(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 September 2004 (02.09.2004)

PCT

(10) International Publication Number
WO 2004/074518 A1

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US2004/005287

(22) International Filing Date: 19 February 2004 (19.02.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/448,991 20 February 2003 (20.02.2003) US

(71) Applicant (*for all designated States except US*): **GENOMIC HEALTH, INC.** [US/US]; 301 Penobscot Drive, Redwood City, CA 94063 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **SCOTT, Randal, W.** [US/US]; 24032 Oak Knoll Circle, Los Altos Hills, CA 94022 (US). **KIEFER, Michael, C.** [US/US]; 401 Wright Court, Clayton, CA 94517 (US). **BAKER, Joffre, B.** [US/US]; P.O. Box 371212, Montara,CA 94937 (US).

(74) Agent: **DREGER, Ginger, R.**; Heller, Ehrman White & McAuliffe LLP, 275 Middlefield Road, Menlo Park, CA 94025-3506 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: USE OF INTRONIC RNA TO MEASURE GENE EXPRESSION

(57) Abstract: The invention is based on the use of intronic RNA in monitoring gene expression. Accordingly, the present invention concerns methods of gene expression profiling using intronic RNA, the expression of which correlates with the expression of corresponding exonic RNA, and diagnostic and prognostic methods based on the results of such gene expression studies.

WO 2004/074518 A1

USE OF INTRONIC RNA TO MEASURE GENE EXPRESSION

Background of the Invention

Field of the Invention

It is well recognized that gene expression within cells and tissues can indicate the physiologic and or pathologic status of the cell, tissue or patient. For several decades, gene expression, as measured by immunohistochemical analysis of protein markers, has been used to make treatment decisions. For example, levels of estrogen receptor and progesterone receptor measured this way are now routinely used to select breast cancer patients for treatment with anti-estrogen drugs.

More recent research literature provides evidence that tissue levels of mRNA species have diagnostic and prognostic value. This is a promising development because the technologies for measurement of cellular RNA levels, as exemplified by multiple RT-PCR and DNA array platforms, can be very sensitive, specific and quantitative. RT-PCR is recognized as generally more sensitive than DNA array technology. However, RT-PCR probe/primer design and selection can be challenging, because multiple criteria exist for optimal performance. This challenge is particularly great when the sample RNA to be studied comes from fixed, wax-embedded tissue, because such RNA tends to be highly fragmented (K.Specht *et al., Am. J. Pathol* 158: 419-29 [2001]; T.E. Godfrey *et al., J. Mol. Diagnostics* 2:84-91 [2000]).

It is accepted practice to measure the expression of any given gene by assaying the level of any of its transcribed, spliced, mature mRNA sequences (exon, as opposed to intron, sequence). In theory, an exon is defined as any segment of an interrupted gene that is represented in the mature RNA product, and an intron is defined as a segment of DNA that is transcribed but removed from within the transcript by splicing together the exons on either side of it [B. Lewin. *Genes IV*, Cell Press, Cambridge Mass. 1990]. The rationale for the accepted practice of using exon sequences is theoretically straightforward because the mature RNAs [mRNAs] encode proteins, which define cell phenotypes, whereas intronic RNA is considered to have comparatively little influence on cell phenotype. Moreover, the prevailing view is that introns are rapidly degraded and therefore more difficult to detect than exon sequences {see introductions of the following articles: Thomas *et al., J. Virol.* 76:532-40

[2002]; Clement *et al., J. Biol. Chem.* 276:16919-30 [2001]; Sharp *et al., Ann. Rev. Biochem.* 55:1119-1150 [1986]}.

The present invention concerns the use of intronic RNA for measuring gene expression. It will be shown that intronic RNA sequences tend to be readily detected by RT-PCR, even using extensively degraded RNA from fixed tissues. Furthermore, they tend to correlate in their expression with their respective exons. The latter point is particularly unexpected because little or no evidence exists that the ratio of the overall rate constants for synthesis and turnover of transcribed intron and exon sequences are similar. In fact, the scientific literature provides evidence for the complexity of pre-mRNA and spliced intron turnover. For example, pre-mRNA can exist in multiple kinetic pools (Elliott and Rosbash, *Exp. Cell Res.* 229:181-8 [1996]), with subpopulations containing intron RNAs that are not efficiently spliced out and are transported to the cytoplasm as "immature" mRNA species, where they can decay at rates different than nuclear intron RNA sequences (Wang *et al., Proc. Natl. Acad. Sci. USA* 94:4360-5 [1997]). Furthermore, certain spliced intron RNAs seem to enter the cytoplasm in lariat structure (Clement *et al., RNA* 5:206-20 [1999]).

Summary of the Invention

The present invention is based on experimental evidence demonstrating that transcribed intron sequences, which by definition are present in heterogeneous nuclear RNA but typically are not incorporated into mRNA, have diagnostic and prognostic utility. This is a significant discovery for several reasons. Typically, intron sequences are longer than exon sequences, by twenty fold or more. Thus, introns, given their much greater average length, provide proportionally increased opportunity for optimal gene expression probe design, for example, in the case of RT-PCR, creation of probe/primer sets that possess better technical performance. Independently, because intron sequences evolve more rapidly than exon sequences, intronic RNAs are well-suited to monitor the expression of different closely related members of a gene family.

In one aspect, the invention concerns a method for monitoring gene expression in a biological sample, comprising:

(a) providing a polynucleotide complementary to an intronic RNA sequence within a target gene, wherein the expression of such intronic RNA sequence correlates with the expression of an exonic mRNA sequence within the target gene;

(b) hybridizing the polynucleotide to the intronic RNA sequence to form a polynucleotide-intronic RNA complex; and

(c) detecting the polynucleotide-intronic RNA complex.

In a particular aspect, expression of the target gene is measured by RT-PCR, in which case an intron-based primer/probe set can be used in the above process.

In another aspect, the invention concerns methods of using intron-based sequences to design and create primer-probe sets for RT-PCR. Such primers and probes are particularly suitable to detect and quantify levels of intron RNA in fixed, paraffin-embedded tissue (FPET) specimens, for high sensitivity gene expression analysis. Accordingly, in a further aspect, the invention concerns using intron-based primer-probe sets in gene expression profiling assays, such as gene expression analysis of FPET samples to diagnose and/or predict the prognosis of various pathologic conditions.

In particular, the invention concerns a method of preparing a single-stranded oligonucleotide molecule for amplification of a target gene, and measuring the level of an intronic RNA species comprising:

(a) identifying at least one intron sequence within the target gene, wherein the expression of the intron sequence correlates with the expression of an exon sequence within the target gene;

(b) preparing a single-stranded oligonucleotide molecule that corresponds to at least a portion of the transcribed intron sequence; and

(c) using the oligonucleotide molecule to measure gene expression.

Just as before, gene expression can be measured, for example, by RT-PCR, in which case an intron-based primer/probe set (consisting of two primers and a probe) is used to measure gene expression.

If the oligonucleotide is a forward primer, it is typically designed to comprise 5'-sequences of a target sequence within the transcribed intron sequence. If the oligonucleotide is a reverse primer, it is typically designed to complement 5'-sequences of a target sequence downstream of the forward primer within the transcribed intron sequence. It is important to identify and use a sufficiently long target sequence for PCR amplification. The target sequence generally should be at least about 50 nucleotide bases long, in particular at least 55 nucleotide bases long, in some embodiments at least about 60 nucleotide bases long. The PCR primers and probes are designed following well known principles. Thus, the PCR primer is typically 17-30 nucleotide bases in length, and usually contains about 20% to 80%

G+C bases. It is desirable to design PCR primers with a melting temperature (Tm) between about 50°C and about 70°C.

When the single-stranded oligonucleotide molecule is a PCR probe, it is usually designed to comprise or complement an internal portion of a target sequence within the transcribed intron sequence. For TaqMan® amplification, the PCR probe is labeled with a reporter fluorescent dye and a quencher moiety.

In another aspect, the invention concerns a method for measuring the expression of a gene by amplifying a target gene by polymerase chain reaction (PCR) comprising:

(a) identifying at least one target intron sequence within the target gene, wherein the expression of the intron sequence correlates with the expression of a corresponding exon sequence within the target gene; and

(b) amplifying the transcribed target intron sequence using an intron-specific PCR primer/probe set.

The target intron sequence is typically at least about 50 bases long, and the PCR primer and probe set is designed to correspond to unique sequences within the transcribed target intron sequence.

In yet another aspect, the invention concerns a method for amplifying RNA fragments in a sample representing at least one gene of interest, comprising the steps of:

(a) contacting the sample with at least one set of PCR primers and probe; and

(b) performing PCR amplification,

wherein the PCR primers and probe are designed based upon an intron sequence identified within the gene of interest, and wherein the expression of the intron sequence correlates with the expression of an exon sequence within the gene of interest.

In particular embodiment, the PCR primers and probe are typically designed based upon a unique sequence within the intron identified. In another embodiment, the sample comprises fragmented RNA representing multiple genes of interest, and is contacted with a pool of PCR primers and probes designed based upon unique sequences within introns present in the genes of interest.

In a preferred embodiment, the amplification is performed on a fixed, paraffin-embedded tissue (FPET) sample, which can, for example, originate from a tumor biopsy obtained from a human patient. The tumor can be any kind of solid tumor, such as, for example, breast cancer, lung cancer, or colorectal cancer. The tumor tissue can be harvested by a variety of methods, including fine needle biopsy, core biopsy or resection.

In a particular embodiment, the invention concerns methods using intron-based PCR primer-probe sets in gene expression analysis to predict the likelihood of recurrent disease for patients with early breast cancer.

In a further aspect, the invention concerns an array comprising a plurality of polynucleotides hybridizing to target genes of interest, wherein preferably at least 70% of the polynucleotides comprises intron sequences.

In yet another aspect, the invention concerns intron-based amplicon sequences, and their use in gene expression analysis.

In a particular embodiment the invention concerns gene expression analysis of a biological sample representative of invasive breast cancer based on determining the expression levels of the RNA transcripts or expression products of a gene or gene set selected from the group consisting of:

(a) Bcl2, cyclinG1, NFKBp65, NME1, EPHX1, TOP2B, DR5, TERC, Src, DIABLO;

(b) Ki67, XIAP, hENT1, TS, CD9, p27, cyclinG1, pS2, NFKBp65, CYP3A4;

(c) GSTM1, XIAP, Ki67, TS, cyclinG1, p27, CYP3A4, pS2, NFKBp65, ErbB3;

(d) PR, NME1, XIAP, upa, cyclinG1, Contig51037, TERC, EPHX1, ALDH1A3, CTSL;

(e) CA9, NME1, TERC, cyclinG1, EPHX1, DPYD, Src, TOP2B, NFKBp65, VEGFC;

(f) TFRC, XIAP, Ki67, TS, cyclinG1, p27, CYP3A4, pS2, ErbB3, NFKBp65;

(g) Bcl2, PRAME, cyclinG1, FOXM1, NFKBp65, TS, XIAP, Ki67, CYP3A4, p27;

(h) FOXM1, cyclinG1, XIAP, Contig51037, PRAME, TS, Ki67, PDGFRa, p27, NFKBp65;

(i) PRAME, FOXM1, cyclinG1, XIAP, Contig51037, TS, Ki6, PDGFRa, p27, NFKBp65;

(j) Ki67, XIAP, PRAME, hENT1, contig51037, TS, CD9, p27, ErbB3, cyclinG1;

(k) STK15, XIAP, PRAME, PLAUR, p27, CTSL, CD18, PREP, p53, RPS6KB1;

(l) GSTM1, XIAP, PRAME, p27, Contig51037, ErbB3, GSTp, EREG, ID1, PLAUR;

(m) PR, PRAME, NME1, XIAP, PLAUR, cyclinG1, Contig51037, TERC, EPHX1, DR5;

(n) CA9, FOXM1, cyclinG1, XIAP, TS, Ki67, NFKBp65, CYP3A4, GSTM3, p27;

(o) TFRC, XIAP, PRAME, p27, Contig51037, ErbB3, DPYD, TERC, NME1, VEGFC; and

(p) CEGP1, PRAME, hENT1, XIAP, Contig51037, ErbB3, DPYD, NFKBp65, ID1, TS,

including the use of intron-based sequences.

In another embodiment, the invention concerns gene expression analysis of a biological sample representative of ER-positive breast cancer based on determining the expression levels of the RNA transcripts or expression products of a gene or gene set selected from the group consisting of:

(a) PRAME, p27, IGFBP2, HIF1A, TIMP2, ILT2, CYP3A4, ID1, EstR1, DIABLO;

(b) Contig51037, EPHX1, Ki67, TIMP2, cyclinG1, DPYD, CYP3A4, TP, AIB1, CYP2C8;

(c) Bcl2, hENT1, FOXM1, Contig51037, cyclinG1, Contig46653, PTEN, CYP3A4, TIMP2, AREG;

(d) HIF1A, PRAME, p27, IGFBP2, TIMP2, ILT2, CYP3A4, ID1, EstR1, DIABLO;

(e) IGF1R, PRAME, EPHX1, Contig51037, cyclinG1, Bcl2, NME1, PTEN, TBP, TIMP2;

(f) FOXM1, Contig51037, VEGFC, TBP, HIF1A, DPYD, RAD51C, DCR3, cyclinG1, BAG1;

(g) EPHX1, Contig51037, Ki67, TIMP2, cyclinG1, DPYD, CYP3A4, TP, AIB1, CYP2C8;

(h) Ki67, VEGFC, VDR, GSTM3, p27, upa, ITGA7, rhoC, TERC, Pin1;

(i) CDC25B, Contig51037, hENT1, Bcl2, HLAG, TERC, NME1, upa, ID1, CYP;

(j) VEGFC, Ki67, VDR, GSTM3, p27, upa, ITGA7, rhoC, TERC, Pin1;

(k) CTSB, PRAME, p27, IGFBP2, EPHX1, CTSL, BAD, DR5, DCR3, XIAP;

(l) DIABLO, Ki67, hENT1, TIMP2, ID1, p27, KRT19, IGFBP2, TS, PDGFB;

(m) p27, PRAME, IGFBP2, HIF1A, TIMP2, ILT2, CYP3A4, ID1, EstR1, DIABLO;

(n) CDH1; PRAME, VEGFC; HIF1A; DPYD, TIMP2, CYP3A4, EstR1, RBP4, p27;

(o) IGFBP3, PRAME, p27, Bcl2, XIAP, EstR1, Ki67, TS, Src, VEGF;

(p) GSTM3, PRAME, p27, IGFBP3, XIAP, FGF2, hENT1, PTEN, EstR1, APC;

(q) hENT1, Bcl2, FOXM1, Contig51037, CyclinG1, Contig46653, PTEN, CYP3A4, TIMP2, AREG;

(r) STK15, VEGFC, PRAME, p27, GCLC, hENT1, ID1, TIMP2, EstR1, MCP1;

(s) NME1, PRAM, p27, IGFBP3, XIAP, PTEN, hENT1, Bcl2, CYP3A4, HLAG;

(t) VDR, Bcl2, p27, hENT1, p53, PI3KC2A, EIF4E, TFRC, MCM3, ID1;

(u) EIF4E, Contig51037, EPHX1, cyclinG1, Bcl2, DR5, TBP, PTEN, NME1, HER2;

(v) CCNB1, PRAME, VEGFC, HIF1A, hENT1, GCLC, TIMP2, ID1, p27, upa;

(w) ID1, PRAME, DIABLO, hENT1, p27, PDGFRa, NME1, BIN1, BRCA1, TP;

(x). FBXO5, PRAME, IGFBP3, p27, GSTM3, hENT1, XIAP, FGF2, TS, PTEN;

(y) GUS, HIA1A, VEGFC, GSTM3, DPYD, hENT1, EBXO5, CA9, CYP, KRT18; and

(z) Bclx, Bcl2, hENT1, Contig51037, HLAG, CD9, ID1, BRCA1, BIN1, HBEGF,

including the use of intron-based sequences.

In a further embodiment, the cancer is breast cancer, and the gene(s) analyzed is/are selected from the group consisting of: FOXM1; PRAME; SKT15, Ki-67; CA9; NME1; SURV; TFRC; YB-1; RPS6KB1; Src; Chk1; CCNB1; Chk2; CDC25B; CYP3A4; EpCAM; VEGFC; hENT1; BRCA2; EGFR; TK1; VDR; Blc12; CEGP1; GSTM1; PR; BBC3; GATA3; DPYD; GSTM3; ID1; EstR1; p27; XIAP; IGF1R; AK055699; P13KC2A; TGFB3; BAGI1; pS2; WISP1; HNF3A; and NFKBp65.

In a still further embodiment, invention concerns gene expression analysis of a biological sample representative of invasive breast cancer, based on determining the expression levels of the RNA transcripts or expression products of a gene or gene set selected from the group consisting of:

(a) p53BP2, Bcl2, BAD, EPHX1, PDGFRβ, DIABLO, XIAP, YB1, CA9, and KRT8;

(b) GRB7, CD68, TOP2A, Bcl2, DIABLO, CD3, ID1, PPM1D, MCM6, and WISP1;

(c) PR, p53BP2, PRAME, DIABLO, CTSL, IGFBP2, TIMP1, CA9, MMP9, and COX2;

(d) CD68, GRB7, TOP2A, Bcl2, DIABLO, CD3, ID1, PPM1D, MCM6, and WISP1;

(e) Bcl2, p53BP2, BAD, EPHX1, PDGFRβ, DIABLO, XIAP, YB1, CA9, and KRT8;

(f) KRT14, KRT5, PRAME, p53BP2, GUS1, AIB1, MCM3, CCNE1, MCM6, and ID1;

(g) PRAME, p53BP2, EstR1, DIABLO, CTSL, PPM1D, GRB7, DAPK1, BBC3, and VEGFB;

(h) CTSL2, GRB7, TOP2A, CCNB1, Bcl2, DIABLO, PRAME, EMS1, CA9, and EpCAM;

(i) EstR1, p53BP2, PRAME, DIABLO, CTSL, PPM1D, GRB7, DAPK1, BBC3, and VEGFB;

(j) Chk1, PRAME, p53BP2, GRB7, CA9, CTSL, CCNB1, TOP2A, tumor size, and IGFBP2;

(k) IGFBP2, GRB7, PRAME, DIABLO, CTSL, β-Catenin, PPM1D, Chk1, WISP1, and LOT1;

(l) HER2, p53BP2, Bcl2, DIABLO, TIMP1, EPHX1, TOP2A, TRAIL, CA9, and AREG;

(m) BAG1, p53BP2, PRAME, IL6, CCNB1, PAI1, AREG, tumor size, CA9, and Ki67;

(n) CEGP1, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, STK15, and AKT2, and FGF18;

(0) STK15, p53BP2, PRAME, IL6, CCNE1, AKT2, DIABLO, cMet, CCNE2, and COX2;

(p) KLK10, EstR1, p53BP2, PRAME, DIABLO, CTSL, PPM1D, GRB7, DAPK1, and BBC3;

(q) AIB1, p53BP2, Bcl2, DIABLO, TIMP1, CD3, p53, CA9, GRB7, and EPHX1

(r) BBC3, GRB7, CD68, PRAME, TOP2A, CCNB1, EPHX1, CTSL GSTM1, and APC;

(s) CD9, GRB7, CD68, TOP2A, Bcl2, CCNB1, CD3, DIABLO, ID1, and PPM1D;

(t) EGFR, KRT14, GRB7, TOP2A, CCNB1, CTSL, Bcl2, TP, KLK10, and CA9;

(u) HIF1α, PR, DIABLO, PRAME, Chk1, AKT2, GRB7, CCNE1, TOP2A, and CCNB1;

(v) MDM2, p53BP2, DIABLO, Bcl2, AIB1, TIMP1, CD3, p53, CA9, and HER2;

(w) MYBL2, p53BP2, PRAME, IL6, Bcl2, DIABLO, CCNE1, EPHX1, TIMP1, and CA9;

(x) p27, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, STK15, AKT2, and ID1;

(y) RAD51, GRB7, CD68, TOP2A, CIAP2, CCNB1, BAG1, IL6, FGFR1, and p53BP2;

(z) SURV, GRB7, TOP2A, PRAME, CTSL, GSTM1, CCNB1, VDR, CA9; and CCNE2;

(aa) TOP2B, p53BP2, DIABLO, Bcl2, TIMP1, AIB1, CA9, p53, KRT8, and BAD;

(ab) ZNF217, GRB7, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, APC4, and β-Catenin.

In a different embodiment, the invention concerns gene expression analysis of a biological sample, using intron-based polynucleotide sequences hybridizing to at least one genes selected from the group consisting of: CD68; CTSL; FBXO5; SURV; CCNB1; MCM2; Chk1; MYBL2; HIF1A; cMET; EGFR; TS; STK15, IGFR1; BCl2; HNF3A; TP53BP2; GATA3; BBC3; RAD51C; BAG1; IGFBP2; PR; CD9; RB1; EPHX1; CEGP1; TRAIL; DR5; p27; p53; MTA; RIZ1; ErbB3; TOP2B; EIF4E, CD68; CTSL; FBXO5; SURV; CCNB1; MCM2; Chk1; MYBL2; HIF1A; cMET; EGFR; TS; and STK15.

Gene expression analysis may be performed in an array format, and the array preferably is a high-density array, comprising at least 100, more preferably at least 150, even more preferably, 200 sequences in a 5-10 μ section.

Brief Description of the Drawings

Figures 1A-M show masked intron sequences for the CEGP1, FOXM1, PRAME, and STK15 genes. Amplicons used for RT-PCR are shown in italics.

Figure 2 shows primer/probe sets for CEGP1, FOXM1, PRAME, and STK15. Sequences of forward and reverse primers are indicated by "F" and "R," respectively. Sequences of primers are designated with "P."

Figure 3 shows correlation coefficients [R] for co-expression of CEGP1 exon RNA with 47 other RNA sequences. Symbols: diamond = CEGP1 exon self vs. self (=1.0 by definition); squares=CEGP1 introns; triangles = sequences of other genes.

Figure 4 shows correlation coefficients [R] for co-expression of PRAME exon RNA with 47 other RNA sequences. Symbols: diamond – PRAME exon self vs. self (=1.0 by definition); squares=PRAME introns; triangles = sequences of other genes.

Figure 5 shows correlation coefficients [R] for co-expression of STK15 exon RNA with 47 other RNA sequences. Symbols: diamond – STK15 exon self vs. self (=1.0 by definition); squares=STK15 introns; triangles = sequences of other genes.

Figure 6 shows an exemplary set of genes, the expression of which can be analyzed by the methods of the present invention.

Detailed Description of the Preferred Embodiment

A. Definitions

Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Singleton *et al.*, Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, NY 1994), provide one skilled in the art with a general guide to many of the terms used in the present application.

One skilled in the art will recognize many methods and materials similar or equivalent to those described herein, which could be used in the practice of the present invention. Indeed, the present invention is in no way limited to the methods and materials described. For purposes of the present invention, the following terms are defined below.

The terms "splicing" and "RNA splicing" are used interchangeably and refer to RNA processing that removes introns and joins exons to produce mature mRNA with continuous coding sequence that moves into the cytoplasm of an eukaryotic cell.

In theory, the term "exon" refers to any segment of an interrupted gene that is represented in the mature RNA product (B. Lewin. Genes IV Cell Press, Cambridge Mass. 1990). In theory the term "intron" refers to any segment of DNA that is transcribed but removed from within the transcript by splicing together the exons on either side of it. Operationally, exon sequences occur in the mRNA sequence of a gene as defined by Ref. Seq ID numbers. Operationally, intron sequences are the intervening sequences within the genomic DNA of a gene, bracketed by exon sequences and having GT and AG splice consensus sequences at their 5' and 3' boundaries.

The term "microarray" refers to an ordered arrangement of hybridizable array elements, preferably polynucleotide probes, on a substrate.

The term "polynucleotide," when used in singular or plural, generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. Thus, for instance, polynucleotides as defined herein include, without limitation, single- and double-stranded DNA, DNA including single- and double-stranded regions, single- and double-stranded RNA, and RNA including single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or include single- and double-stranded regions. In addition, the term "polynucleotide" as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide. The term "polynucleotide" specifically includes cDNAs. The term includes DNAs (including cDNAs) and RNAs that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotides" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritiated bases, are included within the term “polynucleotides” as defined herein. In general, the term “polynucleotide” embraces all chemically, enzymatically and/or metabolically modified forms of unmodified polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including simple and complex cells.

The term "oligonucleotide" refers to a relatively short polynucleotide, including, without limitation, single-stranded deoxyribonucleotides, single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs. Oligonucleotides, such as single-stranded DNA probe oligonucleotides, are often synthesized by chemical methods, for example using automated oligonucleotide synthesizers that are commercially available. However, oligonucleotides can be made by a variety of other methods, including *in vitro* recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms.

The terms “differentially expressed gene,” “differential gene expression” and their synonyms, which are used interchangeably, refer to a gene whose expression is at a higher or lower level in one patient or test subject relative to another, for example, in a subject suffering from a disease, specifically cancer, such as breast cancer, relative to its expression in a normal or control subject. The terms also include genes whose expression is activated to a higher or lower level at different stages of the same disease. It is also understood that a differentially expressed gene may be either activated or inhibited at the nucleic acid level or protein level, or may be subject to alternative splicing to result in a different polypeptide product. Such differences may be evidenced by a change in mRNA levels, surface expression, secretion or other partitioning of a polypeptide, for example. Differential gene expression may include a comparison of expression between two or more genes or their gene products, or a comparison of the ratios of the expression between two or more genes or their gene products, or even a comparison of two differently processed products of the same gene, which differ between normal subjects and subjects suffering from a disease, specifically cancer, or between various stages of the same disease. Differential expression includes both quantitative, as well as qualitative, differences in the temporal or cellular expression pattern in a gene or its expression products among, for example, normal and diseased cells, or among cells which have undergone different disease events or disease stages. For the purpose of this invention, “differential gene expression” is considered to be present when there is at least an about two-fold, preferably at least about four-fold, more preferably at least about six-fold, most preferably at least about ten-fold difference between the expression of a given gene in normal and diseased subjects, or in various stages of disease development in a diseased subject.

The term “over-expression” with regard to an RNA transcript is used to refer to the level of the transcript determined by normalization to the level of reference mRNAs, which might be all measured transcripts in the specimen or a particular reference set of mRNAs.

The phrase "gene amplification" refers to a process by which multiple copies of a gene or gene fragment are formed in a particular cell or cell line. The duplicated region (a stretch of amplified DNA) is often referred to as "amplicon." Often, the amount of the messenger RNA (mRNA) produced, *i.e.,* the level of gene expression, also increases in the proportion of the number of copies made of the particular gene expressed.

The term "prognosis" is used herein to refer to the prediction of the likelihood of cancer-attributable death or progression, including recurrence, metastatic spread, and drug resistance, of a neoplastic disease, such as breast cancer. The term “prediction” is used herein to refer to the likelihood that a patient will respond either favorably or unfavorably to a drug or set of drugs, and also the extent of those responses, or that a patient will survive, following surgical removal or the primary tumor and/or chemotherapy for a certain period of time without cancer recurrence. The predictive methods of the present invention can be used clinically to make treatment decisions by choosing the most appropriate treatment modalities for any particular patient. The predictive methods of the present invention are valuable tools in predicting if a patient is likely to respond favorably to a treatment regimen, such as surgical intervention, chemotherapy with a given drug or drug combination, and/or radiation therapy, or whether long-term survival of the patient, following sugery and/or termination of chemotherapy or other treatment modalities is likely.

The term “long-term” survival is used herein to refer to survival for at least 3 years, more preferably for at least 5 years, most preferably for at least 10 years following surgery or other treatment.

The term "increased resistance" to a particular drug or treatment option, when used in accordance with the present invention, means decreased response to a standard dose of the drug or to a standard treatment protocol.

The term "decreased sensitivity" to a particular drug or treatment option, when used in accordance with the present invention, means decreased response to a standard dose of the drug or to a standard treatment protocol, where decreased response can be compensated for (at least partially) by increasing the dose of drug, or the intensity of treatment.

"Patient response" can be assessed using any endpoint indicating a benefit to the patient, including, without limitation, (1) inhibition, to some extent, of tumor growth, including slowing down and complete growth arrest; (2) reduction in the number of tumor cells; (3) reduction in tumor size; (4) inhibition (i.e., reduction, slowing down or complete stopping) of tumor cell infiltration into adjacent peripheral organs and/or tissues; (5) inhibition (i.e. reduction, slowing down or complete stopping) of metastasis; (6) enhancement of anti-tumor immune response, which may, but does not have to, result in the regression or rejection of the tumor; (7) relief, to some extent, of one or more symptoms associated with the tumor; (8) increase in the length of survival following treatment; and/or (9) decreased mortality at a given point of time following treatment.

The term "treatment" refers to both therapeutic treatment and prophylactic or preventative measures, wherein the object is to prevent or slow down (lessen) the targeted pathologic condition or disorder. Those in need of treatment include those already with the disorder as well as those prone to have the disorder or those in whom the disorder is to be prevented. In tumor (*e.g.*, cancer) treatment, a therapeutic agent may directly decrease the pathology of tumor cells, or render the tumor cells more susceptible to treatment by other therapeutic agents, *e.g.*, radiation and/or chemotherapy.

The term "tumor," as used herein, refers to all neoplastic cell growth and proliferation, whether malignant or benign, and all pre-cancerous and cancerous cells and tissues.

The terms "cancer" and "cancerous" refer to or describe the physiological condition in mammals that is typically characterized by unregulated cell growth. Examples of cancer include but are not limited to, breast cancer, colon cancer, lung cancer, prostate cancer, hepatocellular cancer, gastric cancer, pancreatic cancer, cervical cancer, ovarian cancer, liver cancer, bladder cancer, cancer of the urinary tract, thyroid cancer, renal cancer, carcinoma, melanoma, and brain cancer.

