

(12) FASCÍCULO DE PATENTE DE INVENÇÃO

(22) Data de pedido: 2004.02.19	(73) Titular(es): GENOMIC HEALTH, INC. 301 PENOBSCOT DRIVE REDWOOD CITY, CA 94063 US
(30) Prioridade(s): 2003.02.20 US 448991 P	
(43) Data de publicação do pedido: 2005.11.23	
(45) Data e BPI da concessão: 2008.10.29 247/2008	(72) Inventor(es): MICHAEL C. KIEFER US RANDAL W. SCOTT US JOFFRE B. BAKER US
	(74) Mandatário: PEDRO DA SILVA ALVES MOREIRA RUA DO PATROCÍNIO, N.º 94 1399-019 LISBOA PT

(54) Epígrafe: **UTILIZAÇÃO DE ARN INTRÓNICO PARA MEDIR A EXPRESSÃO GENÉTICA**

(57) Resumo:

DESCRIÇÃO

"UTILIZAÇÃO DE ARN INTRÓNICO PARA MEDIR A EXPRESSÃO GENÉTICA"

Antecedentes da Invenção

Campo da Invenção

É bem reconhecido que a expressão genética nas células e tecidos pode indicar o estado fisiológico e/ou patológico da célula, tecido ou doente. Durante várias décadas, a expressão genética, como medida por análise imuno-histoquímica de marcadores de proteína, foi utilizada para a tomada de decisões de tratamento. Por exemplo, os níveis de receptor de estrogénio e de receptor de progesterona medidos deste modo são agora utilizados em rotina para seleccionar doentes de cancro para tratamento com fármacos antiestrogénio.

A literatura de investigação mais recente proporciona evidências de que os níveis de espécies de ARNm têm valor de diagnóstico e prognóstico. Isto é um desenvolvimento promissor porque as tecnologias para medição dos níveis de ARN celular, como exemplificado por múltiplas plataformas de RT-PCR e *arrays* de ADN, podem ser sensíveis, específicas e quantitativas. A RT-PCR é reconhecida como sendo geralmente mais sensível do que a tecnologia de *array* de ADN. Contudo, a concepção e selecção de sonda/iniciador para RT-PCR pode ser um desafio, porque existem múltiplos critérios para um desempenho óptimo. Este desafio é particularmente grande quando o ARN da amostra a ser estudada

provém de tecido fixo, embebido em cera, porque esse ARN tende a estar altamente fragmentado (K. Specht *et al.*, *Am. J. Pathol* 158: 419-29 [2001]; T.E. Godfrey *et al.*, *J. Mol. Diagnosis* 2: 84-91 [2001]).

É uma prática aceite medir a expressão de qualquer dado gene através do ensaio do nível de qualquer das suas sequências de ARNm maduro transcritas, excisadas (sequência do exão, em oposição à do intrão). Em teoria, um exão é definido como qualquer segmento de um gene interrompido que é representado no produto do ARN maduro e um intrão é definido como um segmento de ADN que é transcrito, mas removido do transcrito através de excisão-união dos exões em qualquer dos lados deste [B. Lewin. *Genes IV*, Cell Press, Cambridge Mass. 1990]. A razão para a prática aceite de utilizar sequências de exões é teoricamente linear, porque os ARN maduros [ARNm] codificam proteínas que definem fenótipos celulares, enquanto que o ARN intrónico é considerado como possuindo comparativamente pouca influência no fenótipo celular. Além disso, a visão prevalente é que os intrões são rapidamente degradados e por isso mais difíceis de detectar do que as sequências dos exões {ver introduções dos seguintes artigos: Thomas *et al.*, *J. Virol.* 76:532-40 [2002]; Clement *et al.*, *J. Biol. Chem.* 276:16919-30 [2001]; Sharp *et al.*, *Ann. Ver. Biochem* 55:1119-1150 [1986]}.

Além disso, Kapranov *et al.*, *Science* 296:916-919 [2002] descrevem arrays de oligonucleótidos contendo sondas espaçadas em média a cada 35 pares de bases ao longo dos cromossomas humanos 21 e 22. Utilizando ARN poliadenilado citossólico obtido a partir de 11 linhas celulares humanas, são relatados mapas identificando áreas activas de transcrição de ARN nestes cromossomas e os autores concluem que até uma ordem de grandeza

é transcrita mais da sequência genómica do que esperado a partir dos exões previamente previstos e caracterizados.

A presente invenção refere-se à utilização de ARN intrónico para medir a expressão genética. Será demonstrado que as sequências de ARN intrónicas tendem a ser rapidamente detectadas por RT-PCR, mesmo utilizando ARN extensivamente degradado, a partir de tecidos fixos. Além disso, estes tendem a estar correlacionados na sua expressão com os seus respectivos exões. O último ponto é particularmente inesperado, porque existe pouca ou nenhuma evidência de que a proporção das constantes de velocidade global para a síntese e o processamento das sequências de intrão e exão transcritas sejam semelhantes. De facto, a literatura científica proporciona evidência para a complexidade de pré-ARNm e processamento de intrões excisados. Por exemplo, pode existir pré-ARNm em múltiplas misturas cinéticas (Elliott e Rosbash, *Exp. Cell Res.* 229:181-8 [1996]), com sub-populações contendo ARN de intrão que não são excisados eficientemente e são transportados para o citoplasma como espécies de ARNm "imaturas" em que podem decair a velocidades diferentes daquelas das sequências de ARN de intrão nuclear (Wang *et al.*, *Proc. Natl. Acad. Sci. USA* 94:4360-5 [1997]). Além disso, determinados ARNm de intrão excisados parecem entrar no citoplasma em estrutura de laço (Clement *et al.*, *RNA* 5:206-20 [1999]).

