCAGCC	1740
AGGGCTTACC TGTACACTGA CTTGAGACCA GTTGAATAAA AGTGACACACC TT	1792

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2762 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 1..2762
- (D) OTHER INFORMATION: /standard_name= "RARB2 retinoic acid receptor, beta (RARB), transcript variant 2, mRNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

GTGACAGAAG TAGTAGGAAG TGAGCTGTTC AGAGGCAGGA GGGTCTATTC TTTGCCAAAG	60
GGGGGACCAG AATTCCCCAT GCGAGCTGTT TGAGGACTGG GATGCCGAGA ACGCGAGCGA	120
TCCGAGCAGG GTTTGTCTGG GCACCGTCGG GGTAGGATCC GGAACGCATT CGGAAGGCTT	180
TTTGCAAGCA TTTACTTGGA AGGAGAACTT GGGATCTTTC TGGGAACCCC CCGCCCCGGC	240
TGGATTGGCC GAGCAAGCCT GGAAAATGCA ATTGAAACAC AGAGCACCAG CTCTGAGGAA	300
CTCGTCCCAA GCCCCCATC TCCACTTCCT CCCCTCGAG TGTACAAACC CTGCTTCGTC	360
TGCCAGGACA AATCATCAGG GTACCACTAT GGGGTCAGCG CCTGTGAGGG ATGTAAGGGC	420
TTTTTCCGCA GAAGTATTCA GAAGAATATG ATTTACACTT GTCACCGAGA TAAGAACTGT	480
GTTATTAATA AAGTCACCAG GAATCGATGC CAATACTGTC GACTCCAGAA GTGCTTTGAA	540
GTGGGAATGT CCAAAGAATC TGTCAGGAAT GACAGGAACA AGAAAAAGAA GGAGACTTCG	600
AAGCAAGAAT GCACAGAGAG CTATGAAATG ACAGCTGAGT TGGACGATCT CACAGAGAAG	660
ATCCGAAAAG CTCACCAGGA AACTTTCCCT TCACTCTGCC AGCTGGGTAA ATACACCACG	720
AATTCCAGTG CTGACCATCG AGTCCGACTG GACCTGGGCC TCTGGGACAA ATTCAGTGAA	780
CTGGCCACCA AGTGCATTAT TAAGATCGTG GAGTTTGCTA AACGTCTGCC TGGTTTCACT	840
GGCTTGACCA TCGCAGACCA AATTACCCTG CTGAAGGCCG CCTGCCTGGA CATCCTGATT	900
CTTAGAATTT GCACCAGGTA TACCCAGAA CAAGACACCA TGACTTTCTC AGACGGCCTT	960
ACCCTAAATC GAACTCAGAT GCACAATGCT GGATTTGGTC CTCTGACTGA CCTTGTGTTC	1020
ACCTTTGCCA ACCAGCTCCT GCCTTTGGAA ATGGATGACA CAGAAACAGG CCTTCTCAGT	1080
GCCATCTGCT TAATCTGTGG AGACCGCCAG GACCTTGAGG AACCGACAAA AGTAGATAAG	1140
CTACAAGAAC CATTGCTGGA AGCACTAAAA ATTTATATCA GAAAAAGACG ACCCAGCAAG	1200

CCTCACATGT TTCCAAAGAT CTTAATGAAA ATCACAGATC TCCGTAGCAT CAGTGCTAAA	1260
GGTGCAGAGC GTGTAATTAC CTTGAAAATG GAAATTCCTG GATCAATGCC ACCTCTCATT	1320
CAAGAAATGC TGGAGAATTC TGAAGGACAT GAACCCTTGA CCCCAAGTTC AAGTGGGAAC	1380
ACAGCAGAGC ACAGTCCTAG CATCTCACCC AGCTCAGTGG AAAACAGTGG GGTCAGTCAG	1440
TCACCACTCG TGCAATAAGA CATTTTCTAG CTACTTCAAA CATTCCCCAG TACCTTCAGT	1500
TCCAGGATTT AAAATGCAAG AAAAAACATT TTTACTGCTG CTTAGTTTTT GGACTGAAAA	1560
GATATTAAAA CTCAAGAAGG ACCAAGAAGT TTTCATATGT ATCAATATAT ATACTCCTCA	1620
CTGTGTA ACT TACCTAGAAA TACAACTTT TCCAATTTTA AAAAATCAGC CATTTTCATGC	1680
AACCAGAAAC TAGTTAAAAG CTTCTATTTT CCTCTTTGAA CACTCAAGAT GCATGGCAAA	1740
GACCCAGTCA AAATGATTTA CCCCTGGTTA AGTTTCTGAA GACTTTGTAC ATACAGAAGT	1800
ATGGCTCTGT TCTTTCTATA CTGTATGTTT GGTGCTTTCC TTTTGTCTTG CATACTCAA	1860
ATAACCATGA CACCAAGGTT ATGAAATAGA CTACTGTACA CGTCTACCTA GGTTCAAAAA	1920
GATAACTGTC TTGCTTTCAT GGAATAGTCA AGACATCAAG GTAAGGAAAC AGGACTATTG	1980
ACAGGACTAT TGTACAGTAT GACAAGATAA GGCTGAAGAT ATTCTACTTT AGTTAGTATG	2040
GAAGCTTGTC TTTGCTCTTT CTGATGCTCT CAAACTGCAT CTTTTATTTC ATGTTGCCCA	2100
GTAAAAGTAT ACAAATTCCC TGCACTAGCA GAAGAGAATT CTGTATCAGT GTAAGTGCCA	2160
GTTCAGTTAA TCAAATGTCA TTTGTTCAAT TGTTAATGTC ACTTTAAATT AAAAGTGGTT	2220
TATTACTTGT TTAATGACAT AACTACACAG TTAGTTAAAA AAAATTTTTT TACAGTAATG	2280
ATAGCCTCCA AGGCAGAAAC ACTTTTCAGT GTTAAGTTTT TGTTTACTTG TTCACAAGCC	2340
ATTAGGGAAA TTTCATGGGA TAATTAGCAG GCTGGTCTAC CACTGGACCA TGTAAGTCTA	2400
GTGTCCTTCC TGATTCATGC CTGATATTGG GATTTTTTTC CAGCCCTTCT TGATGCCAAG	2460
GGCTAATTAT ATTACATCCC AAAGAAACAG GCATAGAATC TGCCTCCTTT GACCTTGTTT	2520
AATCACTATG AAGCAGAGTG AAAGCTGTGG TAGAGTGGTT AACAGATACA AGTGTCAGTT	2580
TCTTAGTTCT CATTTAAGCA CTACTGGAAT TTTTTTTTTT GATATATTAG CAAGTCTGTG	2640
ATGTACTTTC ACTGGCTCTG TTTGTACATT GAGATTGTTT GTTTAACAAT GCTTTCTATG	2700
TTCATATACT GTTTACCTTT TTCCATGGAC TCTCCTGGCA AAGAATAAAA TATATTTATT	2760
TT	2762

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5487 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(ix) FEATURE:

(A) NAME/KEY: misc_feature

(B) LOCATION: 1..5487

(D) OTHER INFORMATION: /standard_name= "Homo sapiens tissue inhibitor of metalloproteinase 3 (Sorsby fundus dystrophy, pseudoinflammatory) (TIMP3), mRNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

TCTGTCGACT TGCCCCAGAG CTGATCCTTG TCTTTGTCCA CTTCTCAGCG AGGATGGCAC	60
TTCAGGGAGC CCTTCCCTTA CTATCGCAGA GAGAGCAGGC CCTCCCCAGT CATGTCCAAC	120
CCAGAACTCT GTTTTGTTTT CTTCATAGCC CTAGCATCAC AGAAAATCAC CCTGTGCATT	180
CATGGATGTC CACGGGGGCA AGGGCTTTGT GTTGCTTAAC CCAGCATCCT GAACCGTGTT	240
TGTTGAATGA ATACAGAACC CCGTTTGCTC TGGGAGAGCA CAGAAAACAG TCTTCTATCA	300
TATATCATAG CCAGCTGCAA ACAGCAGATG GCTTCCCATA TCCCAGAGAG TAAGAACCAG	360
AGAGAGAGAG AAAGAGAGAG AGTTTGGGTC TTTCTCCTCT GTGCCTGCTC TCTCCAGAGA	420
AACTGGAGGG GTAGCAGTTA GCATTCCCCC GCTGGTTCCA CCAAGCACAG TCAAGGTCTC	480
TAGGACATGG CCACCCCTCA CCTGTGGAAG CGGTCCTGCT GGGGTGGGTG GGTGTTAGTT	540
GGTTCTGGTT TGGGTCAGAG ACACCCAGTG GCCCAGGTGG GCGTGGGGCC AGGGCGCAGA	600
CGAGAAGGGG CACGAGGGCT CCGCTCCGAG GACCCAGCGG CAAGCACCGG TCCCGGGCGC	660
GCCCCAGCCC ACCCACTCGC GTGCCCACGG CGGCATTATT CCCTATAAGG ATCTGAACGA	720
TCCGGGGGCG GCCCCGCCCC GTTACCCCTT GCCCCCGGCC CCGCCCCCTT TTTGGAGGGC	780
CGATGAGGTA ATGCGGCTCT GCCATTGGTC TGAGGGGGCG GGCCCCAACA GCCCGAGGCG	840
GGGTCCCCGG GGGCCCAGCG CTATATCACT CGGCCGCCCA GGCAGCGGCG CAGAGCGGGC	900
AGCAGGCAGG CGGCGGGCGC TCAGACGGCT TCTCCTCCTC CTCTTGCTCC TCCAGCTCCT	960
GCTCCTTCGC CGGGAGGCCG CCCGCCGAGT CCTGCGCCAG CGCCGAGGCA GCCTCGCTGC	1020
GCCCCATCCC GTCCCGCCGG GCACTCGGAG GGCAGCGCGC CGGAGGCCAA GGTGCCCCG	1080
CACGGCCCCG CGGGCGAGCG AGCTCGGGCT GCAGCAGCCC CGCCGGCGGC GCGCACGGCA	1140
ACTTTGGAGA GGCGAGCAGC AGCCCCGGCA GCGGCGGCAG CAGCGGCAAT GACCCCTTGG	1200
CTCGGGCTCA TCGTGCTCCT GGGCAGCTGG AGCCTGGGGG ACTGGGGCGC CGAGGCGTGC	1260
ACATGCTCGC CCAGCCACCC CCAGGACGCC TTCTGCAACT CCGACATCGT GATCCGGGCC	1320