The "pathology" of cancer includes all phenomena that compromise the well-being of the patient. This includes, without limitation, abnormal or uncontrollable cell growth, metastasis, interference with the normal functioning of neighboring cells, release of cytokines or other secretory products at abnormal levels, suppression or aggravation of inflammatory or immunological response, neoplasia, premalignancy, malignancy, invasion of surrounding or distant tissues or organs, such as lymph nodes, etc.

"Stringency" of hybridization reactions is readily determinable by one of ordinary skill in the art, and generally is an empirical calculation, dependent upon probe length, washing temperature, and salt concentration. In general, longer probes require higher temperatures for proper annealing, while shorter probes need lower temperatures. Hybridization generally depends on the ability of denatured DNA to reanneal when complementary strands are present in an environment below their melting temperature. The higher the degree of desired homology between the probe and hybridizable sequence, the higher the relative temperature which can be used. As a result, it follows that higher relative temperatures would tend to make the reaction conditions more stringent, while lower temperatures less so. For additional details and explanation of stringency of hybridization reactions, see Ausubel et al., Current Protocols in Molecular Biology, Wiley Interscience Publishers, (1995).

"Stringent conditions" or "high stringency conditions", as defined herein, typically: (1) employ low ionic strength and high temperature for washing, for example 0.015 M sodium chloride/0.0015 M sodium citrate/0.1% sodium dodecyl sulfate at 50°C; (2) employ during hybridization a denaturing agent, such as formamide, for example, 50% (v/v) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50mM sodium phosphate buffer at pH 6.5 with 750 mM sodium chloride, 75 mM sodium citrate at 42°C; or (3) employ 50% formamide, 5 x SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 μg/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2 x SSC (sodium chloride/sodium citrate) and 50% formamide at 55°C, followed by a high-stringency wash consisting of 0.1 x SSC containing EDTA at 55°C.

"Moderately stringent conditions" may be identified as described by Sambrook et al., Molecular Cloning: A Laboratory Manual, New York: Cold Spring Harbor Press, 1989, and include the use of washing solution and hybridization conditions (e.g., temperature, ionic strength and %SDS) less stringent that those described above. An example of moderately stringent conditions is overnight incubation at 37°C in a solution comprising: 20% formamide, 5 x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5 x Denhardt's solution, 10% dextran sulfate, and 20 mg/ml denatured sheared salmon sperm DNA, followed by washing the filters in 1 x SSC at about 37-50°C. The skilled artisan will recognize how to adjust the temperature, ionic strength, etc. as necessary to accommodate factors such as probe length and the like.

In the context of the present invention, reference to "at least one," "at least two," "at least five," etc. of the genes listed in any particular gene set means any one or any and all combinations of the genes listed.

B. Detailed Description

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, and biochemistry, which are within the skill of the art. Such techniques are explained fully in the literature, such as, "Molecular Cloning: A Laboratory Manual", 2nd edition (Sambrook et al., 1989); "Oligonucleotide Synthesis" (M.J. Gait, ed., 1984); "Animal Cell Culture" (R.I. Freshney, ed., 1987); "Methods in Enzymology" (Academic Press, Inc.); "Handbook of Experimental Immunology", 4th edition (D.M. Weir & C.C. Blackwell, eds., Blackwell Science Inc., 1987); "Gene Transfer Vectors for Mammalian Cells" (J.M. Miller & M.P. Calos, eds., 1987); "Current Protocols in Molecular Biology" (F.M. Ausubel et al., eds., 1987); and "PCR: The Polymerase Chain Reaction", (Mullis et al., eds., 1994).

1. *Polymerase Chain Reaction (PCR)*

The purpose of the polymerase chain reaction (PCR) is to make copies of a gene in order to provide larger amounts of nucleic acid for further use. PCR is a process based on a specialized polymerase enzyme (e.g. Taq DNA polymerase), which can synthesize a complementary strand to a given DNA strand in a mixture containing the four dNTP's (sATP, dCTP, dGTP, dTTP) and two oligonucleotide primers flanking the target sequence to be amplified. The two oligonucleotide primers are used to generate an amplicon typical of a PCR reaction. A third oligonucleotide, or probe, is designed to detect the nucleotide sequence located between the two PCR primers. Although the probe design might differ, in the TaqMan® PCR method probe signals are controlled by the proximity of a reporter fluorescent dye and a quencher fluorescent dye. Any laser-induced emission from the reporter dye is quenched by the quenching dye when the two dyes are located close together as they are on the probe. During the amplification reaction, the polymerase enzyme (e.g. Taq DNA polymerase) cleaves the probe in a template-dependent manner. The resultant probe fragments disassociate in solution, and signal from the released reporter dye is free from the quenching effect of the second fluorophore. One molecule of reporter dye is liberated for each new molecule synthesized, and detection of the unquenched reporter dye provides the basis for quantitative interpretation of the data.

The starting material for PCR can be DNA, cDNA, mRNA or any other polynucleotide that needs to be amplified. Since the PCR requires single-stranded DNA as template, if the starting material is double-stranded DNA, it needs to be denatured in order to produce single-stranded DNA.

As RNA cannot serve as a template for PCR, if the starting material is RNA, the first step is the reverse transcription of the RNA template into cDNA, followed by its exponential amplification in a PCR reaction. This version of PCR is generally referred to as reverse transcriptase PCR (RT-PCR). The two most commonly used reverse transcriptases are avian myeloblastosis virus reverse transcriptase (AMV-RT) and Moloney murine leukemia virus reverse transcriptase (MMLV-RT). The reverse transcription step is typically primed using specific primers, random hexamers, or oligo-dT primers, depending on the circumstances. For example, RNA extracted from a tissue sample (e.g. FPET) can be reverse-transcribed using a GeneAmp RNA PCR kit (Perkin Elmer, CA, USA), following the manufacturer's instructions. The derived cDNA can then be used as a template in the subsequent PCR reaction.

Although the PCR step can use a variety of thermostable DNA-dependent DNA polymerases, it typically employs the Taq DNA polymerase, which has a 5'-3' nuclease activity but lacks a 3'-5' proofreading endonuclease activity. Thus, TaqMan® PCR typically utilizes the 5'-nuclease activity of Taq or Tth polymerase to hydrolyze a hybridization probe bound to its target amplicon, but any enzyme with equivalent 5' nuclease activity can be used. In this case, the probe is designed to be non-extendible by Taq DNA polymerase enzyme. TaqMan® RT-PCR can be performed using commercially available equipment, such as, for example, ABI PRISM 7700™ Sequence Detection System™ (Perkin-Elmer-Applied Biosystems, Foster City, CA, USA), or Lightcycler (Roche Molecular Biochemicals, Mannheim, Germany). In a preferred embodiment, the 5' nuclease procedure is run on a real-time quantitative PCR device such as the ABI PRISM 7700™ Sequence Detection System™. The system consists of a thermocycler, laser, charge-coupled device (CCD), camera and computer. The system amplifies samples in a 96-well format on a thermocycler. During amplification, laser-induced fluorescent signal is collected in real-time through fiber optics cables for all 96 wells, and detected at the CCD. The system includes software for running the instrument and for analyzing the data.

5'-Nuclease assay data are initially expressed as Ct, or the threshold cycle. As discussed above, fluorescence values are recorded during every cycle and represent the amount of product amplified to that point in the amplification reaction. The point when the fluorescence signal is first recorded as statistically significant is the threshold cycle (C_t).

To minimize errors and the effect of sample-to-sample variation, RT-PCR is usually performed using an internal standard. The ideal internal standard is expressed at a constant level among different tissues, and is unaffected by the experimental treatment. RNAs frequently used to normalize patterns of gene expression are mRNAs for the housekeeping genes glyceraldehyde-3-phosphate-dehydrogenase (GAPDH) and β-actin.

For further details of real time quantitative PCR see also Held *et al.*, *Genome Research* 6:986-994 (1996). PCR is described in U.S. Patent Nos. 4,683,202, 4,683,195; 4,965,188; and 5,075,216, the entire disclosures of which are hereby expressly incorporated by reference.

2. *Introns and RNA Splicing*

Most genes in higher eukaryotes contain more than 100,000 nucleotide pairs, some containing more than 2 million nucleotide pairs. This is significantly longer than the nucleotide sequence required to encode an average size protein (300-400 amino acids), which is in the order of about 1000 nucleotides. Most of the extra length consists of noncoding (intron) sequences that interrupt the coding (exon) sequences within the gene sequence. Most of higher eukaryotic genes coding for mRNA, tRNA and some coding for rRNA are interrupted by intron sequences. Genes for mRNA typically have 0 to 60 introns; while genes for tRNA typically include 0 or 1 intron.

When mRNA is transcribed from DNA, at first both exon and intron sequences are transcribed into the so-called heterogeneous nuclear RNA (hnRNA) or immature RNA or pre-mRNA. However, before the RNA exits the nucleus, intron sequences are often deleted from the transcribed mRNA as a result of a process known as RNA splicing. The process of intron removal involves a precise looping process controlled by a specific nucleotide sequence abutting the exons. Almost all introns can be identified by specific consensus sequences. The first two bases of an intron are always GU, while the last two bases are always AG, but the 5' and 3' splice sites typically have consensus sequences that extend beyond the GU and AG motifs. Splicing of mRNA takes place on a particle called spliceosome, while tRNA and rRNA are spliced by mechanisms that do not involve spliceosomes.

Introns are typically much longer than exons (sequences that are present in the mRNA). An average eukaryotic exon is about 150 nucleotides long, while a single human intron can be as long as close to 500,000 nucleotides, but typically are about 2000-4000 nucleotides. In general, a eukaryotic gene contains much more intron than exon sequences, as illustrated by the following table (Molecular Biology of the Cell, Bruce Alberts et al., eds., 3rd edition, Garland Publishing Company, New York, N.Y., 1994, p. 340):

Table 1

Gene	Gene Size (x 10^3 nucleotides)	MRNA Size (x 10^3 nucleotides)	Number of Introns
β-globin	1.5	0.6	2
Insulin	1.7	0.4	2
Proteinase C	11	1.4	7
Albumin	25	2.1	14
Catalase	34	1.6	12
LDL receptor	45	5.5	17
Factor VIII	186	9	25
Thyroglobulin	300	8.7	36

In a particular embodiment of the present invention, intron sequences within a gene of interest are subjected to a selection process to identify intronic RNA sequence or sequences that co-express with exon RNA (i.e., mRNA) sequences of the same gene. Such selected intron sequences, the expression of which correlates with the expression of exon sequences, have especially desirable properties as potential diagnostic markers: (1) because of their favorable technical performance (specifically, optimizing assay specificity and sensitivity); and, (2) whatever biomedical importance attaches to the mRNA level of the gene is also attached to the cellular levels of intronic sequences. For example, high levels of an mRNA species that encodes a potent growth factor are likely to correlate with high rate of growth of a cell. Intronic sequences having cellular levels that correlate with mRNA levels of this same gene have the same likelihood to correlate with high growth rate of a cell. Such selected intronic sequences can then be used to screen valuable tissue specimens to search for clinical correlations and diagnostic, predictive or prognostic significance.

An exemplary process for selecting intron sequences that co-express with the mRNA of the same gene is as follows. Briefly, for any gene of interest, a set of relevant tissues from a population of patients of interest are assayed to measure the levels of a set of intronic and mRNA sequences. The intronic sequences found to have the highest Pearson correlation coefficient for co-expression with exon RNA (mRNA) sequences are then selected. The number of patients studied in this process is preferably at least above 50 and more preferably at least about 100.

In a specific example, the biomedical issue of interest regards patients with breast cancer and the gene of interest can be the tumor growth marker Ki-67. In this case, tumors from 50 or more breast cancer patients are used for measurement of Ki-67 mRNA levels and the levels of sequences from multiple Ki-67 introns, and the introns having the highest Pearson correlation coefficient for co-expressing with exon RNA are selected.

An advantage of this process is that the selection of the preferred intronic sequence can be carried out with tissue specimens that are relatively easily obtained and abundant (for example, specimens that lack valuable attached clinical records). Because such tissue can provide large amounts of RNA to screen, it will be possible to detect gene expression signals from even suboptimal probes. The highly sensitive and specific assays based on the selected intronic sequences then can be used to screen valuable tissue specimens, for example, specimens attached to important clinical information, such as disease recurrence, death, or response to defined therapeutic drugs or treatment regimens.

3. *Gene Expression Profiling Using Intron-based PCR Primer/Probe Sets*

At present, PCR primers and probes are designed based upon the mRNA or cDNA sequence, without considering the intron sequences. Indeed, introns are usually regarded as "packaging" material that is removed during splicing and. generally rapidly degraded.

The present invention is based on the unanticipated experimental finding that intron RNAs can be readily detected by RT-PCR, even using highly degraded RNA from fixed, paraffin-embedded tissue specimens. In particular, it has been found that in gene expression profiling for a given gene RT-PCR signals from intron-based probe/primer sets can be as large, or larger, than the signals from exon-based RT-PCR signals. While this finding is supported by a few recent findings with certain mRNA species, it is not in accord with the prevailing view that introns are very rapidly degraded following splicing (Thomas *et al.*, *J.*

Virol. 76:532-40 [2002]; Clement *et al., J. Biol. Chem.* 276:16919-30 [2001]; Sharp *et al., Ann. Rev. Biochem.* 55:1119-1150 [1986])..

Also unexpected, the experimental findings underlying the present invention indicate that intronic RNA can be used for gene expression profiling, because the tissue amounts of expressed intron and exon sequences tend to be correlated. This result is unanticipated because scant or no evidence exists that the ratio of the overall rate constants for synthesis and turnover of transcribed intron and exon sequences are similar. In fact, the scientific literature provides evidence for the complexity of pre-mRNA and spliced intron turnover. For example, pre-mRNA can exist in multiple kinetic pools (Elliott and Rosbash, *Exp. Cell Res.* 229:181-8 [1996]), with subpopulations containing intron RNAs that are not efficiently spliced out and are transported to the cytoplasm in "immature" mRNA species, where they can decay at rates different than nuclear intron RNA sequences (Wang *et al., Proc. Natl. Acad. Sci. USA* 94:4360-5 [1997]). Evidence exists that certain spliced intron RNAs enter the cytoplasm in lariat structure (Clement *et al., RNA* 5:206-20 [1999]).

Finally, data presented here indicate that intron sequences can serve as diagnostic or prognostic molecular markers. Examining four mRNAs previously demonstrated to be prognostic in cancer, it is shown that their corresponding intron sequences are also prognostic, and in the same directions as the parent transcribed exon sequences (i.e., either positively or negatively prognostic).

In brief, the approach of the invention has been demonstrated as follows. Co-pending application Serial No.60/412,049, filed on September 18, 2002, the entire disclosure of which is hereby expressly incorporated by reference, describes a set of genes that predict likelihood of breast cancer recurrence. In that study, the levels of transcribed exon sequences in fixed paraffin-embedded breast cancer tissue specimens from 146 patients were measured by RT-PCR using exon-based PCR primer/probe sets. In the study described here, RT-PCR assays were created to measure the levels of transcribed intron sequences within four of the previously identified marker genes, and then used to screen RNA from 60 fixed paraffin-embedded biopsy specimens (representing 60 different patients, a subset of the patients evaluated in the previous study). The data presented in the examples below show that for each gene the introns and exons are co-expressed, and that the introns predict risk of disease recurrence as predicted by the previous exon-based data.

3. Design of Intron-Based PCR Primers and Probes

According to one aspect of the present invention, PCR primers and probes are designed based upon intron sequences present in the gene to be amplified. Accordingly, the first step in the primer/probe design is the delineation of intron sequences within the genes. This can be done by publicly available software, such as the DNA BLAT software developed by Kent, W.J., *Genome Res.* 12(4):656-64 (2002), or by the BLAST software including its variations. Subsequent steps follow well established methods of PCR primer and probe design.

In order to avoid non-specific signals, it is important to mask repetitive sequences within the introns when designing the primers and probes. This can be easily accomplished by using the Repeat Masker program available on-line through the Baylor College of Medicine, which screens DNA sequences against a library of repetitive elements and returns a query sequence in which the repetitive elements are masked. The masked intron sequences can then be used to design primer and probe sequences using any commercially or otherwise publicly available primer/probe design packages, such as Primer Express (Applied Biosystems); MGB assay-by–design (Applied Biosystems); Primer3 (Steve Rozen and Helen J. Skaletsky (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology.* Humana Press, Totowa, NJ, pp 365-386).

The most important factors considered in PCR primer design include primer length, melting temperature (Tm), and G/C content, specificity, complementary primer sequences, and 3'-end sequence. In general, optimal PCR primers are generally 17-30 bases in length, and contain about 20-80%, such as, for example, about 50-60% G+C bases. Tm's between 50 and 80 °C, e.g. about 50 to 70 °C are typically preferred.

For further guidelines for PCR primer and probe design see, e.g. Dieffenbach, C.W. *et al.*, "General Concepts for PCR Primer Design" in: *PCR Primer, A Laboratory Manual,* Cold Spring Harbor Laboratory Press, New York, 1995, pp. 133-155; Innis and Gelfand, "Optimization of PCRs" in: *PCR Protocols, A Guide to Methods and Applications,* CRC Press, London, 1994, pp. 5-11; and Plasterer, T.N. Primerselect: Primer and probe design. *Methods Mol. Biol.* 70:520-527 (1997), the entire disclosures of which are hereby expressly incorporated by reference.

4. Applications

The methods of the present invention, and specifically, the intron-based PCR primers and probes herein, utility in all fields where amplification of a nucleic acid (including RNA, DNA and, in general, all oligo- and poly nucleotides) representing a gene or a gene fragment is required. Thus the PCR primers and probes designed in accordance with the present invention can be used to amplify individual genes, or multiple genes present in a biological sample for the purpose of gene expression profiling by any methodology including, without limitation, gene expression profiling relying on quantitative PCR (e.g. quantitative RT-PCR), and microarray analysis, and bead-based assays.

For example, in a specific embodiment of the microarray technique, PCR amplified inserts of cDNA clones are applied to a substrate in a dense array. Preferably at least 10,000 nucleotide sequences are applied to the substrate. The microarrayed genes, immobilized on the microchip at at least 10,000 elements each, are suitable for hybridization under stringent conditions. Fluorescently labeled cDNA probes may be generated through incorporation of fluorescent nucleotides by reverse transcription of RNA extracted from tissues of interest. Labeled cDNA probes applied to the chip hybridize with specificity to each spot of DNA on the array. After stringent washing to remove non-specifically bound probes, the chip is scanned by confocal laser microscopy or by another detection method, such as a CCD camera. Quantitation of hybridization of each arrayed element allows for assessment of corresponding mRNA abundance. With dual color fluorescence, separately labeled cDNA probes generated from two sources of RNA are hybridized pairwise to the array. The relative abundance of the transcripts from the two sources corresponding to each specified gene is thus determined simultaneously. The miniaturized scale of the hybridization affords a convenient and rapid evaluation of the expression pattern for large numbers of genes. Such methods have been shown to have the sensitivity required to detect rare transcripts, which are expressed at a few copies per cell, and to reproducibly detect at least approximately two-fold differences in the expression levels (Schena *et al.*, *Proc. Natl. Acad. Sci. USA* 93(2):106-149 (1996)). Microarray analysis can be performed by commercially available equipment, following manufacturer's protocols, such as by using the Affymetrix GenChip technology, or Agilent's microarray technology.

An important aspect of the present invention is to use intron-based gene amplification as part of gene expression profiling to match patients to best drugs or drug combinations, and to provide prognostic information. For example, the measured expression of genes in cancer tissue (e.g. biopsied breast cancer tissue) can be used to predict the likelihood of long-term, disease-free survival of patients following surgery and/or other cancer therapy, or to predict patient response to a particular therapeutic approach. For this purpose it is typically necessary to correct for (normalize away) both differences in the amount of RNA assayed and variability in the quality of the RNA used. Therefore, the assays of the invention usually measure and incorporate the expression of certain normalizing genes, including well known reference genes, such as GAPDH and Cyp1. Alternatively, normalization can be based on the mean or median signal (Ct) of all of the assayed genes or a large subset thereof (global normalization approach). On a gene-by-gene basis, the measured normalized amount of a patient tumor mRNA is compared to the amount found in a cancer, e.g. breast cancer tissue reference set. The number (N) of cancer, e.g. breast cancer, tissues in this reference set should be sufficiently high to ensure that different reference sets (as a whole) behave essentially the same way. If this condition is met, the identity of the individual breast cancer tissues present in a particular set will have no significant impact on the relative amounts of the genes assayed. Usually, the breast cancer tissue reference set consists of at least about 30, preferably at least about 40 different fixed, paraffin-embedded (FPE) breast cancer tissue specimens. Unless noted otherwise, normalized expression levels for each mRNA/tested tumor/patient will be expressed as a percentage of the expression level measured in the reference set. More specifically, the reference set of a sufficiently high number (e.g., 40) tumors yields a distribution of normalized levels of each mRNA species. The level measured in a particular tumor sample to be analyzed falls at some percentile within this range, which can be determined by methods well known in the art.

In a Phase II study of gene expression in paraffin-embedded, fixed tissue samples of invasive breast carcinoma, the overexpression of any of the following genes in the breast cancer tissue was found to indicate a reduced likelihood of survival without cancer recurrence following surgery: FOXM1; PRAME; SKT15, Ki-67; CA9; NME1; SURV; TFRC; YB-1; RPS6KB1; Src; Chk1; CCNB1; Chk2; CDC25B; CYP3A4; EpCAM; VEGFC; hENT1; BRCA2; EGFR; TK1; VDR.

In the same study, the overexpression of any of the following genes in breast cancer indicates a better prognosis for survival without cancer recurrence following surgery: Blc12; CEGP1; GSTM1; PR; BBC3; GATA3; DPYD; GSTM3; ID1; EstR1; p27; XIAP; IGF1R; AK055699; P13KC2A; TGFB3; BAGI1; pS2; WISP1; HNF3A; NFKBp65.

In this same Phase II study of gene expression in paraffin-embedded, fixed tissue samples of ER-positive breast cancer, overexpression of the following genes was indicative of a reduced likelihood of survival without cancer recurrence following surgery: PRAME; FOXM1; EPHX1; HIF1A; VEGFC; Ki-67; VDR; NME1. Some of these genes (PRAME; FOXM1; VEGFC; Ki-67; VDR; and NME1) were also identified as indicators of poor prognosis in the previous analysis, not limited to ER-positive breast cancer. The overexpression of the remaining genes (EPHX1 and HIF1A) was found to be negative indicator of disease free survival in ER-positive breast cancer only. Overexpression of the following genes in ER-positive cancer was found to be indicative of a better prognosis for survival without cancer recurrence following surgery: Bcl-2; DIABLO; IGF1R; GSTM3. Of the latter genes, Bcl-2; IGFR1; and GSTM3 have also been identified as indicators of good prognosis in the previous analysis, not limited to ER-positive breast cancer. The overexpression of DIABLO appeared to be positive indicator of disease free survival in ER-positive breast cancer only. For further details see, co-pending application Serial No. 60/427090, filed on November 15, 2002, the entire disclosure of which is hereby expressly incorporated by reference.

The studies described above were performed essentially as described in Example 2 below, except gene amplification was studied using exon-based amplicons. For further details, see copending application Serial No. 60/364,890 filed on March 13, 2002, the entire disclosure of which is hereby expressly incorporated by reference. As attested by the data set forth in Example 2, the data obtained using intron-based amplicons show excellent correlation with the earlier data, and typically provide the added benefit of increased sensitivity.

The findings of the previous Phase II study of invasive breast ductal carcinoma were subjected to multivariate stepwise analysis, using the Cox Proportional Hazards Model using the following equation:

$$RR=\exp[\mathrm{coef(geneA)} \times \mathrm{Ct(geneA)} + \mathrm{coef(geneB)} \times \mathrm{Ct(geneB)} + \mathrm{coef(geneC)} \times \mathrm{Ct(geneC)} + \ldots\ldots\ldots\ldots].$$

In this equation, coefficients for genes that are predictors of beneficial outcome are positive numbers and coefficients for genes that are predictors of unfavorable outcome are negative numbers. The "Ct" values in the equation are ΔCts, i.e. reflect the difference between the average normalized Ct value for a population and the normalized Ct measured for the patient in question. The convention used in the analysis has been that ΔCts below and above the population average have positive signs and negative signs, respectively (reflecting greater or lesser mRNA abundance). The relative risk (RR) calculated by solving this equation indicated if the patient has an enhanced or reduced chance of long-term survival without cancer recurrence.

In a multivariate analysis, using an interrogation set including a reduced number of genes, the following ten-gene sets have been identified as having particularly strong predictive value of patient survival without cancer recurrence following surgical removal of primary tumor.

1. Bcl2, cyclinG1, NFKBp65, NME1, EPHX1, TOP2B, DR5, TERC, Src, DIABLO;
2. Ki67, XIAP, hENT1, TS, CD9, p27, cyclinG1, pS2, NFKBp65, CYP3A4;
3. GSTM1, XIAP, Ki67, TS, cyclinG1, p27, CYP3A4, pS2, NFKBp65, ErbB3;
4. PR, NME1, XIAP, upa, cyclinG1, Contig51037, TERC, EPHX1, ALDH1A3, CTSL;
5. CA9, NME1, TERC, cyclinG1, EPHX1, DPYD, Src, TOP2B, NFKBp65, VEGFC;
6. TFRC, XIAP, Ki67, TS, cyclinG1, p27, CYP3A4, pS2, ErbB3, NFKBp65.

In a multivariate analysis, using an interrogation set including all genes identified, the following ten-gene sets have been identified as having particularly strong predictive value of patient survival without cancer recurrence following surgical removal of primary tumor.

1. Bcl2, PRAME, cyclinG1, FOXM1, NFKBp65, TS, XIAP, Ki67, CYP3A4, p27;
2. FOXM1, cyclinG1, XIAP, Contig51037, PRAME, TS, Ki67, PDGFRa, p27, NFKBp65;
3. PRAME, FOXM1, cyclinG1, XIAP, Contig51037, TS, Ki6, PDGFRa, p27, NFKBp65;
4. Ki67, XIAP, PRAME, hENT1, contig51037, TS, CD9, p27, ErbB3, cyclinG1;
5. STK15, XIAP, PRAME, PLAUR, p27, CTSL, CD18, PREP, p53, RPS6KB1;

6. GSTM1, XIAP, PRAME, p27, Contig51037, ErbB3, GSTp, EREG, ID1, PLAUR;
7. PR, PRAME, NME1, XIAP, PLAUR, cyclinG1, Contig51037, TERC, EPHX1, DR5;
8. CA9, FOXM1, cyclinG1, XIAP, TS, Ki67, NFKBp65, CYP3A4, GSTM3, p27;
9. TFRC, XIAP, PRAME, p27, Contig51037, ErbB3, DPYD, TERC, NME1, VEGFC;
10. CEGP1, PRAME, hENT1, XIAP, Contig51037, ErbB3, DPYD, NFKBp65, ID1, TS.

Using the same multivariate analysis approach for ER-positive breast cancer, the following ten-gene sets have been identified as having particularly strong predictive value of patient survival without cancer recurrence following surgical removal of primary tumor.

1. PRAME, p27, IGFBP2, HIF1A, TIMP2, ILT2, CYP3A4, ID1, EstR1, DIABLO;
2. Contig51037, EPHX1, Ki67, TIMP2, cyclinG1, DPYD, CYP3A4, TP, AIB1, CYP2C8;
3. Bcl2, hENT1, FOXM1, Contig51037, cyclinG1, Contig46653, PTEN, CYP3A4, TIMP2, AREG;
4. HIF1A, PRAME, p27, IGFBP2, TIMP2, ILT2, CYP3A4, ID1, EstR1, DIABLO;
5. IGF1R, PRAME, EPHX1, Contig51037, cyclinG1, Bcl2, NME1, PTEN, TBP, TIMP2;
6. FOXM1, Contig51037, VEGFC, TBP, HIF1A, DPYD, RAD51C, DCR3, cyclinG1, BAG1;
7. EPHX1, Contig51037, Ki67, TIMP2, cyclinG1, DPYD, CYP3A4, TP, AIB1, CYP2C8;
8. Ki67, VEGFC, VDR, GSTM3, p27, upa, ITGA7, rhoC, TERC, Pin1;
9. CDC25B, Contig51037, hENT1, Bcl2, HLAG, TERC, NME1, upa, ID1, CYP;
10. VEGFC, Ki67, VDR, GSTM3, p27, upa, ITGA7, rhoC, TERC, Pin1;
11. CTSB, PRAME, p27, IGFBP2, EPHX1, CTSL, BAD, DR5, DCR3, XIAP;
12. DIABLO, Ki67, hENT1, TIMP2, ID1, p27, KRT19, IGFBP2, TS, PDGFB;

13. p27, PRAME, IGFBP2, HIF1A, TIMP2, ILT2, CYP3A4, ID1, EstR1, DIABLO;
14. CDH1; PRAME, VEGFC; HIF1A; DPYD, TIMP2, CYP3A4, EstR1, RBP4, p27;
15. IGFBP3, PRAME, p27, Bcl2, XIAP, EstR1, Ki67, TS, Src, VEGF;
16. GSTM3, PRAME, p27, IGFBP3, XIAP, FGF2, hENT1, PTEN, EstR1, APC;
17. hENT1, Bcl2, FOXM1, Contig51037, CyclinG1, Contig46653, PTEN, CYP3A4, TIMP2, AREG;
18. STK15, VEGFC, PRAME, p27, GCLC, hENT1, ID1, TIMP2, EstR1, MCP1;
19. NME1, PRAM, p27, IGFBP3, XIAP, PTEN, hENT1, Bcl2, CYP3A4, HLAG;
20. VDR, Bcl2, p27, hENT1, p53, PI3KC2A, EIF4E, TFRC, MCM3, ID1;
21. EIF4E, Contig51037, EPHX1, cyclinG1, Bcl2, DR5, TBP, PTEN, NME1, HER2;
22. CCNB1, PRAME, VEGFC, HIF1A, hENT1, GCLC, TIMP2, ID1, p27, upa;
23. ID1, PRAME, DIABLO, hENT1, p27, PDGFRa, NME1, BIN1, BRCA1, TP;
24. FBXO5, PRAME, IGFBP3, p27, GSTM3, hENT1, XIAP, FGF2, TS, PTEN;
25. GUS, HIA1A, VEGFC, GSTM3, DPYD, hENT1, FBXO5, CA9, CYP, KRT18;
26. Bclx, Bcl2, hENT1, Contig51037, HLAG, CD9, ID1, BRCA1, BIN1, HBEGF.