Sumário da Invenção

A presente invenção baseia-se na evidência experimental que demonstra que as sequências de intrão transcritas que, por definição, estão presentes no ARN nuclear heterogéneo, mas

tipicamente não são incorporadas no ARNm, têm utilidade em diagnóstico e prognóstico. Esta é uma constatação significativa por diversas razões. Tipicamente, as sequências de intrão são mais compridas do que as sequências do exão, em cerca de 20 vezes ou mais. Assim, os intrões, dado o seu comprimento médio muito maior, proporcionam oportunidade proporcionalmente aumentada para a concepção de sondas para a expressão genética óptima, por exemplo, no caso de RT-PCR, a criação de conjuntos de sonda/iniciador que possuam melhor desempenho técnico. Independentemente, devido às sequências de intrão evoluírem mais rapidamente do que as sequências de exão, os ARN intrónicos são bem adaptados para monitorizar a expressão de diferentes membros intimamente relacionados de uma família genética.

Num aspecto, a invenção refere-se a um método para monitorizar a expressão genética numa amostra biológica, compreendendo:

- (a) proporcionar um polinucleótido complementar a uma sequência de ARN intrónica num gene alvo, em que a expressão dessa sequência cistronica está correlacionada com a expressão de uma sequência de ARNm exónica num gene alvo;
- (b) hibridar o polinucleótido com a sequência de ARN intrónico para formar um complexo polinucleótido-ARN intrónico; e
- (c) detectar o complexo polinucleótido-ARN intrónico.

Num aspecto particular, a expressão do gene alvo é medida por RT-PCR e, nesse caso pode ser utilizado um conjunto de iniciador/sonda à base do intrão no processo acima.

Noutro aspecto, a invenção refere-se a métodos de utilização de sequências à base de intrão para conceber e criar conjuntos de sonda-iniciador para RT-PCR. Esses iniciadores e sondas são particularmente adequados para detectar e quantificar os níveis de ARN de intrão em espécimes de tecido fixo embebido em parafina (FPET), para análise de expressão genética de elevada sensibilidade. Consequentemente, noutro aspecto, a invenção refere-se à utilização de conjuntos de iniciador-sonda à base do intrão em ensaios de perfis de expressão genética, tal como análise de expressão genética de amostras FPET para diagnosticar e/ou prever o prognóstico de vários estados patológicos.

Em particular, a invenção refere-se a um método para preparar uma molécula de oligonucleótido de cadeia simples para amplificação de um gene alvo e medição do nível de uma espécie de ARN intrónico compreendendo:

- (a) identificar, pelo menos, uma sequência de intrão no gene alvo, em que a expressão da sequência do intrão está correlacionada com a expressão de uma sequência de exão no gene alvo;
- (b) preparar um molécula de oligonucleótido de cadeia simples que corresponda a, pelo menos, uma porção da sequência de intrão transcrita; e
- (c) utilizar a molécula de oligonucleótido para medir a expressão genética.

Tal como anteriormente, a expressão genética pode ser medida, por exemplo, por RT-PCR e nesse caso é utilizado um

conjunto de iniciador/sonda à base do intrão (consistindo em dois iniciadores e uma sonda) para medir a expressão genética.

Se o oligonucleótido for um iniciador em sentido directo, é tipicamente concebido de modo a compreender sequências 5' de uma sequência alvo na sequência de intrão transcrita. Se o oligonucleótido for um iniciador reverso, é tipicamente concebido para complementar as sequências a 5' de uma sequência alvo a jusante do iniciador em sentido directo na sequência de intrão transcrita. É importante identificar e utilizar uma sequência alvo suficientemente longa para amplificação por PCR. A sequência alvo deve ter, pelo menos, cerca de 50 bases de nucleótidos de comprimento, em particular, pelo menos, 55 bases de nucleótidos de comprimento, em algumas formas de realização, pelo menos, cerca de 60 bases de nucleótidos de comprimento. Os iniciadores e as sondas de PCR são concebidos seguindo princípios bem conhecidos. Assim, o iniciador de PCR tem tipicamente 17-30 bases de nucleótidos de comprimento e normalmente contém cerca de 20% a 80% de bases G+C. É desejável conceber iniciadores de PCR com uma temperatura de fusão (T_m) entre cerca de 50 °C e cerca de 70 °C.

Quando a molécula de oligonucleótido de cadeia simples é uma sonda de PCR é, normalmente, concebida para compreender ou complementar uma porção interna de uma sequência alvo na sequência de intrão transcrita. Para amplificação pela TaqMan[®], a sonda de PCR é marcada com um corante fluorescente repórter e uma unidade de inibição.

Noutro aspecto, a invenção refere-se a um método para medir a expressão de um gene por amplificação de um gene alvo pela reacção da polimerase em cadeia (PCR) compreendendo:

(a) identificar, pelo menos, uma sequência de intrão alvo no gene alvo, em que a expressão da sequência de intrão está correlacionada com a expressão de uma sequência de exão correspondente no gene alvo; e

(b) amplificar a sequência de intrão alvo transcrita utilizando um conjunto iniciador de PCR específico para intrão/sonda.