AAGGTGGTGG	GGAAGAAGCT	GGTAAAGGAG	GGGCCCTTCG	GCACGCTGGT	CTACACCATC	1380
AAGCAGATGA	AGATGTACCG	AGGCTTCACC	AAGATGCCCC	ATGTGCAGTA	CATCCACACG	1440
GAAGCTTCCG	AGAGTCTCTG	TGGCCTTAAG	CTGGAGGTCA	ACAAGTACCA	GTACCTGCTG	1500
ACAGGTCGCG	TCTATGATGG	CAAGATGTAC	ACGGGGCIGT	GCAACTTCGT	GGAGAGGTGG	1560
GACCAGCTCA	CCCTCTCCCA	GCGCAAGGGG	CTGAACTATC	GGTATCACCT	GGGTTGTAAC	1620
TGCAAGATCA	AGTCCTGCTA	CTACCTGCCT	TGCTTTGTGA	CTTCCAAGAA	CGAGTGTCTC	1680
TGGACCGACA	TGCTCTCCAA	TTTCGGTTAC	CCTGGCTACC	AGTCCAAACA	CTACGCCTGC	1740
ATCCGGCAGA	AGGGCGGCTA	CTGCAGCTGG	TACCGAGGAT	GGGCCCCCCC	GGATAAAAGC	1800
ATCATCAATG	CCACAGACCC	CTGAGCGCCA	GACCCTGCCC	CACCTCACTT	CCCTCCCTTC	1860
CCGCTGAGCT	TCCCTTGGAC	ACTAACTCTT	CCCAGATGAT	GACAATGAAA	TTAGTGCCTG	1920
TTTTCTTGCA	AATTTAGCAC	TTGGAACATT	TAAAGAAAGG	TCTATGCTGT	CATATGGGGT	1980
TTATTGGGAA	CTATCCTCCT	GGCCCCACCC	TGCCCCTTCT	TTTTGGTTTT	GACATCATTC	2040
ATTTCCACCT	GGGAATTTCT	GGTGCCATGC	CAGAAAGAAT	GAGGAACCTG	TATTCCTCTT	2100
CTTCGTGATA	ATATAATCTC	TATTTTTTTA	GGAAAACAAA	AATGAAAAAC	TACTCCATTT	2160
GAGGATTGTA	ATTCCCACCC	CTCTTGCTTC	TTCCCCACCT	CACCATCTCC	CAGACCCTCT	2220
TCCCTTTGCC	CTTCTCCTCC	AATACATAAA	GGACACAGAC	AAGGAACTTG	CTGAAAGGCC	2280
AACCATTTCA	GGATCAGTCA	AAGGCAGCAA	GCAGATAGAC	TCAAGGTGTG	TGAAAGATGT	2340
TATACACCAG	GAGCTGCCAC	TGCATGTCCC	AACCAGACTG	TGTCTGTCTG	TGTCTGCATG	2400
TAAGAGTGAG	GGAGGGAAGG	AAGGAACTAC	AAGAGAGTCG	GAGATGATGC	AGCACACACA	2460
CAATTCCCCA	GCCCAGTGAT	GCTTGTGTTG	ACCAGATGTT	CCTGAGTCTG	GAGCAAGCAC	2520
CCAGGCCAGA	ATAACAGAGC	TTTCTTAGTT	GGTGAAGACT	TAAACATCTG	CCTGAGGTCA	2580
GGAGGCAATT	TGCCTGCCTT	GTACAAAAGC	TCAGGTGAAA	GACTGAGATG	AATGTCTTTC	2640
CTCTCCCTGC	CTCCCACCAG	ACTTCCTCCT	GGAAAACGCT	TTGGTAGATT	TGGCCAGGAG	2700
CTTTCTTTTA	TGTAAATTGG	ATAAATACAC	ACACCATACA	CTATCCACAG	ATATAGCCAA	2760
GTAGATTTGG	GTAGAGGATA	CTATTICCAG	AATAGTGTTT	AGCTCACCTA	GGGGGATATG	2820
TTTGTATACA	CATTTGCATA	TACCCACATG	GGGACATAAG	CTAATTTTTT	TACAGGACAC	2880
AGAATTCTGT	TCAATGCTGT	TAAATATGCC	AATAGTTTAA	TCTCTTCTAT	TTTGTGTGCG	2940
TTGCTTGTTT	GAAGAAAATC	ATGACATTCC	AAGTTGACAT	TTTTTTTTCA	TTTTAATTAA	3000
AATTTGAAAT	TCTGAACACC	GTCAGCACCC	TCTCTTCCCT	ATCATGGGTC	ATCTGACCCC	3060
TGTCCGTCTC	CTTGTCCCTG	CTTCAIGTTT	GGGGGCCTTT	CTTTAACTGC	CTTCCTGGCT	3120
TAGCTCAGAT	GGCAGATGAG	AGTGTAGTCA	AGGGCCTGGG	CACAGGAGGG	AGAGCTGCAG	3180

AGTGTCTTGC	CTGCCTTGGC	TGGAGGGACA	CCTCTCCTGG	GTGTGGAGAC	AGCTTGGTTC	3240
CCTTTCCCTA	GCTCCCTGGT	GGGTGAATGC	CACCTCCTGA	GATCCTCACC	TCTTGGAATT	3300
AAAATTGTTG	GTCACTGGGG	AAAGCCTGAG	TTTGCAACCA	GTTGTAGGGT	TTCTGTTGTG	3360
TTTTTTTTTT	TTTTTTTGAA	ATAAACTAT	AATATAAATT	CTCCTATTAA	ATAAAATTAT	3420
TTTAAGTTTT	AGTGTCAAAA	GTGAGATGCT	GAGAGTAGGT	GATAATGTAT	ATTTTACAGA	3480
GTGGGGGTTG	GCAGGATGGT	GACATTGAAC	ATGATTGCTC	TCTGTCTCTT	TTTTCAGCTT	3540
ATGGGTATTT	ATCTTCTATT	AGTATTTGTA	TCTTCAGTTC	ATTCCACTTT	AGGAAACAGA	3600
GCTGCCAATT	GAAACAGAAG	AAGAAAAAAA	AAAAAAGCAG	CAGACAACAC	ACTGTAGAGT	3660
CTTGACACAC	CACAAGTGCC	CAGGCAAGGT	GCTTGGCAGA	ACCGCAGAGT	GGGAAGAGAG	3720
TACCGGCATC	GGGTTTCCTT	GGGATCAATT	TCATTACCGT	GTACCTTTCC	CATTGTGGTC	3780
ATGCCATTTG	GCAGGGGGAG	AATGGGAGGC	TTGGCCTTCT	TTGTGAGGCA	GTGTGAGCAG	3840
AAGCTGATGC	CAGCATGTCA	CTGGTTTTGA	AGGGATGAGC	CCAGACTTGA	TGTTTTGGGA	3900
TTGTCCTTAT	TTTAACCTCA	AGGTCTCGCA	TGGTGGGGCC	CCTGACCAAC	CTACACAAGT	3960
TCCCTCCCAC	AAGTGGACAT	CAGTGTCTTC	TCTGTGAGGC	ATCTGGCCAT	TCGCACTCCC	4020
TGGTGTGGTC	AGCCTCTCTC	ACACAAGGAG	GAAGTGGGT	GAAGGCTGAG	TGTGAGGCAC	4080
CTGAAGTTTC	CCTGCGGAGT	CGATAAATTA	GCAGAACCAC	ATCCCCATCT	GTTAGGCCTT	4140
GGTGAGGAGG	CCCTGGGCAA	AGAAGGGTCT	TTCGCAAAGC	GATGTCAGAG	GGCGGTTTTG	4200
AGCTTTCTAT	AAGCTATAGC	TTTGTTTATT	TCACCCGTTC	ACTTACTGTA	TAATTTAAAA	4260
TCATTTATGT	AGCTGAGACA	CTTCTGTATT	TCAATCATAT	CATGAACATT	TTATTTTGCT	4320
AAATCTTGTTG	TCATGTGTAG	GCTGTAATAT	GTGTACATTG	TGTTTAAGAG	AAAAATGAAA	4380
CCCACATGCC	GCCATTTTCC	TGAATCAAAT	TCTGCAGTGG	AATGGAGAGG	AAAATACTTC	4440
TAGGCAAGCA	GCTAGACTGG	TGAATTGGGG	GAAATAGAAG	GAAGTAGTAA	CTGAGACTCC	4500
TCCAGCCTCC	TCCCTATTGG	AATCCCAATG	GCTCCTGGAG	TAGGAAAAAA	GTTTAAACTA	4560
CATTCATGTT	CTTGTTCTGT	GTCACCTGGC	CCTGGGTAGT	CTACCATTTA	CTTCACCCCA	4620
AGTCCTGCTG	CCCATCCAGT	TGGGAAGCCA	TGATTTTCCT	AAGAATCCAG	GGCCATGGGA	4680
GATACAATTC	CAAGTTCTCG	CTTCCTCCTT	TGGGCATCTC	TTCTGCCTCC	CAATCAAGGA	4740
AGCTCCATGC	TCAGGCTCTC	AGCTCTCGGG	CCAGTGCTCT	GCTCTGTCCA	GGGTAGGTAA	4800
TACTGGGAGA	CTCCTGTCTT	TTACCCTCCC	CTCGTTCCAG	ACCTGCCTCA	TGGTGGCAAC	4860
ATGGTTCTTG	AACAATTAAA	GAAACAAATG	ACTTTTTGGA	ATAGCCCTGT	CTAGGGCAAA	4920
CTGTGGCCCC	CAGGAGACAC	TACCCTTCCA	TGCCCCAGAC	CTCTGTCTTG	CATGTGACAA	4980

```

TTGACAATCT GGACTACCCC AAGATGGCAC CCAAGTGTTT GGCTTCTGGC TACCTAAGGT      5040
TAACATGTCA CTAGAGTATT TTTATGAGAG ACAAACATTA TAAAAATCTG ATGGCAAAAG      5100
CAAAACAAAA TGGAAAGTAG GGGAGGTGGA TGTGACAACA ACTTCCAAAT TGGCTCTTTG      5160
GAGGCGAGAG GAAGGGGAGA ACTTGGAGAA TAGTTTTTGC TTTGGGGGTA GAGGCTTCTT      5220
AGATTCTCCC AGCATCCGCC TTTCCCTTTA GCCAGTCTGC TGTCTGAAA CCCAGAAGTG      5280
ATGGAGAGAA ACCAACAAGA GATCTCGAAC CCTGTCTAGA AGGAATGTAT TTGTTGCTAA      5340
ATTTTCGTAGC ACTGTTTACA GTTTTCCTCC ATGTTATTTA TGAATTTTAT ATTCCGTGAA      5400
TGTATATTGT CTTGTAATGT TGCATAATGT TCACTTTTTA TAGTGTGTCC TTTATTCTAA      5460
ACAGTAAAGT GGTTTTATTT CTATCAC                                     5487

```

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 866 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 315..315
- (D) OTHER INFORMATION: /standard_name= "N is any of A, C, G or T"

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 1..866
- (D) OTHER INFORMATION: /standard_name= "HSU02509 Human adenomatous polyposis coli (APC) gene, promoter sequence"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

```

ACTTATATAT CTGACAGTTG ATTTGTCCTC ACCTCTAAAT TGGAATTTAA GCATCACCTG      60
GTTTCGATTTA ATGCAATGTA GAATTTCAT TAAATACTA CATTAAAGCC TCAGATTTGT      120
AGTAGCTAAC AGCACTTCTA TGTATGTGTC AGGGACTGCT CTAAATACTT CATATATATT      180
AACTCCTCTA TTCTGTACTT CTGTTCCCGT TTTATACAGC AGGAAATTGA AACACTGAGA      240
GGTTAAGTAA CTAAAGTTAC AGAGCTAGAG TGACAGGAGT AAAGCTTCAA CTCAGGCAAC      300
CCAGACGTCC AGAGNTCTGA TCTCCACTAC TAAGCTGCTA GCATAGCTTT TCTGGTAACT      360
ATTTTTAATT CAATATAATT CGAATGATCT ATCTAACAAG TCATCACTCT GACAACTCAG      420
TGACTTGTA TGTAAAATTA TTCATTGTAA TCACTTAAT ATTATTGTTT CTCTGTGCTG      480

```

CAAAAATCAT AGCAATCGAG ATGTAATTTA TTACTCTCCC TCCCACCTCC GGCATCTTGT	540
GCTAATCCTT CTGCCCTGCG GACCTCCCCC GACTCTTTAC TATGCGTGTC AACTGCCATC	600
AACTTCCTTG CTTGCTGGGG ACTGGGGCCG TGAGGGCATA CCCCCGAGGG GTACGGGGCT	660
AGGGCTAGGC AGGCTGTGCG GTTGGGCGGG GCCCTGTGCC CCACTGCGGA GTGCGGGTCG	720
GGAAGCGGAG AGAGAAGCAG CTGTGTAATC CGCTGGATGC GGACCAGGGC GCTCCCCATT	780
CCCGTCGGGA GCCCGCCGAT TGGCTGGGTG TGGGCGCACG TGACCGACAT GTGGCTGTAT	840
TGGTGCAGCC CGCCAGGGTG TCACTG	866

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10386 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 9521..9521
- (D) OTHER INFORMATION: /standard_name= "N is A, C, G or T"

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 1..10386
- (D) OTHER INFORMATION: /standard_name= "Homo sapiens adenomatous polyposis coli (APC), mRNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

ATTGAGGACT CGGAAATGAG GTCCAAGGGT AGCCAAGGAT GGCTGCAGCT TCATATGATC	60
AGTTGTTAAA GCAAGTTGAG GCACTGAAGA TGGAGAACTC AAATCTTCGA CAAGAGCTAG	120
AAGATAATTC CAATCATCTT ACAAACCTGG AACTGAGGC ATCTAATATG AAGGAAGTAC	180
TTAAACAACCT ACAAGGAAGT ATTGAAGATG AAGCTATGGC TTCTTCTGGA CAGATTGATT	240
TATTAGAGCG TCTTAAAGAG CTTAACTTAG ATAGCAGTAA TTTCCCTGGA GTAAACCTGC	300
GGTCAAAAAT GTCCCTCCGT TCTTATGGAA GCCGGGAAGG ATCTGTATCA AGCCGTTCTG	360
GAGAGTGCAG TCCTGTTTCT ATGGGTTCAT TTCCAAGAAG AGGGTTTGTA AATGGAAGCA	420
GAGAAAGTAC TGGATATTTA GAAGAACTTG AGAAAGAGAG GTCATTGCTT CTTGCTGATC	480
TTGACAAAGA AGAAAAGGAA AAAGACTGGT ATTACGCTCA ACTTCAGAAT CTCACTAAAA	540