In view of the excellent correlation between exon-based and intron-based gene expression profiling results (see Example 2), the same gene sets are expected to have similar prognostic value when gene expression profiling is based on the quantitation of RT-PCR signals from intron-based primer/probe sets.

Further details of the invention will be apparent from the following non-limiting examples.

Example 1

Design and Use of Intron-Specific PCR Primer/Probe Sets

RNA was extracted from formalin-fixed, paraffin-embedded (FPET) breast cancer biopsy specimens (Clinomics Biosciences Inc., Pittsfield, MA) as follows. Three 10 μM sections were cut and placed in a 1.5 ml tube. Paraffin was removed by xylene extraction (1ml, 3 times) followed by ethanol wash (1ml, twice). RNA was isolated from sectioned tissue blocks using the MasterPure^{TM} Purification kit (Epicentre, Madison, WI). RNA was quantitated by the RiboGreen Fluorescence method (Molecular Probes). Twenty FPET RNA samples were then pooled and used as described below.

First-strand cDNA was synthesized using Qiagen's Omniscript Reverse Transcriptase with pooled gene specific primers (reverse primers shown in Figure 2) random hexamers and RNase Inhibitor, using pooled FPET RNA (400 ng). A no reverse transcriptase (RT) reaction was also performed with 150 ng of pooled FPET RNA, sufficient RNA to perform the Taqman amplification at 5 ng/well.

Table 2

Reagents	RT Vol (μl)	No RT Vol (μl)	Final conc
10X Buffer RT	4	2	1X
dNTP mix, 5mM each dNTP	4	2	500μM each
ABI Random hexamer, 50μM	1	0.5	1.25μM
GSP pool,1μM	2	1	50nM
ABI RNase Inhibitor, 20U/μl	1	1	20U/rxn
Omniscript RT, 4U/μl	2	0	8U or 0U/rxn
Nuclease free water	10	5.5	
Pooled FPET RNA (164 ng/μl)	16	8	65.6 ng/μl
Total vol	40	20	

Reaction conditions: 37 °C, 60 min, 93°C, 5 min

TaqMan Assay

TaqMan assays for the 48-gene panel were carried out in triplicate wells with reaction volume of 25 μl and RNA input of 5 ng per assay. A "no RT " reaction for each gene was carried out in a single well as a control to verify that RNA rather than DNA signals were being measured. Real time quantitation was performed on the ABI 7700 using the following parameters:

Cycling conditions: 95°C, 10 min for one cycle, 95°C, 20 sec followed by 60°C, 45 sec, 40 cycles.

Volume reaction: 25 µl.

Dye layer setting: FAM, (the passive reference is ROX)

Results

Intron specific Taqman primer-probe sets were designed based on masked introns of CEGP1, FOXM1, PRAME and STK15.genes, To delineate intron sequences within the genes, the NCBI reference sequence for each mRNA (NM_XXXXXX) was aligned to the human genome using the BLAST-like alignment tool (BLAT) program available at the University of Santa Cruz on-line genome resource site (http://genome.ucsc.edu). Intron sequences were then searched for repetitive sequences using the Repeat Masker program available on-line through the Baylor College of Medicine (http://searchlauncher.bcm.tmc.edu/seq-util/seq-util.html). Repeat sequences, such as Alu repeats, are identified by this program and masked. It is important to exclude these sequences prior to designing primer-probes because they yield strong, non-specific signals. The masked intron sequences (Figures 1A-M) were then used to design Taqman primer-probe sets using Primer Express (ABI). Other programs suitable for primer-probe sets include, for example, the newer primer probe design program for MGB assays-by-design (ABI). The amplicons for each primer-probe set are delineated in bold font in Figure 1. Each specific primer-probe set is shown in Figure 2.

The intron-specific primer-probe sets (test genes) were used together with their corresponding exon-specific primer-probe set (references gene) in standard Taqman gene expression profile experiments using pooled FPET RNA. Normalized expression was calculated by the formula $2^{\Delta Ct}$ where ΔCt is the difference between the Cts of the test gene primer-probe set and the reference gene primer-probe sets [Ct (reference)-Ct (test)].

Example 2

A Phase II Study of Gene Expression in Premalignant and Malignant Breast Tumors

A gene expression study was designed and conducted with the primary goal to molecularly characterize gene expression in paraffin-embedded, fixed tissue samples of invasive breast ductal carcinoma, and to explore the correlation between such molecular profiles and disease-free survival.

Study design

Molecular assays were performed on paraffin-embedded, formalin-fixed primary breast tumor tissues obtained from 60 individual patients diagnosed with breast cancer. All patients underwent surgery with diagnosis of invasive carcinoma of the breast. Patients were included in the study only if histopathologic assessment, performed as described in the Materials and Methods section, indicated adequate amounts of tumor tissue and homogeneous pathology.

Materials and Methods

Each representative tumor block was characterized by standard histopathology for diagnosis, semi-quantitative assessment of amount of tumor, and tumor grade. A total of 6 sections (10 microns in thickness each) were prepared and placed in two Costar Brand Microcentrifuge Tubes (Polypropylene, 1.7 mL tubes, clear; 3 sections in each tube). If the tumor constituted less than 30% of the total specimen area, the sample may have been crudely dissected by the pathologist, using gross microdissection, putting the tumor tissue directly into the Costar tube.

If more than one tumor block was obtained as part of the surgical procedure, all tumor blocks were subjected to the same characterization, as described above, and the block most representative of the pathology was used for analysis.

Gene Expression Analysis

mRNA was extracted and purified from fixed, paraffin-embedded tissue samples, and prepared for gene expression analysis as described above.

Molecular assays of quantitative gene expression were performed by RT-PCR, using the ABI PRISM 7900™ Sequence Detection System™ (Perkin-Elmer-Applied Biosystems, Foster City, CA, USA). ABI PRISM 7900™ consists of a thermocycler, laser, charge-coupled device (CCD), camera and computer. The system amplifies samples in a 384-well format on a thermocycler. During amplification, laser-induced fluorescent signal is collected in real-time through fiber optics cables for all 384 wells, and detected at the CCD. The system includes software for running the instrument and for analyzing the data.

Analysis and Results

Tumor tissue was analyzed for expression of 48 different RNA sequences representing products of 37 different genes. The threshold cycle (Ct) values for each patient were normalized based on the median of all genes for that particular patient. Clinical outcome data were available for all patients from a review of registry data and selected patient charts.

Outcomes were classified as:

0 died due to breast cancer or to unknown cause or alive with breast cancer recurrence;

1 alive without breast cancer recurrence or died due to a cause other than breast cancer

Analysis was performed by:

Analysis of the relationship between normalized gene expression and the time to outcome (0 or 1 as defined above) where patients who were alive without breast cancer recurrence or who died due to a cause other than breast cancer were censored. This approach was used to evaluate the prognostic impact of individual genes and also sets of multiple genes.

For each gene a Cox Proportional Hazards model (see, e.g. Cox, D. R., and Oakes, D. (1984), *Analysis of Survival Data*, Chapman and Hall, London, New York) was defined with time to recurrence or death as the dependent variable, and the expression level of the gene as the independent variable. The genes that have a p-value < 0.05 in the Cox model were identified. For each gene, the Cox model provides the relative risk (RR) of recurrence or death for a unit change in the expression of the gene. One can choose to partition the patients into subgroups at any threshold value of the measured expression (on the Ct scale), where all patients with expression values above the threshold have higher risk, and all patients with expression values below the threshold have lower risk, or vice versa, depending on whether the gene is an indicator of poor (RR>1.01) or good (RR<1.01) prognosis. Thus, any threshold value will define subgroups of patients with respectively increased or decreased risk.

Table 3, below, shows pairwise correlation of expression (presented by correlation coefficients) between the tested introns and exons for the genes CEGP1, FOXM1, PRAME, and STK15. For two of the four genes, CEGP1 and PRAME, introns were found that yielded correlation coefficients [for co-expression with their respective exons] above 0.90. In the case of STK15, one intron correlated with exon expression with a correlation coefficient ~0.80. For FOXM1, intron:exon expression correlations were significantly lower. In this last case, however, it seems likely that actual expression may be highly correlated but not detectable for a technical reason. Expression of the FOXM1 exon in many patients was beneath the detection threshold of the assay, which potentially prevents detection of high correlations that may exist. If this hypothesis is correct, FOXM1 introns would still register as negative clinical prognostic markers as previously demonstrated for FOXM1. As shown later, this result occurs.

Figures 3, 4 and 5 show that the pairwise correlation of expression of the tested RNAs against CEGP1, PRAME and STK15 exon RNAs. As shown, respective introns of these genes yielded the highest correlations. It is noteworthy that the panel of 48 genes included genes that we selected, by several bioinformatics-based strategies, as particularly likely to correlate in expression with CEGP1, PRAME, STK15, and FOXM1. Those non-intron-based strategies were most successful in the case of STK15, as several candidate genes had expression correlation coefficients in the range of 0.6-0.7.

Table 3. Correlations between Intron and Exon Expression for Four Genes

	ExpressionCorrelation Coefficient {R}				
	CEGP1intron1.1	*CEGP1intron3.1*	*CEGP1intron4.1*	*CEGP1intron5.1*	*CEGP1.2*
CEGP1intron1.1	1.00				
CEGP1intron3.1	0.89	1.00			
CEGP1intron4.1	0.97	0.82	1.00		
CEGP1intron5.1	0.91	0.87	0.88	1.00	
CEGP1.2	0.91	0.80	0.90	0.87	1.00
	FOXM1intron3.3	FOXM1intron5.1	FOXM1intron7.1	FOXM1.1	
FOXM1intron3.3	1.00				
FOXM1intron5.1	0.48	1.00			
FOXM1intron7.1	0.54	0.73	1.00		
FOXM1.1	0.44	0.33	0.38	1.00	
	STK15intron1.1	STK15intron2.1	STK15intron4.1	STK15.2	
STK15intron1.1	1.00				
STK15intron2.1	0.78	1.00			
STK15intron4.1	0.69	0.74	1.00		
STK15.2	0.63	0.70	0.78	1.00	
	PRAMEintron2.1	PRAME.3			
PRAMEintron2.1	1.00				
PRAME.3	0.97	1.00			

Table 4, below, shows the impact upon patient survival of expression of CEGP1, FOXM1, PRAME, and STK15, exons and introns. The parent exons all had statistically significant impact on relative risk [RR], as we previously determined, except in the case of FOXM1. Because the present study evaluated 60 patients from the original 146 patient group, the FOXM1 marker may have fallen from significance because the statistical hazard of examining a reduced data set. Very notably, for all four tested genes, intron expression significantly impacted RR, and in the same direction as the parent exons.

Table 4. Cox Model Results for 60 Patients with Breast Cancer

Prognostic Correlations

Gene	Coef	RR=exp(coef)	se(coef)	z	p
CEGP1.2	-0.202	0.817	0.050	-4.024	0.00006
CEGP1intron1.1	-0.329	0.720	0.087	-3.771	0.00016
CEGP1intron3.1	-0.261	0.770	0.078	-3.335	0.00085
CEGP1intron4.1	-0.275	0.760	0.073	-3.774	0.00016
CEGP1intron5.1	-0.312	0.732	0.082	-3.817	0.00014
FOXM1.1	0.175	1.192	0.136	1.289	0.19700
FOXM1intron3.3	0.304	1.355	0.120	2.523	0.01160
FOXM1intron5.1	0.514	1.673	0.195	2.639	0.00832
FOXM1intron7.1	0.546	1.726	0.182	2.993	0.00276
PRAME.3	0.125	1.133	0.054	2.294	0.02180
PRAMEintron2.1	0.125	1.133	0.052	2.397	0.01650
STK15.2	0.692	1.998	0.201	3.450	0.00056
STK15intron1.1	0.357	1.429	0.149	2.400	0.01640
STK15intron2.1	0.391	1.479	0.154	2.536	0.01120
STK15intron4.1	0.410	1.506	0.133	3.084	0.00204

A common perception exists that steady state levels of transcribed exon sequences greatly exceed those of transcribed intron sequences (Sharp *et al. Ann. Rev. Biochem.* 55: 1119-50 [1986]). Nevertheless, our examination of CEGP1, FoxM1, PRAME and STK15 exon and intron expression, using TaqMan[TM] RT-PCR to assay RNA from fixed paraffin-embedded breast cancer tissue, demonstrated that intron and exon signal intensities were in the same range, and in all cases in the useful detection range of the assay [data not shown]. The detection of intronic RNA in this study is all the more notable because the tissue used was fixed in formalin, which degrades RNA, and thus substantially limits the ability to detect RNA (T.E. Godfrey et al. J. Mol. Diag. 2: 84-91 [2000]). In the case of CEGP1 three of the tested introns yielded lower signals and one a higher signal than the exon. In the case of FOXM1, five of nine tested introns yielded higher signals than the exon. In the case of PRAME signal intensities from the tested intron and exon were nearly identical. Finally, for STK15 all introns had signal intensities that were 1/4 to 1/20 those of the exon, but were still in the useful range of the assay. Thus, these results indicate that steady state levels of expressed introns are adequate for use of intron RNAs as molecular markers.

All references cited throughout the disclosure, including the examples, are hereby expressly incorporated by reference for their entire disclosure.

While the present invention has been described with reference to what is considered to be specific embodiments, it is to be understood that the invention is not so limited. To the contrary, the invention is intended to cover various modifications and equivalents included within the spirit and scope of the appended claims. For example, while the disclosure includes various breast cancer-associated genes and gene sets, similar genes and gene sets and methods concerning other types of cancer are specifically within the scope herein.

WHAT IS CLAIMED IS:

1. A method for monitoring gene expression in a biological sample, comprising:

(a) providing a polynucleotide complementary to an intronic RNA sequence within a target gene, wherein the expression of said intronic RNA sequence correlates with the expression of an exonic mRNA sequence within said gene;

(b) hybridizing said polynucleotide to said intronic RNA sequence to form a polynucleotide-intronic RNA complex; and

(c) detecting the polynucleotide-intronic RNA complex.

2. The method of claim 1 wherein said intronic RNA sequence is selected by identifying intronic sequences which are co-expressed with the mRNA of said target gene, and selecting an intronic RNA sequence having the highest correlation coefficient for said co-expression.

3. The method of claim 1 wherein said intronic RNA sequence is at least 50 nucleotide bases long.

4. The method of claim 1 wherein said intronic RNA sequence is at least 55 nucleotide bases long.

5. The method of claim 1 wherein said intronic RNA sequence is at least 60 nucleotide bases long.

6. The method of claim 1 wherein said biological sample is a tissue sample.

7. The method of claim 6 wherein said tissue is a tumor tissue.

8. The method of claim 7 wherein said tumor is cancer.

9. The method of claim 8 wherein said cancer is selected from the group consisting of breast cancer, colon cancer, lung cancer, prostate cancer, hepatocellular cancer, gastric cancer, pancreatic cancer, cervical cancer, ovarian cancer, liver cancer, bladder cancer, cancer of the urinary tract, thyroid cancer, renal cancer, carcinoma, melanoma, and brain cancer.

10. The method of claim 6 wherein said tissue sample is a fixed, wax-embedded tissue sample.

11. The method of claim 10 wherein said tissue sample is a formalin-fixed, paraffin-embedded tissue sample.

12. The method of claim 10 wherein said exonic RNA is fragmented.

13. The method of claim 1 wherein said biological sample is a biological fluid,

14. The method of claim 1 wherein said hybridization is performed under stringent conditions.

15. The method of claim 1 further comprising the step of quantifying the expression of said intronic RNA.

16. The method of claim 1 wherein said polynucleotide is a single-stranded oligonucleotide.

17. The method of claim 16 wherein said single-stranded oligonucleotide is a PCR probe.

18. The method of claim 17 wherein said PCR probe is designed based upon a unique sequence within said intronic RNA sequence.

19. The method of claim 16 wherein said single-stranded oligonucleotide is a PCR primer.

20. The method of claim 19 wherein said PCR primer is designed based upon a unique sequence within said intronic RNA sequence.

21. The method of claim 1 wherein the expression of more than one target gene is monitored.

22. The method of claim 21 comprising simultaneous monitoring of a least 50 target genes.

23. The method of claim 21 comprising simultaneous monitoring of at least 100 target genes.

24. The method of claim 21 comprising simultaneous monitoring of at least 500 target genes.

25. The method of claim 21 comprising simultaneous monitoring of at least 10,000 target genes.

26. The method of claim 25 wherein intronic RNA sequences corresponding to a plurality of said target genes are displayed as an array immobilized on a solid surface.

27. The method of claim 1 wherein said target gene is selected from the genes listed in Figure 6.

28. A method of preparing a single-stranded oligonucleotide molecule for amplification of a target gene comprising:

(a) identifying at least one intron sequence within said target gene, the expression of which correlates with the expression of an exonic mRNA sequence within said target gene; and

(b) preparing a single-stranded oligonucleotide molecule that corresponds to at least a portion of the transcribed intron sequence.

29. The method of claim 28 comprising identifying repeat sequences within said intron sequence prior to preparing said single-stranded oligonucleotide molecule.

30. The method of claim 29 wherein said repeat sequences are masked prior to preparing said oligonucleotide molecule.

31. The method of claim 28 wherein said single-stranded oligonucleotide molecule is a PCR primer.

32. The method of claim 31 wherein said PCR primer is a forward primer designed to comprise 5'-sequences of a target sequence within said transcribed intron sequence.

33. The method of claim 31 herein said PCR primer is a reverse primer designed to complement 5'-sequences of a target sequence downstream of the forward primer within said transcribed intron sequence.

34. The method of claim 32 or 33 wherein said target sequence is at least 50 nucleotide bases long.

35. The method of claim 32 or 33 wherein said target sequence is at least 55 nucleotide bases long.

36. The method of claim 32 or 33 wherein said target sequence is at least 60 nucleotide bases long.

37. The method of claim 31 wherein said PCR primer is 17-30 nucleotide bases in length.

38. The method of claim 31 wherein said PCR primer contains about 20% to 80% G+C bases.

39. The method of claim 31 wherein the melting temperature (Tm) of said PCR primer is between about 50 °C and about 70 °C.

40. The method of claim 28 wherein said single-stranded oligonucleotide molecule is a PCR probe.

41. The method of claim 40 wherein said PCR probe is designed to comprise or complement an internal portion of a target sequence within the transcribed intron sequence.

42. The method of claim 41 wherein said PCR probe is labeled with a reporter fluorescent dye and a quencher fluorescent dye.

43. The method of claim 28 wherein said target gene is selected from the genes listed in Figure 6.

44. A method for amplifying intronic RNA in a fixed paraffin-embedded tissue sample representing at least one gene of interest, comprising the steps of:

(a) contacting DNA obtained by reverse transcription of intronic RNA, the expression of which correlates with the expression of a corresponding exonic RNA, with at least one set of PCR primers and probe corresponding to said intronic RNA; and

(b) performing PCR amplification.

45. The method of claim 44 wherein said PCR primers and probe are designed based upon a unique sequence within said intronic RNA.

46. The method of claim 44 wherein said sample comprises fragmented RNA representing multiple genes of interest.

47. The method of claim 46 wherein said sample is contacted with a pool of PCR primers and probes designed based upon unique sequences within introns, the expression of which correlates with the expression of corresponding exons, present in said genes of interest.

48. The method of claim 47 wherein said pool comprises at least one of the intron-based primer/probe sets set forth in Figure 2.

49. The method of claim 47 wherein said pool comprises at least one forward or reverse primer or probe set forth in Figure 2.

50. The method of claim 46 wherein said tissue sample is from a tumor biopsy.

51. The method of claim 50 wherein said tumor biopsy is obtained from a human patient.

52. The method of claim 51 wherein said tumor is breast cancer.

53. The method of claim 51 wherein said tumor is lung cancer.

54. The method of claim 51 wherein said tumor is colorectal cancer.

55. The method of claim 51 further comprising the step of determining the expression levels of the RNA transcripts of said genes of interest or their expression products.

56. The method of claim 55 wherein differential expression of said RNA transcripts or their products is correlated with predicted patient response to treatment or patient survival.

57. The method of claim 44 wherein said gene of interest is selected from the genes listed in Figure 6.

58. The method of claim 52 wherein the tumor is invasive breast cancer, and the method comprises

(1) determining the expression levels of the RNA transcripts or expression products of a gene or gene set selected from the group consisting of:

(a) Bcl2, cyclinG1, NFKBp65, NME1, EPHX1, TOP2B, DR5, TERC, Src, DIABLO;

(b) Ki67, XIAP, hENT1, TS, CD9, p27, cyclinG1, pS2, NFKBp65, CYP3A4;

(c) GSTM1, XIAP, Ki67, TS, cyclinG1, p27, CYP3A4, pS2, NFKBp65, ErbB3;

(d) PR, NME1, XIAP, upa, cyclinG1, Contig51037, TERC, EPHX1, ALDH1A3, CTSL;

(e) CA9, NME1, TERC, cyclinG1, EPHX1, DPYD, Src, TOP2B, NFKBp65, VEGFC;

(f) TFRC, XIAP, Ki67, TS, cyclinG1, p27, CYP3A4, pS2, ErbB3, NFKBp65;

(g) Bcl2, PRAME, cyclinG1, FOXM1, NFKBp65, TS, XIAP, Ki67, CYP3A4, p27;

(h) FOXM1, cyclinG1, XIAP, Contig51037, PRAME, TS, Ki67, PDGFRa, p27, NFKBp65;

(i) PRAME, FOXM1, cyclinG1, XIAP, Contig51037, TS, Ki6, PDGFRa, p27, NFKBp65;

(j) Ki67, XIAP, PRAME, hENT1, contig51037, TS, CD9, p27, ErbB3, cyclinG1;

(k) STK15, XIAP, PRAME, PLAUR, p27, CTSL, CD18, PREP, p53, RPS6KB1;

(l) GSTM1, XIAP, PRAME, p27, Contig51037, ErbB3, GSTp, EREG, ID1, PLAUR;

(m) PR, PRAME, NME1, XIAP, PLAUR, cyclinG1, Contig51037, TERC, EPHX1, DR5;

(n) CA9, FOXM1, cyclinG1, XIAP, TS, Ki67, NFKBp65, CYP3A4, GSTM3, p27;

(o) TFRC, XIAP, PRAME, p27, Contig51037, ErbB3, DPYD, TERC, NME1, VEGFC; and

(p) CEGP1, PRAME, hENT1, XIAP, Contig51037, ErbB3, DPYD, NFKBp65, ID1, TS

in said sample;

(2) subjecting the data obtained in step (a) to statistical analysis; and

(3) determining whether the likelihood of long-term survival of said patient, without the recurrence of breast cancer has increased or decreased.

59. The method of claim 58 wherein the expression levels of said RNA transcripts or their expression products are normalized against the expression levels of all RNA transcripts or their expression products in said breast cancer tissue sample, or of a reference set of RNA transcripts or their products.

60. The method of claim 52 wherein the tumor is estrogen receptor (ER)-positive invasive breast cancer, and the method comprises

(1) determining the expression levels of the RNA transcripts or expression products of a gene or gene set selected from the group consisting of:

(a) PRAME, p27, IGFBP2, HIF1A, TIMP2, ILT2, CYP3A4, ID1, EstR1, DIABLO;

(b) Contig51037, EPHX1, Ki67, TIMP2, cyclinG1, DPYD, CYP3A4, TP, AIB1, CYP2C8;

(c) Bcl2, hENT1, FOXM1, Contig51037, cyclinG1, Contig46653, PTEN, CYP3A4, TIMP2, AREG;

(d) HIF1A, PRAME, p27, IGFBP2, TIMP2, ILT2, CYP3A4, ID1, EstR1, DIABLO;

(e) IGF1R, PRAME, EPHX1, Contig51037, cyclinG1, Bcl2, NME1, PTEN, TBP, TIMP2;

(f) FOXM1, Contig51037, VEGFC, TBP, HIF1A, DPYD, RAD51C, DCR3, cyclinG1, BAG1;

(g) EPHX1, Contig51037, Ki67, TIMP2, cyclinG1, DPYD, CYP3A4, TP, AIB1, CYP2C8;

(h) Ki67, VEGFC, VDR, GSTM3, p27, upa, ITGA7, rhoC, TERC, Pin1;

(i) CDC25B, Contig51037, hENT1, Bcl2, HLAG, TERC, NME1, upa, ID1, CYP;

(j) VEGFC, Ki67, VDR, GSTM3, p27, upa, ITGA7, rhoC, TERC, Pin1;

(k) CTSB, PRAME, p27, IGFBP2, EPHX1, CTSL, BAD, DR5, DCR3, XIAP;

(l) DIABLO, Ki67, hENT1, TIMP2, ID1, p27, KRT19, IGFBP2, TS, PDGFB;

(m) p27, PRAME, IGFBP2, HIF1A, TIMP2, ILT2, CYP3A4, ID1, EstR1, DIABLO;

(n) CDH1; PRAME, VEGFC; HIF1A; DPYD, TIMP2, CYP3A4, EstR1, RBP4, p27;

(o) IGFBP3, PRAME, p27, Bcl2, XIAP, EstR1, Ki67, TS, Src, VEGF;

(p) GSTM3, PRAME, p27, IGFBP3, XIAP, FGF2, hENT1, PTEN, EstR1, APC;

(q) hENT1, Bcl2, FOXM1, Contig51037, CyclinG1, Contig46653, PTEN, CYP3A4, TIMP2, AREG;

(r) STK15, VEGFC, PRAME, p27, GCLC, hENT1, ID1, TIMP2, EstR1, MCP1;

(s) NME1, PRAM, p27, IGFBP3, XIAP, PTEN, hENT1, Bcl2, CYP3A4, HLAG;

(t) VDR, Bcl2, p27, hENT1, p53, PI3KC2A, EIF4E, TFRC, MCM3, ID1;

(u) EIF4E, Contig51037, EPHX1, cyclinG1, Bcl2, DR5, TBP, PTEN, NME1, HER2;

(v) CCNB1, PRAME, VEGFC, HIF1A, hENT1, GCLC, TIMP2, ID1, p27, upa;

(w) ID1, PRAME, DIABLO, hENT1, p27, PDGFRa, NME1, BIN1, BRCA1, TP;

(x). FBXO5, PRAME, IGFBP3, p27, GSTM3, hENT1, XIAP, FGF2, TS, PTEN;

(y) GUS, HIA1A, VEGFC, GSTM3, DPYD, hENT1, EBXO5, CA9, CYP, KRT18; and

(z) Bclx, Bcl2, hENT1, Contig51037, HLAG, CD9, ID1, BRCA1, BIN1, HBEGF;

(2) subjecting the data obtained in step (1) to statistical analysis; and

(3) determining whether the likelihood of long-term survival of said patient, without the recurrence of breast cancer has increased or decreased.

61. The method of claim 60 wherein the expression levels of said RNA transcripts or their expression products are normalized against the expression levels of all RNA transcripts or their expression products in said breast cancer tissue sample, or of a reference set of RNA transcripts or their products.

62. The method of claim 52 wherein the tumor is breast cancer, and the method comprises

(1) determining the expression levels of the RNA transcripts or expression products of a gene or gene set selected from the group consisting of: FOXM1; PRAME; SKT15, Ki-67; CA9; NME1; SURV; TFRC; YB-1; RPS6KB1; Src; Chk1; CCNB1; Chk2; CDC25B; CYP3A4; EpCAM; VEGFC; hENT1; BRCA2; EGFR; TK1; VDR; Blc12; CEGP1; GSTM1; PR; BBC3; GATA3; DPYD; GSTM3; ID1; EstR1; p27; XIAP; IGF1R; AK055699; P13KC2A; TGFB3; BAGI1; pS2; WISP1; HNF3A; and NFKBp65, normalized against the expression levels of all RNA transcripts or their products in said sample, or of a reference set of RNA transcripts or their expression products;

(2) subjecting the data obtained in step (a) to statistical analysis; and

(3) determining whether the likelihood of long-term survival of said patient, without the recurrence of breast cancer has increased or decreased.

63. The method of claim 62 wherein the expression levels of said RNA transcripts or their expression products are normalized against the expression levels of all RNA transcripts or their expression products in said breast cancer tissue sample, or of a reference set of RNA transcripts or their products.