A sequência de intrão alvo tem tipicamente, pelo menos, cerca de 50 bases de comprimento e o conjunto de iniciador de PCR e sonda é concebido para corresponder a sequências únicas na sequência de intrão alvo transcrita.

Ainda noutro aspecto, a invenção refere-se a um método para amplificar fragmentos de ARN numa amostra representando, pelo menos, um gene de interesse, compreendendo os passos de:

(a) contactar a amostra com, pelo menos, um conjunto de iniciadores de PCR e sonda; e

(b) realizar a amplificação por PCR,

em que os iniciadores de PCR e a sonda são concebidos com base numa sequência de intrão identificada no gene de interesse e em que a expressão da sequência de intrão está correlacionada com a expressão de uma sequência de exão no gene de interesse.

Numa forma de realização particular, os iniciadores de PCR e a sonda são tipicamente concebidos com base numa sequência única no intrão identificado. Noutra forma de realização, a amostra

compreende ARN fragmentado, representando múltiplos genes de interesse e está em contacto com uma mistura dos iniciadores de PCR e sondas concebidas com base em sequências únicas nos intrões presentes nos genes de interesse.

Numa forma de realização preferida, a amplificação é realizada numa amostra de tecido fixo, embebido em parafina (FPET) que pode, por exemplo, ter origem numa biopsia tumoral obtida a partir de um doente humano. O tumor pode ser qualquer tipo de tumor sólido, tais como, por exemplo, cancro da mama, cancro do pulmão ou cancro colorrectal. O tecido tumoral pode ser recolhido através de uma variedade de métodos, incluindo biopsia com agulha fina, biopsia do núcleo ou ressecção.

Numa forma de realização particular, a invenção refere-se a métodos utilizando conjuntos de iniciador de PCR à base de intrão-sonda numa análise de expressão genética para prever a probabilidade de doença recorrente para doentes com cancro da mama precoce.

Ainda noutro aspecto, a invenção refere-se a um *array* compreendendo uma pluralidade de polinucleótidos que hibridam com genes alvo de interesse em que, de um modo preferido, pelo menos, 70% dos polinucleótidos compreende sequências de intrão.

Ainda noutro aspecto, a invenção refere-se a sequências de amplificação à base de intrão e sua utilização na análise de expressão genética.

Numa forma de realização particular, a invenção refere-se a análise de expressão genética de uma amostra biológica representativa de cancro da mama invasivo com base na

determinação dos níveis de expressão dos transcritos de ARN ou dos produtos de expressão de um gene ou conjuntos de genes seleccionados do grupo consistindo em:

- (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,

incluindo a utilização de sequências à base de intrões.

Ainda noutra forma de realização, a invenção refere-se à análise de expressão genética de uma amostra biológica representativa de cancro da mama positivo para ER com base na determinação dos níveis de expressão dos transcritos de ARN ou dos produtos de expressão de um gene ou conjunto de genes seleccionados do grupo consistindo em:

- (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, in1;
- (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; e

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

incluindo a utilização de sequências à base do intrão.

Noutra forma de realização, o cancro é cancro da mama e o(s) gene(s) analisado(s) é/são seleccionado(s) do grupo consistindo de: 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; Bcl2; CEGP1; GSTM1; PR; BBC3; GATA3; DPYD; GSTM3; ID1; EstR1; p27; XIAP; IGF1R; AK055699; P13KC2A; TGFB3; BAG1; pS2; WISP1; HNF3A; e NFKBp65.

Ainda noutra forma de realização, a invenção refere-se à análise da expressão genética de uma amostra biológica representativa de cancro da mama invasivo, com base na determinação dos níveis de expressão dos transcritos de ARN ou dos produtos de expressão de um gene ou conjunto de genes seleccionados do grupo consistindo em:

- (a) p53BP2, Bcl2, BAD, EPHX1, PDGFR, DIABLO, XIAP, YB1, CA9 e KRT8;
- (b) GRB7, CD68, TOP2A, Bcl2, DIABLO, CD3, ID1, PPM1D, MCM6 e WISP1;
- (c) PR, p53BP2, PRAME, DIABLO, CTSL, IGFBP2, TIMP1, CA9, MMP9 e COX2;
- (d) CD68, GRB7, TOP2A, Bcl2, DIABLO, CD3, ID1, PPM1D, MCM6 e WISP1;
- (e) Bcl2, p53BP2, BAD, EPHX1, PDGFR β , DIABLO, XIAP, YB1, CA9 e KRT8;
- (f) KRT14, KRT5, PRAME, p53BP2, GUS1, AIB1, MCM3, CCNE1, MCM6 e ID1;