GAATAGATAG	TCTTCCTTTA	ACTGAAAATT	TTTCCTTACA	AACAGATATG	ACCAGAAGGC	600
AATTGGAATA	TGAAGCAAGG	CAAATCAGAG	TTGCGATGGA	AGAACAATA	GGTACCTGCC	660
AGGATATGGA	AAAACGAGCA	CAGCGAAGAA	TAGCCAGAAT	TCAGCAAATC	GAAAAGGACA	720
TACTTCGTAT	ACGACAGCTT	TTACAGTCCC	AAGCAACAGA	AGCAGAGAGG	TCATCTCAGA	780
ACAAGCATGA	AACCGGCTCA	CATGATGCTG	AGCGGCAGAA	TGAAGGTCAA	GGAGTGGGAG	840
AAATCAACAT	GGCAACTTCT	GGTAATGGTC	AGGGTTCAAC	TACACGAATG	GACCATGAAA	900
CAGCCAGTGT	TTTGAGTTCT	AGTAGCACAC	ACTCTGCACC	TCGAAGGCTG	ACAAGTCATC	960
TGGGAACCAA	GGTGGAAATG	GTGTATTCAT	TGTTGTCAAT	GCTTGGTACT	CATGATAAGG	1020
ATGATATGTC	GCGAACTTTG	CTAGCTATGT	CTAGCTCCCA	AGACAGCTGT	ATATCCATGC	1080
GACAGTCTGG	ATGTCTTCCT	CTCCTCATCC	AGCTTTTACA	TGGCAATGAC	AAAGACTCTG	1140
TATTGTTGGG	AAATTCCCGG	GGCAGTAAAG	AGGCTCGGGC	CAGGGCCAGT	GCAGCACTCC	1200
ACAACATCAT	TCACTCACAG	CCTGATGACA	AGAGAGGCAG	GCGTGAAATC	CGAGTCCTTC	1260
ATCTTTTGGA	ACAGATACGC	GCTTACTGTG	AAACCTGITG	GGAGTGGCAG	GAAGCTCATG	1320
AACCAGGCAT	GGACCAGGAC	AAAAATCCAA	TGCCAGCTCC	TGTTGAACAT	CAGATCTGTC	1380
CTGCTGTGTG	TGTTCTAATG	AAACTTTCAT	TTGATGAAGA	GCATAGACAT	GCAATGAATG	1440
AACTAGGGGG	ACTACAGGCC	ATTGCAGAAT	TATTGCAAGT	GGACTGTGAA	ATGTACGGGC	1500
TTACTAATGA	CCACTACAGT	ATTACACTAA	GACGATAIGC	TGGAATGGCT	TTGACAAACT	1560
TGACTTTTGG	AGATGTAGCC	AACAAGGCTA	CGCTATGCTC	TATGAAAGGC	TGCATGAGAG	1620
CACTTGTGGC	CCAATAAAA	TCTGAAAGTG	AAGACTTACA	GCAGGTTATT	GCAAGTGTTT	1680
TGAGGAATTT	GTCTTGGCGA	GCAGATGTAA	ATAGTAAAAA	GACGTTGCGA	GAAGTTGGAA	1740
GTGTGAAAGC	ATTGATGGAA	TGTGCTTTAG	AAGTTAAAAA	GGAATCAACC	CTCAAAAGCG	1800
TATTGAGTGC	CTTATGGAAT	TTGTCAGCAC	ATTGCACTGA	GAATAAAGCT	GATATATGTG	1860
CTGTAGATGG	TGCACTTGCA	TTTTTGGTTG	GCACTCTTAC	TTACCGGAGC	CAGACAAACA	1920
CTTTAGCCAT	TATTGAAAGT	GGAGGTGGGA	TATTACGGAA	TGTGTCCAGC	TTGATAGCTA	1980
CAAATGAGGA	CCACAGGCAA	ATCCTAAGAG	AGAACAATA	TCTACAAACT	TTATTACAAC	2040
ACTTAAAATC	TCATAGTTTG	ACAATAGTCA	GTAATGCATG	TGGAACTTTG	TGGAATCTCT	2100
CAGCAAGAAA	TCCTAAAGAC	CAGGAAGCAT	TATGGGACAT	GGGGGCAGTT	AGCATGCTCA	2160
AGAACCTCAT	TCATTCAAAG	CACAAAATGA	TTGCTATGGG	AAGTGCTGCA	GCTTTAAGGA	2220
ATCTCATGGC	AAATAGGCCT	GCGAAGTACA	AGGATGCCAA	TATTATGTCT	CCTGGCTCAA	2280
GCTTGCCATC	TCTTCATGTT	AGGAAACAAA	AAGCCCTAGA	AGCAGAATTA	GATGCTCAGC	2340
ACTTATCAGA	AACTTTTGAC	AATATAGACA	ATTTAAGTCC	CAAGGCATCT	CATCGTAGTA	2400

AGCAGAGACA	CAAGCAAAGT	CTCTATGGTG	ATTATGTTTT	TGACACCAAT	CGACATGATG	2460
ATAATAGGTC	AGACAATTTT	AATACTGGCA	ACATGACTGT	CCTTTCACCA	TATTTGAATA	2520
CTACAGTGTT	ACCCAGCTCC	TCTTCATCAA	GAGGAAGCTT	AGATAGTTCT	CGTTCTGAAA	2580
AAGATAGAAG	TTTGGAGAGA	GAACGCGGAA	TTGGTCTAGG	CAACTACCAT	CCAGCAACAG	2640
AAAATCCAGG	AACTTCTTCA	AAGCGAGGTT	TGCAGATCTC	CACCACTGCA	GCCCAGATTG	2700
CCAAAGTCAT	GGAAGAAGTG	TCAGCCATTC	ATACCTCTCA	GGAAGACAGA	AGTTCTGGGT	2760
CTACCACTGA	ATTACATTGT	GTGACAGATG	AGAGAAATGC	ACTTAGAAGA	AGCTCTGCTG	2820
CCCATACACA	TTCAAACACT	TACAATTTCA	CTAAGTCGGA	AAATTCAAAT	AGGACATGTT	2880
CTATGCCTTA	TGCCAAATTA	GAATACAAGA	GATCTTCAAA	TGATAGTTTA	AATAGTGTCA	2940
GTAGTAGTGA	TGGTTATGGT	AAAAGAGGTC	AAATGAAACC	CTCGATTGAA	TCCTATTCTG	3000
AAGATGATGA	AAGTAAGTTT	TGCAGTTATG	GTCAATACCC	AGCCGACCTA	GCCCATAAAA	3060
TACATAGTGC	AAATCATATG	GATGATAATG	ATGGAGAACT	AGATACACCA	ATAAATTATA	3120
GTCTTAAATA	TTCAGATGAG	CAGTTGAACT	CTGGAAGGCA	AAGTCCTTCA	CAGAATGAAA	3180
GATGGGCAAG	ACCCAAACAC	ATAATAGAAG	ATGAAATAAA	ACAAAGTGAG	CAAAGACAAT	3240
CAAGGAATCA	AAGTACAAC	TATCCTGTTT	ATACTGAGAG	CACTGATGAT	AAACACCTCA	3300
AGTTCCAACC	ACATTTTGGA	CAGCAGGAAT	GTGTTTCTCC	ATACAGGTCA	CGGGGAGCCA	3360
ATGGTTCAGA	AACAAATCGA	GTGGGTTCTA	ATCATGGAAT	TAATCAAAAT	GTAAGCCAGT	3420
CTTTGTGTCA	AGAAGATGAC	TATGAAGATG	ATAAGCCTAC	CAATTATAGT	GAACGTTACT	3480
CTGAAGAAGA	ACAGCATGAA	GAAGAAGAGA	GACCAACAAA	TTATAGCATA	AAATATAATG	3540
AAGAGAAACG	TCATGTGGAT	CAGCCTATTG	ATTATAGTTT	AAAATATGCC	ACAGATATTC	3600
CTTCATCACA	GAAACAGTCA	TTTTCATTTCT	CAAAGAGTTC	ATCTGGACAA	AGCAGTAAAA	3660
CCGAACATAT	GTCTTCAAGC	AGTGAGAATA	CGTCCACACC	TTCATCTAAT	GCCAAGAGGC	3720
AGAATCAGCT	CCATCCAAGT	TCTGCACAGA	GTAGAAGTGG	TCAGCCTCAA	AAGGCTGCCA	3780
CTTGCAAAGT	TTCTTCTATT	AACCAAGAAA	CAATACAGAC	TTATTGTGTA	GAAGATACTC	3840
CAATATGTTT	TTCAAGATGT	AGTTCATTAT	CATCTTTGTC	ATCAGCTGAA	GATGAAATAG	3900
GATGTAATCA	GACGACACAG	GAAGCAGATT	CTGCTAATAC	CCTGCAAATA	GCAGAAATAA	3960
AAGAAAAGAT	TGGAAC TAGG	TCAGCTGAAG	ATCCTGTGAG	CGAAGTTCCA	GCAGTGTAC	4020
AGCACCCTAG	AACCAAATCC	AGCAGACTGC	AGGGTTCTAG	TTTATCTTCA	GAATCAGCCA	4080
GGCACAAAGC	TGTTGAATTT	TCTTCAGGAG	CGAAATCTCC	CTCCAAAAGT	GGTGCTCAGA	4140
CACCCAAAAG	TCCACCTGAA	CACTATGTTC	AGGAGACCCC	ACTCATGTTT	AGCAGATGTA	4200