64. The method of claim 52 wherein the tumor is invasive breast cancer, and the method comprises determining the expression levels of RNA transcripts or expression products of a gene or gene set from the group consisting of:

(a) p53BP2, Bcl2, BAD, EPHX1, PDGFRβ, DIABLO, XIAP, YB1, CA9, and KRT8;

(b) GRB7, CD68, TOP2A, Bcl2, DIABLO, CD3, ID1, PPM1D, MCM6, and WISP1;

(c) PR, p53BP2, PRAME, DIABLO, CTSL, IGFBP2, TIMP1, CA9, MMP9, and COX2;

(d) CD68, GRB7, TOP2A, Bcl2, DIABLO, CD3, ID1, PPM1D, MCM6, and WISP1;

(e) Bcl2, p53BP2, BAD, EPHX1, PDGFRβ, DIABLO, XIAP, YB1, CA9, and KRT8;

(f) KRT14, KRT5, PRAME, p53BP2, GUS1, AIB1, MCM3, CCNE1, MCM6, and ID1;

(g) PRAME, p53BP2, EstR1, DIABLO, CTSL, PPM1D, GRB7, DAPK1, BBC3, and VEGFB;

(h) CTSL2, GRB7, TOP2A, CCNB1, Bcl2, DIABLO, PRAME, EMS1, CA9, and EpCAM;

(i) EstR1, p53BP2, PRAME, DIABLO, CTSL, PPM1D, GRB7, DAPK1, BBC3, and VEGFB;

(j) Chk1, PRAME, p53BP2, GRB7, CA9, CTSL, CCNB1, TOP2A, tumor size, and IGFBP2;

(k) IGFBP2, GRB7, PRAME, DIABLO, CTSL, β-Catenin, PPM1D, Chk1, WISP1, and LOT1;

(l) HER2, p53BP2, Bcl2, DIABLO, TIMP1, EPHX1, TOP2A, TRAIL, CA9, and AREG;

(m) BAG1, p53BP2, PRAME, IL6, CCNB1, PAI1, AREG, tumor size, CA9, and Ki67;

(n) CEGP1, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, STK15, and AKT2, and FGF18;

(o) STK15, p53BP2, PRAME, IL6, CCNE1, AKT2, DIABLO, cMet, CCNE2, and COX2;

(p) KLK10, EstR1, p53BP2, PRAME, DIABLO, CTSL, PPM1D, GRB7, DAPK1, and BBC3;

(q) AIB1, p53BP2, Bcl2, DIABLO, TIMP1, CD3, p53, CA9, GRB7, and EPHX1

(r) BBC3, GRB7, CD68, PRAME, TOP2A, CCNB1, EPHX1, CTSL GSTM1, and APC;

(s) CD9, GRB7, CD68, TOP2A, Bcl2, CCNB1, CD3, DIABLO, ID1, and PPM1D;

(t) EGFR, KRT14, GRB7, TOP2A, CCNB1, CTSL, Bcl2, TP, KLK10, and CA9;

(u) HIF1α, PR, DIABLO, PRAME, Chk1, AKT2, GRB7, CCNE1, TOP2A, and CCNB1;

(v) MDM2, p53BP2, DIABLO, Bcl2, AIB1, TIMP1, CD3, p53, CA9, and HER2;

(w) MYBL2, p53BP2, PRAME, IL6, Bcl2, DIABLO, CCNE1, EPHX1, TIMP1, and CA9;

(x) p27, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, STK15, AKT2, and ID1;

(y) RAD51, GRB7, CD68, TOP2A, CIAP2, CCNB1, BAG1, IL6, FGFR1, and p53BP2;

(z) SURV, GRB7, TOP2A, PRAME, CTSL, GSTM1, CCNB1, VDR, CA9; and CCNE2;

(aa) TOP2B, p53BP2, DIABLO, Bcl2, TIMP1, AIB1, CA9, p53, KRT8, and BAD;

(ab) ZNF217, GRB7, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, APC4, and β-Catenin,

in a breast cancer tissue sample obtained from said patient,

(2) subjecting the data obtained in step (a) to statistical analysis; and

(3) determining whether the likelihood of said long-term survival has increased or decreased.

65. The method of claim 64 wherein the expression levels of said RNA transcripts or their expression products are normalized against the expression levels of all RNA transcripts or their expression products in said breast cancer tissue sample, or of a reference set of RNA transcripts or their products.

66. A method for measuring gene expression using an array comprising a plurality of polynucleotides hybridizing to target genes of interest immobilized on a solid surface, wherein at least one of the said polynucleotides comprises an intron-based sequence the expression of which correlates with the expression of a corresponding exon sequence.

67. The method of claim 66 wherein all of said polynucleotides comprise intron sequences.

68. The method of claim 66 comprising at least one of the amplicons shown in Figures 1A-M, or the complement thereof.

69. The method of claim 66 comprising two or more of the amplicons shown in Figures 1A-M, or the complement thereof.

70. The method of claim 66 comprising all of the amplicons shown in Figures 1A-M, or the complement thereof.

71. The method of claim 66 comprising using intron-based polynucleotide sequences hybridizing to at least one gene of interest selected from the group consisting of: FOXM1, PRAME, Bcl2, STK15, CEGP1, Ki-67, GSTM1, PR, BBC3, NME1, SURV, GATA3, TFRC, YB-1, DPYD, CA9, Contig51037, RPS6K1 and Her2, wherein at least 80% of the sequences on said .array are intron-based.

72. The method of claim 71 comprising using intron-based polynucleotide sequences hybridizing to at least 5 of said genes.

73. The method of claim 71 comprising using intron-based polynucleotide sequences hybridizing to at least 10 of said genes.

74. The method of claim 71 comprising using intron-based polynucleotide sequences hybridizing to all of said genes.

75. The method of claim 66 comprising using intron-based polynucleotide sequences hybridizing to at least one gene of interest selected from the group consisting of: FOXM1, PRAME, Bcl2, STK15, CEGP1, Ki-67, GSTM1, CA9, PR, BBC3, NME1, SURV, GATA3, TFRC, YB-1, DPYD, GSTM3, RPS6KB1, Src, Chk1, ID1, EstR1, p27, CCNB1, XIAP, Chk2, CDC25B, IGF1R, AK055699, P13KC2A, TGFB3, BAGI1, CYP3A4, EpCAM, VEGFC, pS2, hENT1, WISP1, HNF3A, NFKBp65, BRCA2, EGFR, TK1, VDR, Contig51037, pENT1, EPHX1, IF1A, CDH1, HIF1α, IGFBP3, CTSB, Her2 and DIABLO.

76. The method of claim 75 comprising using intron-based polynucleotide sequences hybridizing to at least 5 of said genes.

77. The method of claim 75 comprising using intron-based polynucleotide sequences hybridizing to at least 10 of said genes.

78. The arramethod of claim 75 comprising using intron-based polynucleotide sequences hybridizing to all of said genes.

79. The method of claim 66 comprising using intron-based polynucleotide sequences hybridizing to at least one gene set selected from the group consisting of:

(a) Bcl2, cyclinG1, NFKBp65, NME1, EPHX1, TOP2B, DR5, TERC, Src, DIABLO;

(b) Ki67, XIAP, hENT1, TS, CD9, p27, cyclinG1, pS2, NFKBp65, CYP3A4;

(c) GSTM1, XIAP, Ki67, TS, cyclinG1, p27, CYP3A4, pS2, NFKBp65, ErbB3;

(d) PR, NME1, XIAP, upa, cyclinG1, Contig51037, TERC, EPHX1, ALDH1A3, CTSL;

(e) CA9, NME1, TERC, cyclinG1, EPHX1, DPYD, Src, TOP2B, NFKBp65, VEGFC;

(f) TFRC, XIAP, Ki67, TS, cyclinG1, p27, CYP3A4, pS2, ErbB3, NFKBp65;

(g) Bcl2, PRAME, cyclinG1, FOXM1, NFKBp65, TS, XIAP, Ki67, CYP3A4, p27;

(h) FOXM1, cyclinG1, XIAP, Contig51037, PRAME, TS, Ki67, PDGFRa, p27, NFKBp65;

(i) PRAME, FOXM1, cyclinG1, XIAP, Contig51037, TS, Ki6, PDGFRa, p27, NFKBp65;

(j) Ki67, XIAP, PRAME, hENT1, contig51037, TS, CD9, p27, ErbB3, cyclinG1;

(k) STK15, XIAP, PRAME, PLAUR, p27, CTSL, CD18, PREP, p53, RPS6KB1;

(l) GSTM1, XIAP, PRAME, p27, Contig51037, ErbB3, GSTp, EREG, ID1, PLAUR;

(m) PR, PRAME, NME1, XIAP, PLAUR, cyclinG1, Contig51037, TERC, EPHX1, DR5;

(n) CA9, FOXM1, cyclinG1, XIAP, TS, Ki67, NFKBp65, CYP3A4, GSTM3, p27;

(o) TFRC, XIAP, PRAME, p27, Contig51037, ErbB3, DPYD, TERC, NME1, VEGFC; and

(p) CEGP1, PRAME, hENT1, XIAP, Contig51037, ErbB3, DPYD, NFKBp65, ID1, TS.

80. The method of claim 66 comprising using intron-based polynucleotide sequences hybridizing to at least one gene set selected from the group consisting of:

(a) PRAME, p27, IGFBP2, HIF1A, TIMP2, ILT2, CYP3A4, ID1, EstR1, DIABLO;

(b) Contig51037, EPHX1, Ki67, TIMP2, cyclinG1, DPYD, CYP3A4, TP, AIB1, CYP2C8;

(c) Bcl2, hENT1, FOXM1, Contig51037, cyclinG1, Contig46653, PTEN, CYP3A4, TIMP2, AREG;

(d) HIF1A, PRAME, p27, IGFBP2, TIMP2, ILT2, CYP3A4, ID1, EstR1, DIABLO;

(e) IGF1R, PRAME, EPHX1, Contig51037, cyclinG1, Bcl2, NME1, PTEN, TBP, TIMP2;

(f) FOXM1, Contig51037, VEGFC, TBP, HIF1A, DPYD, RAD51C, DCR3, cyclinG1, BAG1;

(g) EPHX1, Contig51037, Ki67, TIMP2, cyclinG1, DPYD, CYP3A4, TP, AIB1, CYP2C8;

(h) Ki67, VEGFC, VDR, GSTM3, p27, upa, ITGA7, rhoC, TERC, Pin1;

(i) CDC25B, Contig51037, hENT1, Bcl2, HLAG, TERC, NME1, upa, ID1, CYP;

(j) VEGFC, Ki67, VDR, GSTM3, p27, upa, ITGA7, rhoC, TERC, Pin1;

(k) CTSB, PRAME, p27, IGFBP2, EPHX1, CTSL, BAD, DR5, DCR3, XIAP;

(l) DIABLO, Ki67, hENT1, TIMP2, ID1, p27, KRT19, IGFBP2, TS, PDGFB;

(m) p27, PRAME, IGFBP2, HIF1A, TIMP2, ILT2, CYP3A4, ID1, EstR1, DIABLO;

(n) CDH1; PRAME, VEGFC; HIF1A; DPYD, TIMP2, CYP3A4, EstR1, RBP4, p27;

(o) IGFBP3, PRAME, p27, Bcl2, XIAP, EstR1, Ki67, TS, Src, VEGF;

(p) GSTM3, PRAME, p27, IGFBP3, XIAP, FGF2, hENT1, PTEN, EstR1, APC;

(q) hENT1, Bcl2, FOXM1, Contig51037, CyclinG1, Contig46653, PTEN, CYP3A4, TIMP2, AREG;

(r) STK15, VEGFC, PRAME, p27, GCLC, hENT1, ID1, TIMP2, EstR1, MCP1;

(s) NME1, PRAM, p27, IGFBP3, XIAP, PTEN, hENT1, Bcl2, CYP3A4, HLAG;

(t) VDR, Bcl2, p27, hENT1, p53, PI3KC2A, EIF4E, TFRC, MCM3, ID1;

(u) EIF4E, Contig51037, EPHX1, cyclinG1, Bcl2, DR5, TBP, PTEN, NME1, HER2;

(v) CCNB1, PRAME, VEGFC, HIF1A, hENT1, GCLC, TIMP2, ID1, p27, upa;

(w) ID1, PRAME, DIABLO, hENT1, p27, PDGFRa, NME1, BIN1, BRCA1, TP;

(x). FBXO5, PRAME, IGFBP3, p27, GSTM3, hENT1, XIAP, FGF2, TS, PTEN;

(y) GUS, HIA1A, VEGFC, GSTM3, DPYD, hENT1, EBXO5, CA9, CYP, KRT18; and

(z) Bclx, Bcl2, hENT1, Contig51037, HLAG, CD9, ID1, BRCA1, BIN1, HBEGF.

81. The method of claim 66 comprising using intron-based polynucleotide sequences hybridizing to at least one gene set selected from the group consisting of:

(a) p53BP2, Bcl2, BAD, EPHX1, PDGFRβ, DIABLO, XIAP, YB1, CA9, and KRT8;

(b) GRB7, CD68, TOP2A, Bcl2, DIABLO, CD3, ID1, PPM1D, MCM6, and WISP1;

(c) PR, p53BP2, PRAME, DIABLO, CTSL, IGFBP2, TIMP1, CA9, MMP9, and COX2;

(d) CD68, GRB7, TOP2A, Bcl2, DIABLO, CD3, ID1, PPM1D, MCM6, and WISP1;

(e) Bcl2, p53BP2, BAD, EPHX1, PDGFRβ, DIABLO, XIAP, YB1, CA9, and KRT8;

(f) KRT14, KRT5, PRAME, p53BP2, GUS1, AIB1, MCM3, CCNE1, MCM6, and ID1;

(g) PRAME, p53BP2, EstR1, DIABLO, CTSL, PPM1D, GRB7, DAPK1, BBC3, and VEGFB;

(h) CTSL2, GRB7, TOP2A, CCNB1, Bcl2, DIABLO, PRAME, EMS1, CA9, and EpCAM;

(i) EstR1, p53BP2, PRAME, DIABLO, CTSL, PPM1D, GRB7, DAPK1, BBC3, and VEGFB;

(j) Chk1, PRAME, p53BP2, GRB7, CA9, CTSL, CCNB1, TOP2A, tumor size, and IGFBP2;

(k) IGFBP2, GRB7, PRAME, DIABLO, CTSL, β-Catenin, PPM1D, Chk1, WISP1, and LOT1;

(l) HER2, p53BP2, Bcl2, DIABLO, TIMP1, EPHX1, TOP2A, TRAIL, CA9, and AREG;

(m) BAG1, p53BP2, PRAME, IL6, CCNB1, PAI1, AREG, tumor size, CA9, and Ki67;

(n) CEGP1, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, STK15, and AKT2, and FGF18;

(o) STK15, p53BP2, PRAME, IL6, CCNE1, AKT2, DIABLO, cMet, CCNE2, and COX2;

(p) KLK10, EstR1, p53BP2, PRAME, DIABLO, CTSL, PPM1D, GRB7, DAPK1, and BBC3;

(q) AIB1, p53BP2, Bcl2, DIABLO, TIMP1, CD3, p53, CA9, GRB7, and EPHX1

(r) BBC3, GRB7, CD68, PRAME, TOP2A, CCNB1, EPHX1, CTSL GSTM1, and APC;

(s) CD9, GRB7, CD68, TOP2A, Bcl2, CCNB1, CD3, DIABLO, ID1, and PPM1D;

(t) EGFR, KRT14, GRB7, TOP2A, CCNB1, CTSL, Bcl2, TP, KLK10, and CA9;

(u) HIF1α, PR, DIABLO, PRAME, Chk1, AKT2, GRB7, CCNE1, TOP2A, and CCNB1;

(v) MDM2, p53BP2, DIABLO, Bcl2, AIB1, TIMP1, CD3, p53, CA9, and HER2;

(w) MYBL2, p53BP2, PRAME, IL6, Bcl2, DIABLO, CCNE1, EPHX1, TIMP1, and CA9;

(x) p27, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, STK15, AKT2, and ID1;

(y) RAD51, GRB7, CD68, TOP2A, CIAP2, CCNB1, BAG1, IL6, FGFR1, and p53BP2;

(z) SURV, GRB7, TOP2A, PRAME, CTSL, GSTM1, CCNB1, VDR, CA9; and CCNE2;

(aa) TOP2B, p53BP2, DIABLO, Bcl2, TIMP1, AIB1, CA9, p53, KRT8, and BAD; and (ab) ZNF217, GRB7, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, APC4, and β-Catenin.

82. The method of claim 66 comprising intron-based polynucleotide sequences hybridizing to at least one genes selected from the group consisting of: CD68; CTSL; FBXO5; SURV; CCNB1; MCM2; Chk1; MYBL2; HIF1A; cMET; EGFR; TS; STK15, IGFR1; BCl2; HNF3A; TP53BP2; GATA3; BBC3; RAD51C; BAG1; IGFBP2; PR; CD9; RB1; EPHX1; CEGP1; TRAIL; DR5; p27; p53; MTA; RIZ1; ErbB3; TOP2B; EIF4E, CD68; CTSL; FBXO5; SURV; CCNB1; MCM2; Chk1; MYBL2; HIF1A; cMET; EGFR; TS; and STK15.

83. The method of claim 66 comprising using intron-based polynucleotide sequences corresponding to at least one gene selected from the genes listed in Figure 6.

84. The method of claim 66 comprising using intron-based polynucleotide sequences corresponding to a plurality of genes selected from the genes listed in Figure 6.

85. The method of claim 66 comprising using both intron-based and exon-based polynucleotide sequences.

86. The method of claim 66 comprising using both intron-based and exon-based polynucleotide sequences hybridizing to the same target gene of interest.

87. The method of claim 66 wherein said array comprises at least 100 genes.

88. The method of claim 87 wherein said array comprising at least 100 genes in a 100 μ section.

89. The method of claim 87 wherein said array comprises at lest 150 genes in a 100 μ section.

>CEGP1 intron 1, 1566 bases
GTGAGTGTCCGGCCGCGGGGGCGCACCTGGCACAGCAGGCAGGGCCAGGA
AGAGTGTTTAGGTCCCCGGCGGAGTCCAGAGCCGGGCGCGCGGGGCTCGG
GGCTGGCGGCTGCAGCTCCGCGGGGGCCTCTGCTCCCCCCGGGACCTCAC
CCGCCGGCCGGGCCAAGGCGCCACGACCGCTGGGGCCCTGAGTCCTTCGG
CCCGGCCTCGGACCCGGAGCTGCTGACGGTTCCCGCCCCGGTCCGGATGC
CTCCAGAGCGCCTGCTAGTCAGACCGTCGCCGGCGAGCAGGCAGGAGGGT
GCGGACCCTGGCCTTGGGGTCCCGCGCCTCAGCGTAGGCGGGGAAACTGA
GGGCCGGGCCGGGCACATCCGCGAGGCGGTGGCAGCTTTGCCGTTTCTTT
CTTTGGGGGCCGGCAAGTTCTGCTGATGGCTTCGGGGTGGGCTCCAGAGA
CTTTTCTGTCAGCGGAAC*AGCGCCTGTTCCGATCTGGGAATTACCCTGAA*
*GCAGCAACAAGCCTAGGTTTTCAGCAGAGAACTTTGGTT*TCCAGAGAGGA
CTCTGGACGTGCTGTGCTTACTGGACTTGCAATACTTTCAAAATGCTTTT
GTTTTTAATTAATATCCTGGAGTAGTGTCAACCCAGGAAATACTTCTGCC
AAGGCGGGTTTCCAGGTTGAGAGGATGGGCAGGGGTGGGAGTGCAGGGGG
CCGGCCATGGGGACACCATCCCCGCTTCGCAGCATCTGAGAGCCCTGGAT
GACATCTGCTCCGATCCCGGGGCAGACTTCCCATAAATACTCTAAACCAG
CNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNCTCTGAGCTCCGAGAAAGCTGAC
AGACAGCTGCTTGGTGTTCAGAGCTTGTCTGTCCGTTTGGTCCTTTCCTC
CTTTAGCGGGCATGTAGGTACTATTNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNCACATGGCCTGGGAGCCTGTACCAGGTGTCAG
CTGTGCTCTTTTGCAG

MGB-CEGP1 int1.1

FIG.1A

>CEGP1 intron 4, 4985 bases
GTACCTCTGCCCAGCTGTGGATGGGGGCAGAGCCACATCTGAGACCCTCT
CCCTTGCACGCGCACACACACACTGACTCTAGNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNATCTTTACATAGAATACATTTCAAACATGACTAGATGTCTCAGGAGC
AATATAGTGGATGATCTGCCAAGTTTTTCAAAAAGGTGCTGAAAACCACA
GCACCAGTATGAGCCTGCTCCCTGCTCTGGGTGGGTAGGGAGGAGGCTGG
ATCCTTCCCATGCAGACTTTCAATGAAGTGCCCTGTTTTCAGCCCCAAGC
TAGATCCGGCCCTTCCATGTTTTGCATTTTTGAGCTCCGAGGGGCAGAAG
GGCTCCCTCCCTGGACTTTCCGTGCTGTGGTTTCCTTCGCCTACGTCACC
ATTTATCATTCCTCTGTAAATTTGCCGGAAACTCTTCTCTTCTGATGTCC
TTCTCTTCATTCTCTTTGCTTTGAGTTTATACCTTTTTTCATTCCTCTGT
TACTTAGTAGATTCTTGAGAGGAAGGGGCATTAAGTACATGTGGCCAATC
AGTTATTTTTAACTGAATGTCATCCTTTTAACTCTTCCCTGCTCTTTCTT

FIG. 1B

AAGCTAAAGAGTCACATTTTGGTGGCTGTGTTCCTCTTGGAGTTGCATCT
GCCTATTTTTAGGGGAAGTGCCCTAAATACTAGCCTATTAACCCCTTTGG
CCATGTGCTGCTTATTCTTTCCCATTACTTAAGAATGAGGTCATTTTAAT
TTCTTCTACTATTTAATCACAAATTTATAGATTGTTTTAATCCTGGTCTT
GGTAACTTTTCAAGGGTTTCTTCATGGAAGATGATTTTTGTCTCATTTTC
CAAGGATGGCAGCTCACACCTTATACTTAACTAGAATACCTGTTTGGGTA
CCAAGAAAAATTGTCAGAGGAACCCCCAGGGGCCAATGGGTTTGATGGCT
ATCATCACCCAGAGCCTGCTCATTCTCAGCGTTTGGGGCGGGGAAGTCAC
ACATACTGGCTTTGATCAGGCAGATTTCCTATCTTGTGCCAGGTGTGGCC
CTTGATAAAGTAGCAGTTGGGTTTCATTTTCCTGCCAGGTTCTCTGGGGT
CATTGGTGTGCCCTGCACTCTTGTCCAATGTAGGCCAAATTCGAGATGGG
AATGAATTAGGAGGCCAGTGGCACAGAGTGATCCGAATCTCAGGGCATCT
CTCCTTTTGATTGCTCAAAGCTGCTTCCTGGGAAGTCACTTTGGCTTCCT
CTGCAGGTGGCTGGGGAGGGATGTGGGAACTGCAGGTTAAAGCCATCGCT
TGAGCCCTCACGGTCTGGGTCCCACCCAGTTACAAAGCAGCTGGTAGCGA
TTAAGATCACCTCTTATCCCTGTACTTCCAGAGCCCTGGCTCAGCCCCAC
TCTCCCCTCCTGCAAGCCCCCGGACTGATTAGAGACACAGGCTCCTCATA
CCAGAAGCAAATACAAATGCAGTTCCTTTCTGCAAACTGTGTTTTCTAAA
TTTTCTACAATTCAGACATTCTTGGATCCCCTAAAGAGTATTTGAAGTGA
ACATTTTTGTCTGGAACTAAAACCAAAATCTAAGAATTTGCGTTGTGGTC
TGGAAGTGCTCTCTGTGATTTTCTGTTGTGTTTCAACCTGATTGCTTGGC
AAATTCATGGGAGTGTCAGCCAACAGATTATAGCAATTGGTAACGGAGAA
CCTTTGCATCCTAGGGTTTTGATTCTTCAAATAGAACAGCCTGTAAAAAG
TTTTCTTCTAGGATTTCCTCTCTGATATGCACATTAAACTCTATGAAACT
GTAGGCTTAAAAACCCACAGTGGTNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCAC
AGTGGATACCTTCAAAGTGATTAAAAGAAGGTAACACAGGAAGCTAGTAT
TTTCTATTGCTGTTGTTTTAATAATTATTTACCAAATGTTCTTTAATAT
AGGGCATCATAATCATTGACTCTGAGGGAAAGCTCAAGATACTGNNNNNN
NNNNNNNNNNNNNNNNNNNCCTTAGAGACTCCAAAGCTGTGATAAAGAGNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NCCCGGCCCGCCTTTGTTCCTATTCATGGGTGCTCAGGCTCTCAGAATGA
GCACTCCTCTTTTGTTTTGTGTGTTCTGAGAATATTTAGATGGTGTACTG
ATGCCTTTTCAGGGCAACAGGGAAGGTGTCAGGGTGGCAAAGTGGAGGCT
GTGCTTTCAGCAGGACCTGTTACCCGTTTTATGTCATGTTTTCCTCCCAA
TTCACAAGGCATATTTTTGTTTGGTTTCCAGAAATAATCTTCAGTGGAGC
CCTGATCTTGGGGTGCACCAGAATGGGGGATTTCCAATGTTTCTGAGCTG
TTTCCCTTCTGGTGAACGAACCATCCTGGACGTGACAACCAGACCAATTT
TGGAAAGAGCTAGGGCCATTTGCTGGGCTGCCTAGTTTGGAACAGATTAA
TCTGCTCACCCCAGCAGTGGTCTTGCATTAAGTCAGAGTGCTACAAAGGC
TTTGAGGTCACTTCTTGAAAAGCTGTCAGCGTTTCCAGAGCCATTTAAGT
CTCTATTATGTCTTGGTAACTTCAGGTGTAGCTTGATGTGGTAGGACATT
AGGTGGTAGGTTCTCTGTGTATCACAATGGCATCTGGCATACAGGCATTC
TTACGAAATATTTCTTGTGTAGGTGAATTACTCTGAGGCAGTAAAGGTCA
CTTTGCAAATGTCTTAACAGTCTTGTAAACAGAGTGAAAAAGCAGCAGCA
GCTGGCCTGTTTGGGAGTGTACTTTCCAGGTGTTCCTGCCCCCATTTCTT
GGGCAGTATTATATTTACCCCCGAGCACTAGTTACTTCCCATGCTCGGCT
GACCCAAGGACAAACACAACGCTTTCTGGGCCTTCTCAGACAGGACACTG
CTTCTAGAGGCAGCTGTCACCTCCCGCGCCATCTCAGTACTGGGGTGCAA
ATCACATCTTCGGAATTACCAGCCAGAGCAAGAGAAAGCTTTCCACCAAT
CCAGTGCAAGTCTCTTTCTGTGTTAATTGACAGCCACCCTTGGCATGGAT

GAATGAATCCCAGCAACCAGCAGACTGAGTGCTGGAGTGCAGGCAGCTCA
TAACTGTCAGGCAAAAGAGCAAGAGGGTTTTAAGAGAGACTCCAGAAAGT
ATGGGATATATTAACCCTTGCACTGTCTTCTGGAATAGGAATGACATCTG
TTTGTATTAAAACAATTGTTCCGTTTAAGCACAGTTTGACAGCTCTGGAG
TGGGAGCTGGAGAGAGAACTTTGACTTCACTAGAACCTGTTGGCTAAGGT
TTTAGGGGCACAATATAGAAGGGTGTTGGATTCTAGAGAAGTGAAAGCAA
CCTTTTTGTACTCGTGTTGAAAACAGTGCCCTACTAGTATTAGAGTGTCT
CATTGATAGAGAGCCAATGACAACCAAGTCCCTACTCTCAGAGATGTTTT
AGAGTTACATTGCACGAATGCAAAGAAGCAACATAGGAACAGGTAATTAA
TAATAAAGTATAAACTGAGCAGATGTCTTGAAAGTATTCTAGGGTATGAA
AAGAATTCCTTCAGGATGCTGGTAGGCAGCAGGATCTCAAAGAATTAGTT
TTGAGATGAGGCAGAATGCTGGTAAACCACACGGGCAGTTACCTTGCTGT
GCCCCCTCATTTAGATGTGTGCCGAGCCCTGCAAGAACAGAAGCAGCTGT
TCCCCTTCCCACCATCATACTACAAGGTTAAGCCTAATCAGAATTTACTG
TATACCTCAAAAAAATTGTACAGCAGCTACCACACACGAGCACANNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNGTTCTTGTTG*CTGTTGCTGTGTGATGCTGTCAGAGGCTTATGCCCTGA*
*GGGAGGGATCAAGGGAGTGGCTGAGG*GTGGTCACAGAAGACAGATTCCGG
GGCATGTGGCCCGTACGAGGATGCCAAAATGCCACAGTCACACTCACCTC
AGAAGGGTGGGATTGGTGGGGCAGAGAGGGGCGTTGAAATGTTTTGAAA
ATTATCTTCAAGAGTATGTGAAAAAATTGAGAATCTTGATCATTCTATCT
GAACATTTTCTTAGGAGGATTCTCCTTTTCTCTTTACATTCTTGATCAGC
TCTTGGGTAAAGACATGGCAGAGATAAGAGCGTGAGTACCAGTTCCTGGG
GTCAGCAGGCTCTGATCCTGCATGCAATAGAGAGCTCCAGTGTATTGGGA
AGGCTCCCAACTCGTTAGGAGAGTTGAGACATCGTATCTCTTGGGTGACA
GAATAAATTTTTCATGTCTATTAATTGGCCTAGGTTGACTTTAATGACAT
ATACTTTTCAAATGTGGGGCTGATGGAGACCTAAGCAGACAGATCTGTGG
GCCACCCCTTAGCCCTTTGCCGCTCTCCCAGGGCTCAGGATTCTGACCAC
AGCCTAGTCACCTGTCGCACACTGCTGTTTTTCAG

MGB-CEGP1 int 3.1

>CEGP1 intron 5, 2556 bases
GTAAGTATGGGCCAGTGCACACCTGCCATGGGAACCGTCGTATTCCACAG
GCTGCCTTCTGTGGCCCAGCTCAGAAGCACCACCTCATGGCACGGCTGCA
GCAGCAGGGAAGGCAGTTAGCACGGGATACCGACCTCTACCAAGTACTTG
TTCACTGCAGAAGGGTGGTCTCCCTTAGGGAAGGGAAATGATATTTTAAA
AAGGAACTCATCAGGAGGAAATGAAATTCAGGAGTAAGGAGTGTGAATGT
TGGGGGGCAGTTCTCCCTGTTCCCACAGAATAAAACCAAATGTCCTCATC
TGGCAATCACAGCTCTTTGCCACCAGGTCCTGCTTCCCCTATAAACCTCA
TCTGCCTCCTTTCCGCAGACACTAC*TCCCCTTGCCTTTGGAGAACAGCCC*
*AAATCCTTTGATGCCTCCAGGCCTTT*CCCAAGCCCTCCTGCCTTCCTGGC
GTGGTGGACTCTCACTCAACCTTCAATATTCTGTTTAACTTCTAATAAGG
ATAAGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCATCCT
CTGCTATCAGAAGCCTCCTGGGTGCTTCAGACAGGGCAGCCATCTTGTAC
TTTGGCTCCCACAGCACTTTCCTCAGCTGTATAGCTCTGGGTTGACTTGT
GTGTTGATGTGTCTGTCTCCCCAGGTATGAGCCCCCTCCAAGTCAGGGAC
CTTGCCTCATTTTTCCTCTCAGTCCTCCCCTGGTACCTGCTATGGGATAT
GCTCAGTACACTTGTGTTTAATGAGTGGGTAAATGGGTGGCCTACACCAT
CGGGCCGCAGCTCCTGCACCACGATTGTAGTAACAAAACTCCACCTGGGA
ACAGGAAACCACTGGCAATTCATGGTGTTCCTAAACCACGATTTATGCCA
GGGGAAGCACTGAGGAGTTCCCTTTAGGAACCTTCCCAAAGCCATGGACA
GAAGACCCCTGCCATTTGGTGGGGATGGTGGTTTATGGTGAGTAGGAGAT