- (g) PRAME, p53BP2, EstR1, DIABLO, CTSL, PPM1D, GRB7, DAPK1, BBC3 e VEGFB;
- (h) CTSL2, GRB7, TOP2A, CCNB1, Bcl2, DIABLO, PRAME, EMS1, CA9 e EpCAM;
- (i) EstR1, p53BP2, PRAME, DIABLO, CTSL, PPM1D, GRB7, DAPK1, BBC3 e VEGFB;
- (j) Chk1, PRAME, p53BP2, GRB7, CA9, CTSL, CCNB1, TOP2A, tamanho de tumor e IGFBP2;
- (k) IGFBP2, GRB7, PRAME, DIABLO, CTSL, β -Catenina, PPM1D, Chk1, WISP1 e LOT1;
- (l) HER2, p53BP2, Bcl2, DIABLO, TIMP1, EPHX1, TOP2A, TRAIL, CA9 e AREG;
- (m) BAG1, p53BP2, PRAME, IL6, CCNB1, PAI1, AREG, tamanho de tumor, CA9 e Ki67;
- (n) CEGP1, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, STK15 e AKT2 e FGF18;
- (o) STK15, p53BP2, PRAME, IL6, CCNE1, AKT2, DIABLO, cMet, CCNE2 e COX2;
- (p) KLK10, EstR1, p53BP2, PRAME, DIABLO, CTSL, PPM1D, GRB7, DAPK1 e BBC3;
- (q) AIB1, p53BP2, Bcl2, DIABLO, TIMP1, CD3, p53, CA9, GRB7 e EPHX1
- (r) BBC3, GRB7, CD68, PRAME, TOP2A, CCNB1, EPHX1, CTSL GSTM1 e APC;
- (s) CD9, GRB7, CD68, TOP2A, Bcl2, CCNB1, CD3, DIABLO, ID1 e PPM1D;
- (t) EGFR, KRT14, GRB7, TOP2A, CCNB1, CTSL, Bcl2, TP, KLK10 e CA9;
- (u) HIF1 α , PR, DIABLO, PRAME, Chk1, AKT2, GRB7, CCNE1, TOP2A e CCNB1;
- (v) MDM2, p53BP2, DIABLO, Bcl2, AIB1, TIMP1, CD3, p53, CA9 e HER2;

- (w) MYBL2, p53BP2, PRAME, IL6, Bcl2, DIABLO, CCNE1, EPHX1, TIMP1 e CA9;
- (x) p27, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, STK15, AKT2 e ID1;
- (y) RAD51, GRB7, CD68, TOP2A, CIAP2, CCNB1, BAG1, IL6, FGFR1 e p53BP2;
- (z) SURV, GRB7, TOP2A, PRAME, CTSL, GSTM1, CCNB1, VDR, CA9 e CCNE2;
- (aa) TOP2B, p53BP2, DIABLO, Bcl2, TIMP1, AIB1, CA9, p53, KRT8 e BAD;
- (ab) ZNF217, GRB7, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, APC4 e β -Catenina.

Numa forma de realização diferente, a invenção refere-se à análise de expressão genética de uma amostra biológica, utilizando sequências de polinucleótidos à base de intrões que hibridam com, pelo menos, um gene seleccionado do grupo consistindo em: CD68; CTSL; FBX05; SURV; CCNB1; MCM2; Chk1; MYBL2; HIF1A; cMET; EGFR; TS; STK15, IGFR1; BC12; HNF3A; TP53BP2; GATA3; BBC3; RAD51C; BAG1; IGFBP2; PR; CD9; RB1; EPHX1; CEGP1; TRAIL; DR5; p27; p53; MTA; RIZ1; ErbB3; TOP2B; EIF4E, CD68; CTSL; FBX05; SURV; CCNB1; MCM2; Chk1; MYBL2; HIF1A; cMET; EGFR; TS; e STK15.

A análise da expressão genética pode ser realizada num formato de *array* e o *array* é, de um modo preferido, um *array* de elevada densidade, compreendendo, pelo menos, 100, de um modo mais preferido, pelo menos, 150, de um modo ainda mais preferido, 200 sequências numa secção de 5-10 μ .

Finalmente, a presente invenção refere-se à utilização de ARN intrónico no prognóstico do cancro, como definido nas reivindicações.

Breve Descrição dos Desenhos

As Figuras 1A-M apresentam sequências de intrão mascaradas para os genes CEGP1, FOXM1, PRAME e STK15. Os amplicões utilizados para RT-PCR são apresentados em itálico.

A Figura 2 apresenta conjuntos de iniciador/sonda para CEGP1, FOXM1, PRAME e STK15. As sequências dos iniciadores em sentido directo e reverso estão indicadas por "F" e "R", respectivamente. As sequências de iniciadores são designadas com "P".

A Figura 3 apresenta coeficientes de correlação [R] para co-expressão do ARN do exão CEGP1 com 47 outras sequências de ARN. Símbolos: losango = exão CEGP1 próprio vs. próprio (= 1,0 por definição); quadrados = intrões de CEGP1; triângulos = sequências de outros genes.

A Figura 4 apresenta coeficientes de correlação [R] para co-expressão de ARN do exão PRAME com 47 outras sequências de ARN. Símbolos: losango - exão PRAME próprio vs. próprio (= 1,0 por definição); quadrados=intrões PRAME; triângulos = sequências de outros genes.

A Figura 5 apresenta coeficientes de correlação [R] para co-expressão de ARN do exão STK15 com 47 outras sequências de ARN. Símbolos: losango - exão STK15 próprio vs. próprio (= 1,0 por

definição); quadrados=intrões STK15; triângulos = sequências de outros genes.

A Figura 6 apresenta um conjunto exemplar de genes, cuja expressão pode ser analisada pelos métodos da presente invenção.

Descrição Detalhada da Forma de Realização Preferida

A. Definições

Salvo definição em contrário, os termos técnicos e científicos aqui utilizados possuem o mesmo significado que o normalmente entendido por um especialista na técnica á qual esta invenção pertence. Singleton *et al.*, *Dictionary of Microbiology and Molecular Biology*, 2^a ed., J. Wiley & Sons (Nova Iorque, NI 1994), proporcionam a um especialista na técnica, um guia geral para muitos dos termos utilizados no presente pedido.