CTTCTGTCAG	TTCACTTGAT	AGTTTTGAGA	GTCGTTCGAT	TGCCAGCTCC	G TTCAGAGTG	4260
AACCATGCAG	TGGAATGGTA	AGTGGCATT	TAAGCCCCAG	TGATCTTCCA	GATAGCCCTG	4320
GACAAACCAT	GCCACCAAGC	AGAAGTAAAA	CACCTCCACC	ACCTCCTCAA	ACAGCTCAAA	4380
CCAAGCGAGA	AGTACCTAAA	AATAAAGCAC	CTACTGCTGA	AAAGAGAGAG	AGTGGACCTA	4440
AGCAAGCTGC	AGTAAATGCT	GCAGTTCAGA	GGGTCCAGGT	TCTTCCAGAT	GCTGATACTT	4500
TATTACATTT	TGCCACGGAA	AGTACTCCAG	ATGGATTTTC	TTGTTTCATCC	AGCCTGAGTG	4560
CTCTGAGCCT	CGATGAGCCA	TTTATACAGA	AAGATGTGGA	ATTAAGAATA	ATGCCTCCAG	4620
TTCAGGAAAA	TGACAATGGG	AATGAAACAG	AATCAGAGCA	GCCTAAAGAA	TCAAATGAAA	4680
ACCAAGAGAA	AGAGGCAGAA	AAAAC TATTG	ATTCTGAAAA	GGACCTATTA	GATGATTCAG	4740
ATGATGATGA	TATTGAAATA	CTAGAAGAAT	GTATTATTTT	TGCCATGCCA	ACAAAGTCAT	4800
CACGTAAAGC	AAAAAAGCCA	GCCCAGACTG	CTTCAAAATT	ACCTCCACCT	GTGGCAAGGA	4860
AACCAAGTCA	GCTGCCTGTG	TACAAACTTC	TACCATCACA	AAACAGGTTG	CAACCCCAAA	4920
AGCATGTTAG	TTTTACACCG	GGGGATGATA	TGCCACGGGT	GTATTGTGTT	GAAGGGACAC	4980
CTATAAACTT	TTCCACAGCT	ACATCTCTAA	GTGATCTAAC	AATCGAATCC	CCTCCAAATG	5040
AGTTAGCTGC	TGGAGAAGGA	GTTAGAGGAG	GAGCACAGTC	AGGTGAATTT	GAAAAACGAG	5100
ATACCATTCC	TACAGAAGGC	AGAAGTACAG	ATGAGGCTCA	AGGAGGAAAA	ACCTCATCTG	5160
TAACCATACC	TGAATTGGAT	GACAATAAAG	CAGAGGAAGG	TGATATTCTT	GCAGAATGCA	5220
TTAATTCTGC	TATGCCCAAA	GGGAAAAGTC	ACAAGCCTTT	CCGTGTGAAA	AAGATAATGG	5280
ACCAGGTCCA	GCAAGCATCT	GCGTCGTCTT	CTGCACCCAA	CAAAAATCAG	TTAGATGGTA	5340
AGAAAAAGAA	ACCAACTTCA	CCAGTAAAAC	CTATACCACA	AAATACTGAA	TATAGGACAC	5400
GTGTAAGAAA	AAATGCAGAC	TCAAAAAATA	ATTTAAATGC	TGAGAGAGTT	TTCTCAGACA	5460
ACAAAGATTC	AAAGAAACAG	AATTTGAAAA	ATAATTCCAA	GGACTTCAAT	GATAAGCTCC	5520
CAAATAATGA	AGATAGAGTC	AGAGGAAGTT	TTGCTTTTGA	TTCACCTCAT	CATTACACGC	5580
CTATTGAAGG	AACTCCTTAC	TGTTTTTTCAC	GAAATGATTC	TTTGAGTTCT	CTAGATTTTG	5640
ATGATGATGA	TGTTGACCTT	TCCAGGGAAA	AGGCTGAATT	AAGAAAGGCA	AAAGAAAATA	5700
AGGAATCAGA	GGCTAAAGTT	ACCAGCCACA	CAGAACTAAC	CTCCAACCAA	CAATCAGCTA	5760
ATAAGACACA	AGCTATTGCA	AAGCAGCCAA	TAAATCGAGG	TCAGCCTAAA	CCCATACTTC	5820
AGAAACAATC	CAC TTTTCCC	CAGTCATCCA	AAGACATACC	AGACAGAGGG	GCAGCAACTG	5880
ATGAAAAGTT	ACAGAATTTT	GCTATTGAAA	ATACTCCAGT	TTGCTTTTCT	CATAATTCCT	5940
CTCTGAGTTC	TCTCAGTGAC	ATTGACCAAG	AAAACAACAA	TAAAGAAAAT	GAACCTATCA	6000
AAGAGACTGA	GCCCCCTGAC	TCACAGGGAG	AACCAAGTAA	ACCTCAAGCA	TCAGGCTATG	6060

CTCCTAAATC	ATTTTCATGTT	GAAGATACCC	CAGTTTGTTT	CTCAAGAAAC	AGTTCTCTCA	6120
GTTCTCTTAG	TATTGACTCT	GAAGATGACC	TGTTGCAGGA	ATGTATAAGC	TCCGCAATGC	6180
CAAAAAAGAA	AAAGCCTTCA	AGACTCAAGG	GTGATAATGA	AAAACATAGT	CCCAGAAATA	6240
TGGGTGGCAT	ATTAGGTGAA	GATCTGACAC	TTGATTTGAA	AGATATACAG	AGACCAGATT	6300
CAGAACATGG	TCTATCCCCT	GATTCAGAAA	ATTTTGATTG	GAAAGCTATT	CAGGAAGGTG	6360
CAAATTCAT	AGTAAGTAGT	TTACATCAAG	CTGCTGCTGC	TGCATGTTTA	TCTAGACAAG	6420
CTTCGTCTGA	TTCAGATTCC	ATCCTTTCCC	TGAAATCAGG	AATCTCTCTG	GGATCACCAT	6480
TTCATCTTAC	ACCTGATCAA	GAAGAAAAAC	CCTTTACAAG	TAATAAAGGC	CCACGAATTC	6540
TAAAACCAGG	GGAGAAAAGT	ACATTGGAAA	CTAAAAAGAT	AGAATCTGAA	AGTAAAGGAA	6600
TCAAAGGAGG	AAAAAAAGTT	TATAAAAGTT	TGATTACTGG	AAAAGTTCGA	TCTAATTCAG	6660
AAATTTTCAGG	CCAAATGAAA	CAGCCCCTTC	AAGCAAACAT	GCCTTCAATC	TCTCGAGGCA	6720
GGACAATGAT	TCATATTCCA	GGAGTTCGAA	ATAGCTCCTC	AAGTACAAGT	CCTGTTTCTA	6780
AAAAAGGCCC	ACCCCTTAAG	ACTCCAGCCT	CCAAAAGCCC	TAGTGAAGGT	CAAACAGCCA	6840
CCACTTCTCC	TAGAGGAGCC	AAGCCATCTG	TGAAATCAGA	ATTAAGCCCT	GTTGCCAGGC	6900
AGACATCCCA	AATAGGTGGG	TCAAGTAAAG	CACCTTCTAG	ATCAGGATCT	AGAGATTCTGA	6960
CCCCTTCAAG	ACCTGCCCAG	CAACCATTAA	GTAGACCTAT	ACAGTCTCCT	GGCCGAAACT	7020
CAATTTCCCC	TGGTAGAAAT	GGAATAAGTC	CTCCTAACAA	ATTATCTCAA	CTTCCAAGGA	7080
CATCATCCCC	TAGTACTGCT	TCAACTAAGT	CCTCAGGTTC	TGGAAAAATG	TCATATACAT	7140
CTCCAGGTAG	ACAGATGAGC	CAACAGAACC	TTACCAAACA	AACAGGTTTA	TCCAAGAATG	7200
CCAGTAGTAT	TCCAAGAAGT	GAGTCTGCCT	CCAAAGGACT	AAATCAGATG	AATAATGGTA	7260
ATGGAGCCAA	TAAAAAGGTA	GAACTTTCTA	GAATGTCTTC	AACTAAATCA	AGTGGAAGTG	7320
AATCTGATAG	ATCAGAAAGA	CCTGTATTAG	TACGCCAGTC	AACTTTCATC	AAAGAAGCTC	7380
CAAGCCCAAC	CTTAAGAAGA	AAATTGGAGG	AATCTGCTTC	ATTTGAATCT	CTTTCTCCAT	7440
CATCTAGACC	AGCTTCTCCC	ACTAGGTCCC	AGGCACAAAC	TCCAGTTTTA	AGTCCTTCCC	7500
TTCCTGATAT	GTCTCTATCC	ACACATTTCGT	CTGTTTCAGGC	TGGTGGATGG	CGAAAACCTCC	7560
CACCTAATCT	CAGTCCCCT	ATAGAGTATA	ATGATGGAAG	ACCAGCAAAG	CGCCATGATA	7620
TTGCACGGTC	TCATTCTGAA	AGTCCTTCTA	GACTTCCAAT	CAATAGGTCA	GGAACCTGGA	7680
AACGTGAGCA	CAGCAAACAT	TCATCATCCC	TTCCTCGAGT	AAGCACTTGG	AGAAGAACTG	7740
GAAGTTCATC	TTCAATTCTT	TCTGCTTCAT	CAGAATCCAG	TGAAAAAGCA	AAAAGTGAGG	7800
ATGAAAAACA	TGTGAACTCT	ATTTTCAGGAA	CCAAACAAAG	TAAAGAAAAC	CAAGTATCCG	7860

CAAAAGGAAC	ATGGAGAAAA	ATAAAAGAAA	ATGAATTTTC	TCCCACAAAT	AGTACTTCTC	7920
AGACCGTTTC	CTCAGGTGCT	ACAAATGGTG	CTGAATCAAA	GACTCTAATT	TATCAAATGG	7980
CACCTGCTGT	TTCTAAAACA	GAGGATGTTT	GGGTGAGAAT	TGAGGACTGT	CCCATTAACA	8040
ATCCTAGATC	TGGAAGATCT	CCCACAGGTA	ATACTCCCCC	GGTGATTGAC	AGTGTTTCAG	8100
AAAAGGCAAA	TCCAAACATT	AAAGATTCAA	AAGATAATCA	GGCAAAACAA	AATGTGGGTA	8160
ATGGCAGTGT	TCCCATGCGT	ACCGTGGGTT	TGGAAAAICG	CCTGAACTCC	TTTATTCAGG	8220
TGGATGCCCC	TGACCAAAAA	GGAAGTGAAG	TAAAACCAGG	ACAAAATAAT	CCTGTCCCTG	8280
TATCAGAGAC	TAATGAAAGT	TCTATAGTGG	AACGTACCCC	ATTCAGTTCT	AGCAGCTCAA	8340
GCAAACACAG	TTCACCTAGT	GGGACTGTTG	CTGCCAGAGT	GACTCCTTTT	AATTACAACC	8400
CAAGCCCTAG	GAAAAGCAGC	GCAGATAGCA	CTTCAGCTCG	GCCATCTCAG	ATCCCAACTC	8460
CAGTGAATAA	CAACACAAAG	AAGCGAGATT	CCAAAACCTGA	CAGCACAGAA	TCCAGTGGAA	8520
CCCAAAGTCC	TAAGCGCCAT	TCTGGGTCTT	ACCTTGTTGAC	ATCTGTTTAA	AAGAGAGGAA	8580
GAATGAAACT	AAGAAAATTC	TATGTTAATT	ACAACTGCTA	TATAGACATT	TTGTTTCAAA	8640
TGAAACTTTA	AAAGACTGAA	AAATTTTGTA	AATAGGTTTG	ATTCTTGTTA	GAGGGTTTTT	8700
GTTCTGGAAG	CCATATTTGA	TAGTATACTT	TGTCTTCACT	GGTCTTATTT	TGGGAGGCAC	8760
TCTTGATGGT	TAGGAAAAAA	ATAGTAAAGC	CAAGTATGTT	TGTACAGTAT	GTTTTACATG	8820
TATTTAAAGT	AGCATCCCAT	CCCAACTTCC	TTTAATTATT	GCTTGTCTTA	AAATAATGAA	8880
CACTACAGAT	AGAAAATATG	ATATATTGCT	GTTATCAATC	ATTTCTAGAT	TATAAACTGA	8940
CTAAACTTAC	ATCAGGGAAA	AATTGGTATT	TATGCAAAAA	AAAATGTTTT	TGTCCTTGTC	9000
AGTCCATCTA	ACATCATAAT	TAATCATGTG	GCTGTGAAAT	TCACAGTAAT	ATGGTTCCCG	9060
ATGAACAAGC	TTTACCCAGC	CTGTTTGCTT	TACTGCATGA	ATGAAACTGA	TGGTTCAATT	9120
TCAGAAGTAA	TGATTAACAG	TTATGTGGTC	ACATGATGTG	CATAGAGATA	GCTACAGTGT	9180
AATAATTTAC	ACTATTTTGT	GCTCCAAACA	AAACAAAAAT	CTGTGTAACT	GTAAAACATT	9240
GAATGAAACT	ATTTTACCTG	AACTAGATTT	TATCTGAAAG	TAGGTAGAAT	TTTTGCTATG	9300
CTGTAATTTG	TTGTATATTC	TGGTATTTGA	GGTGAGATGG	CTGCTCTTTT	ATTAATGAGA	9360
CATGAATTGT	GTCTCAACAG	AACTAAATG	AACATTTTCAG	AATAAATTAT	TGCTGTATGT	9420
AACTGTTTAC	TGAAATTGGT	ATTTGTTTGA	AGGGTCTTGT	TTCACATTTG	TATTAATAAT	9480
TGTTTAAAAT	GCCTCTTTTA	AAAGCTTATA	TAAATTTTTT	NCTTCAGCTT	CTATGCATTA	9540
AGAGTAAAAT	TCCTCTTACT	GTAATAAAAA	CAATTGAAGA	AGACTGTTGC	CACTTAACCA	9600
TTCCATGCGT	TGGCACTTAT	CTATTCCTGA	AATTCCTTTA	TGTGATTAGC	TCATCTTGAT	9660
TTTTAACATT	TTTCCACTTA	AACTTTTTTT	TCTTACTCCA	CTGGAGCTCA	GTAAAAGTAA	9720

```

ATTCATGTAA TAGCAATGCA AGCAGCCTAG CACAGACTAA GCATTGAGCA TAATAGGCCC      9780
ACATAATTTT CTCTTTCTTA ATATTATAGA AATTCTGTAC TTGAAATTGA TTCTTAGACA      9840
TTGCAGTCTC TTCGAGGCTT TACAGTGTA ACTGTCTTGC CCCTTCATCT TCTTGTTGCA      9900
ACTGGGTCTG ACATGAACAC TTTTATCAC CCTGTATGTT AGGGCAAGAT CTCAGCAGTG      9960
AAGTATAATC AGCACTTTGC CATGCTCAGA AAATTCAAAT CACATGGAAC TTTAGAGGTA     10020
GATTTAATAC GATTAAGATA TTCAGAAGTA TATTTTAGAA TCCCTGCCTG TTAAGGAAAC     10080
TTTATTTGTG GTAGGTACAG TTCTGGGGTA CATGTAAAGT GTCCCCTTAT ACAGTGGAGG     10140
GAAGTCTTCC TTCCTGAAGG AAAATAAACT GACACTTATT AACTAAGATA ATTTACTTAA     10200
TATATCTTCC CTGATTTGTT TTAAAAGATC AGAGGGTGAC TGATGATACA TGCATACATA     10260
TTTGTTGAAT AAATGAAAAT TTATTTTGTAG TGATAAGATT CATACTCTCT GTATTTGGGG     10320
AGAGAAAACC TTTTAAAGCA TGGTGGGGCA CTCAGATAGG AGTGAATACA CCTACCTGGT     10380
GGTCAT                                           10386