MGB-CEGP1 int 4.1

FIG. 1C

GAGGGGACAGTTTCACTGGTGAGGGACTTCTCTCCATTGTCTCCCTCACA
AAGCAGACTGCCACCCCAAAGCTGTCCAAGCCAAGGCTGGTGCCACCATC
ACACTCAAGCAACAGGTTCTGACATGCTCTTAGGGCCCCTCGAAGTCAGG
CTGTCCCTGAGGGCTTCCAGTGAGCTAGCAGAGTGGAGACCATTTTCCCA
CCTCCAGATCTTCGGAAGGAAGACCCAGACCCTCCAAGACTCACCTGCGG
GGCGAGACCCTCAACATTTCATAGTCTTTCAGGGAACAGTTGCTGAAGGG
GGCGGGGGGGTGGGCACCTGTAAGCTTGTTTTTAAAGATTTTAAATGTCT
TTAAGATATCACTGCTCAAATAATATTGTTCTGNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNCGTTTGAGGAATTAACAAAGAA
AAAAACTAAGACCTAGAATCTCACCACATAACCAGCTGTTTCAATTTTTC
CATATTCCTATTTAGTTGTTGTTCATATGCATACACAATTTTTACATAGC
TATAATCACAGGACAACACAAATATGTAATTAGTTCTTTTGAATTAGAAA
AATTACAAAGGGCCTATGTAAAATGCAAACACTCCAAAGCATATAAAGAA
AACATGCAGTTTCCCGCCTCCCGTTTCCCTTGCCAGAGGTAACCACGGTT
AGCAGTTTGATGAATAGATAGTTTTGTAGTTGGCTTTTTTTCTTTTTGGC
CTATCATCAATACATTCATATATAGTCTTGATAATTACCAGTTACTGTCA
CGTTAATTGTGTGCAGAATCATCCTGTGATTATCCTTCCTTCTAACTAAT
CTAGATTGAATCTGATGAGAGAAATTCTGACATATATGTACAAATTAAAT
ATTGTCTGTTTTATTCCAGCATAAAGTGCTATAGCATTTCCCAAAGCCCC
AGTACAGCTGTATTAATAGGTAAACTTCTCTAGATAGAACAAAGCAGTAG
TCTAGAATCTCTTGGTATAATTTCCCTTATATAATAAAAGTCTCTCCCCC
AACTCTCCCATCTCCCTCTTCCTGTATGACTTTGTTTAAACCCATGTTTC
AGCATTTCTACAATTTGTATTGTAACTATCTGCATACACAGACACCACAG
GGTCTGACTTGGAGTTATGTCTTTCGTNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNAATCAGAATTTCTCTGGAGCAAACAC
AGCCCTGTGTTTGTGGAAATCTCAGTGCTTTATGTATTGATTCATTTTGC
TGTCAG

>CEGP1 intron 6, 716 bases
GTGAGTGGCAACCCCAACACTGAGTGAGGGTCTGCACCAGCCTGCCTGTC
CCTACCCCTACCC*CTTAATGGTGTTTAGCACAGATGCAGGCTGTTTCCTG*
*TGCATTTGCCCCCCCAGCAGGCCCTGTGCTGCTTCGCATGCTACAGTGG*G
AGTGGTCTAGGCCTGTGGGGAAGGCCCCTCTCTCCCTGTGTGACCTTGGG
AAGCCCTTCCTCCTCTCCTGGACTAGGCTGCTCCTAACGCTGGTATTCCA
GAGACTGGCACAACACCTCCCAGGAGGCCAGGGCAGCACGAAGTTAGAGC
TGTTTATAATGATGCGGCACTTCTGGCCAGCAGGAGCCAGGGCCGTATAT
TTCTGGCGGGATGCCTGCCTTGCCCTTCACGGTGTGTCCTTCACTAGCTC
CATTTTAGAGGTTTCCAGGCCCAAGGCTCTTTTTCTCCTCGACTCAGGGG
ACTGAAGCTTGCATTCCCTAGTGTCTCTTTGGTCAGTGCAATATACCTCC
AAAATCTTTTCCATGTTTAATGTTTGCTAAGGATCTGTGGCCCTTTAACG
GGCTGTGTCTCCCACAGAGCCTCATTACAACACATTTTTATTGCGTGAAC
AGAGTCACATATCTTTCATTCCTCTTATGTCTGGGATTTCAGCAAACACA
GTTGTATGGGGATGAGCAATCTAACTCATTCAGTCTGAGAACCGTGCTCT
TTTGCTTCTCTTGTAG

CEGP1 int 5.1

FIG. 1D

>FOXM1 intron 3, 2041 bases
GTAATGTGTCCCACAGCAACCAAAATCAAGGTCAGCCCAGCCTGACAGTC
TCTCCAGTGCTGTACTGCAACTTGTATCTGGGACAGCAGTTAAGTGCAAA
GGACACTAGAATGATAAACAAATGTATCTTTTAGATTGTGACTCAATCTT
ATTGAATCCAGGCAAAATCATTAAGAAGAGCTCCTTAACTACTTCATGTG
TTACTACCTAAAGTCCATGGAGGGTCTTCAATGTAGCACTCAAGCCCACT
TTTCTGCTACACTCAACAGCCGTCCTAGATGCCAGCAGCTAGAGTGGCTA
AGTAGTTTTATGAAAATGTCTTGATTAAAAAAA*AAAATGCTGTCTGTGAG*

FOXM1 int 3.2

CCTCATGACCCAAGATGTCATCTCCTGTAGCGTCACATAGCATTTCTAGT
*GGGCAGGGGTT*TTCCTTTCACTTCATTCATGGAAAGACCGAGATGCCTGT
GAGTCAACATAGCTCACGCAGTTGGTCGGTGTCAGAGCCACAAATGAGGT
CTTCTGACGGGTGCTCAATTCCAAGTCAAGTGTGCTTTGTTTTCCTCATG
GTAGAACTCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNCCTCATGGTAGAACTATGNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNCATGGTAGAACTTTTAATTTACTCCCTTCCATCAGCTTACTTTCCTA
GNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTAATTTCCTAGNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTAATTCCC
TAGTTTCTTAATTTCTCTGAGCCACCTTTCTTGCTATTGATCACTACCTC
ACAGCCTTACTCTGCTTTTCTAGCCCCTGACAGCTATCTAGGTCTTTTCT
TTATCACAATCTAAGGTTGGCATCAGTCTTTATTCCCGTAG*AATAGATGG*
*GTTTATGGCTGAAGGTGAC****GCTCTGCGGTGTGGAGTGTCAGGAGAGTTG***
CCAAGAGGGCTGCAAAGACACCAGACGAAGCCTGTGCTGAGCACAGTGGG
AGGGGCCTGAGGCTGGTTTCCCCATGTGTTTGAAGGGTGATGTTTCTGAA
TCTAAAGTAGCTGATAACCAGTTGTCTTGCTCTTCTTCCAG

FIG. 1E

MGB-FOXM1 int 3.2 (underlined)
FOXM1 int 3.1 (bold)

>FOXM1 intron 4, 993 bases
GTGAATGCCCTGCTTTCCTCTAAATAGGGCCTAAGTTGGAGGTTGTCATA
G*CCATCTCAAAAGGAAACAAGTTCTGCTAGTGATGCTTTCATTTGATCAG*
*GGGAGAGTT****AGAAGCCAGCCACCCAATTAGTGACTTGCACAAAACCCAGT***
GAATTAAGTACACTTGACAAATACCAAATGACACATTTTTGTGCCAGACC
AGAGCAAGGAGAAGGCTGTTCTGACCCAACAGAAAGGGCTCCCCAGGGNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTAC
AAGAAATTCTGGGAATGCTTGCTCTAAAAAAAGCCCTTCCTNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNG
CCTGCCTTAGGCTGGAGACCAGAAGCTGAGCTACCAGAACGTCTTTTCAG
AAAGAAGTTATTTGGTTTTTCAGAGTGCCCATAAGGCTGCTGGTAGCTG
TAACCATTCTCCTGGGAGGGGCAGTTGTCTGGGGTGTCTTTTGTCATCAG
TCAGGAATAAGTGTTTTTCCCAATCCGGTCAAATTGACCACGTTGGTGGT

FOXM1 int 4.1 (underlined)
FOXM1 int 4.2 (bold)

FIG. 1F

AACTTCATCTCATTTCTCTCCCACAATGCCTGGCCGCCACCAG

>FOXM1 intron 5, 602 bases
GTAAGGTTCTTTCCCTCTGGCTCGGGGCTTGGCCTTGTTTTCCTTTCACT
GCTCAGCATGGCTTTAG*TGGACAGAGACAAGATGTGATGTGGGGAAGGGT*
*CCCTATGGCCATGTTTTGTCTAGGTGCCAGC*CCTAGACACAGAACACCCT
GAGGGTCAGGCACACACCCACTTCCCTCCCCTTCCATGGGCATCACAAGG
GCACACTGAGCAGAGCAGGGCACAGCAGGGGAGCATGCTGCAGCAGCCAC
AAGCGCATGGCACCAGCCTCAGGGGCGGCAGTTCGTTCGCTCACTTTTGT
GCCTAGCTTTTCTTTGCCACGCATATAGCTACCTGCTCTGGCATCCCCCA
GGGGTGTTGAGGACACGTGGGTGAAGCGGTAGTGCCACTCTGCCATCATG
TGTCTGTAGGCCACCCACCTGCCCACTCATCACAGTTTTGGAGACTGCTC
GCCTACGTCCATCCCCTCAGGTTGGCCTCCTCTCTCTGGGCTGTCATTAA
CTCAAGCACACACCACCAGAGCAGCTGGTGGGGTTTTGCCATCCCCTCTT
TACCTTATTGTGTTAACATAGGTTTCTTTCTCTCCCCATCTGCCACAAGC
AG

MGB-FOXM1 int 5.1

FIG. 1G

>FOXM1 intron 7, 4656 bases
GTG*GGTGTCCTATTTTCCTCTGAAGAGAGATTCTGGCCAATTAAGAATGT*
*TGGACCTTCAGCTTGCA*AAGCACTCTGATAAGTGTTCCTTGAGAGCTTAT
AAATCTAGTTGGGTAGAAAAGGCATAAAAACATAGGGAAGTGTAATAGCA
TTAGAAGAGCTAAAAAGGTATTTGGATTACAATGTAAGTGGTGTCAGAAG
GCCCATAAATACCTGATGAGCTTGTAAGAATTCAGACAAAAGTGATTGTG
ATAGATGGGCTAGGATTATTAAGGAAGATACACAAGGGAGGCAGGCCTTA
GAAAGAGATGGATTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNGTGGATTTGANNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTAGATTTGGGTAAGCAAAC
AGGTGTAGAGAGAGCATGCTAATGGGCAGTGCCATGGAGGCGGGAAATGC
AGTTCGTACCTGGCAGTAGTAAAGTGACTGGGTCAGACTAACTNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNGTCTAGCTTGAGGGGAAGGTGAGAAGGGTAAATTCAGAGCC
AACTTGGATCAGCCATCAGATCTGCACTTAACACTGTTAAAGGGTTCTGT
GAGTACGGGCTGACATGTAACCAAAGTGAAAAGCTTCCCCCATCCCCTTC
AGAGAGATGAAAATAGCATAGAGTCTGGAGTTTAGAGCGACTTGGGTTTG
CNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

MGB-FOXM1 int 7.1

FIG. 1H

MGB-FOXM1 int 7.2

NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNCATATTACACACAAAATTATACCACACATACATAATTTAGCGTAAAT*T*
CATTCATGTGGCCGTAGCATGTGCCCTGTTTGGGTTTTCATGCAGTGGGT
*TTTCTCCCCTTTCCTTTTTGGCTCCCTCTCCACC*CTACCATCACCCACAT
CACCCCTACTCCCAAGATAACTGGTTGATAATTTATGATGCTTTCTTGCA
TATTTTATCAATGCTCTTAGTTATACTATACATGTATAGCGATAGCCATT
TTATATGTACACATACAACACACAGAACATTGATNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNGTTATCAATTTGTGAGAGCTCNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGTGAGAGNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNCCCTGTGAGAGCTCTTTGTTGT
TAAAATAATCTTCTTTCTTTTATGCTGAAGATATTTTTCTACTTCTATTG
TTTATCTCTTTACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGTTTGTTGTTTCTTAACT
TTGTTTATGGTATCTCTTGCCACAGTAAAATTTTAAAGTTTTATGTAGTC
AAATGTCTCTCTTCTCTTTTACAGTTTCTGGGTTTCCAGTCTTGGTTAAG
AAGGTCACCCGCACCCTCAGATTGTATATGTAGTCTCCTAGATTCTCTTT
CAGGATTTGTATGATTTTAAGGTTTTCATTTTTTTTTTANNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNTAAGGTTCTTATTTTCATGCATTAAATCT
TTGTATACAGTGTAAGACAAGCATGCAATTTTATTTCCTCTCGGATGAAT
GCTATTATAATTATGCCACTACATACTACATACCCGCATCTTTTACCCCC
AGAATTGAACTACCAACTTCAACATACATCGTATTCTCATATTTAATAGA

TTTTAAGACTTCAAAACGACACAA*AGAGGATCAGAACCCGTATGTGATAT*
TTTTGTGCGTCCTGTCTGGTGACCGTTGGTTCACCTTATCTCTGTTTCCC
TTTCAG

MGB-FOXM1 int 7.3

>PRAME intron 2, 614 bases
GTAAGTTCGAGCCCTGATTCCTCCGCTTCCCCGCA*GGGTGACCTTGGGCT*
TGTGCCCCCGGCACCACCCCTGTCCCGGGTCCCTGTTTTCTCTCTGGAAA
*TGGGTTGAAG*ACCAAAGAAAATAATGTGCGCCACTTGGGTCACCCCGGGC
CGCCTGCCCCGGAAAATTGGCCCCAGTTGAGGAGTTGTGGCTGTAAGGAT
GCCTTGAACCGAGGCGGCGGTGCTCGTGGTTGGAGCTCTCCAGGGTGGGT
GCGCATTTGTAATGCGGTGGATGCTCTGGGACTCGGCCCCTCTGAAGGTG
CTGGGGGTTGGGGACGGCCCAGGCAGTGGCGTAGGCGTCCTAGGAAGGCG
GGAGCAGAGGCAGAAATGTCGCTGCAAGACCGTAGTCAGGGTCCTTGACC
ACAGGGGTCACTTGTGACCAACCACATGGTCTGTTGTTCCTCCTGCCCCC
TGGTTCAGCCCAGGAAACACTGGTGCTCAGGTTTGGAGCCAGAGATTTGC
ACTGAAAGGGCGGGATTGAGTCGCCAGTTGTCAGTTTCCTCAGCAGTATT
TGCGGAGGTTTTCACAGGAGGCCGTTGCTTCGTAAATATTATACATGTAT
TCTTCTTTTTGGAG

MGB-PRAME int 2.1

FIG. 1I

>PRAME intron 4, 432 bases
GTAA*GGGTGACCTAGCAGCTTGGTGTGGGGCCCTGGGAACCTGAGCAGGA*
*TGCAGCTGGGGTCAGGGAGCATGGAGCGCCTAAGGCTGGGCCAGAGGC*TC
TGATGGTTGCCAGCAAGGAAGTTCAGGGAGGCCTTGGGGCTACTGCAGGG
GTCACTCTTGGAATGGGCTTCTGGACATGGGGCACTGATTAAAATGCAGA
GGTGTCTGAAGGAACATGCACCTGCTTCCTCCTGGTGGGGTGGGAATTGG
GGACCAGGAAGGATCCCAGGATCCTAGTGGGAAAGGGAGCAGCTGATGCC
TGAAGTACGAAGTAAAAGTGCAGATCTAAGGTGGATGTCTGTTTGGTTCT
TACCTACATTATGAGACTCATGGTCTTATTTTGAGTTGATCTTAAAGCAT
CATCTCAGCTAATTACCTGTTTTTCCCCACAG

MGB-PRAME int 4.1

FIG. 1J

FIG. 1K

MGB-STK15 int 1.1

```
>STK15 intron 1, 3740 bases
GTACAAGGGGTTTGTTGAGTGGTGTTGACATGCGCGGGAGGGGTGGGTGG
GCTTCAGATTGGATTTTGTCCTCCGAGATCACCNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NGGTAAGCGTACGGAGAACTTGCAGCTGGGGTGGGTGTTACAGAGGAAAA
GCAGGAGTGCGGTTTAACGGGGGCCGCTTTAGATAGAATAGCCTAAGAAG
GCCCTTGTCCTGGCTGGATGAGTGGGTGAATTGATGAATGAGAACCTCCT
TGCAGAGGCCTTCCCGGTCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGGATGCAG
ACCGGTGCATACAAATCGTCTGGGGACGTTAAAATGNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNCACTGTCCTTAACTCTCGTAATGTCTCTTCCTCTT
CCGTAACCTTCCTTGTCCCTTGAATTAAACGTTTTTCAGCAACCTACTCA
GTTCGTCCTTCCCTTCATCTCTGCAGACATGCACAGGTCTGAGGGAGGAA
GGAATAAACCGTATAAACCTCCTGCGCTATTAGCCTAACAGCTTTTCTAT
TCAAAATAGTAGGACTTCTGGTTTGAACTGAATGGATCCTGTGAAAGTCA
TCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAGTCTC
CTTGGCGTTGTCTCCAGAATTCTGGATTAGAATCTTATTCCATTCTGCTT
GTTATTCAATTTCCCTAGAAAGAAAGGTAGAATAAATTGGAGCAAATGCC
TGTAGCTTCTGTCAGAAGAATGTTGAATAAATGTTGTTAGGCCTATGTGA
TCTCATTAGACTGCTACTTAGAATTGTAAGGGAAGTAAAGCATTAGAGCA
TGTGTGAAATTAAATATTTGATTAACACAAGTGTGCATTTCCTTGTTGCT
GTTTATCAACTTTTACTTACCCACTGTTTTTTATAAGGGCTGCAGCCTG
TAGTCTGGGCCTGGCTTCATCATGGAATTATTTGCTTAATTGTAAAATGG
TAATCTTAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNGGATATTTGATAAGAAACTTCAGTGAANNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNGGGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
```

NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNGCAGGGAAAGCACATGCCTGTCCCTCCCTCATT
AGCTTCATTTGGACAAAACATGTAAAATCCGGTGTGTTGTGGAGGCCTTT
TGATTGGGGAACTGTAACGCTGCCTATCGAGCAACAGCACTTTAAGCAGG
TGGCTTTGTTCAAATTAAAGGTTCTTCTTTTTCTTTTCAG

>STK15 intron 2, 1622 bases

MGB-STK15 int 2.1

GTAAATTGAATAATCTGTAATCT*CATTCACATTTATAAACCCACATGGAG*
*GTTGGTCTTGTCGGGAATTCTTTCCGCCTTTACTTTGGATT*TAAATTTAG
ATCCCTTACTGTGATCCTGGATATGAATTAGTCACTTTTCTCGTGTTCAG
TAACATTTTGCTGCTTCTTAGAGTAGCTTTTTTGTTCTGCTTTGTCTTAT
AATCGGCTGCTTAAGTTTCTATATCCCTCCACTGTATGCAGGATAATAGT
AATAATGCATCTGGCAGGAGTTCAAAACTTTTAAAATTGGCCATAAATAT
AAAATAATTAGAAAAAGGCTACCTTGAATTACTGTATTTGATTCTAAGTT
CCTATGATAACGGCCATTTAAAAAATTGCTCTATATTTAAAATGTTTCTT
TTTATTTGTCTTTGTCTGAATGCCTGCTGCGTTGTGGACAGTGTGCTAAT
TTCAGGAGTAACTGACTTTGTATTTGGAAGTCTTAACACCCTCTCTTTGT
AGAGCACTCATACCGTTGAGCTGGGGATGGACTTTGAGGCTTTCATTTCT
AGCACTTGTCCCTCACTTACAATGAGCTGTTGAAGCTGAAGGAAATCTCA
TCCCTCCTACCCCTTTTAGTTTGATTAGCTGAGGGTGTTAGAGTTAACTT
AACAATTTAAGGTTGTAATACAGTACTTACAGGCGTATAAATAATACATT
TCAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTAATAATAATA
ATACATTTTAGTAGTAACTTTGTGAAGTGTCTACATTTGTTTCCTCTTTG
TCAGTTTTTTGCTCAATTCCATTTTGTCAATACTTGGAAAATGAAACATT
GGTTAATCAATAGTACAGTAATAAGCTTATTGTGGAAAATCTTCGATATA
TGAAAACTTAGACTCTTCTAAAACTTCATGAAGATAATACCACTGTTGAA
CGTTTTGACGTATTTTTTTTTGGTCTTTTTCTTAAACGTATATTATCAAA
GAAATTTCAATGGAACTGAGATTTTGGCATAAAGTTTTTGTATCATAGCT
TTTTGCCAAATAGCAATGTAGTGTCTATTTCCAAATTATTGAGAAATTTT
AGAAAGTGTCTCCTTCATTAATGGATATTTGTTAATAAAGCATGATTTTT
AGGGGTGAGGAATTGGAGGGGATAGAAGGTATCATTCAGGTATTCTTAGC
CACATACTAACTATCCTCTGGAGGTACTGATTAAAATACCTTTTCACCTT
CCATCTCTTATCAGTGACATTCATTATTTGCTATACTAGAGAACAAACT
TTGTGAAATTCTCAATATATTCATCTTTTGCTTTCATGAATGCCAGAAAG
TTTATTTTCTCTTCCATTCTAG

FIG. 1L

>STK15 intron 4, 1093 bases
GTAAGCTTTCTTATTTACAAAGTTCTGTACTGTTCTACTAGAATATATTA
TTTCGTTGCAAATTTCGTTGTGGAACTCTGGGGAAAAAAATGAGGCCTT
TATTTGCATTTAGAGGATATAAATGTTTCCAGATTTCCAATCTTAAAAAA
AATGGAATTTTGTGTAATGAGGTATTTTACTAGGAACTCAAGTGCTTTAA
AAAATGGCTTTCAAATTTAGAAAAAGCTTGTATGAATCTTTTATAGAAAT
GTGTGGAAGTTCCTCTCTGTCCTTAGAAATAACCACTACATATGGTTTAT
GCGTCTGTACTTTTTTATTGTACAAAAGTGCAAGTTTTTAAAAAATAGAA
TATGTTGCAGAACTATATACTCATATATGACTGAGGGTTTTGACAGTATT
ATAGTTTTAGTTCTTTATTGTAAAGGTTGGCTGTAATGTCTTCCCCAGGG
CTTTTCTAAAAGCCTCCTCTCAGTCTCTGAACTATCTGGACTCTAGAATG
TACCGGGAGGA*GCGAGGAATGAACCCACAGACTCTTTTGCTTTTAGCGGT*
*CTAACAGAGGCTAAGAGTCTAAATCCACTGGTTCTCATGC*CCCAGCTAGC
CTGTGGGCTCCATCCCGCTTCCATTAGTAACAGTGGCTCTGTCTCCACCA
CCAGAGTGGTTCTCCACCCAGAGAGAATTAGCACCTCTGGGACTGGAGGG
AGCAGCTGGGGTTAGTTTGAAACATGCCCCCAGATGGTCTGGAAGCATTC
CTCCCTCTCTGGTCACTTATCCTTTTTGTGGTCTTCAGCGTTGTCATGGC
CCTGTTCCTCTGAGCATAGTACGGGCTTGGGACATTTCCCATAGAGTGCT
TCAGGTCTAAAACCCGAGACTGCTCCTTGTCACTGACTCTCACACCTGAC
GGCAGCTAGGGACGTCAGGGTTTCATGTCGTGGCAGCTCTTTGATAGTGG
TTATTGCCTTGGTTCTTGCTGAGGATGCATATTGAGTGAAGTTGGAATAC
GAAATTATTTGTAGAATGTGTCTGCTACTCATTGAAAATTTGTTAGAAAA
GCTTTGTTTTCTTCACATTCTAAAGTGTTCAAATTCCTCCTAG

FIG. 1M

MGB-STK15 int 4.1

Name	Sequence	SEQ ID NO:
MGB-CEGP1 int 1.F1	AGCGCCTGTTCCGATCTG	14
MGB-CEGP1 int 1.R1	AACCAAAGTTCTCTGCTGAAAACC	15
MGB-CEGP1 int 1.P1	CCCTGAAGCAGCAAC	16
MGB-CEGP1 int 3.F1	CTGTTGCTGTGTGATGCTGTCA	17
MGB-CEGP1 int 3.R1	CCTCAGCCACTCCCTTGATC	18
MGB-CEGP1 int 3.P1	TCAGGGCATAAGCCT	19
MGB-CEGP1 int 4.F1	TCCCCTTGCCTTTGGAGAA	20
MGB-CEGP1 int 4.R1	AAAGGCCTGGAGGCATCAA	21
MGB-CEGP1 int 4.P1	CAGCCCAAATCCT	22
MGB-CEGP1 int 5.F1	CTTAATGGTGTTTAGCACAGATGCA	23
MGB-CEGP1 int 5.R1	CCACTGTAGCATGCGAAGCA	24
MGB-CEGP1 int 5.P1	CAAATGCACAGGAAAC	25
FOXM1 int 3.F1	GCTCTGCGGTGTGGAGTGT	26
FOXM1 int 3.R1	CACAGGCTTCGTCTGGTGTCT	27
FOXM1 int 3.P1	TGCAGCCCTCTTGGCAACTCTCCT	28
FOXM1 int 3.F2	AAAATGCTGTCTGTGAGCCTCAT	29
FOXM1 int 3.R2	AACCCCTGCCCACTAGAAATG	30
FOXM1 int 3.P2	ACCCAAGATGTCATCTCCTGTAGCGTCACA	31
MGB-FOXM1 int 3.F2	AATAGATGGGTTTATGGCTGAAGGT	32
MGB-FOXM1 int 3.R2	CTCTTGGCAACTCTCCTGACACT	33
MGB-FOXM1 int 3.P2	CCGCAGAGCCGTC	34
FOXM1 int 4.F1	CCATCTCAAAAGGAAACAAGTTCTG	35
FOXM1 int 4.R1	GGGTGGCTGGCTTCTAACTCT	36
FOXM1 int 4.P1	CCCTGATCAAATGAAAGCATCACT	37
FOXM1 int 4.F2	AGAAGCCAGCCACCCAATTA	38
FOXM1 int 4.R2	TGTGTCATTTGGTATTTGTCAAGTGT	39
FOXM1 int 4.P2	TGACTTGCACAAAACCCAGTGAATTA	40
MGB-FOXM1 int 5.F1	TGGACAGAGACAAGATGTGATGTG	41
MGB-FOXM1 int 5.R1	GCTGGCACCTAGACAAAACATG	42
MGB-FOXM1 int 5.P1	CCATAGGGACCCTTC	43
MGB-FOXM1 int 7.F1	GGTGTCCTATTTTCCTCTGAAGAGA	44
MGB-FOXM1 int 7.R1	TGCAAGCTGAAGGTCCAACAT	45
MGB-FOXM1 int 7.P1	TTCTGGCCAATTAAG	46
MGB-FOXM1 int 7.F2	TCATTCATGTGGCCGTAGCAT	47
MGB-FOXM1 int 7.R2	GGTGGAGAGGGAGCCAAAA	48
MGB-FOXM1 int 7.P2	CCTGTTTGGGTTTTCA	49
MGB-FOXM1 int 7.F3	AGAGGATCAGAACCCGTATGTGA	50
MGB-FOXM1 int 7.R3	GGGAAACAGAGATAAGGTGAACCA	51
MGB-FOXM1 int 7.P3	TGTGCGTCCTGTCTG	52
MGB-PRAME int 2.F1	GGGTGACCTTGGGCTTGTG	53
MGB-PRAME int 2.R1	CTTCAACCCATTTCCAGAGAGAA	54
MGB-PRAME int 2.P1	CCCGGGTCCCTGTT	55
MGB-PRAME int 4.F1	AGGGTGACCTAGCAGCTTGGT	56
MGB-PRAME int 4.R1	GCCTCTGGCCCAGCCTTA	57
MGB-PRAME int 4.P1	TCCCTGACCCCAGCTG	58
MGB-STK15 int 1.F1	CGTAATGTCTCTTCCTCTTCCGTAA	59
MGB-STK15 int 1.R1	ACGAACTGAGTAGGTTGCTGAAAA	60
MGB-STK15 int 1.P1	TCAAGGGACAAGGAAG	61
MGB-STK15 int 2.F1	CATTCACATTTATAAACCCACATGGA	62
MGB-STK15 int 2.R1	AATCCAAAGTAAAGGCGGAAAGA	63
MGB-STK15 int 2.P1	TGGTCTTGTCGGGAAT	64
MGB-STK15 int 4.F1	GCGAGGAATGAACCCACAGA	65
MGB-STK15 int 4.R1	GCATGAGAACCAGTGGATTTAGACT	66
MGB-STK15 int 4.P1	CGCTAAAAGCAAAAGA	67