Um especialista na técnica irá reconhecer muitos métodos e materiais semelhantes ou equivalentes aqueles aqui descritos que podem ser utilizados na prática da presente invenção. Na realidade, a presente invenção não está, de forma alguma, limitada aos métodos e materiais descritos. Para efeitos da presente invenção, os seguintes termos são definidos abaixo.

Os termos “excisão” e “excisão de ARN” são utilizados indistintamente e referem-se ao processamento de ARN que remove intrões e une os exões para produzir ARNm maduro com sequência codificante contínua que se desloca para o citoplasma de uma célula eucariótica.

Em teoria, o termo “exão” refere-se a qualquer segmento de um gene interrompido que é representado no produto do ARN maduro (B. Lewin. *Genes IV* Cell Press, Cambridge Mass. 1990). Em teoria, o termo “intrão” refere-se a qualquer segmento de ADN que é transcrito, mas removido do transcrito através de excisão-união dos exões em qualquer dos seus lados. Operacionalmente, as sequências de exão ocorrem na sequência de ARNm de um gene como definido pelos Números de Ref. De Seq ID. Operacionalmente, as sequências de intrão são as sequências intervenientes no ADN genómico de um gene, flanqueadas por sequências de exão e possuindo sequências de consenso de excisão GT e AG nas suas fronteiras 5' e 3'.

O termo *microarray* refere-se a um arranjo ordenado de elementos de *array* hibridáveis, de um modo preferido, sondas de polinucleótido, num substrato.

O termo “polinucleótido,” quando utilizado no singular ou plural, refere-se geralmente a qualquer polirribonucleótido ou polidesoxirribonucleótido que pode ser ARN ou ADN não modificado ou ARN ou ADN modificado. Assim, por exemplo, polinucleótidos como aqui definido incluem, sem limitação, ADN de cadeia simples e cadeia dupla, ADN incluindo regiões de cadeia simples e dupla, ARN de cadeia simples e dupla e ARN incluindo regiões de cadeia simples e dupla, moléculas híbridas compreendendo ADN e ARN que podem ser de cadeia simples ou, mais tipicamente, cadeia dupla ou incluir regiões de cadeia simples e dupla. Adicionalmente, o termo “polinucleótido”, como aqui utilizado, refere-se a regiões de cadeia tripla compreendendo ARN ou ADN ou ambos ARN e ADN. As cadeias nessas regiões podem ser da mesma molécula ou de moléculas diferentes. As regiões podem incluir a totalidade de uma ou mais das moléculas, mas mais tipicamente envolvem apenas

uma região de algumas dessas moléculas. Uma das moléculas de uma região de tripla hélice é frequentemente um oligonucleótido. O termo "polinucleótido" inclui especificamente ADNc. O termo inclui ADN (incluindo ADNc) e ARN que contêm uma ou mais bases modificadas. Assim, ADN ou ARN com estruturas modificadas para a estabilidade ou por outras razões são "polinucleótidos" no sentido em que o termo é aqui utilizado. Além disso, os ADN ou ARN compreendendo bases pouco usuais, tal como inosina, ou bases modificadas, tal como bases tritiadas, estão incluídas no termo "polinucleótidos", como aqui definido. Em geral, o termo "polinucleótido" abarca todas as formas quimicamente, enzimaticamente e/ou metabolicamente modificadas de polinucleótidos não modificados, assim como as formas químicas de ADN e ARN de vírus e células, incluindo células simples e complexas.

O termo "oligonucleótido" refere-se a um polinucleótido relativamente curto incluindo, sem limitação, desoxirribonucleótidos de cadeia simples, ribonucleótidos de cadeia simples ou dupla, híbridos ARN:ADN e ADN de cadeia dupla. Oligonucleótidos, tal como oligonucleótidos de sonda de ADN de cadeia simples, são frequentemente sintetizados por métodos químicos, por exemplo, utilizando sintetizadores de oligonucleótido automatizados que estão disponíveis comercialmente. Contudo, os oligonucleótidos podem ser produzidos através de uma variedade de outros métodos, incluindo técnicas mediadas por ADN recombinante *in vitro* e através da expressão de ADN em células e organismos.

Os termos "gene expresso diferencialmente", "expressão de gene diferencial" e os seus sinónimos que são utilizados indistintamente, referem-se a um gene cuja expressão está a um

nível superior ou inferior num doente ou indivíduo de teste, em relação a outro, por exemplo, num indivíduo que sofre de uma doença, especificamente cancro, tal como cancro da mama, em relação à sua expressão num indivíduo normal ou de controlo. Os termos também incluem genes cuja expressão é activada a um nível superior ou inferior a estádios diferentes da mesma doença. É também entendido que um gene expresso diferencialmente pode ser ou activado ou inibido ao nível do ácido nucleico ou da proteína ou pode ser submetido a excisão alternativa para resultar num produto polipeptídico diferente. Essas diferenças podem ser evidenciadas por, por exemplo, uma alteração nos níveis de ARNm, expressão de superfície, secreção ou outra partição de um polipéptido. A expressão de gene diferencial pode incluir uma comparação da expressão entre dois ou mais genes ou os seus produtos genéticos ou uma comparação das proporções da expressão entre dois ou mais genes ou seus produtos genéticos ou mesmo uma comparação de dois produtos processados diferentemente, do mesmo gene que diferem entre indivíduos normais e indivíduos que sofrem de uma doença, especificamente cancro ou entre vários estádios da mesma doença. A expressão diferencial inclui diferenças quantitativas, como qualitativas no padrão de expressão temporal ou celular num gene ou seus produtos de expressão entre, por exemplo, células normais e doentes, ou entre células que sofreram diferentes eventos de doença ou estádios de doença. Para o objectivo desta invenção, a "expressão genética diferencial" é considerada como estando presente quando há uma diferença de cerca de duas vezes, de um modo preferido, pelo menos, cerca de quatro vezes, de um modo muito preferido, pelo menos, cerca de seis vezes, de um modo muito preferido, pelo menos, cerca de dez vezes de diferença entre a expressão de um dado gene em indivíduos normais e