```

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7273 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 1..7273
- (D) OTHER INFORMATION: /standard_name= "Homo sapiens cyclooxygenase (COX-2) gene, promoter and exon 1"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

```

GGTACCCAGG CTGGAGTGCA CTGGTGTGAT CATAGCTCAC TAACCTCGAA CTCCTGGGCT      60
TAGGCAATCC TCTTGCCTTG GCCTCCCAA GTGCCAGGAT TACAGGCATG AGCCACCACA     120
GTGGAGCTCT CAATTCTGAT ACTAATAATT TGTGTCTTCT CTTTTTTTCC TTAGCCTGAC     180
TAGAGTAATT AACTTTATGT CTTTAAAG AACCACCTTT TTGGTTTAC CCATTTTCTT      240
TTTTGATTTT CTGTTTTTGA TTTGATTGAT ATCTACTCTA ATTTTTTATT ATTTCTTTTC     300
CTCTGCTTAC TTTGAATTTA ATTACTTTTC TTTTTGTAG TCTCCTAAAA TAGAAGCTTA     360
TATTATTGAT TTTAGATCTT TCTTCTTTTC TATTACAGCA CTCAATGCTA TAAATTTCCC     420

```

TCTAAGCATT GCTTTCACTG CATCCTACAA TATTTCAACT CTATTGTTAT TTAGCTCAAA	480
AGAGGTTCTT AATTTCTATT GGGATTTCTC TTTGACCCAT GTGTTATTCA GAAGTGTTC	540
GTGTGATCTC CAAATATTTG GGAGTTTTTC AGCTATCTTT CTATTAATCA TTTCTTGTTT	600
AATTCTATTG TGGCCTGAGA GCATATATTG TATGATTTAT ATTCTTGTA ATGTGTTAAG	660
GTGTGTCTTA TGGTGCAGAA TCGGGTTTAT CTTGCTATAT GTTCCTTAGA GAATAATGTA	720
TGTTCTGCTG TTATTGGATA AAGTAGTCTA TAGATGTCAG TTACATCTCG TTGATTAATG	780
GTGCTGTTGA GTTCAGCTAT GTCCTAAATG ATTTTCTGTC TGCTGTATCT GTCTATTTCT	840
GACACAAGGC TGTGAAGTC TCCAACCATA ATAATGAATT AATCTATTTT TCTTTGCAGT	900
TTTATCAATT TTGTCTTATA TATATTGATG CTCCATTGTT TGGCACATAC ACATTAAGAA	960
TTGTTATGTC TTCTTGGAGA ATTTACCTTT CCATAACATG TAACATTTCC CTTTATTCCT	1020
GATAATTTTT CTTGCTCAAA AGTTTGCCCT GTTGGAATT ACCAGAACTA CTCTGGCTTT	1080
ATTTGATTAG TGTTAGCATG CTCTCTCTTT CTCTATTCTT AACTTTTTAA TGTATACTTG	1140
ACTTTGTATT TAAAGTGGGG TTCTTATAGA AAACATATAC TTGGTAGGGT GGAAGTAAA	1200
ATAAAAAGAA ATACTTGGGT ATTGGTTTGA TCCACTCIAA CAATCTCTAT GTTTTAATTG	1260
ATGTATTTAG ACCATTGATA CTTATTTTTT TATCCTCATC CCTGTGATTA CCCAGAGAGC	1320
TGCTTAAATT GATTATTGAT ATAGACAAAT TAATAATTAA TATCTACCGT TTGTTACTGT	1380
TTTCTATTTT TCATTGCCCT TACTTTCTGC TCCTATTTTT TGCTCCTTTT TCTGTTAATT	1440
TAGGTTTTGA GTTATTTTAT ATCATTCTAT TTTCTCTCCC TTCTCAGCAT ATGAATTATC	1500
TTTCTTTTTG ACTTTTTTAG TGGCTGCCCT GAAGGTTGCA ATGTACATTT ACAACCAGTC	1560
CCAATCTCCT TTCAAAAAC ACAATACTGT TTCATGGCTA GTGCAAGTAC CTAATAATAA	1620
GAAGTCACTC CTAATTTCTT TCTCTCATTC TTTGTATCTT TACTGTTATT CATTTCACTT	1680
GTACATAAGC TGTAATCTTT CAATACATTA TTGCTATTAT TATTTCAAAA CATGTTATCT	1740
ATTATATCTA TTAAATAA GAAAATAGG CCAGGTGCAG TGGCTTACTC ATGTAATCCC	1800
AGCACTTTGG GAGACCGATG GATTGCTAGA GCTCAGGAAT TCGAGACCAG CCTGGGCAAC	1860
ATAGTGAAAC CCTGTCTCTA CTAATAATAC AAAAAAAAAA ATTGCTGGGC ATGGTGGCAT	1920
GGCCTGTGG TCACAGCTAC TCGGGAGGCT GAGGTGAGAG GATTGCTTGA GCCTGGGAGG	1980
CAGAGGTTGC AGTGAACCAA AATCAAGCTA CTGCACTCCA GCCTAAGTGA CAGAGTGAGA	2040
CCCTGTCTCA AAAAAAAAAA GAAAGAATTA TTTTATTTA TCTTCACTTA TTTCTTCTCT	2100
AATGCTCTTT GTTTCTTTAG TATGTAGATC CAAGTTTCTA ACCTGTATCA TTTTCTTAT	2160
CTCAATAACT TCTTTTAACA TTTCTCACAA AGCAGATCTA CTGGCCACAG AATGCCTCAA	2220

TTTTCATTTG	TCTGAGAAAA	CCTTATTTCT	CCTTCACTTT	TGAAAGATAA	TTTTGTAGGG	2280
TACAGAATTC	TAGGTTGTAG	GTTTTTTCCC	CTCAAAGTGA	AATATTTTCAT	TCCACTCTTT	2340
TCTTCTTTGT	ATGGTATCTG	AGAAGAAGTC	AGATGTAATT	CTTATCATTA	TTACTTAAAA	2400
GATTGCTTCT	GTTCCTTTCT	CTCTTCTCCT	TCCCTTCTTT	CCTTCTCTGT	ATATTACACC	2460
TTTTATAGTT	GCCCCATATT	TCTTAGATAT	TATGTTTTGG	TTTTCTTCTG	TGTTTTTTTC	2520
TTTGATTCTC	AGTTTTAGAA	GTCTCTATTT	ATATATCTGC	AATCGCAGGG	ATTCTTTCCT	2580
CTGCCATGTC	CAGTCTACTA	ATAAGCCCTT	ACAGACATTG	TTGACTTCTG	TTCCAGTGTT	2640
TTTGATCTCT	AGCATTTCTC	TGATTATTTT	TTGGAATTGC	CATCTGTCTA	CTTACATTAC	2700
CAACCTATTC	TTGTGTGTTG	TCTTATCATA	GTAATTGCAG	TTGTTTTAAT	TTCATAGGTA	2760
TTGTAATTTT	AACATCTCTA	CCATATTTGA	CATTGATTCT	GATGCTTGCT	CTGTCTTATC	2820
AAGCTATGTT	TTTGTCTTTT	AGTGTGACTT	CTAATTTTTT	GTTGAAAGCC	AGGCATGATG	2880
TACTGAGTGA	AAGAAACTCA	ATACATTGTA	ATGTGACGAT	AAGAGTTCAG	GGGAAGTGAA	2940
GCATTCTATA	GTCCTATAGC	AGGTCTCGGC	CTTTTAGTGA	GCCTGTGCCT	ATGAACGGTG	3000
ACTTTCAACA	AGTGCTTTTC	ATTCCACTCT	TTTCCTGTCC	TTAAGTGGGA	CAAGATCACT	3060
GGGGGGGGGC	TAGAATTGGG	TATTTCCCTT	CTCCAATGTA	GAAGCTAAAG	AGAGGGCTGG	3120
AGTTGGGTAT	TTTTCTTCCC	CTGTATGGAA	AGCTAGAGGC	AGTTAAATTT	GGATATTTTC	3180
CTTCTTCTAA	TTCAGTTAGG	CTGCGACAAA	AATCCCGACA	GTTTAGGCTC	TAATATTATA	3240
AAATAATTTT	TCTTGAGTAT	AGGCCTTATT	AAGAACACTA	TACTCTGATG	GAGCTGAGGG	3300
GGAGTTTTCT	CTGATATTCA	CTGCGAGAAC	CTCGTAGAGC	TCCAGGAAGC	AAAAC TCACA	3360
AAAGTGTGGG	AGTCTTCCAG	AATTTTTTCT	TTGCAGACTT	ATCTGCACTG	AACCTCCAGA	3420
AATTCATCAA	TTACAGTTCA	GGTTTTCTTA	CCCAGGTACT	GGTTTTCATG	GAGGTTTCTG	3480
CCTGTGCATT	TCTGCTCCAG	TAAGTTGTTC	TTCTTGATATG	GTCTGTCTTT	CAAATTTTTT	3540
AAGTAGGGTT	ATGACCTGTC	GCCTCACTTC	TCTGACAGTT	CTGAGAGTGT	TGATTTTTCA	3600
GTTTGCTTAG	ATTTTTACTT	GTTTTTAGGA	TGAAGTGACA	ATTTCCAAGC	TCCTCCCTGA	3660
CATGCCAGAT	CAGAAACTGA	AAGTCCTAAG	CCTCATATTC	TGTGCGTGGG	TATGTT CACA	3720
TCCTGCCTGC	TCCAGTGCCC	CCACCTCACA	CTCTCTTTCC	CTTCCTTGTC	CCCTTG TGAG	3780
ATTTCTAGGT	CCAATACAAA	GACTGTGTTC	AACTCATTCA	ACTACTTGGC	TCATCTGAGT	3840
ATTATAATGA	ACAATCACAA	AAAAAAATGA	AGTAAAAGAA	AAATCCATCA	AAGAATTGAG	3900
ATATTTGAGA	AAAAGAAAGG	AGATCAGTGT	TTTATAAAAC	TTAGAAATAG	ATTTTTTAAG	3960
TGTTTCTTCA	TTGACTTATG	TGAAAGGACT	TTTCTTAATT	TAACAAATTA	TGTGCTTTTCG	4020
TTTATAGCCT	CAAAACTTCT	TGTGTAGCTA	AGAATGGGTA	AATAATCAGG	CTTTACTAAA	4080