FIG. 2

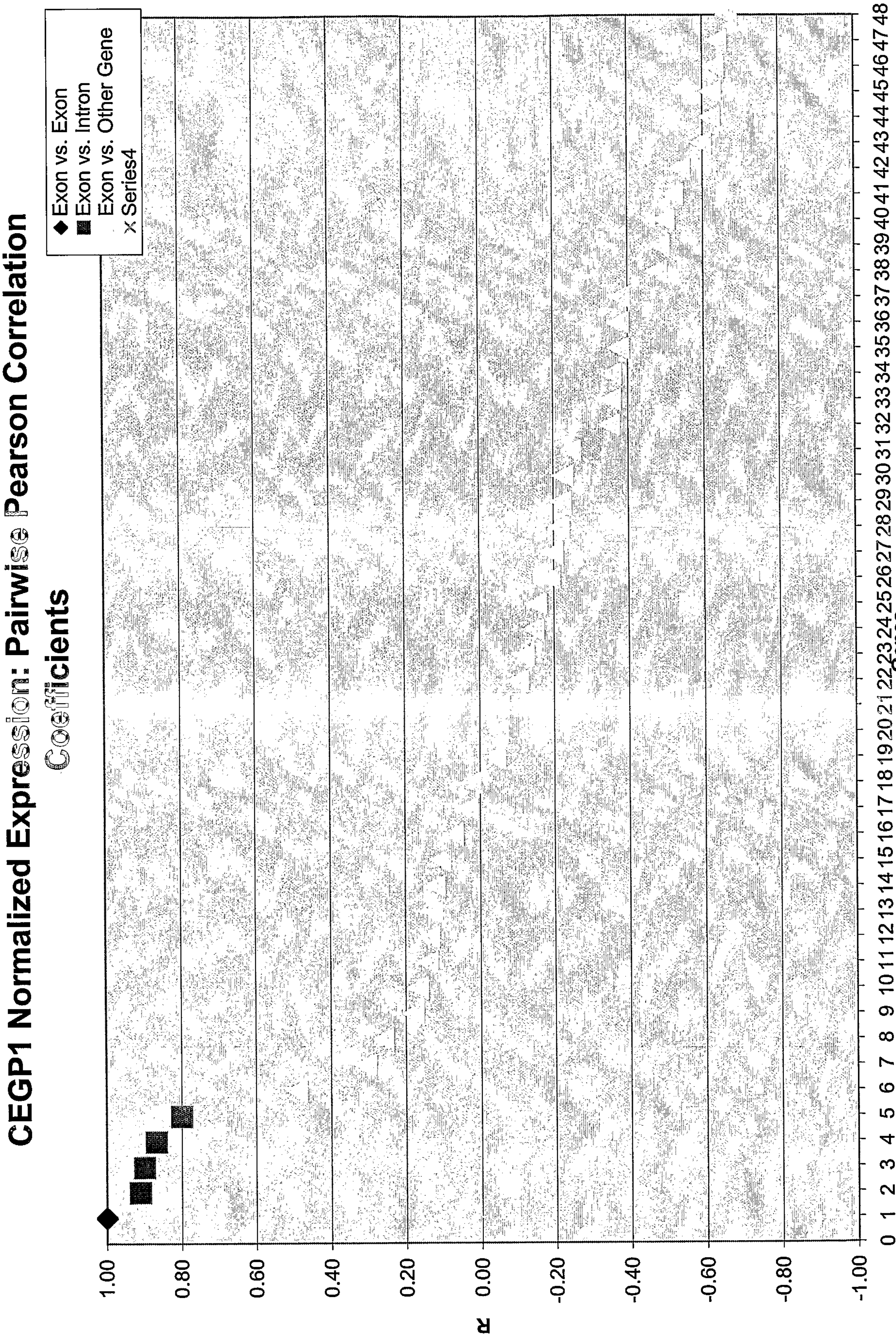
CEGP1 Normalized Expression: Pairwise Pearson Correlation Coefficients
Exon vs. Exon
Exon vs. Intron
Exon vs. Other Gene
Series4
R
1.00
0.80
0.60
0.40
0.20
0.00
-0.20
-0.40
-0.60
-0.80
-1.00
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48
Genes

FIG 3

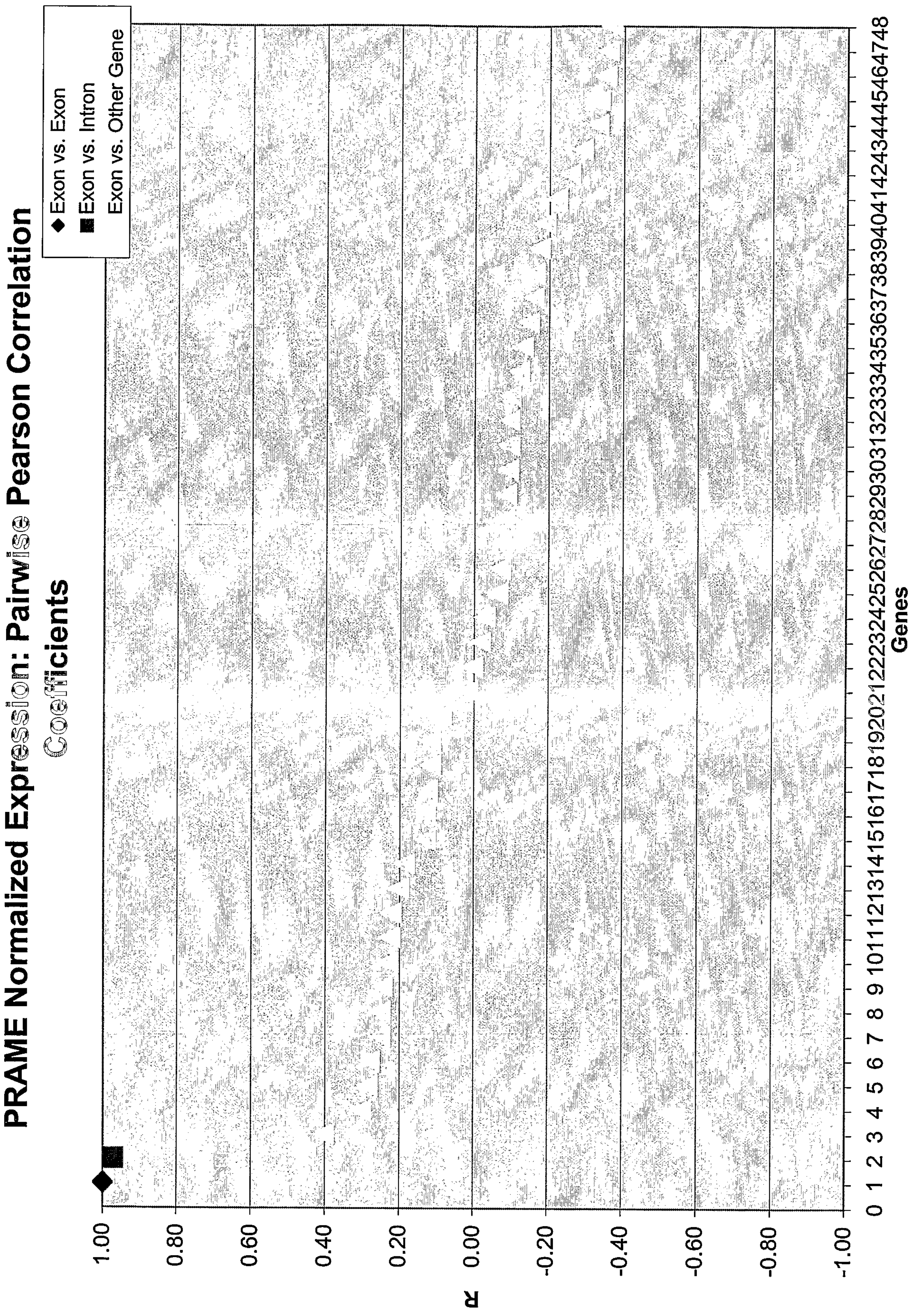
PRAME Normalized Expression: Pairwise Pearson Correlation Coefficients
◆ Exon vs. Exon
■ Exon vs. Intron
Exon vs. Other Gene
R
1.00
0.80
0.60
0.40
0.20
0.00
-0.20
-0.40
-0.60
-0.80
-1.00
0 1 2 3 4 5 6 7 8 9 101112131415161718192021222324252627282930313233343536373839404142434445464748
Genes

FIG. 4

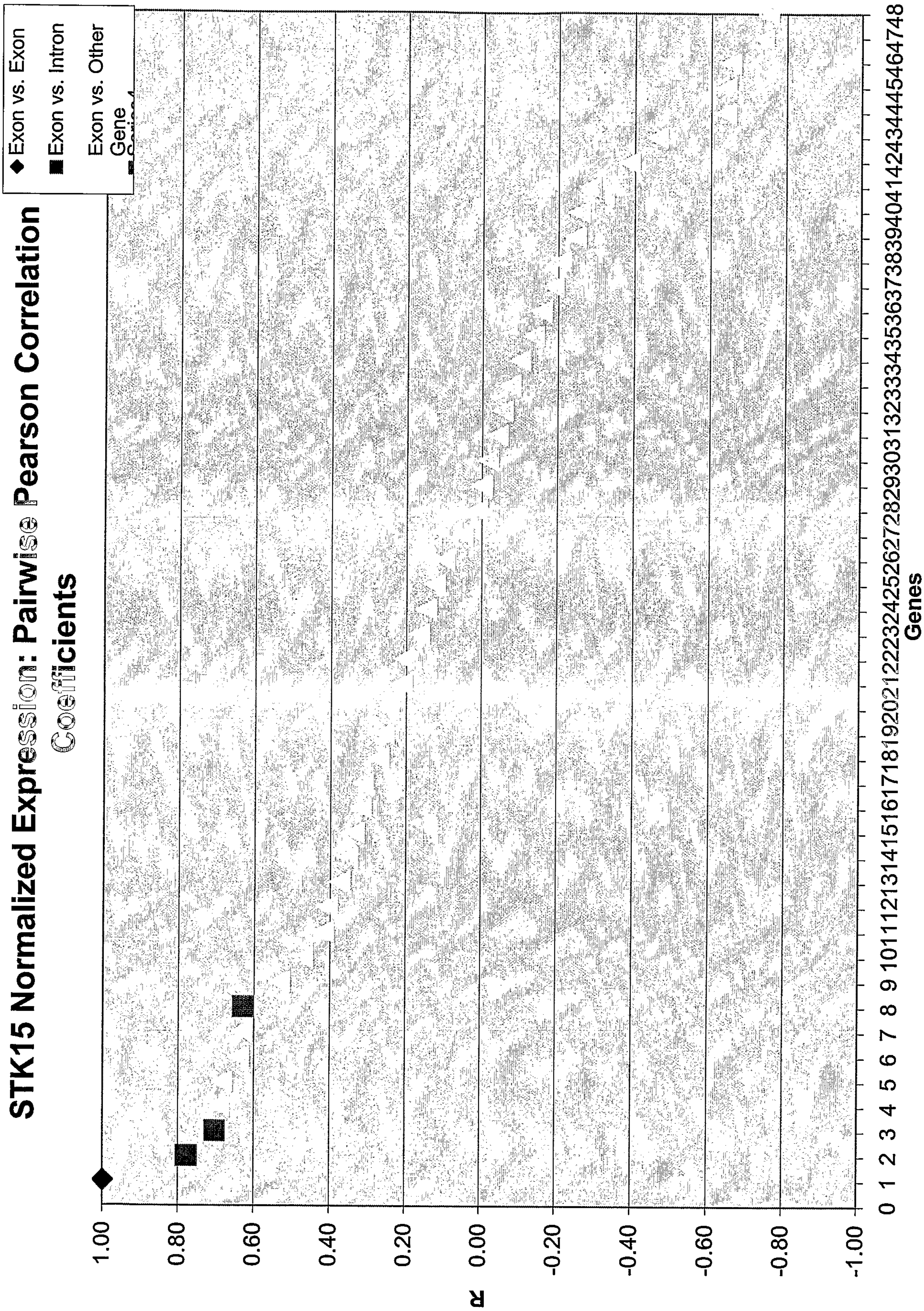
STK15 Normalized Expression: Pairwise Pearson Correlation Coefficients
◆ Exon vs. Exon
■ Exon vs. Intron
Exon vs. Other Gene
R
1.00
0.80
0.60
0.40
0.20
0.00
-0.20
-0.40
-0.60
-0.80
-1.00
Genes
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48

FIG. 5

Gene Name	Sequence ID	Sequence Source	GHIallgenes2-03.txt Accession Number
18srRNA	M10098	NCBI	M10098
28srRNA	M11167	NCBI	M11167
A-Catenin	NM_001903	NCBI	NM_001903
ABCB1	NM_000927	NCBI	NM_000927
ACTG2	NM_001615	NCBI	NM_001615
AD024	NM_020675	NCBI	NM_020675
AIB1	NM_006534	NCBI	NM_006534
AK055699	AK055699	NCBI	AK055699
AKAP-2	NM_007203	NCBI	NM_007203
AKAP12	NM_005100	NCBI	NM_005100
AKT1	NM_005163	NCBI	NM_005163
AKT2	NM_001626	NCBI	NM_001626
AKT3	NM_005465	NCBI	NM_005465
ALDH10	U75289	NCBI	U75289
ALDH12	AK074266	NCBI	AK074266
ALDH1A1	NM_000689	NCBI	NM_000689
ALDH1A3	NM_000693	NCBI	NM_000693
ALDH2	NM_000690	NCBI	NM_000690
ALDH3	BC004370	NCBI	BC004370
ALDH3A1	NM_000691	NCBI	NM_000691
ALDH3B1	NM_000694	NCBI	NM_000694
ALDH3B2	NM_000695	NCBI	NM_000695
ALDH4	NM_003748	NCBI	NM_003748
ALDH5	NM_000692	NCBI	NM_000692
ALDH7	U10868	NCBI	U10868
ALDH8A1	NM_022568	NCBI	NM_022568
AluSx	NM_AluSx	GHI Custom	NM_AluSx
AMFR	NM_001144	NCBI	NM_001144
annexin I	NM_000700	NCBI	NM_000700
annexin II	NM_004039	NCBI	NM_004039
AP2B1	NM_001282	NCBI	NM_001282
APC	NM_000038	NCBI	NM_000038
APN	NM_001150	NCBI	NM_001150
APP	AY011354	NCBI	AY011354
APP	M33112	NCBI	M33112
AREG	NM_001657	NCBI	NM_001657
ARG	NM_005158	NCBI	NM_005158
ASH1	NM_018489	NCBI	NM_018489
B-actin	NM_001101	NCBI	NM_001101
B-Catenin	NM_001904	NCBI	NM_001904

FIG. 6

Gene Name	Sequence ID	Sequence Source	GHIallgenes2-03.txt Accession Number
B2M	NM_004048	NCBI	NM_004048
BAD	NM_004322	NCBI	NM_004322
BAD	AK023420	NCBI	AK023420
BAD	NM_032989	NCBI	NM_032989
BAG1	NM_004323	NCBI	NM_004323
Bak	NM_001188	NCBI	NM_001188
Bax	NM_004324	NCBI	NM_004324
BBC3	NM_014417	NCBI	NM_014417
BCAR1	NM_014567	NCBI	NM_014567
BCAR3	NM_003567	NCBI	NM_003567
Bcl2	NM_000633	NCBI	NM_000633
BCL2-related	BC017197	NCBI	BC017197
BCL2L10	NM_020396	NCBI	NM_020396
BCL2L11	NM_006538	NCBI	NM_006538
BCL2L12	NM_052842	NCBI	NM_052842
BOX	NM_001191	NCBI	NM_001191
BCRP	NM_004827	NCBI	NM_004827
BECN1	NM_003766	NCBI	NM_003766
BFGF	NM_007083	NCBI	NM_007083
BG675392	BG675392	NCBI	BG675392
BID	NM_001196	NCBI	NM_001196
BIK	NM_001197	NCBI	NM_001197
BIN1	NM_004305	NCBI	NM_004305
BIM	NM_000057	NCBI	NM_000057
BNIP1	NM_013979	NCBI	NM_013979
BNIP2	NM_004330	NCBI	NM_004330
BNIP3	NM_004052	NCBI	NM_004052
BNIP3L	NM_004331	NCBI	NM_004331
BPAG1	NM_015548	NCBI	NM_015548
BRAF	NM_004333	NCBI	NM_004333
BRCA1	NM_007295	NCBI	NM_007295
BRCA2	NM_000059	NCBI	NM_000059
BRK	NM_005975	NCBI	NM_005975
BRMS1	AF159141	NCBI	AF159141
BRS3	NM_001727	NCBI	NM_001727
BTC	NM_001729	NCBI	NM_001729
BTF3	NM_001207	NCBI	NM_001207
BUB1	NM_004336	NCBI	NM_004336
c-abl	NM_005157	NCBI	NM_005157
c-kit	NM_000222	NCBI	NM_000222

FIG. 6

Gene Name	Sequence ID	Sequence Source	GHIallgenes2-03.txt Accession Number
c-myb	NM_005375	NCBI	NM_005375
C20 orf1	NM_012112	NCBI	NM_012112
C20orf103	NK-012261	NCBI	NM_012261
c20orflO8	NM_080821	NCBI	NM_080821
CA9	NM_001216	NCBI	NM_001216
CACNA2D2	NM_006030	NCBI	NM_006030
Cad17	NM_004063	NCBI	NM_004063
CASP8	NM_033357	NCBI	NM_033357
CCNA2	NM_001237	NCBI	NM_001237
CCNB1	NM_031966	NCBI	NM_031966
CCNB2	NM_004701	NCBI	NM_004701
CCND1	NM_001758	NCBI	NM_001758
CCND3	NM_001760	NCBI	NM_001760
CCNE1	NM_001238	NCBI	NM_001238
CCNE2	NM_057749	NCBI	NM_057749
CCNE2 variant 1	NM_057749var1	NCBI	NM_057749
CCNE2 variant 3	NM_004702	NCBI	NM_004702
CD105	NM_000118	NCBI	NM_000118
CD134	NM_003327	NCBI	NM_003327
CD18	NM_000211	NCBI	NM_000211
CD31	NM_000442	NCBI	NM_000442
CD3z	NM_000734	NCBI	NM_000734
CD40	NM_000074	NCBI	NM_000074
CD44E	X55150	NCBI	X55150
CD44s	M59040	NCBI	M59040
CD44v3	AJ251595v3	NCBI	AJ251595v3
CD44v6	AJ251595v6	NCBI	AJ251595v6
CD68	NM_001251	NCBI	NM_001251
CD82	NM_002231	NCBI	NM_002231
CD9	NM_001769	NCBI	NK-001769
CDC20	NM_001255	NCBI	NM_001255
cdc25A	NM_001789	NCBI	NM_001789
CDC25B	NM_021874	NCBI	NM_021874
CDH1	NM_004360	NCBI	NM_004360
CDX2	NM_001265	NCBI	NM_001265
CEACAM6	NM_002483	NCBI	NM_002483
CEGP1	NM_020974	NCBI	NM_020974
CEGP1 intron 1	NM_020974int1	NCBI	NM_020974int1
CEGP1 intron 2	NM_020974int2	NCBI	NM_020974int2
CEGP1 intron 3	NM_020974int3	NCBI	NM_020974int3

FIG. 6

Gene Name	Sequence ID	Sequence Source	GHIallgenes2-03.txt Accession Number
CEGP1 intron 4	NM_020974int4	NCBI	NM_020974int4
CEGP1 intron 5	NM_020974int5	NCBI	NM_020974int5
CEGP1 intron 7	NM_020974int7	NCBI	NM_020974int7
CGA	NM_001275	NCBI	NM_001275
CHAF1B	NM_005441	NCBI	NM_005441
chk1	NM_001274	NCBI	NM_001274
chk2	NM_007194	NCBI	NM_007194
CIAP1	NM_001166	NCBI	NM_001166
CIAP2	NM_001165	NCBI	NM_001165
CKAP4	NM_006825	NCBI	NM_006825
Claudin	4 NM_001305	NCBI	NM_001305
cMet	NM_000245	NCBI	NM_000245
cMYC	NM_002467	NCBI	NM_002467
CNN	NM_001299	NCBI	NM_001299
COL1A1	NM_000088	NCBI	NM_000088
COL1A2	NM_000089	NCBI	NM_000089
Contig 27882	AK000618	NCBI	AK000618
Contig 36744	XM_087225	NCBI	XM-087225
Contig 51037	XM_058945	NCBI	XM_058945
Contig38438	AI744123	NCBI	AI744123
Contig44799	Contig44799	Rosetta Contigs	Contig44799
Contig46653	Contig46653	Rosetta Contigs	Contig46653
Contig47405	Contig47405	Rosetta Contigs	Contig47405
COX2	NM_000963	NCBI	NM_000963
CRBP	NM_002899	NCBI	NM_002899
cripto	NM_003212	NCBI	NM_003212
CRK7	NM_016507	NCBI	NM_016507
CRMP1	NM_001313	NCBI	NM_001313
CSF1	NM_000757	NCBI	NM_000757
CSF1R	NM_005211	NCBI	NM_005211
CSF3	NM_000759	NCBI	NM_000759
CSNK1D	NM_001893	NCBI	NM_001893
CSTF1	NM_001324	NCBI	NM_001324
CTSB	NM_001908	NCBI	NM_001908
CTSD	NM_001909	NCBI	NM_001909
CTSH	NM_004390	NCBI	NM_004390
CTSL	NM_001912	NCBI	NM_001912
CTSL2	NM_001333	NCBI	NM_001333
Cyclin C	NM_005190	NCBI	NM_005190
Cyclin G1	NM_004060	NCBI	NM_004060

FIG. 6

Gene Name	Sequence ID	Sequence Source	GHIallgenes2-03.txt Accession Number
Cyclin G2	NM_004354	NCBI	NM_004354
Cyclin K	BC015935	NCBI	BC015935
CYP	NM_006347	NCBI	NM_006347
CYP11B1	NM_000497	NCBI	NM_000497
CYP17	NM_000102	NCBI	NM_000102
CYP1A1	NM_000499	NCBI	NM_000499
CYP1A2	NM_000761	NCBI	NM_000761
CYP181	NM_000104	NCBI	NM_000104
CYP21A2	NM_000500	NCBI	NM_000500
CYP24	NM_000782	NCBI	NM_000782
CYP27A1	NM_000784	NCBI	NM_000784
CYP27B1	NM_000785	NCBI	NM_000785
CYP2A13	NM_000766	NCBI	NM_000766
CYP2A6	NM_000762	NCBI	NM_000762
CYP2A7	NM_000764	NCBI	NM_000764
CYP2A7	NM_030589	NCBI	NM_030589
CYP2C18	NM_000772	NCBI	NM_000772
CYP2C19	NM_000769	NCBI	NM_000769
CYP2C8	NM_030878	NCBI	NM_030878
CYP2C8	NM_000770	NCBI	NM_000770
CYP2C9	NM_000771	NCBI	NM_000771
CYP2D6	NM_000106	NCBI	NM_000106
CYP2E	NM_000773	NCBI	NM_000773
CYP2F1	NM_000774	NCBI	NM_000774
CYP2S1	NM_030622	NCBI	NM_030622
CYP39A1	NM_016593	NCBI	NM_016593
CYP3A4	NM_017460	NCBI	NM_017460
CYP3A43	NM_022820	NCBI	NM_022820
CYP4F12	NM_023944	NCBI	NM_023944
CYP4F2	NM_001082	NCBI	NM_001082
CYP51	NM_000786	NCBI	NM_000786
CYP7A1	NM_000780	NCBI	NM_000780
CYP7B1	NM_004820	NCBI	NM_004820
CYP8B1	NM_004391	NCBI	NM_004391
DAPK1	NM_004938	NCBI	NM_004938
DBC2	NM_015178	NCBI	NM_015178
DCC_exons18-23	X76132-18-23	NCBI	X76132_18-23
DCC_exons6-7	X76132_6-7	NCBI	X76132_6-7
DCK	NM_000788	NCBI	NM_000788
DCR3	NM_016434	NCBI	NM_016434

FIG. 6

Gene Name	Sequence ID	Sequence Source	GHIallgenes2-03.txt Accession Number
DHFR	NM_000791	NCBI	NM_000791
DHPS	NM_013407	NCBI	NM_013407
DIABLO	NM_019887	NCBI	NM_019887
DKFZp564	XM_047080	NCBI	XM_047080
DKFZp586M0723	AL050227	NCBI	AL050227
DNMT3A	NM_022552	NCBI	NM_022552
DPYD	NM_000110	NCBI	NM_000110
DR4	NM_003844	NCBI	NM_003844
DR5	NM_003842	NCBI	NM_003842
E2F1	NM_005225	NCBI	NM_005225
EDN1 endothelin	NM_001955	NCBI	NM_001955
EGF	NM_001963	NCBI	NM_001963
EGFR	NM_005228	NCBI	NM_005228
EGFRd27	EGFRd27	GHI Custom	EGFRd27
EIF4E	NM_001968	NCBI	NM_001968
EIF4EL3	NM_004846	NCBI	NM_004846
EMP1	NM_001423	NCBI	NM_001423
EMS1	NM_005231	NCBI	NM_005231
EN01	NM_001428	NCBI	NM_001428
EpCAM	NM_002354	NCBI	NM_002354
EPHX1	NM_000120	NCBI	NM_000120
ER2	NM_001437	NCBI	NM_001437
ErbB3	NM_001982	NCBI	NM_001982
ERBB4	NM_005235	NCBI	NM_005235
ERCC1	NM_001983	NCBI	NM_001983
EREG	NM_001432	NCBI	NM_001432
ERK1	Z11696	NCBI	Z11696
ERK2	NM_002745	NCBI	NM_002745
ERRa	NM_004451	NCBI	NM_004451
ESM1	NM_007036	NCBI	NM_007036
EstR1	NM_000125	NCBI	NM_000125
F2	NM_000506	NCBI	NM_000506
F2R	NM_001992	NCBI	NM_001992
F2RL2	NM_004101	NCBI	NM_004101
fas	NM_000043	NCBI	NM_000043
fasl	NM_000639	NCBI	NM_000639
FBXO5	NM_012177	NCBI	NM_012177
FGF 19	NM_005117	NCBI	NM_005117
FGF1	NM_000800	NCBI	NM_000800
FGF18	NM_003862	NCBI	NM_003862

FIG. 6

Gene Name	Sequence ID	Sequence Source	GHIallgenes2-03.txt Accession Number
FGF2	NM_002006	NCBI	NM_002006
FGF8	NM_033163	NCBI	NM_033163
FGFR 2	NM_023028	NCBI	NM_023028
FGFR 3	NM_000142	NCBI	NM_000142
FGFR1	NM_023109	NCBI	NM_023109
FGFR4	NM_002011	NCBI	NM_002011
FHIT	NM_002012	NCBI	NM_002012
FKBP4	NM_002014	NCBI	NM_002014
FL310713	NM_018189	NCBI	NM_018189
FLJ20354	NM_017779	NCBI	NM_017779
FLT1	NM_002019	NCBI	NM_002019
FN1	NM_002026	NCBI	NM_002026
FOLR1	NM_016730	NCBI	NM_016730
FOXM1	NM_021953	NCBI	NM_021953
FOXM1 intron 3	NM_021953 int3	NCBI	NM_021953 int3
FOXM1 intron 4	NM_021953 int4	NCBI	NM_021953 int4
FOXM1 intron 5	NM_021953 int5	NCBI	NM_021953 int5
FOXM1 intron 7	NM_021953 int7	NCBI	NM_021953 int7
FPPS	NM_002004	NCBI	NM_002004
FRPI	NM_003012	NCBI	NM_003012
Furin	NM_002569	NCBI	NM_002569
FUS	NM_004960	NCBI	NM_004960
FUT1	NM_000148	NCBI	NM_000148
FUT2	NM_000511	NCBI	NM_000511
G-Catenin	NM_002230	NCBI	NM_002230
GAMMA	NM_003890	NCBI	NM_003890
GAPDH	NM_002046	NCBI	NM_002046
GATA3	NM_002051	NCBI	NM_002051
GCLC	NM_001498	NCBI	NM_001498
GCLM	NM_002061	NCBI	NM_002061
GGPS1	NM_004837	NCBI	NM_004837
GGT	X98922	NCBI	X98922
gpl30	NM_002184	NCBI	NM_002184
GPC1	NM_002081	NCBI	NM_002081
GPC3	NM_004484	NCBI	NM_004484
GPX1	NM_000581	NCBI	NM_000581
GPX2	NM_002083	NCBI	NM_002083
GPX3	NM_002084	NCBI	NM_002084
GPX4	NM_002085	NCBI	NM_002085
GRB7	NM_005310	NCBI	NM_005310