doentes, ou em vários estádios do desenvolvimento da doença num indivíduo doente.

O termo “sobre expressão” em relação a um transcrito de ARN é utilizado para referir o nível do transcrito determinado pela normalização para o nível de ARNm de referência que pode ser todos os transcritos medidos no espécimen ou um conjunto de referência particular de ARNm.

A frase “amplificação de gene” refere-se a um processo através do qual são formadas múltiplas cópias de um gene ou fragmento de um gene numa célula ou linha celular particular. A região duplicada (uma porção de ADN amplificada) é frequentemente referida como “amplicão”. Frequentemente, a quantidade do ARN mensageiro (ARNm) produzida, *i. e.*, o nível de expressão genética, também aumenta na proporção do número de cópias produzidas do gene particular expressado.

O termo “prognóstico” é aqui utilizado para se referir à previsão da probabilidade de morte atribuível ao cancro, ou sua progressão, incluindo recorrência, disseminação metastática e resistência ao fármaco, de uma doença neoplásica, tal como cancro da mama. O termo “previsão” é aqui utilizado para referir a probabilidade de um doente responder favoravelmente ou desfavoravelmente a um fármaco ou conjunto de fármacos e, também, a extensão dessas respostas ou que um doente sobreviva, após remoção cirúrgica e/ou quimioterapia de um tumor primário, durante um certo período de tempo sem recorrência do cancro. Os métodos previstos da presente invenção podem ser utilizados clinicamente para tomar decisões sobre o tratamento, escolhendo as modalidades de tratamento mais apropriadas para qualquer doente em particular. Os métodos de previsão da presente

invenção são ferramentas valiosas na previsão se for provável que um doente responda favoravelmente a um regime de tratamento, tais como intervenção cirúrgica, quimioterapia com um dado fármaco ou combinação de fármacos e/ou terapia por radiação, ou se é provável uma sobrevivência do doente a longo termo, após cirurgia e/ou paragem da quimioterapia ou outras modalidades de tratamento.

O termo sobrevivência a “longo termo” é aqui utilizado para se referir à sobrevivência durante, pelo menos, 3 anos, de um modo mais preferido durante, pelo menos, 5 anos, de um modo muito preferido durante, pelo menos, 10 anos após cirurgia ou outro tratamento.

O termo “resistência aumentada” em relação a um fármaco particular ou opção de tratamento, quando utilizado de acordo com a presente invenção, significa resposta diminuída em relação a uma dose padrão do fármaco ou a um protocolo de tratamento padrão.

O termo “sensibilidade diminuída” a um fármaco particular ou opção de tratamento, quando utilizado de acordo com a presente invenção, significa resposta diminuída a uma dose padrão do fármaco ou a um protocolo de tratamento padrão, em que a resposta diminuída pode ser compensada (pelo menos parcialmente) aumentando a dose do fármaco ou a intensidade de tratamento.

A “resposta do doente” pode ser avaliada utilizando qualquer ponto final indicando um benefício ao doente, incluindo, sem limitação, (1) inibição, em alguma medida, do crescimento tumoral, incluindo retardamento e paragem completa do crescimento; (2) redução no número de células tumorais;

(3) redução do tamanho do tumor; (4) inibição (*i. e.*, redução, retardamento ou paragem completa) da infiltração da célula tumoral em órgãos periféricos adjacentes e/ou tecidos; (5) inibição (*i. e.* redução, retardamento ou paragem completa) das metástases; (6) aumento da resposta imunitária anti-tumoral que pode, mas não tem obrigatoriamente que, resultar na regressão ou rejeição do tumor; (7) alívio, em alguma medida, de um ou mais sintomas associados ao tumor; (8) aumento do período de sobrevivência após tratamento; e/ou (9) diminuição da mortalidade num dado ponto de tempo após tratamento.

O termo "tratamento" refere-se a tratamento terapêutico e profiláctico ou medidas preventivas, em que o objectivo é prevenir ou retardar (diminuir) o estado patológico ou distúrbio a combater. Aqueles que têm necessidade de tratamento incluem os que já têm o distúrbio, assim como aqueles predispostos a terem o distúrbio ou aqueles em que o distúrbio deve ser prevenido. No tratamento tumoral (*e. g.*, cancro), um agente terapêutico pode diminuir directamente a patologia das células tumorais ou tornar as células tumorais mais susceptíveis ao tratamento por outros agentes terapêuticos, *e. g.*, radiação e/ou quimioterapia.

O termo "tumor," como aqui utilizado, refere-se a todo o crescimento e proliferação celular neoplásica, seja maligna ou benigna e a todas as células e tecidos pré-cancerosos e cancerosos.