GGACTAACGT AAAGATCTTC TGTAAGTAAC ATTTCTGCTA CTCAAGGAAG AGATAAACTT	4140
CATGGCATAA CCTTGCCAAA GTATACTAAG AATAACCCTG ACACAAAGCT CTTTTTTCAG	4200
CCAACATGCC ATGAAAGAAA GAAGACAAGG GGTGATCTCC ACTCTCTAAG TGAACCACTA	4260
AACCCACCAA AGAAGAAACG AGGGAAATAG AAAGAGGACC CTTGCCTGAG ATAATGGATC	4320
TGTATGTATG AGTAGTAGAA CCCTGCTCAA AGTACAAGGA AGGGAAAAAA AAGTTAGTTT	4380
ATTTGGAATT TTGGACATTA AGAGTCTTTA TTGTTCAATT TCTTTTAACT CACATGAATG	4440
GCTTATCACT TCAATTAATA AATATTTTCAT TTCTTTTCAA TCTATATTCA TGAAACAAAT	4500
CTGAAATGAA CAGTGCAACA TGTGAATGTT TAGAACATTA TAAAATTAAA CACAAAATCT	4560
GTCTGGCAAT CTTCTAGCA TCTTAGGAAA AAAGTTGACA AAATTTCAAG CAGCAGAAGG	4620
GGGCAGTAAA ACTCAACAGA AAGCTCTGGA AGATTTTTAA GATTCTTCCT TATTTTCTTT	4680
TCATGTAGAG TATTTCCCAA CAAATTTTCAG ACGCTAATAG AAATTTTGTA CAACAGATCC	4740
ATATATTTGC CTAAAATAGA CACAGAAACA TTGATATATG CAAACATGAG AGCTATAAGT	4800
TTTACATGAT CAAAACCTTT TTTTATGGT ACACAATAGT CACAGTACTT TTCCATATAA	4860
AACAGGTTTA GTGGTCTTAA TTTAGTTTGG CACATTTAAT ACACTCCCAT GACCAGCATC	4920
CCAAATGTAC CTATCCGTTT TATTTIATTG TCTCAGAATT GTCAGTTATT TAATAAATTA	4980
TGTAACTTTT TTCCTTATGC TCAGATTGTC ACTTCTTTCT AAAACTCTGC CCATCCTTAA	5040
AGTCCCAGAT TCTCCTTGAA CTTTTTTTTT TGACTTTCCA AGTACATGGA ACTCTTCACT	5100
CTATCCTGCT ATATAAGTGA CAGAATTTCC ACTATGGGAT AGATGGAGTT CAATTCCTTT	5160
GAGTTTAAAA TAATCTAAAT ATAATTATTC CTTATGCCCT GTTTTTCCCT CACTTTTGTA	5220
TCCAAATCTC TTTTCAGACA ACAGAACAAT TAATGTCTGA TAAGGAAGAC AATGATGATG	5280
ATCACTTCAA AATAAGCTTG AATTCAGGAT TGTAATGTAA AATTTTAGTA CTCTCTCACA	5340
GTATGGATTC TAACATGGCT TCTAACCCAA ACTAACATTA GTAGCTCTAA CTATAAACTT	5400
CAAATTTTCAG TAGATGCAAC CTACTCCTTT AAAATGAAAC AGAAGATTGA AATTATTAAA	5460
TTATCAAAAA GAAAATGATC CACGCTCTTA GTTGAAATTT CATGTAAGAT TCCATGCAAT	5520
AAATAGGAGT GCCATAAATG GAATGATGAA ATATGACTAG AGGAGGAGAA AGGCTTCCTA	5580
GATGAGATGG AATTTTAGTC ATCCGTGTCT CATGAAGAAT CAGATGTGTA CACTAAGCAA	5640
AACAGTTAAA AAAAAACCT CCAAGTGAGT CTCTTATTTA TTTTTTCTT ATAAGACTTC	5700
TACAAATTGA GGTACCTGGT GTAGTTTTAT TTCAGGTTTT ATGCTGTCAT TTTCCTGTAA	5760
TGCTAAGGAC TTAGGACATA ACTGAATTTT CTATTTTCCA CTTCTTTTCT GGTGTGTGTG	5820
TATATATATA TGTATATATA CACACACACA TATACATATA TATATTTTTA GTATCTCACC	5880

CTCACATGCT CCTCCCTGAG CACTACCCAT GATAGATGTT AAACAAAAGC AAAGATGAAA	5940
TTCCAACGTGT TAAAATCTCC CTTCCATCTA ATTAATTCCT CATCCAATA TGTTCCAAAA	6000
CGAGAATAGA AAATTAGCCC CAATAAGCCC AGGCAACTGA AAAGTAAATG CTATGTTGTA	6060
CTTTGATCCA TGGTCACAAC TCATAATCTT GGAAAAGTGG ACAGAAAAGA CAAAAGAGTG	6120
AACTTTAAAA CTCGAATTTA TTTTACCAGT ATCTCCTATG AAGGGCTAGT AACCAAAATA	6180
ATCCACGCAT CAGGGAGAGA AATGCCTTAA GGCATACGTT TTGGACATTT AGCGTCCCTG	6240
CAAATTCTGG CCATCGCCGC TTCCTTTGTC CATCAGAAGG CAGGAACTT TATATTGGTG	6300
ACCCGTGGAG CTCACATTAA CTATTTACAG GGTAAGTCT TAGGACCAGT ATTATGAGGA	6360
GAATTTACCT TTCCCGCCTC TCTTTCCAAG AAACAAGGAG GGGGTGAAGG TACGGAGAAC	6420
AGTATTTCTT CTGTTGAAAG CAACTTAGCT ACAAAGATAA ATTACAGCTA TGTACACTGA	6480
AGGTAGCTAT TTCATTCCAC AAAATAAGAG TTTTTTAAAA AGCTATGTAT GTATGTCCTG	6540
CATATAGAGC AGATATACAG CCTATTAAGC GTCGTCAC TA AACATAAAA CATGTCAGCC	6600
TTTCTTAACC TTAATCGCCC CAGTCTGTCC CGACGTGACT TCCTCGACCC TCTAAAGACG	6660
TACAGACCAG ACACGGCGGC GGCGGCGGGA GAGGGGATTC CCTGCGCCCC CGGACCTCAG	6720
GGCCGCTCAG ATTCCTGGAG AGGAAGCCAA GTGTCCTTCT GCCCTCCCCC GGTATCCCAT	6780
CCAAGGCGAT CAGTCCAGAA CTGGCTCTCG GAAGCGCTCG GGCAAAGACT GCGAAGAAGA	6840
AAAGACATCT GGCGGAAACC TGTGCGCCTG GGGCGGTGGA ACTCGGGGAG GAGAGGGAGG	6900
GATCAGACAG GAGAGTGGGG ACTACCCCCT CTGCTCCCAA ATTGGGGCAG CTCCTGGGT	6960
TTCCGATTTT CTCATTTCCG TGGGTAAAAA ACCCTGCCCC CACCGGGCTT ACGCAATTTT	7020
TTTAAGGGGA GAGGAGGGAA AAATTTGTGG GGGGTACGAA AAGGCGGAAA GAAACAGTCA	7080
TTTCGTCACA TGGGCTTGGT TTTCAGTCTT ATAAAAAGGA AGGTTCTCTC GGTTAGCGAC	7140
CAATTGTCAT ACGACTTGCA GTGAGCGTCA GGAGCACGTC CAGGAACTCC TCAGCAGCGC	7200
CTCCTTCAGC TCCACAGCCA GACGCCCTCA GACAGCAAAG CCTACCCCCC GCGCCGCGCC	7260
CTGCCCCGAAG CTT	7273

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4465 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION: 1..4465
 (D) OTHER INFORMATION: /standard_name="Homo sapiens prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase) (PTGS2), mRNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NC:62:

CAATTGTCAT ACGACTTGCA GTGAGCGTCA GGAGCACGTC CAGGAACTCC TCAGCAGCGC	60
CTCCTTCAGC TCCACAGCCA GACGCCCTCA GACAGCAAAG CCTACCCCCG CGCCGCGCCC	120
TGCCCCGCCGC TCGGATGCTC GCCCGCGCCC TGCTGCTGTG CGCGGTCCTG GCGCTCAGCC	180
ATACAGCAAA TCCTTGCTGT TCCCACCCAT GTCAAAACCG AGGTGTATGT ATGAGTGTGG	240
GATTTGACCA GTATAAGTGC GATTGTACCC GGACAGGATT CTATGGAGAA AACTGCTCAA	300
CACCGGAATT TTTGACAAGA ATAAAATTAT TTCTGAAACC CACTCCAAAC ACAGTGCACT	360
ACATACTTAC CCACTTCAAG GGATTTTGGA ACGTTGTGAA TAACATTCCC TTCCTTCGAA	420
ATGCAATTAT GAGTTATGTC TTGACATCCA GATCACATTT GATTGACAGT CCACCAACTT	480
ACAATGCTGA CTATGGCTAC AAAAGCTGGG AAGCCTTCTC TAACCTCTCC TATTATACTA	540
GAGCCCTTCC TCCTGTGCCT GATGATTGCC CGACTCCCTT GGGTGTCAAA GGTA AAAAGC	600
AGCTTCCTGA TTCAAATGAG ATTGTGGAAA AATTGCTTCT AAGAAGAAAG TTCATCCCTG	660
ATCCCCAGGG CTCAAACATG ATGTTTGCAT TCTTTGCCCA GCACTTCACG CATCAGTTTT	720
TCAAGACAGA TCATAAGCGA GGGCCAGCTT TCACCAACGG GCTGGGCCAT GGGGTGGACT	780
TAAATCATAT TTACGGTGAA ACTCTGGCTA GACAGCGTAA ACTGCGCCTT TTCAAGGATG	840
GAAAAATGAA ATATCAGATA ATTGATGGAG AGATGTATCC TCCCACAGTC AAAGATACTC	900
AGGCAGAGAT GATCTACCCT CCTCAAGTCC CTGAGCATCT ACGGTTTGCT GTGGGGCAGG	960
AGGTCTTTGG TCTGGTGCCT GGTCTGATGA TGTATGCCAC AATCTGGCTG CGGGAACACA	1020
ACAGAGTATG CGATGTGCTT AAACAGGAGC ATCCTGAATG GGGTGATGAG CAGTTGTTCC	1080
AGACAAGCAG GCTAATACTG ATAGGAGAGA CTATTAAGAT TGTGATTGAA GATTATGTGC	1140
AACACTTGAG TGGCTATCAC TTCAAACCTGA AATTTGACCC AGAACTACTT TTCAACAAAC	1200
AATTCCAGTA CCAAATCGT ATTGCTGCTG AATTTAACAC CCTCTATCAC TGGCATCCCC	1260
TTCTGCCTGA CACCTTTCAA ATTCATGACC AGAAATACAA CTATCAACAG TTTATCTACA	1320
ACAACTCTAT ATTGCTGGAA CATGGAATTA CCCAGTTTGT TGAATCATTC ACCAGGCAAA	1380
TTGCTGGCAG GGTGCTGGT GGTAGGAATG TTCCACCCGC AGTACAGAAA GTATCACAGG	1440
CTTCCATTGA CCAGAGCAGG CAGATGAAAT ACCAGTCTTT TAATGAGTAC CGCAAACGCT	1500
TTATGCTGAA GCCCTATGAA TCATTTGAAG AACTTACAGG AGAAAAGGAA ATGTCTGCAG	1560