FIG. 6

Gene Name	Sequence ID	Sequence Source	GHIallgenes2-03.txt Accession Number
GRO1	NM_001511	NCBI	NM_001511
GRP	NM_002091	NCBI	NM_002091
GRPR	NM_005314	NCBI	NM_005314
GSN	NM_000177	NCBI	NM_000177
GSTM1	NM_000561	NCBI	NM_000561
GSTM3	NM_000849	NCBI	NM_000849
GSTp	NM_000852	NCBI	NM_000852
GSTT1	NM_000853	NCBI	NM_000853
GSTT2	NM_000854	NCBI	NM_000854
GUS	NM_000181	NCBI	NM_000181
H2AFZ	NM_002106	NCBI	NM_002106
Ha-Ras	NM_005343	NCBI	NM_005343
HB-EGF	NM_001945	NCBI	NM_001945
hCRA a	U78556	NCBI	U78556
hENT1	NM_004955	NCBI	NM_004955
Hepsin	NM_002151	NCBI	NM_002151
HER2	NM_004448	NCBI	NM_004448
her2P	J05264	NCBI	J05264
HGF	M29145	NCBI	M29145
HIF1A	NM_001530	NCBI	NM_001530
HLA-DPB1	NM_002121	NCBI	NM_002121
HLA-G	NM_002127	NCBI	NM_002127
HNF3A	NM_004496	NCBI	NM_004496
HNRPAB	NM_004499	NCBI	NM_004499
HOXA5	NM_019102	NCBI	NM_019102
HOXB7	NM_004502	NCBI	NM_004502
IBSP	NM_004967	NCBI	NM_004967
Id-3	NM_002167	NCBI	NM_002167
ID1	NM_002165	NCBI	NM_002165
ID2	NM_002166	NCBI	NM_002166
IER3	NM_052815	NCBI	NM_052815
IGF1	NM_000618	NCBI	NM_000618
IGF1R	NM_000875	NCBI	NM_000875
IGF2	NM_000612	NCBI	NM_000612
IGFBP1	XM004688	NCBI	XM_004688
IGFBP2	NM_000597	NCBI	NM_000597
IGFBP2(XM)	XM_002636	NCBI	XM_002636
IGFBP3	NM_000598	NCBI	NM_000598
IGFBP4	XM_049937	NCBI	XM_049937
IGFBP5	AY052629	NCBI	AY052629

FIG. 6

Gene Name	Sequence ID	Sequence Source	GHIallgenes2-03.txt Accession Number
IGFBP5	NM_000599	NCBI	NM_000599
IGFBP6	NM_002178	NCBI	NM_002178
IGFBP7	NM_001553	NCBI	NM_001553
IL10	NM_000572	NCBI	NM_000572
IL6	NM_000600	NCBI	NM_000600
ILT-2	NM_006669	NCBI	NM_006669
ING1	NM_005537	NCBI	NM_005537
INSM1	NM_002196	NCBI	NM_002196
IRS1	NM_005544	NCBI	NM_005544
ITGA3	NM_002204	NCBI	NM_002204
ITGA7	NM_002206	NCBI	NM_002206
ITGB3	NM_000212	NCBI	NM_000212
k-ras	NM_033360	NCBI	NM_033360
KB1527G9 8q23	KB1527G9	NCBI	KB1527G9
KDR	NM_002253	NCBI	NM_002253
Ki-67	NM_002417	NCBI	NM_002417
KIAA1209	AJ420468	NCBI	AJ420468
Kitlng	NM_000899	NCBI	NM_000899
KLK10	NM_002776	NCBI	NM_002776
KLK11	NM_006853	NCBI	NM_006853
KNSL2	BC000712	NCBI	BC000712
KRT 4	NM_002272	NCBI	NM_002272
KRT14	NM_000526	NCBI	NM_000526
KRT17	NM_000422	NCBI	NM_000422
KRT18	NM_000224	NCBI	NM_00.0224
KRT19	NM_002276	NCBI	NM_002276
KRT5	NM_000424	NCBI	NM_000424
KRT8	NM_002273	NCBI	NM_002273
LAMC2	NM_005562	NCBI	NM_005562
LAMP2	NM_013995	NCBI	NM_013995
LMNB1	NM_005573	NCBI	NM_005573
LMYC	NM_012421	NCBI	NM_012421
LOC51038	NM_015858	NCBI	NM_015858
LOT1 variant 1	NM_002656	NCBI	NM_002656
Lot1 variant 2	NM_006718	NCBI	NM_006718
LPL	NM_000237	NCBI	NM_000237
LTA	NM_000595	NCBI	NM_000595
M20259	M20259	NCBI	M20259
MAGEE 1	NM_016249	NCBI	NM_016249
MAPK4	NM_002747	NCBI	NM_002747

FIG. 6

Gene Name	Sequence ID	Sequence Source	GHIallgenes2-03.txt Accession Number
Maspin	NM_002639	NCBI	NM_002639
MCJ	NM_013238	NCBI	NM_013238
MCL1	NM_021960	NCBI	NM_021960
MCM2	NM_004526	NCBI	NM_004526
MCM3	NM_002388	NCBI	NM_002388
MCM6	NM_005915	NCBI	NM_005915
MCM7	NM_005916	NCBI	NM_005916
MCP1	NM_002982	NCBI	NM_002982
MDM2	NM_002392	NCBI	NM_002392
MDS028	NM_018463	NCBI	NM_018463
MEL	NM_005370	NCBI	NM_005370
MELK	NM_014791	NCBI	NM_014791
Mgb1	NM_002411	NCBI	NM_002411
MGMT	NM_002412	NCBI	NM_002412
MGST1	NM_020300	NCBI	NM_020300
MLH1	NM_000249	NCBI	NM_000249
MLL	NM_005933	NCBI	NM_005933
MMP12	NM_002426	NCBI	NM_002426
MMP2	NM_004530	NCBI	NM_004530
MMP9	NM_004994	NCBI	NM_004994
MRP1	NM_004996	NCBI	NM_004996
MRP2	NM_000392	NCBI	NM_000392
MRP3	NM_003786	NCBI	NM_003786
MRP4	NM_005845	NCBI	NM_005845
MSH2	NM_000251	NCBI	NM_000251
MSH3	NM_002439	NCBI	NM_002439
MSH6	NM_000179	NCBI	NM_000179
MT3	NM_005954	NCBI	NM_005954
MTA1	NM_004689	NCBI	NM_004689
MUC1	NM_002456	NCBI	NM_002456
MUC1	J05582	NCBI	J05582
MUC3	AF007194	NCBI	AF007194
MVP	NM_017458	NCBI	NM_017458
MYBL2	NM_002466	NCBI	NM_002466
MYH11	NM_002474	NCBI	NM_002474
MYLK	NM_053025	NCBI	NM_053025
MYRIP	NM_015460	Incyte Diagnostic	NM_015460
NADPHP450	AF258341	NCBI	AF258341
NCAM1	NM_000615	NCBI	NM_000615
NEK2	NM_002497	NCBI	NM_002497

FIG. 6

Gene Name	Sequence ID	Sequence Source	GHIallgenes2-03.txt Accession Number
NFKBp50	NM_003998	NCBI	NM_003998
NFKBp65	NM_021975	NCBI	NM_021975
NMB	NM_021077	NCBI	NM_021077
NMBR	NM_002511	NCBI	NM_002511
NME1	NM_000269	NCBI	NM_000269
NMU	NM_006681	NCBI	NM_006681
NMYC	NM_005378	NCBI	NM_005378
NPDO09	NM_020686	NCBI	NM_020686
NR4A1	NM_002135	NCBI	NM_002135
NRG1	NM_013957	NCBI	NM_013957
NRP1	NM_003873	NCBI	NM_003873
NRP2	NM_003872	NCBI	NM_003872
ODC1	NM_002539	NCBI	NM_002539
OPN, osteopontin	NM_000582	NCBI	NM_000582
ORC3	NM_012381	NCBI	NM_012381
Osteonectin	NM_003118	NCBI	NM_003118
P14ARF	S78535	NCBI	S78535
p14ARF	NM_000077P14	NCBI	NM_000077
p16-INK4	L27211	NCBI	L27211
P161NK4	NM_000077	NCBI	NM_000077
p21	NM_000389	NCBI	NM_000389
p27	NM_004064	NCBI	NM_004064
P40	NM_006824	NCBI	NM_006824
P53	NM_000546	NCBI	NM_000546
p53R2	AB036063	NCBI	AB036063
p63	NM_003722	NCBI	NM_003722
P66	AB032976	NCBI	AB032976
PAI1	NM_000602	NCBI	NM_000602
PBGD	X04217	NCBI	X04217
PCNA	NM_002592	NCBI	NM_002592
PCP4	NM_006198	NCBI	NM_006198
PDGFA	NM_002607	NCBI	NM_002607
PDGFB	NM_002608	NCBI	NM_002608
PDGFC	NM_016205	NCBI	NM_016205
PDGFD	NM_025208	NCBI	NM_025208
PDGFRa	NM_006206	NCBI	NM_006206
PDGFRb	NM_002609	NCBI	NM_002609
PECI	NM_006117	NCBI	NM_006117
PGK1	NM_000291	NCBI	NM_000291
P13K	NM_002646	NCBI	NM_002646

FIG. 6

Gene Name	Sequence ID	Sequence Source	GHIallgenes2-03.txt Accession Number
P13Kc2A	NM_002645	NCBI	NM_002645
PIK3R1	M61906	NCBI	M61906
Pin1	NM_006221	NCBI	NM_006221
PKCe	NM_005400	NCBI	NM_005400
PKLR	NM_000298	NCBI	NM_000298
PLAUR	NM_002659	NCBI	NK-002659
PLK	NM_005030	NCBI	NM_005030
PMS1	NM_000534	NCBI	NM_000534
PMS2	NM_000535	NCBI	NM_000535
PPARG	NM_005037	NCBI	NM_005037
PPM1D	NM_003620	NCBI	NM_003620
PR	NM_000926	NCBI	NM_000926
PRAME	NM_006115	NCBI	NM_006115
PRAME intron 1	NM_006115int1	NCBI	NM_006115int1
PRAME intron 2	NM_006115int2	NCBI	NM_006115int2
PRAME intron 3	NM_006115int3	NCBI	NM_006115int3
PRAME intron 4	NM_006115int4	NCBI	NM_006115int4
PRAME intron 5	NM_006115int5	NCBI	NM_006115int5
PREP	NM_002726	NCBI	NM_002726
PRKCD	NM_006254	NCBI	NM_006254
PRKCG	NM_002739	NCBI	NM_002739
PR02000	NM_014109	NCBI	NM_014109
pS2	NM_003225	NCBI	NM_003225
PTDO16	NM_016125	NCBI	NM_016125
PTEN	NM_000314	NCBI	NM_000314
PTHLH	NM_002820	NCBI	NM_002820
PTPD1	NM_007039	NCBI	NM_007039
PTTG1	NM_004219	NCBI	NM_004219
PUNC	AK095529	NCBI	AK095529
Q9BQI9	NM_031474	NCBI	NM_031474
Q9BSD3	NM_031465	NCBI	NM_031465
RAB27B	NM_004163	NCBI	NM_004163
RAD51C	NM_058216	NCBI	NM_058216
RAD54L	NM_003579	NCBI	NM_003579
RANBP2	NM_006267	NCBI	NM_006267
RARA	NM_000964	NCBI	NM_000964
RARB	NM_016152	NCBI	NM_016152
RASSF1	NM_007182	NCBI	NM_007182
RB1	NM_000321	NCBI	NM_000321
RBM5	NM_005778	NCBI	NM_005778

FIG. 6

Gene Name	Sequence ID	Sequence Source	GHIallgenes2-03.txt Accession Number
RBP4	NM_006744	NCBI	NM_006744
RERG	NM_032918	NCBI	NM_032918
RFC	NM_003056	NCBI	NM_003056
RFC4	NM_002916	NCBI	NM_002916
rhoc	NM_005167	NCBI	NM_005167
RIZ1	NM_012231	NCBI	NM_012231
RNase P	RNaseP	GHI Custom	RNaseP
RPL19	NM_000981	NCBI	NM_000981
RPLPO	NM_001002	NC5I	NM_001002
RPS6KB1	NM_003161	NCBI	NM_003161
RRM1	NM_001033	NCBI	NM_001033
RUNX1	NM_001754	NCBI	NM_001754
SDC1	NM_002997	NCBI	NM_002997
SEMA3B	NM_004636	NCBI	NM_004636
SEMA3F	NM_004186	NCBI	NM_004186
SERPINC1	NM_000488	NCBI	NM_000488
SIR2	NM_012238	NCBI	NM_012238
SLC19A3	NM_025243	NCBI	NM_025243
SLC20A1	NM_005415	NCBI	NM_005415
SLC2A3	NM_006931	NCBI	NM_-006931
Smad4	NM_005359	NCBI	NM_005359
SNRPF	NM_003095	NCBI	NM_003095
SPRY1	AK026960	NCBI	AK026960
SPRY2	NM_005842	NCBI	NM_005842
Src	NM_004383	NCBI	NM_004383
STAT1	NM_007315	NCBI	NM_007315
STAT3	NM_003150	NCBI	NM_003150
STAT5A	NM_003152	NCBI	NM_0031-52
STAT5B	NM_012448	NCBI	NM_012448
STC1	NM_003155	NCBI	NM_003155
STC2	AK027663	NCBI	AK027663
STC2	NM_003714	NCBI	NM_003714
STK11	NM_000455	NCBI	NM_000455
STK15	NM_003600	NCBI	NM_003600
STK15 intron 1	NM_003600int1	NCBI	NM_003600int1
STK15 intron 2	NM_003600int2	NCBI	NM_003600int2
STK15 intron 3	NM_003600int3	NCBI	NM_003600int3
STK15 intron 4	NM_003600int4	NCBI	NM_003600int4
STK15 intron 6	NM_003600int6	NCBI	NM_003600int6
STMY3	NM_005940	NCBI	NM_005940

FIG. 6

Gene Name	Sequence ID	Sequence Source	GHIallgenes2-03.txt Accession Number
SUHW1	NM_080740	NCBI	NM_080740
SURF6	NM_006753	NCBI	NM_006753
Surfact A1	NM_005411	NCBI	NM_005411
SURV	NM_001168	NCBI	NM_001168
SXR	NM_003889	NCBI	NM_003889
SYK	NM_003177	NCBI	NM_003177
TAGLN	NM_003186	NCBI	NM_003186
TBD	NM_016261	NCBI	NM_016261
TBP	NM_003194	NCBI	NM_003194
TBX2	NM_005994	NCBI	NM_005994
TEK	NM_000459	NCBI	NM_000459
TERC	U86046	NCBI	U86046
TERT	NM_003219	NCBI	NM_003219
TFF3	NM_003226	NCBI	NM_003226
TFRC	NM_003234	NCBI	NM_003234
TGFA	NM_003236	NCBI	NM_003236
TGFbl	NM_000660	NCBI	NM_000660
TGFB3	NM_003239	NCBI	NM_003239
TGFBR2	NM_003242	NCBI	NM_003242
Thrombospondin 1	NM_003246	NCBI	NM_003246
Thymosin B	NM_021992	NCBI	NM_021992
TIE	NM_005424	NCBI	NM_005424
TIMP1	NM_003254	NCBI	NM_003254
TIMP2	NM_003255	NCBI	NM_003255
TIMP3	NM_000362	NCBI	NM_000362
TITF1	NM_003317	NCBI	NM_003317
TJP1	NM_003257	NCBI	NM_003257
TK1	NM_003258	NCBI	NM_003258
TNF	NM_000594	NCBI	NM_000594
TNFRSF11A	NM_003839	NCBI	NM_003839
TNFRSF11B	NM_002546	NCBI	NM_002546
TNFSF11	NM_003701	NCBI	NM_003701
TNFSF11	NM_033012	NCBI	NM_033012
TOP2A	NM_001067	NCBI	NM_001067
TOP2AF	AF071738	NCBI	AF071738
TOP2B	NM_001068	NCBI	NM_001068
TP	NM_001953	NCBI	NM_001953
TP53BP1	NM_005657	NCBI	NM_005657
TP53BP2	NM_005426	NCBI	NM_005426
TRAIL	NM_003810	NCBI	NM_003810

FIG. 6

Gene Name	Sequence ID	Sequence Source	GHIallgenes2-03.txt Accession Number
TS	NM_001071	NCBI	NM_001071
TSC2	NM_000548	NCBI	NM_000548
TULP3	NM_003324	NCBI	NM_003324
upa	NM_002658	NCBI	NM_002658
VCAM1	NM_001078	NCBI	NM_001078
VDR	NM_000376	NCBI	NM_000376
VEGF	NM_003376	NCBI	NM_003376
VEGFB	NM_003317	NCBI	NM_003377
VEGFC	NM_005429	NCBI	NM_005429
vim	NM_003380	NCBI	NM_003380
WISP1	NM_003882	NCBI	NM_003882
Wnt-5a	NM_003392	NCBI	NM_003392
Wnt-5b	NM_032642	NCBI	NM_032642
wwox	NM_016373	NCBI	NM_016373
XIAP	NM_001167	NCBI	NM_001167
XIST	M97168	NCBI	M97168
YB-1	NM_004559	NCBI	NM_004559
ZNF217	NM_006526	NCBI	NM_006526

FIG. 6

SEQUENCE LISTING

<110> GENOMIC HEALTH, INC.
SCOTT, Randal W.
BAKER, Joffre B.
KIEFER, Michael C.

<120> USE OF INTRONIC RNA TO MEASURE GENE EXPRESSION

<130> 39740-0007 PCT

<140> Not Assigned
<141> 2004-02-19

<150> US 60/448,991
<151> 2003-02-20

<160> 67

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 1566
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (802)...(1027)
<223> n= a, t, c, or g

<221> misc_feature
<222> (1126)...(1518)
<223> n= a, t, c, or g

<221> intron
<222> (0)...(0)
<223> MGB-CEGP1 int 1.1

<400> 1
gtgagtgtcc ggccgcgggg gcgcacctgg cacagcaggc agggccagga agagtgttta 60
ggtccccggc ggagtccaga gccgggcgcg cggggctcgg ggctggcggc tgcagctccg 120
cgggggcctc tgctcccccc gggacctcac ccgccggccg ggccaaggcg ccacgaccgc 180
tggggccctg agtccttcgg cccggcctcg gacccggagc tgctgacggt tcccgccccg 240
gtccggatgc ctccagagcg cctgctagtc agaccgtcgc cggcgagcag gcaggagggt 300
gcggaccctg gccttggggt cccgcgcctc agcgtaggcg gggaaactga gggccgggcc 360
gggcacatcc gcgaggcggt ggcagctttg ccgtttcttt ctttgggggc cggcaagttc 420
tgctgatggc ttcggggtgg gctccagaga cttttctgtc agcggaacag cgcctgttcc 480
gatctgggaa ttaccctgaa gcagcaacaa gcctaggttt tcagcagaga actttggttt 540
ccagagagga ctctggacgt gctgtgctta ctggacttgc aatactttca aaatgctttt 600
gtttttaatt aatatcctgg agtagtgtca acccaggaaa tacttctgcc aaggcgggtt 660
tccaggttga gaggatgggc aggggtggga gtgcaggggg ccggccatgg ggacaccatc 720
cccgcttcgc agcatctgag agccctggat gacatctgct ccgatcccgg ggcagacttc 780
ccataaatac tctaaaccag cnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 840
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 900
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 960
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1020
nnnnnnnctc tgagctccga gaaagctgac agacagctgc ttggtgttca gagcttgtct 1080
gtccgtttgg tcctttcctc ctttagcggg catgtaggta ctattnnnnn nnnnnnnnnn 1140
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1200
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1260
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1320
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1380
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1440
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1500
nnnnnnnnnn nnnnnnnnca catggcctgg gagcctgtac caggtgtcag ctgtgctctt 1560

ttgcag 1566

<210> 2
<211> 4985
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (83)...(253)
<223> n= a, t, c, or g

<221> misc_feature
<222> (1975)...(2296)
<223> n= a, t, c, or g

<221> misc_feature
<222> (2445)...(2468)
<223> n= a, t, c, or g

<221> misc_feature
<222> (2499)...(2551)
<223> n= a, t, c, or g

<221> misc_feature
<222> (4295)...(4352)
<223> n= a, t, c, or g

<221> intron
<222> (0)...(0)
<223> MGB-CEGP1 int 3.1

<400> 2
gtacctctgc ccagctgtgg atgggggcag agccacatct gagaccctct cccttgcacg 60
cgcacacaca cactgactct agnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
nnnnnnnnnn nnnatcttta catagaatac atttcaaaca tgactagatg tctcaggagc 300
aatatagtgg atgatctgcc aagtttttca aaaaggtgct gaaaaccaca gcaccagtat 360
gagcctgctc cctgctctgg gtgggtaggg aggaggctgg atccttccca tgcagacttt 420
caatgaagtg ccctgttttc agcccccaagc tagatccggc ccttccatgt tttgcatttt 480
tgagctccga ggggcagaag ggctccctcc ctggactttc cgtgctgtgg tttccttcgc 540
ctacgtcacc atttatcatt cctctgtaaa tttgccggaa actcttctct tctgatgtcc 600
ttctcttcat tctctttgct ttgagtttat accttttttc attcctctgt tacttagtag 660
attcttgaga ggaaggggca ttaagtacat gtggccaatc agttattttt aactgaatgt 720
catcctttta actcttccct gctctttctt aagctaaaga gtcacatttt ggtggctgtg 780
ttcctcttgg agttgcatct gcctattttt aggggaagtg ccctaaatac tagcctatta 840
acccctttgg ccatgtgctg cttattcttt cccattactt aagaatgagg tcattttaat 900
ttcttctact atttaatcac aaatttatag attgttttaa tcctggtctt ggtaactttt 960
caagggtttc ttcatggaag atgatttttg tctcattttc caaggatggc agctcacacc 1020
ttatacttaa ctagaatacc tgtttgggta ccaagaaaaa ttgtcagagg aacccccagg 1080
ggccaatggg tttgatggct atcatcaccc agagcctgct cattctcagc gtttggggcg 1140
gggaagtcac acatactggc tttgatcagg cagatttcct atcttgtgcc aggtgtggcc 1200
cttgataaag tagcagttgg gtttcatttt cctgccaggt tctctggggt cattggtgtg 1260
ccctgcactc ttgtccaatg taggccaaat tcgagatggg aatgaattag gaggccagtg 1320
gcacagagtg atccgaatct cagggcatct ctccttttga ttgctcaaag ctgcttcctg 1380
ggaagtcact ttggcttcct ctgcaggtgg ctggggaggg atgtgggaac tgcaggttaa 1440
agccatcgct tgagccctca cggtctgggt cccacccagt tacaaagcag ctggtagcga 1500
ttaagatcac ctcttatccc tgtacttcca gagccctggc tcagccccac tctcccctcc 1560
tgcaagcccc cggactgatt agagacacag gctcctcata ccagaagcaa atacaaatgc 1620
agttcctttc tgcaaactgt gttttctaaa ttttctacaa ttcagacatt cttggatccc 1680
ctaaagagta tttgaagtga acatttttgt ctggaactaa aaccaaaatc taagaatttg 1740
cgttgtggtc tggaagtgct ctctgtgatt ttctgttgtg tttcaacctg attgcttggc 1800
aaattcatgg gagtgtcagc caacagatta tagcaattgg taacggagaa cctttgcatc 1860
ctagggtttt gattcttcaa atagaacagc ctgtaaaaag ttttcttcta ggatttcctc 1920
tctgatatgc acattaaact ctatgaaact gtaggcttaa aaacccacag tggtnnnnnn 1980
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2040

```
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2100
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2160
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2220
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2280
nnnnnnnnnn nnnnnnccac agtggatacc ttcaaagtga ttaaaagaag gtaacacagg 2340
aagctagtat tttctattgc tgttgttttt aataattatt taccaaatgt tctttaatat 2400
agggcatcat aatcattgac tctgagggaa agctcaagat actgnnnnnn nnnnnnnnnn 2460
nnnnnnnncc ttagagactc caaagctgtg ataaagagnn nnnnnnnnnn nnnnnnnnnn 2520
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn ncccggcccg cctttgttcc tattcatggg 2580
tgctcaggct ctcagaatga gcactcctct tttgttttgt gtgttctgag aatatttaga 2640
tggtgtactg atgccttttc agggcaacag ggaaggtgtc agggtggcaa agtggaggct 2700
gtgctttcag caggacctgt tacccgtttt atgtcatgtt ttcctcccaa ttcacaaggc 2760
atatttttgt ttggtttcca gaaataatct tcagtggagc cctgatcttg gggtgcacca 2820
gaatgggggga tttccaatgt ttctgagctg tttcccttct ggtgaacgaa ccatcctgga 2880
cgtgacaacc agaccaattt tggaaagagc tagggccatt tgctgggctg cctagtttgg 2940
aacagattaa tctgctcacc ccagcagtgg tcttgcatta agtcagagtg ctacaaaggc 3000
tttgaggtca cttcttgaaa agctgtcagc gtttccagag ccatttaagt ctctattatg 3060
tcttggtaac ttcaggtgta gcttgatgtg gtaggacatt aggtggtagg ttctctgtgt 3120
atcacaatgg catctggcat acaggcattc ttacgaaata tttcttgtgt aggtgaatta 3180
ctctgaggca gtaaaggtca ctttgcaaat gtcttaacag tcttgtaaac agagtgaaaa 3240
agcagcagca gctggcctgt ttgggagtgt actttccagg tgttcctgcc cccatttctt 3300
gggcagtatt atatttaccc ccgagcacta gttacttccc atgctcggct gacccaagga 3360
caaacacaac gctttctggg ccttctcaga caggacactg cttctagagg cagctgtcac 3420
ctcccgcgcc atctcagtac tggggtgcaa atcacatctt cggaattacc agccagagca 3480
agagaaagct ttccaccaat ccagtgcaag tctctttctg tgttaattga cagccacccct 3540
tggcatggat gaatgaatcc cagcaaccag cagactgagt gctggagtgc aggcagctca 3600
taactgtcag gcaaaagagc aagagggttt taagagagac tccagaaagt atgggatata 3660
ttaacccttg cactgtcttc tggaatagga atgacatctg tttgtattaa aacaattgtt 3720
ccgtttaagc acagtttgac agctctggag tgggagctgg agagagaact ttgacttcac 3780
tagaacctgt tggctaaggt tttaggggca caatatagaa gggtgttgga ttctagagaa 3840
gtgaaagcaa cctttttgta ctcgtgttga aaacagtgcc ctactagtat tagagtgtct 3900
cattgataga gagccaatga caaccaagtc cctactctca gagatgtttt agagttacat 3960
tgcacgaatg caaagaagca acataggaac aggtaattaa taataaagta taaactgagc 4020
agatgtcttg aaagtattct agggtatgaa aagaattcct tcaggatgct ggtaggcagc 4080
aggatctcaa agaattagtt ttgagatgag gcagaatgct ggtaaaccac acgggcagtt 4140
accttgctgt gccccctcat ttagatgtgt gccgagccct gcaagaacag aagcagctgt 4200
tccccttccc accatcatac tacaaggtta agcctaatca gaatttactg tatacctcaa 4260
aaaaattgta cagcagctac cacacacgag cacannnnnn nnnnnnnnnn nnnnnnnnnn 4320
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nngttcttgt tgctgttgct gtgtgatgct 4380
gtcagaggct tatgccctga gggagggatc aagggagtgg ctgagggtgg tcacagaaga 4440
cagattccgg ggcatgtggc ccgtacgagg atgccaaaat gccacagtca cactcacctc 4500
agaagggtgg gattggtggg ggcagagagg ggcgttgaaa tgttttgaaa attatcttca 4560
agagtatgtg aaaaaattga gaatcttgat cattctatct gaacattttc ttaggaggat 4620
tctccttttc tctttacatt cttgatcagc tcttgggtaa agacatggca gagataagag 4680
cgtgagtacc agttcctggg gtcagcaggc tctgatcctg catgcaatag agagctccag 4740
tgtattggga aggctcccaa ctcgttagga gagttgagac atcgtatctc ttgggtgaca 4800
gaataaattt ttcatgtcta ttaattggcc taggttgact ttaatgacat atacttttca 4860
aatgtggggc tgatggagac ctaagcagac agatctgtgg gccacccctt agccctttgc 4920
cgctctccca gggctcagga ttctgaccac agcctagtca cctgtcgcac actgctgttt 4980
ttcag                                                             4985
```

<210> 3
<211> 2556
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (507)...(743)
<223> n= a, t, c, or g

<221> misc_feature
<222> (1584)...(1678)
<223> n= a, t, c, or g

<221> misc_feature
<222> (2428)...(2474)

<223> n= a, t, c, or g

<221> intron
<222> (0)...(0)
<223> MGB-CEGP1 int 4.1

<400> 3
gtaagtatgg gccagtgcac acctgccatg ggaaccgtcg tattccacag gctgccttct 60
gtggcccagc tcagaagcac cacctcatgg cacggctgca gcagcaggga aggcagttag 120
cacgggatac cgacctctac caagtacttg ttcactgcag aagggtggtc tcccttaggg 180
aagggaaatg atattttaaa aaggaactca tcaggaggaa atgaaattca ggagtaagga 240
gtgtgaatgt tggggggcag ttctccctgt tcccacagaa taaaaccaaa tgtcctcatc 300
tggcaatcac agctctttgc caccaggtcc tgcttcccct ataaacctca tctgcctcct 360
ttccgcagac actactcccc ttgcctttgg agaacagccc aaatcctttg atgcctccag 420
gcctttccca agccctcctg ccttcctggc gtggtggact ctcactcaac cttcaatatt 480
ctgtttaact tctaataagg ataagcnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 720
nnnnnnnnnn nnnnnnnnnn nnnccatcct ctgctatcag aagcctcctg ggtgcttcag 780
acagggcagc catcttgtac tttggctccc acagcacttt cctcagctgt atagctctgg 840
gttgacttgt gtgttgatgt gtctgtctcc ccaggtatga gccccctcca agtcagggac 900
cttgcctcat ttttcctctc agtcctcccc tggtacctgc tatgggatat gctcagtaca 960
cttgtgttta atgagtgggt aaatgggtgg cctacaccat cgggccgcag ctcctgcacc 1020
acgattgtag taacaaaact ccacctggga acaggaaacc actggcaatt catggtgttc 1080
ctaaaccacg atttatgcca ggggaagcac tgaggagttc cctttaggaa ccttcccaaa 1140
gccatggaca gaagacccct gccatttggt ggggatggtg gtttatggtg agtaggagat 1200
gaggggacag tttcactggt gagggacttc tctccattgt ctccctcaca aagcagactg 1260
ccaccccaaa gctgtccaag ccaaggctgg tgccaccatc acactcaagc aacaggttct 1320
gacatgctct tagggcccct cgaagtcagg ctgtccctga gggcttccag tgagctagca 1380
gagtggagac cattttccca cctccagatc ttcggaagga agacccagac cctccaagac 1440
tcacctgcgg ggcgagaccc tcaacatttc atagtctttc agggaacagt tgctgaaggg 1500
ggcggggggg tgggcacctg taagcttgtt tttaaagatt ttaaatgtct ttaagatatc 1560
actgctcaaa taatattgtt ctgnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1620
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnncg 1680
tttgaggaat taacaaagaa aaaaactaag acctagaatc tcaccacata accagctgtt 1740
tcaatttttc catattccta tttagttgtt gttcatatgc atacacaatt tttacatagc 1800
tataatcaca ggacaacaca aatatgtaat tagttctttt gaattagaaa aattacaaag 1860
ggcctatgta aaatgcaaac actccaaagc atataaagaa aacatgcagt ttcccgcctc 1920
ccgtttccct tgccagaggt aaccacggtt agcagtttga tgaatagata gttttgtagt 1980
tggctttttt tctttttggc ctatcatcaa tacattcata tatagtcttg ataattacca 2040
gttactgtca cgttaattgt gtgcagaatc atcctgtgat tatccttcct tctaactaat 2100
ctagattgaa tctgatgaga gaaattctga catatatgta caaattaaat attgtctgtt 2160
ttattccagc ataaagtgct atagcatttc ccaaagcccc agtacagctg tattaatagg 2220
taaacttctc tagatagaac aaagcagtag tctagaatct cttggtataa tttcccttat 2280
ataataaaag tctctccccc aactctccca tctccctctt cctgtatgac tttgtttaaa 2340
cccatgtttc agcatttcta caatttgtat tgtaactatc tgcatacaca gacaccacag 2400
ggtctgactt ggagttatgt ctttcgtnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2460
nnnnnnnnnn nnnnaatcag aatttctctg gagcaaacac agccctgtgt ttgtggaaat 2520
ctcagtgctt tatgtattga ttcattttgc tgtcag 2556