Os termos "cancro" e "canceroso" referem-se ou descrevem o estado fisiológico em mamíferos que é tipicamente caracterizado por um crescimento celular desregulado. Exemplos de cancro incluem, mas não estão limitados a, cancro da mama, cancro do cólon, cancro do pulmão, cancro da próstata, cancro

hepatocelular, cancro gástrico, cancro pancreático, cancro do colo do útero, cancro ovárico, cancro do fígado, cancro da bexiga, cancro do tracto urinário, cancro da tiróide, cancro renal, carcinoma, melanoma e cancro do cérebro.

A “patologia” do cancro inclui todos os fenómenos que comprometem o bem estar do doente. Isto inclui, sem limitação, o crescimento celular anormal ou incontrolável, metástase, interferência com o normal funcionamento das células vizinhas, libertação de citocinas ou outros produtos secretores a níveis anormais, supressão ou agravamento de resposta inflamatória ou imunológica, neoplasia, estado pré-maligno, estado maligno, invasão de tecidos ou órgãos circundantes ou distantes, tal como nódulos linfáticos, etc.

“Restringência” de reacções de hibridação é prontamente determinável por um especialista na técnica e geralmente é um cálculo empírico, dependente do comprimento da sonda, da temperatura de lavagem e da concentração de sais. Em geral, sondas mais compridas requerem temperaturas mais elevadas para um emparelhamento apropriado, enquanto que sondas mais curtas necessitam de temperaturas mais baixas. A hibridação depende geralmente da capacidade do ADN desnaturado reemparelhar quando estão presentes cadeias complementares no ambiente abaixo da sua temperatura de fusão. Quanto mais elevado for o grau de homologia desejado entre a sonda e a sequência hibridável, mais elevada será a temperatura relativa que pode ser utilizada. Como resultado, segue-se que temperaturas relativas elevadas tenderiam a fazer as condições de reacção mais restritivas, enquanto que temperaturas mais baixas as diminuiriam. Para detalhes adicionais e explicação de restringência de reacções de

hibridação, ver Ausubel et al., *Current Protocols in Molecular Biology*, Wiley Interscience Publishers, (1995).

“Condições restritivas” ou “condições de elevada restritência”, como aqui definido, tipicamente: (1) empregam baixa força iônica e temperatura elevada para lavagem, por exemplo cloreto de sódio a 0,015 M/citrato de sódio a 0,0015 M/dodecilsulfato de sódio a 0,1% a 50 °C; (2) empregam durante a hibridação um agente desnaturante, tal como formamida, por exemplo, formamida a 50% (v/v) com albumina de soro bovino a 0,1%/Ficoll a 0,1%/polivinilpirrolidona a 0,1%/tampão fosfato de sódio a 50 mM a pH 6,5 com cloreto de sódio a 750 mM, citrato de sódio a 75 mM a 42 °C; ou (3) empregar formamida a 50%, 5 x SSC (NaCl a 0,75 M, citrato de sódio a 0,075 M), fosfato de sódio a 50 mM (pH 6,8), pirofosfato de sódio a 0,1%, solução de Denhardt a 5 x, ADN de esperma de salmão sonificado (50 µg/mL), SDS a 0,1% e sulfato de dextrano a 10% a 42 °C, com lavagens a 42 °C em SSC a 0,2 x (cloreto de sódio/citrato de sódio) e formamida a 50% a 55 °C, seguido por uma lavagem a restritência elevada, consistindo em SSC a 0,1 x contendo EDTA a 55 °C.

“Condições moderadamente restritivas” podem ser identificadas como descrito por Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Nova Iorque: Cold Spring Harbor Press, 1989 e incluem a utilização de solução de lavagem e condições de hibridação (e. g., temperatura, força iônica e %SDS) menos restritivas do que as descritas acima. Um exemplo de condições moderadamente restritivas são incubação durante a noite a 37 °C numa solução compreendendo: formamida a 20%, SSC a 5 x (NaCl a 150 mM, citrato trissódico a 15 mM), fosfato de sódio a 50 mM (pH 7,6), solução de Denhardt a 5 x, sulfato de dextrano a 10% e 20 mg/mL de ADN de esperma de salmão

fragmentado desnaturado, seguido por lavagem dos filtros em SSC a 1 x a cerca de 37-50 °C. O especialista na técnica reconhecerá como ajustar a temperatura, força iônica, etc. consoante a necessidade, para acomodar factores, tal com o comprimento da sonda e semelhantes.

No contexto da presente invenção, a referência a “pelo menos um”, “pelo menos dois”, “pelo menos cinco”, etc. dos genes listados num conjunto de genes particular significa qualquer uma ou qualquer e todas as combinações dos genes listados.

B. Descrição Detalhada

A prática da presente invenção empregará, salvo indicação em contrário, técnicas convencionais de biologia molecular (incluindo técnicas recombinantes), microbiologia, biologia celular e bioquímica que estão dentro do âmbito do conhecimento na técnica. Essas técnicas são explicadas completamente na literatura, tais como, “*Molecular Cloning: A Laboratory Manual*”, 2ª edição (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*”, 4ª edição (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); e “*PCR: The Polymerase Chain Reaction*”, (Mullis et al., eds., 1994).