AGTTGGAAGC	ACTCTATGGT	GACATCGATG	CTGTGGAGCT	GTATCCTGCC	CTTCTGGTAG	1620
AAAAGCCTCG	GCCAGATGCC	ATCTTTGGTG	AAACCATGGT	AGAAGTTGGA	GCACCATTTCT	1680
CCTTGAAAGG	ACTTATGGGT	AATGTTATAT	GTTCTCCTGC	CTACTGGAAG	CCAAGCACTT	1740
TTGGTGGAGA	AGTGGGTTTT	CAAATCATCA	ACACTGCCTC	AATTCAGTCT	CTCATCTGCA	1800
ATAACGTGAA	GGGCTGTCCC	TTTACTTCAT	TCAGTGTTCC	AGATCCAGAG	CTCATTAATA	1860
CAGTCACCAT	CAATGCAAGT	TCTTCCCGCT	CCGGACTAGA	TGATATCAAT	CCCACAGTAC	1920
TACTAAAAGA	ACGTTCGACT	GAACTGTAGA	AGTCTAATGA	TCATATTTAT	TTATTTATAT	1980
GAACCATGTC	TATTAATTTA	ATTATTTAAT	AATATTTATA	TTAAACTCCT	TATGTTACTT	2040
AACATCTTCT	GTAACAGAAG	TCAGTACTCC	TGTTGCGGAG	AAAGGAGTCA	TACTTGTGAA	2100
GACTTTTATG	TCACTACTCT	AAAGATTTTG	CTGTTGCTGT	TAAGTTTGGA	AAACAGTTTT	2160
TATTCTGTTT	TATAAACCAG	AGAGAAATGA	GTTTTGACGT	CTTTTTACTT	GAATTTCAAC	2220
TTATATTATA	AGAACGAAAG	TAAAGATGTT	TGAATACTTA	AACACTATCA	CAAGATGGCA	2280
AAATGCTGAA	AGTTTTTACA	CTGTCGATGT	TTCCAATGCA	TCTTCCATGA	TGCATTAGAA	2340
GTAAC TAATG	TTTGAAATTT	TAAAGTACTT	TTGGTTATTT	TTCTGTCATC	AAACAAAAC	2400
AGGTATCAGT	GCATTATTAA	ATGAATATTT	AAATTAGACA	TTACCAGTAA	TTTCATGTCT	2460
ACTTTTTAAA	ATCAGCAATG	AAACAATAAT	TTGAAATTTT	TAAATTCATA	GGGTAGAATC	2520
ACCTGTAAAA	GCTTGTTTGA	TTTCTTAAAG	TTATTAAACT	TGTACATATA	CCAAAAAGAA	2580
GCTGTCTTGG	ATTTAAATCT	GTAAAATCAG	ATGAAATTTT	ACTACAATTG	CTTGTTAAAA	2640
TATTTTATAA	GTGATGTTCC	TTTTTCACCA	AGAGTATAAA	CCTTTTTAGT	GTGACTGTTA	2700
AAACTTCCTT	TTAAATCAAA	ATGCCAAATT	TATTAAGGTG	GTGGAGCCAC	TGCAGTGTTA	2760
TCTCAAAATA	AGAATATTTT	GTTGAGATAT	TCCAGAATTT	GTTTATATGG	CTGGTAACAT	2820
GTAAAATCTA	TATCAGCAAA	AGGGTCTACC	TTTAAAATAA	GCAATAACAA	AGAAGAAAAC	2880
CAAATTATTG	TTCAAATTTA	GGTTTAAACT	TTTGAAGCAA	ACTTTTTTTT	ATCCTTGTGC	2940
ACTGCAGGCC	TGGTACTCAG	ATTTTGCTAT	GAGGTTAATG	AAGTACCAAG	CTGTGCTTGA	3000
ATAACGATAT	GTTTTCTCAG	ATTTTCTGTT	GTACAGTTTA	ATTTAGCAGT	CCATATCACA	3060
TTGCAAAAGT	AGCAATGACC	TCATAAAATA	CCTCTTCAAA	ATGCTTAAAT	TCATTTTACA	3120
CATTAATTTT	ATCTCAGTCT	TGAAGCCAAT	TCAGTAGGTG	CATTGGAATC	AAGCCTGGCT	3180
ACCTGCATGC	TGTTCCTTTT	CTTTTCTTCT	TTTAGCCATT	TTGCTAAGAG	ACACAGTCTT	3240
CTCATCACTT	CGTTTCTCCT	ATTTTGTTTT	ACTAGTTTTA	AGATCAGAGT	TCACTTTCTT	3300
TGGACTCTGC	CTATATTTTC	TTACCTGAAC	TTTTGCAAGT	TTTCAGGTAA	ACCTCAGCTC	3360

```

AGGACTGCTA TTTAGCTCCT CTTAAGAAGA TTAAAAGAGA AAAAAAAGG CCCTTTTAAA 3420
AATAGTATAC ACTTATTTTA AGTGAAAAGC AGAGAATTTT ATTTATAGCT AATTTTAGCT 3480
ATCTGTAACC AAGATGGATG CAAAGAGGCT AGTGCCTCAG AGAGAACTGT ACGGGGTTTG 3540
TGACTGGAAA AAGTTACGTT CCCATTCTAA TTAATGCCCT TTCTTATTTA AAAACAAAAC 3600
CAAATGATAT CTAAGTAGTT CTCAGCAATA ATAATAATGA CGATAATACT TCTTTTCCAC 3660
ATCTCATTGT CACTGACATT TAATGGTACT GTATATTACT TAATTTATTG AAGATTATTA 3720
TTTATGTCTT ATTAGGACAC TATGGTTATA AACTGTGTTT AAGCCTACAA TCATTGATTT 3780
TTTTTTGTTA TGTCACAATC AGTATATTTT CTTTGGGGTT ACCTCTCTGA ATATTATGTA 3840
AACAAATCCAA AGAAATGATT GTATTAAGAT TTGTGAATAA ATTTTATAGAA ATCTGATTGG 3900
CATATTGAGA TATTTAAGGT TGAATGTTTG TCCTTAGGAT AGGCCTATGT GCTAGCCCAC 3960
AAAGAATATT GTCTCATTAG CCTGAATGTG CCATAAGACT GACCTTTTAA AATGTTTTGA 4020
GGGATCTGTG GATGCTTCGT TAATTTGTTC AGCCACAATT TATTGAGAAA ATATTCTGTG 4080
TCAAGCACTG TGGGTTTTAA TATTTTAAA TCAAACGCTG ATTACAGATA ATAGTATTTA 4140
TATAAATAAT TGAAAAAAT TTTCTTTTGG GAAGAGGGAG AAAATGAAAT AAATATCATT 4200
AAAGATAACT CAGGAGAATC TTCTTTACAA TTTTACGTTT AGAATGTTTA AGGTTAAGAA 4260
AGAAATAGTC AATATGCTTG TATAAACAC TGTTCCTGT TTTTTTTAAA AAAAAAATT 4320
GATTTGTTAT TAACATTGAT CTGCTGACAA AACCTGGGAA TTTGGGTTGT GTATGCGAAT 4380
GTTTCAGTGC CTCAGACAAA TGTGTATTTA ACTTATGTAA AAGATAAGTC TGGAAATAAA 4440
TGTCTGTTTA TTTTGTACT ATTTA 4465

```

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1336 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 1..1336
- (D) OTHER INFORMATION: /standard_name= "stratifin (SFN),

mRNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

```

GAGAGACACA GAGTCCGGCA TTGGTCCCAG GCAGCAGTTA GCCCGCCGCC CGCCTGTGTG 60

```

```

TCCCCAGAGC CATGGAGAGA GCCAGTCTGA TCCAGAAGGC CAAGCTGGCA GAGCAGGCCG      120
AACGCTATGA GGACATGGCA GCCTTCATGA AAGGCGCCGT GGAGAAGGGC GAGGAGCTCT      180
CCTGCGAAGA GCGAAACCTG CTCTCAGTAG CCTATAAGAA CGTGGTGGGC GGCCAGAGGG      240
CTGCCTGGAG GGTGCTGTCC AGTATTGAGC AGAAAAGCAA CGAGGAGGGC TCGGAGGAGA      300
AGGGGCCCCG GGTGCGTGAG TACCGGGAGA AGGTGGAGAC TGAGCTCCAG GCGGTGTGCG      360
ACACCGTGCT GGGCCTGCTG GACAGCCACC TCATCAAGGA GGCCGGGGAC GCCGAGAGCC      420
GGGTCTTCTA CCTGAAGATG AAGGGTGA CTACTACGCTA CCTGGCCGAG GTGGCCACCG      480
GTGACGACAA GAAGCGCATC ATTGACTCAG CCCGGTCAGC CTACCAGGAG GCCATGGACA      540
TCAGCAAGAA GGAGATGCCG CCCACCAACC CCATCCGCCT GGGCCTGGCC CTGAACTTTT      600
CCGTCTTCCA CTACGAGATC GCCAACAGCC CCGAGGAGGC CATCTCTCTG GCCAAGACCA      660
CTTTCGACGA GGCCATGGCT GATCTGCACA CCCTCAGCGA GGA CTCTCTAC AAAGACAGCA      720
CCCTCATCAT GCAGCTGCTG CGAGACAACC TGACACTGTG GACGGCCGAC AACGCCGGGG      780
AAGAGGGGGG CGAGGCTCCC CAGGAGCCCC AGAGCTGAGT GTTGCCCGCC ACCGCCCCGC      840
CCTGCCCCCT CCAGTCCCCC ACCCTGCCGA GAGGACTAGT ATGGGGTGGG AGGCCCCACC      900
CTTCTCCCCT AGGCGCTGTT CTTGCTCCAA AGGGCTCCGT GGAGAGGGAC TGGCAGAGCT      960
GAGGCCACCT GGGGCTGGGG ATCCCACTCT TCTTGCACT GTTGAGCGCA CCTAACCCT      1020
GGTCATGCCC CCACCCCTGC TCTCCGCACC CGCTTCCTCC CGACCCAGG ACCAGGCTAC      1080
TTCTCCCCTC CTCTTGCTC CCTCCTGCCC CTGCTGCCTC TGATCGTAGG AATTGAGGAG      1140
TGTCCCGCCT TGTGGCTGAG AACTGGACAG TGGCAGGGGC TGGAGATGGG TGTGTGTGTG      1200
TGTGTGTGTG TGTGTGTGTG TGTGCGCGCG CGCCAGTGCA AGACCGAGAT TGAGGGAAAG      1260
CATGTCTGCT GGGTGTGACC ATGTTTCCTC TCAATAAAGT TCCCCTGTGA CACTCAAAAA      1320
AAAAAAAAAA AAAAAA                                     1336

```

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1968 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 1..1968

(D) OTHER INFORMATION: /standard_name= "Homo sapiens Ras association (RalGDS/AF-6) domain family 1 (RASSF1), transcript variant A, mRNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

TCTCCTCAGC TCCTTCCCGC CGCCCAGTCT GGATCCTGGG GGAGGCGCTG AAGTCGGGGC	60
CCGCCCTGTG GCCCCGCCCC GCCCGCGCTT GCTAGCGCCC AAAGCCAGCG AAGCACGGGC	120
CCAACCGGGC CATGTCGGGG GAGCCTGAGC TCATTGAGCT GCGGGAGCTG GCACCCGCTG	180
GGCGCGCTGG GAAGGGCCGC ACCCGGCTGG AGCGTGCCAA CGCGCTGCGC ATCGCGCGGG	240
GCACCGCGTG CAACCCACACA CGGCAGCTGG TCCCTGGCCG TGGCCACCGC TTCCAGCCCG	300
CGGGGCCCCG CACGCACACG TGGTGCGACC TCTGTGGCGA CTTCATCTGG GCGTCGTGC	360
GCAAAGGCCT GCAGTGCGCG CATTGCAAGT TCACCTGCCA CTACCGCTGC CGCGCGCTCG	420
TCTGCCTGGA CTGTTGCGGG CCCCAGGACC TGGGCTGGGA ACCCGCGGTG GAGCGGGACA	480
CGAACGTGGA CGAGCCTGTG GAGTGGGAGA CACCTGACCT TTCTCAAGCT GAGATTGAGC	540
AGAAGATCAA GGAGTACAAT GCCCAGATCA ACAGCAACCT CTTCATGAGC TTGAACAAGG	600
ACGGTTCTTA CACAGGCTTC ATCAAGGTTC AGCTGAAGCT GGTGCGCCCT GTCTCTGTGC	660
CCTCCAGCAA GAAGCCACCC TCCTTGCAGG ATGCCCCGGC GGGCCCAGGA CGGGGCACAA	720
GTGTCAGGCG CCGCACTTCC TTTTACCTGC CCAAGGATGC TGTCAAGCAC CTGCATGTGC	780
TGTCACGCAC AAGGGCACGT GAAGTCATTG AGGCCCTGCT GCGAAAGTTC TTGGTGGTGG	840
ATGACCCCCG CAAGTTTGCA CTCTTTGAGC GCGCTGAGCG TCACGGCCAA GTGTACTTGC	900
GGAAGCTGTT GGATGATGAG CAGCCCCTGC GGCTGCGGCT CCTGGCAGGG CCCAGTGACA	960
AGGCCCTGAG CTTTGTCTTG AAGGAAAATG ACTCTGGGGA GGTGAACTGG GACGCCTTCA	1020
GCATGCCTGA ACTACATAAC TTCCTACGTA TCCTGCAGCG GGAGGAGGAG GAGCACCTCC	1080
GCCAGATCCT GCAGAAGTAC TCCTATTGCC GCCAGAAGAT CCAAGAGGCC CTGCACGCCT	1140
GCCCCCTTGG GTGACCTCTT GTACCCCCAG GTGGAAGGCA GACAGCAGGC AGCGCCAAGT	1200
GCGTGCCGTG TGAGTGTGAC AGGGCCAGTG GGGCCTGTGG AATGAGTGTG CATGGAGGCC	1260
CTCCTGTGCT GGGGGAATGA GCCCAGAGAA CAGCGAAGTA GCTTGCTCCC TGTGTCCACC	1320
TGTGGGTGTA GCCAGGTATG GCTCTGCACC CCTCTGCCCT CATTACTGGG CCTTAGTGGG	1380
CCAGGGCTGC CCTGAGAAGC TGCTCCAGGC CTGCAGCAGG AGTGGTGCAG ACAGAAGTCT	1440
CCTCAATTTT TGTCTCAGAA GTGAAAATCT TGGAGACCCT GCAAACAGAA CAGGGTCATG	1500
TTTGCAGGGG TGACGGCCCT CATCTATGAG GAAAGGTTTT GGATCTTGAA TGTGGTCTCA	1560
GGATATCCTT ATCAGAGCTA AGGGTGGGTG CTCAGAATAA GGCAGGCATT GAGGAAGAGT	1620
CTTGGTTTCT CTCTACAGTG CCAACTCCTC ACACACCCTG AGGTCAGGGA GTGCTGGCTC	1680