<210> 4
<211> 716
<212> DNA
<213> Homo sapiens

<220>
<221> intron
<222> (0)...(0)
<223> MGB-CEGP1 int 5.1

<400> 4
gtgagtggca accccaacac tgagtgaggg tctgcaccag cctgcctgtc cctaccccta 60
ccccttaatg gtgtttagca cagatgcagg ctgtttcctg tgcatttgcc cccccagcag 120
gccctgtgct gcttcgcatg ctacagtggg agtggtctag gcctgtgggg aaggcccctc 180
tctccctgtg tgaccttggg aagcccttcc tcctctcctg gactaggctg ctcctaacgc 240
tggtattcca gagactggca caacacctcc caggaggcca gggcagcacg aagttagagc 300

tgtttataat gatgcggcac ttctggccag caggagccag ggccgtatat ttctggcggg 360
atgcctgcct tgcccttcac ggtgtgtcct tcactagctc catttagag gtttccaggc 420
ccaaggctct ttttctcctc gactcagggg actgaagctt gcattcccta gtgtctcttt 480
ggtcagtgca atatacctcc aaaatctttt ccatgtttaa tgtttgctaa ggatctgtgg 540
ccctttaacg ggctgtgtct cccacagagc ctcattacaa cacattttta ttgcgtgaac 600
agagtcacat atctttcatt cctcttatgt ctgggatttc agcaaacaca gttgtatggg 660
gatgagcaat ctaactcatt cagtctgaga accgtgctct tttgcttctc ttgtag 716

<210> 5
<211> 2041
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (559)...(869)
<223> n= a, t, c, or g

<221> misc_feature
<222> (888)...(1052)
<223> n= a, t, c, or g

<221> misc_feature
<222> (1102)...(1380)
<223> n= a, t, c, or g

<221> misc_feature
<222> (1393)...(1692)
<223> n= a, t, c, or g

<221> intron
<222> (1841)...(1907)
<223> FOXM1 int 3.2

<221> intron
<222> (1871)...(1936)
<223> FOXM1 int 3.1

<400> 5
gtaatgtgtc ccacagcaac caaaatcaag gtcagcccag cctgacagtc tctccagtgc 60
tgtactgcaa cttgtatctg ggacagcagt taagtgcaaa ggacactaga atgataaaca 120
aatgtatctt ttagattgtg actcaatctt attgaatcca ggcaaaatca ttaagaagag 180
ctccttaact acttcatgtg ttactaccta aagtccatgg agggtcttca atgtagcact 240
caagcccact tttctgctac actcaacagc cgtcctagat gccagcagct agagtggcta 300
agtagtttta tgaaaatgtc ttgattaaaa aaaaaaatgc tgtctgtgag cctcatgacc 360
caagatgtca tctcctgtag cgtcacatag catttctagt gggcaggggt tttcctttca 420
cttcattcat ggaaagaccg agatgcctgt gagtcaacat agctcacgca gttggtcggt 480
gtcagagcca caaatgaggt cttctgacgg gtgctcaatt ccaagtcaag tgtgctttgt 540
tttcctcatg gtagaactcn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 720
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 840
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnc ctcatggtag aactatgnnn nnnnnnnnnn 900
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 960
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1020
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nncatggtag aacttttaat tttactccct 1080
tccatcagct tactttccta gnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1140
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1200
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1260
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1320
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1380
ttaatttcct agnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1440
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1500
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1560
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1620
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1680

nnnnnnnnnn nntaattccc tagtttctta atttctctga gccacctttc ttgctattga 1740
tcactacctc acagccttac tctgcttttc tagcccctga cagctatcta ggtcttttct 1800
ttatcacaat ctaaggttgg catcagtctt tattcccgta gaatagatgg gtttatggct 1860
gaaggtgacg gctctgcggt gtggagtgtc aggagagttg ccaagagggc tgcaaagaca 1920
ccagacgaag cctgtgctga gcacagtggg aggggcctga ggctggtttc cccatgtgtt 1980
tgaagggtga tgtttctgaa tctaaagtag ctgataacca gttgtcttgc tcttcttcca 2040
g 2041

<210> 6
<211> 993
<212> DNA
<213> Homo sapiens

<220>
<221> intron
<222> (52)...(124)
<223> FOXM1 int 4.1

<221> intron
<222> (110)...(185)
<223> FOXM1 int 4.2

<221> misc_feature
<222> (249)...(397)
<223> n= a, t, c, or g

<221> misc_feature
<222> (442)...(749)
<223> n= a, t, c, or g

<400> 6
gtgaatgccc tgctttcctc taaatagggc ctaagttgga ggttgtcata gccatctcaa 60
aaggaaacaa gttctgctag tgatgctttc atttgatcag gggagagtta gaagccagcc 120
acccaattag tgacttgcac aaaacccagt gaattaagta cacttgacaa ataccaaatg 180
acacattttt gtgccagacc agagcaagga gaaggctgtt ctgacccaac agaaagggct 240
ccccagggnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnntac aagaaattct gggaatgctt 420
gctctaaaaa aagcccttcc tnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 720
nnnnnnnnnn nnnnnnnnnn nnnnnnnnng cctgccttag gctggagacc agaagctgag 780
ctaccagaac gtcttttcag aaagaagtta ttttggtttt tcagagtgcc cataaggctg 840
ctggtagctg taaccattct cctgggaggg gcagttgtct ggggtgtctt ttgtcatcag 900
tcaggaataa gtgtttttcc caatccggtc aaattgacca cgttggtggt aacttcatct 960
catttctctc ccacaatgcc tggccgccac cag 993

<210> 7
<211> 602
<212> DNA
<213> Homo sapiens

<220>
<221> intron
<222> (0)...(0)
<223> FOXM1 int 5.1

<400> 7
gtaaggttct ttccctctgg ctcggggctt ggccttgttt tcctttcact gctcagcatg 60
gctttagtgg acagagacaa gatgtgatgt ggggaagggt ccctatggcc atgttttgtc 120
taggtgccag ccctagacac agaacaccct gagggtcagg cacacaccca cttccctccc 180
cttccatggg catcacaagg gcacactgag cagagcaggg cacagcaggg gagcatgctg 240
cagcagccac aagcgcatgg caccagcctc aggggcggca gttcgttcgc tcacttttgt 300
gcctagcttt tctttgccac gcatatagct acctgctctg gcatccccca ggggtgttga 360
ggacacgtgg gtgaagcggt agtgccactc tgccatcatg tgtctgtagg ccacccacct 420

gcccactcat cacagttttg gagactgctc gcctacgtcc atcccctcag gttggcctcc 480
tctctctggg ctgtcattaa ctcaagcaca caccaccaga gcagctggtg gggttttgcc 540
atcccctctt taccttattg tgttaacata ggtttctttc tctccccatc tgccacaagc 600
ag 602

<210> 8
<211> 4656
<212> DNA
<213> Homo sapiens

<220>
<221> intron
<222> (0)...(0)
<223> FOXM1 int 7.1

<221> misc_feature
<222> (316)...(615)
<223> n= a, t, c, or g

<221> misc_feature
<222> (626)...(930)
<223> n= a, t, c, or g

<221> misc_feature
<222> (1044)...(1259)
<223> n= a, t, c, or g

<221> misc_feature
<222> (1452)...(1802)
<223> n= a, t, c, or g

<221> misc_feature
<222> (2085)...(3116)
<223> n= a, t, c, or g

<221> misc_feature
<222> (3138)...(3433)
<223> n= a, t, c, or g

<221> misc_feature
<222> (3442)...(3728)
<223> n= a, t, c, or g

<221> misc_feature
<222> (3814)...(4081)
<223> n= a, t, c, or g

<221> misc_feature
<222> (4289)...(4371)
<223> n= a, t, c, or g

<400> 8
gtgggtgtcc tattttcctc tgaagagaga ttctggccaa ttaagaatgt tggaccttca 60
gcttgcaaag cactctgata agtgttcctt gagagcttat aaatctagtt gggtagaaaa 120
ggcataaaaa catagggaag tgtaatagca ttagaagagc taaaaaggta tttggattac 180
aatgtaagtg gtgtcagaag gcccataaat acctgatgag cttgtaagaa ttcagacaaa 240
agtgattgtg atagatgggc taggattatt aaggaagata cacaagggag gcaggcctta 300
gaaagagatg gatttnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
nnnnnnnnnn nnnnngtgga tttgannnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 720
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 840
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 900

```
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn gtagatttgg gtaagcaaac aggtgtagag 960
agagcatgct aatgggcagt gccatggagg cgggaaatgc agttcgtacc tggcagtagt 1020
aaagtgactg ggtcagacta actnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1080
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1140
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1200
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnng 1260
tctagcttga ggggaaggtg agaagggtaa attcagagcc aacttggatc agccatcaga 1320
tctgcactta acactgttaa agggttctgt gagtacgggc tgacatgtaa ccaaagtgaa 1380
aagcttcccc catccccttc agagagatga aaatagcata gagtctggag tttagagcga 1440
cttgggtttg cnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1500
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1560
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1620
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1680
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1800
nncatattac acacaaaatt ataccacaca tacataattt agcgtaaatt cattcatgtg 1860
gccgtagcat gtgccctgtt tgggttttca tgcagtgggt tttctcccct ttcctttttg 1920
gctccctctc caccctacca tcacccacat caccccctact cccaagataa ctggttgata 1980
atttatgatg ctttcttgca tattttatca atgctcttag ttatactata catgtatagc 2040
gatagccatt ttatatgtac acatacaaca cacagaacat tgatnnnnnn nnnnnnnnnn 2100
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2160
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2220
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2280
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2340
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2400
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2460
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2520
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2580
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2640
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2700
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2760
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2820
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2880
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2940
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3000
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3060
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnngtta 3120
tcaatttgtg agagctcnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3420
nnnnnnnnnn nnntgtgaga gnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3600
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3720
nnnnnnnncc ctgtgagagc tctttgttgt taaaataatc ttctttcttt tatgctgaag 3780
atatttttct acttctattg tttatctctt tacnnnnnnn nnnnnnnnnn nnnnnnnnnn 3840
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3900
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3960
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 4020
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 4080
ntgtttgttg tttcttaact ttgtttatgg tatctcttgc cacagtaaaa ttttaaagtt 4140
ttatgtagtc aaatgtctct cttctctttt acagtttctg ggtttccagt cttggttaag 4200
aaggtcaccc gcaccctcag attgtatatg tagtctccta gattctcttt caggatttgt 4260
atgattttaa ggttttcatt ttttttttan nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 4320
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn ntaaggttct 4380
tattttcatg cattaaatct ttgtatacag tgtaagacaa gcatgcaatt ttatttcctc 4440
tcggatgaat gctattataa ttatgccact acatactaca tacccgcatc ttttaccccc 4500
agaattgaac taccaacttc aacatacatc gtattctcat atttaataga ttttaagact 4560
tcaaaacgac acaaagagga tcagaacccg tatgtgatat ttttgtgcgt cctgtctggt 4620
gaccgttggt tcaccttatc tctgtttccc tttcag                           4656
```

<210> 9
<211> 614
<212> DNA

<213> Homo sapiens

<220>
<221> intron
<222> (0)...(0)
<223> MGB-PRAME int 2.1

<400> 9
gtaagttcga gccctgattc ctccgcttcc ccgcagggtg accttgggct tgtgcccccg 60
gcaccacccc tgtcccgggt ccctgttttc tctctggaaa tgggttgaag accaaagaaa 120
ataatgtgcg ccacttgggt caccccgggc cgcctgcccc ggaaaattgg ccccagttga 180
ggagttgtgg ctgtaaggat gccttgaacc gaggcggcgg tgctcgtggt tggagctctc 240
cagggtgggt gcgcatttgt aatgcggtgg atgctctggg actcggcccc tctgaaggtg 300
ctgggggttg gggacggccc aggcagtggc gtaggcgtcc taggaaggcg ggagcagagg 360
cagaaatgtc gctgcaagac cgtagtcagg gtccttgacc acaggggtca cttgtgacca 420
accacatggt ctgttgttcc tcctgccccc tggttcagcc caggaaacac tggtgctcag 480
gtttggagcc agagatttgc actgaaaggg cgggattgag tcgccagttg tcagtttcct 540
cagcagtatt tgcggaggtt ttcacaggag gccgttgctt cgtaaatatt atacatgtat 600
tcttcttttt ggag 614

<210> 10
<211> 432
<212> DNA
<213> Homo sapiens

<220>
<221> intron
<222> (0)...(0)
<223> MGB-PRAME int 4.1

<400> 10
gtaagggtga cctagcagct tggtgtgggg ccctgggaac ctgagcagga tgcagctggg 60
gtcagggagc atggagcgcc taaggctggg ccagaggctc tgatggttgc cagcaaggaa 120
gttcagggag gccttggggc tactgcaggg gtcactcttg gaatgggctt ctggacatgg 180
ggcactgatt aaaatgcaga ggtgtctgaa ggaacatgca cctgcttcct cctggtgggg 240
tgggaattgg ggaccaggaa ggatcccagg atcctagtgg gaaagggagc agctgatgcc 300
tgaagtacga agtaaaagtg cagatctaag gtggatgtct gtttggttct tacctacatt 360
atgagactca tggtcttatt ttgagttgat cttaaagcat catctcagct aattacctgt 420
ttttccccac ag 432

<210> 11
<211> 3740
<212> DNA
<213> Homo sapiens

<220>
<221> intron
<222> (0)...(0)
<223> MGB-STK15 int 1.1

<221> misc_feature
<222> (84)...(151)
<223> n= a, t, c, or g

<221> misc_feature
<222> (321)...(491)
<223> n= a, t, c, or g

<221> misc_feature
<222> (537)...(615)
<223> n= a, t, c, or g

<221> misc_feature
<222> (853)...(894)
<223> n= a, t, c, or g

<221> misc_feature

<222> (1259)...(1309)
<223> n= a, t, c, or g

<221> misc_feature
<222> (1337)...(2516)
<223> n= a, t, c, or g

<221> misc_feature
<222> (2521)...(3567)
<223> n= a, t, c, or g

<400> 11
gtacaagggg tttgttgagt ggtgttgaca tgcgcgggag gggtgggtgg gcttcagatt 60
ggattttgtc ctccgagatc accnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nggtaagcgt acggagaact tgcagctggg 180
gtgggtgtta cagaggaaaa gcaggagtgc ggtttaacgg gggccgcttt agatagaata 240
gcctaagaag gcccttgtcc tggctggatg agtgggtgaa ttgatgaatg agaacctcct 300
tgcagaggcc ttcccggtcc nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
nnnnnnnnnn ngggatgcag accggtgcat acaaatcgtc tggggacgtt aaaatgnnnn 540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
nnnnnnnnnn nnnnncactg tccttaactc tcgtaatgtc tcttcctctt ccgtaacctt 660
ccttgtccct tgaattaaac gtttttcagc aacctactca gttcgtcctt cccttcatct 720
ctgcagacat gcacaggtct gagggaggaa ggaataaacc gtataaacct cctgcgctat 780
tagcctaaca gcttttctat tcaaaatagt aggacttctg gtttgaactg aatggatcct 840
gtgaaagtca tcnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnagtctc 900
cttggcgttg tctccagaat tctggattag aatcttattc cattctgctt gttattcaat 960
ttccctagaa agaaaggtag aataaattgg agcaaatgcc tgtagcttct gtcagaagaa 1020
tgttgaataa atgttgttag gcctatgtga tctcattaga ctgctactta gaattgtaag 1080
ggaagtaaag cattagagca tgtgtgaaat taaatatttg attaacacaa gtgtgcattt 1140
ccttgttgct gtttatcaac ttttacttac ccactgtttt tttataaggg ctgcagcctg 1200
tagtctgggc ctggcttcat catggaatta tttgcttaat tgtaaaatgg taatcttaan 1260
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnng gatatttgat 1320
aagaaacttc agtgaannnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1380
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1440
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1500
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1560
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1620
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1680
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1800
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1860
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1920
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1980
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2040
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2100
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2160
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2220
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2280
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2340
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2400
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2460
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnggga 2520
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2580
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2640
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2700
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2760
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2820
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2880
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2940
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3000
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3060
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3300

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3540
nnnnnnnnnn nnnnnnnnnn nnnnnnngca gggaaagcac atgcctgtcc ctccctcatt 3600
agcttcattt ggacaaaaca tgtaaaatcc ggtgtgttgt ggaggccttt tgattgggga 3660
actgtaacgc tgcctatcga gcaacagcac tttaagcagg tggctttgtt caaattaaag 3720
gttcttcttt ttcttttcag 3740

<210> 12
<211> 1622
<212> DNA
<213> Homo sapiens

<220>
<221> intron
<222> (0)...(0)
<223> MGB-STK15 int 2.1

<221> misc_feature
<222> (705)...(989)
<223> n= a, t, c, or g

<400> 12
gtaaattgaa taatctgtaa tctcattcac atttataaac ccacatggag gttggtcttg 60
tcgggaattc tttccgcctt tactttggat ttaaatttag atcccttact gtgatcctgg 120
atatgaatta gtcacttttc tcgtgttcag taacattttg ctgcttctta gagtagcttt 180
tttgttctgc tttgtcttat aatcggctgc ttaagtttct atatccctcc actgtatgca 240
ggataatagt aataatgcat ctggcaggag ttcaaaactt ttaaaattgg ccataaatat 300
aaaataatta gaaaaaggct accttgaatt actgtatttg attctaagtt cctatgataa 360
cggccattta aaaaattgct ctatatttaa aatgtttctt tttatttgtc tttgtctgaa 420
tgcctgctgc gttgtggaca gtgtgctaat ttcaggagta actgactttg tatttggaag 480
tcttaacacc ctctctttgt agagcactca taccgttgag ctggggatgg actttgaggc 540
tttcatttct agcacttgtc cctcacttac aatgagctgt tgaagctgaa ggaaatctca 600
tccctcctac cccttttagt ttgattagct gagggtgtta gagttaactt aacaatttaa 660
ggttgtaata cagtacttac aggcgtataa ataatacatt tcaannnnnn nnnnnnnnnn 720
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 840
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 900
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 960
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnt aataataata atacatttta gtagtaactt 1020
tgtgaagtgt ctacatttgt ttcctctttg tcagtttttt gctcaattcc attttgtcaa 1080
tacttggaaa atgaaacatt ggttaatcaa tagtacagta ataagcttat tgtggaaaat 1140
cttcgatata tgaaaactta gactcttcta aaacttcatg aagataatac cactgttgaa 1200
cgttttgacg tatttttttt tggtcttttt cttaaacgta tattatcaaa gaaatttcaa 1260
tggaactgag attttggcat aaagtttttg tatcatagct ttttgccaaa tagcaatgta 1320
gtgtctattt ccaaattatt gagaaatttt agaaagtgtc tccttcatta atggatattt 1380
gttaataaag catgattttt aggggtgagg aattggaggg gatagaaggt atcattcagg 1440
tattcttagc cacatactaa ctatcctctg gaggtactga ttaaaatacc ttttcacctt 1500
ccatctctta tcagtgacat tcattatttt gctatactag agaacaaact ttgtgaaatt 1560
ctcaatatat tcatcttttg ctttcatgaa tgccagaaag tttattttct cttccattct 1620
ag 1622

<210> 13
<211> 1093
<212> DNA
<213> Homo sapiens

<220>
<221> intron
<222> (0)...(0)
<223> MGB-STK15 int 4.1

<400> 13
gtaagctttc ttatttacaa agttctgtac tgttctacta gaatatatta tttcgttgca 60
aatttcgttg tgggaactct ggggaaaaaa atgaggcctt tatttgcatt tagaggatat 120
aaatgtttcc agatttccaa tcttaaaaaa aatggaattt tgtgtaatga ggtattttac 180

taggaactca agtgctttaa aaaatggctt tcaaatttag aaaaagcttg tatgaatctt 240
ttatagaaat gtgtggaagt tcctctctgt ccttagaaat aaccactaca tatggtttat 300
gcgtctgtac ttttttattg tacaaaagtg caagtttta aaaaatagaa tatgttgcag 360
aactatatac tcatatatga ctgagggttt tgacagtatt atagttttag ttctttattg 420
taaaggttgg ctgtaatgtc ttccccaggg cttttctaaa agcctcctct cagtctctga 480
actatctgga ctctagaatg taccgggagg agcgaggaat gaacccacag actcttttgc 540
ttttagcggt ctaacagagg ctaagagtct aaatccactg gttctcatgc cccagctagc 600
ctgtgggctc catcccgctt ccattagtaa cagtggctct gtctccacca ccagagtggt 660
tctccaccca gagagaatta gcacctctgg gactggaggg agcagctggg gttagtttga 720
aacatgcccc cagatggtct ggaagcattc ctccctctct ggtcacttat cctttttgtg 780
gtcttcagcg ttgtcatggc cctgttcctc tgagcatagt acgggcttgg gacatttccc 840
atagagtgct tcaggtctaa aacccgagac tgctccttgt cactgactct cacacctgac 900
ggcagctagg gacgtcaggg tttcatgtcg tggcagctct ttgatagtgg ttattgcctt 960
ggttcttgct gaggatgcat attgagtgaa gttggaatac gaaattattt gtagaatgtg 1020
tctgctactc attgaaaatt tgttagaaaa gctttgtttt cttcacattc taaagtgttc 1080
aaattcctcc tag 1093

<210> 14
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> forward primer

<400> 14
agcgcctgtt ccgatctg 18

<210> 15
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> reverse primer

<400> 15
aaccaaagtt ctctgctgaa aacc 24

<210> 16
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> probe

<400> 16
ccctgaagca gcaac 15

<210> 17
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> forward primer

<400> 17
ctgttgctgt gtgatgctgt ca 22

<210> 18
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> reverse primer

<400> 18
cctcagccac tcccttgatc 20

<210> 19
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> probe

<400> 19
tcagggcata agcct 15

<210> 20
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> forward primer

<400> 20
tccccttgcc tttggagaa 19

<210> 21
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> reverse primer

<400> 21
aaaggcctgg aggcatcaa 19

<210> 22
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> probe

<400> 22
cagcccaaat cct 13

<210> 23
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> forward primer

<400> 23
cttaatggtg tttagcacag atgca 25

<210> 24
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> reverse primer

<400> 24
ccactgtagc atgcgaagca 20

<210> 25
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> probe

<400> 25
caaatgcaca ggaaac 16

<210> 26
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> forward primer

<400> 26
gctctgcggt gtggagtgt 19

<210> 27
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> reverse primer

<400> 27
cacaggcttc gtctggtgtc t 21

<210> 28
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> probe

<400> 28
tgcagccctc ttggcaactc tcct 24

<210> 29
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> forward primer

<400> 29
aaaatgctgt ctgtgagcct cat 23

<210> 30
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> reverse primer

<400> 30
aacccctgcc cactagaaat g 21

<210> 31
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> probe

<400> 31
acccaagatg tcatctcctg tagcgtcaca 30

<210> 32
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> forward primer

<400> 32
aatagatggg tttatggctg aaggt 25

<210> 33
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> reverse primer

<400> 33
ctcttggcaa ctctcctgac act 23

<210> 34
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> probe

<400> 34
ccgcagagcc gtc 13

<210> 35
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> forward primer

<400> 35
ccatctcaaa aggaaacaag ttctg 25

<210> 36
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> reverse primer

<400> 36

gggtggctgg cttctaactc t 21

<210> 37
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> probe

<400> 37
ccctgatcaa atgaaagcat cact 24

<210> 38
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> forward primer

<400> 38
agaagccagc cacccaatta 20

<210> 39
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> reverse primer

<400> 39
tgtgtcattt ggtatttgtc aagtgt 26

<210> 40
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> probe

<400> 40
tgacttgcac aaaacccagt gaatta 26

<210> 41
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> forward primer

<400> 41
tggacagaga caagatgtga tgtg 24

<210> 42
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> reverse primer

<400> 42
gctggcacct agacaaaaca tg 22

<210> 43
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> probe

<400> 43
ccatagggac ccttc 15

<210> 44
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> forward primer

<400> 44
ggtgtcctat tttcctctga agaga 25

<210> 45
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> reverse primer

<400> 45
tgcaagctga aggtccaaca t 21

<210> 46
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> probe

<400> 46
ttctggccaa ttaag 15

<210> 47
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> forward primer

<400> 47
tcattcatgt ggccgtagca t 21

<210> 48
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> reverse primer

<400> 48
ggtggagagg gagccaaaa 19

<210> 49
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> probe

<400> 49
cctgtttggg ttttca 16

<210> 50
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> forward primer

<400> 50
agaggatcag aacccgtatg tga 23

<210> 51
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> reverse primer

<400> 51
gggaaacaga gataaggtga acca 24

<210> 52
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> probe

<400> 52
tgtgcgtcct gtctg 15

<210> 53
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> forward primer

<400> 53
gggtgacctt gggcttgtg 19

<210> 54
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> reverse primer

<400> 54
cttcaaccca tttccagaga gaa 23

<210> 55

<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> probe

<400> 55
cccgggtccc tgtt 14

<210> 56
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> forward primer

<400> 56
agggtgacct agcagcttgg t 21

<210> 57
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> reverse primer

<400> 57
gcctctggcc cagcctta 18

<210> 58
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> probe

<400> 58
tccctgaccc cagctg 16

<210> 59
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> forward primer

<400> 59
cgtaatgtct cttcctcttc cgtaa 25

<210> 60
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> reverse primer

<400> 60
acgaactgag taggttgctg aaaa 24

<210> 61
<211> 16

<212> DNA
<213> Artificial Sequence

<220>
<223> probe

<400> 61
tcaagggaca aggaag 16

<210> 62
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> forward primer

<400> 62
cattcacatt tataaaccca catgga 26

<210> 63
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> reverse primer

<400> 63
aatccaaagt aaaggcggaa aga 23

<210> 64
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> probe

<400> 64
tggtcttgtc gggaat 16

<210> 65
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> forward primer

<400> 65
gcgaggaatg aacccacaga 20

<210> 66
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> reverse primer

<400> 66
gcatgagaac cagtggattt agact 25

<210> 67
<211> 16
<212> DNA

<213> Artificial Sequence

<220>
<223> probe

<400> 67
cgctaaaagc aaaaga 16

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2004/005287

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, EMBASE, BIOSIS, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	THOMÁZY VILMOS A ET AL: "Determination of cyclin D1 and CD20 mRNA levels by real-time quantitative RT-PCR from archival tissue sections of mantle cell lymphoma and other non-Hodgkin's lymphomas." THE JOURNAL OF MOLECULAR DIAGNOSTICS : JMD. NOV 2002, vol. 4, no. 4, November 2002 (2002-11), pages 201-208, XP002285735 ISSN: 1525-1578 *Materials and Methods* -----	44-65

-/--

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
25 June 2004	12/07/2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Cornelis, K

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2004/005287

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SCHNEIDER J ET AL: "Expression of LRP and MDR1 in locally advanced breast cancer predicts axillary node invasion at the time of rescue mastectomy after induction chemotherapy." BREAST CANCER RESEARCH : BCR. 2001, vol. 3, no. 3, 2001, pages 183-191, XP002285737 ISSN: 1465-5411 the whole document -----	44-65
A	BURNS C G ET AL: "REMOVAL OF A SINGLE ALPHA-TUBULIN GENE INTRON SUPPRESSES CELL CYCLE ARREST PHENOTYPES OF SPLICING FACTOR MUTATIONS IN SACCHAROMYCES CEREVISIAE" MOLECULAR AND CELLULAR BIOLOGY, AMERICAN SOCIETY FOR MICROBIOLOGY, WASHINGTON, US, vol. 22, no. 3, February 2002 (2002-02), pages 801-815, XP009014911 ISSN: 0270-7306 the whole document -----	1-43, 66-89
A	KAPRANOV PHILIPP ET AL: "Large-scale transcriptional activity in chromosomes 21 and 22." SCIENCE. 3 MAY 2002, vol. 296, no. 5569, 3 May 2002 (2002-05-03), pages 916-919, XP002285738 ISSN: 1095-9203 the whole document -----	1-43, 66-89
A	CLARK TYSON A ET AL: "Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays." SCIENCE. 3 MAY 2002, vol. 296, no. 5569, 3 May 2002 (2002-05-03), pages 907-910, XP002285850 ISSN: 1095-9203 the whole document -----	1-43, 66-89