ACAGTACAGC ATGTGCCTTA ATGCTTCATA TGAGGAGGAT GTCCCTGGGC CAGGGTCTGT 1740
 GTGAATGTGG GCACTGGCCC AGGTTTCATAC CTTATTTGCT AATCAAAGCC AGGGTCTCTC 1800
 CCTCAGGTGT TTTTATGAA GTGCGTGAAT GTATGTAATG TGTGGTGGCC TCAGCTGAAT 1860
 GCCTCCTGTG GGGAAAGGGG TTGGGGTGAC AGTCATCATC AGGGCCTGGG GCCTGAGAGA 1920
 ATTGGCTCAA TAAAGATTTC AAGATCCTCA AAAAAAAAAA AAAAAAAAAA 1968

(2) INFORMATION FOR SEQ ID NO:65:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: artificial

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

AGTTGCGCGG CGATTTC 17

(2) INFORMATION FOR SEQ ID NO:66:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: artificial

- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 1..23
 - (D) OTHER INFORMATION: /standard_name= "probe
 (5'FAM/3'TAMRA) GSTP1"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

CGGTCGACGT TCGGGGTGTA GCG 23

Claims

We Claim:

1. A method of diagnosing the presence or predicting the course of a proliferative disorder comprising determining the methylation status of a Marker in a biological sample with the use of a reagent selected from the group consisting of: Seq. ID NO. 1-35 and 44-53.
2. The method according to claim 1 further comprising measuring the presence of a reference Marker.
3. The method according to claim 2 wherein the reference Marker is selected from the group consisting of B Actin.
4. The method of claim 1 wherein a combination of Markers are used.
5. The method of claim 4 wherein the combination of Markers includes a Marker for GSTP1 and a Marker for APC, TIMP, RASSF1A, RARB2, PTGS2, or 14-3-3.
6. The method according to claim 4 wherein reagents for the Markers include a member of the group consisting of Seq. ID. No. 1-35 and 44-53.
7. The method of claim 1 wherein the sample is prostate tissue.
8. The method of claim 1 wherein the sample is urine, serum, plasma, or circulating cells.
9. A composition comprising a member of the group consisting of Seq. ID No. 1-43.
10. A composition comprising a mixture of more than one composition of claim 10.
11. A kit for conducting an assay according to claim 1, comprising: nucleic acid amplification and detection reagents.
12. The kit of claim 11 wherein the amplification and detection reagents are selected from the group consisting of Seq. ID No. 1-43.

13. The kit of claim 12 wherein the amplification and detection reagents are selected from Seq. ID No. 1-25 together with reagents selected from the group consisting of Seq. ID No. 36- 43.
14. The kit of claim 11 further comprising instructions.
- 5 15. The kit of claim 11 wherein the reagents detect the hypermethylation of a gene selected from the group consisting of GSTP1 and APC, TIMP, RASSF1a, RARB2, PTGS2, or 14-3-3.
16. The kit of claim 11 further comprising reagents for amplifying and detecting the presence of a constitutively expressed gene.
- 10 17. The kit of claim 11 further comprising reagents for amplifying and detecting the presence of constitutively expressed genes.
18. The method of claim 1 further comprising establishing a methylation ratio and determining whether the methylation ratio exceeds a cutoff value.
19. The method of claim 1 for use in therapy monitoring.
- 15 20. The method of claim 1 for use in screening.
21. The method of claim 1 for use in resolving ambiguous or indeterminate results of cancer assays obtained using methods different from claim 1.

Fig. 1 Prior Art MSP Assay for GSTP1

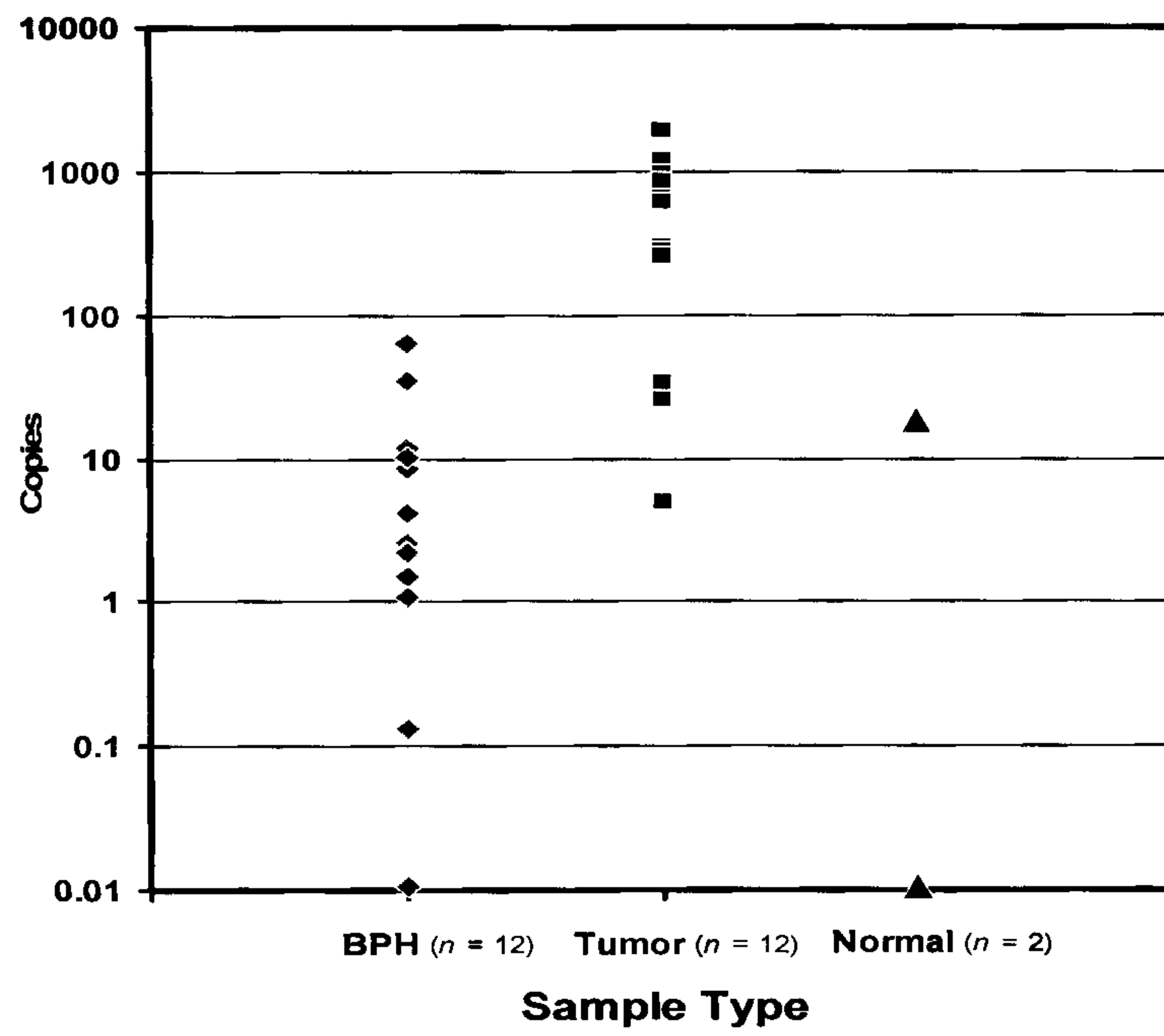
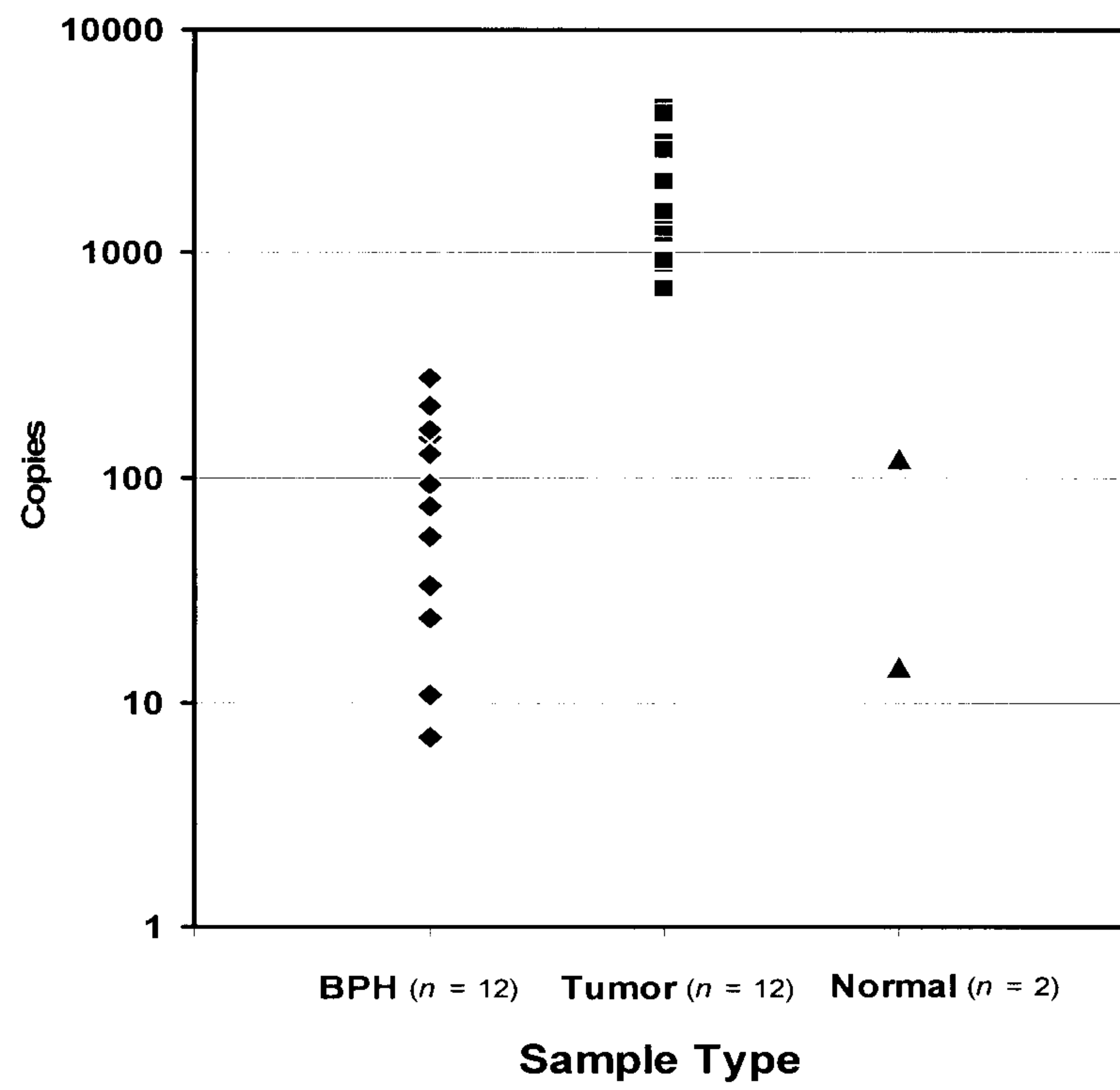


Fig. 2 MSP Assay for GSTP1 According to the Invention



MSP Assay for GSTP1 According to the Invention