1. Reacção da Polimerase em Cadeia (PCR)

O objectivo da reacção da polimerase em cadeia (PCR) é fazer cópias de um gene de modo a proporcionar quantidades maiores de ácido nucleico para posterior utilização. A PCR é um processo baseado numa enzima polimerase especializada (e. g. polimerase de ADN Taq) que pode sintetizar uma cadeia complementar a uma dada cadeia de ADN numa mistura contendo os quatro dNTP (sATP, dCTP, dGTP, dTTP) e dois oligonucleótido iniciadores flanqueando a sequência alvo a ser amplificada. Os dois oligonucleótidos iniciadores são utilizados para criar um amplicão típico de uma reacção de PCR. Um terceiro oligonucleótido ou sonda, é concebido para detectar a sequência de nucleótidos localizado entre os dois iniciadores de PCR. Embora a concepção da sonda possa diferir, no método de PCR TaqMan® os sinais da sonda são controlados pela proximidade de um corante repórter fluorescente e um corante de interrupção fluorescente. Qualquer emissão induzida por laser a partir do corante repórter é interrompida pelo corante de interrupção quando os dois corantes estão localizados em proximidade um do outro, uma vez que se encontram na sonda. Durante a reacção de amplificação, a enzima polimerase (e. g., polimerase de ADN Taq) cliva a sonda de um modo dependente do molde. Os fragmentos da sonda resultantes dissociam-se em solução e o sinal do corante repórter libertado fica livre do efeito de interrupção do segundo fluoróforo. Uma molécula do corante repórter é libertada por cada nova molécula sintetizada e a detecção do corante repórter não interrompido proporciona a base para a interpretação quantitativa dos resultados.

O material de partida para a PCR pode ser ADN, ADNc, ARNm ou qualquer outro polinucleótido que necessite de ser amplificado.

Uma vez que a PCR requer ADN de cadeia simples como molde, se o material de partida é ADN de cadeia dupla, necessita de ser desnaturado de modo a produzir ADN de cadeia simples.

Como o ARN não pode servir como um molde para PCR, se o material de partida é ARN, o primeiro passo é a transcrição reversa do ARN molde em ADNc, seguido pela sua amplificação exponencial numa reacção de PCR. Esta versão de PCR é geralmente referida como PCR de transcriptase reversa (RT-PCR). As duas transcriptases reversas mais normalmente utilizadas são a transcriptase reversa do vírus de mieloblastose de aves (AMV-RT) e transcriptase reversa do vírus da leucemia de murino Moloney (MMLV-RT). O passo de transcrição reversa é tipicamente iniciado utilizando iniciadores específicos, hexâmeros aleatórios, ou iniciadores de oligo-dT, dependendo das circunstâncias. Por exemplo, ARN extraído a partir de uma amostra de tecido (e. g. FPET) pode ser alvo de transcrição reversa utilizando um GeneAmp RNA PCR kit (Perkin Elmer, CA, EUA), seguindo as instruções do fabricante. O ADNc derivado pode ser então utilizado como um molde na reacção de PCR subsequente.

Embora o passo de PCR possa utilizar uma variedade de polimerases de ADN dependentes de ADN termoestável, emprega tipicamente a polimerase de ADN Taq que possui uma actividade de nuclease 5'-3' mas não tem uma actividade de endonuclease de controlo de leitura 3'-5'. Assim, a PCR TaqMan[®] utiliza tipicamente a actividade de nuclease 5' da polimerase Taq ou Tth para hidrolisar uma sonda de hibridação ligada ao seu amplicão alvo, mas pode ser utilizada qualquer enzima com actividade de nuclease 5' equivalente. Neste caso, a sonda é concebida para ser não extensível pela enzima polimerase de ADN Taq. A RT-PCR TaqMan[®] pode ser realizada utilizando equipamento comercialmente

disponível, tal como, por exemplo, ABI PRISM 7700™ Sequence Detection System™ (Perkin-Elmer-Applied Biosystems, Foster City, CA, EUA), ou Lightcycler (Roche Molecular Biochemicals, Mannheim, Alemanha). Numa forma de realização preferida, o processo de nuclease 5' é realizado num dispositivo de PCR em tempo real quantitativo, tal como o ABI PRISM 7700™ Sequence Detection System™. O sistema consiste num termociclador, laser, dispositivo de carga acoplado (CCD), câmara e computador. O sistema amplifica amostras num formato de 96 poços num termociclador. Durante a amplificação, o sinal fluorescente induzido por laser é recolhido em tempo real através de cabos de fibra óptica para todos os 96 poços e detectado no CCD. O sistema inclui software para operar o instrumento e para analisar os dados.

Os dados do ensaio da nuclease 5' são inicialmente expressados como Ct ou o ciclo do limiar. Como discutido acima, os valores de fluorescência são registados durante cada ciclo e representam a quantidade de produto amplificado até esse ponto na reacção de amplificação. O ponto em que o sinal de fluorescência é primeiro registado como estatisticamente significativo é o ciclo do limiar (Ct).

Para minimizar erros e o efeito de variação amostra-para-amostra, a RT-PCR é normalmente realizada utilizando um padrão interno. O padrão interno ideal é expresso a um nível constante entre diferentes tecidos e não é afectado pelo tratamento experimental. Os ARN utilizados frequentemente para normalizar padrões de expressão genética são ARNm para os genes domésticos gliceraldeído-3-fosfato-desidrogenase (GAPDH) e β -actina.

Para mais detalhes sobre PCR em tempo real quantitativo, ver também Held *et al.*, *Genome Research* 6:986-994 (1996). A PCR é descrita nas Patentes U.S. N° 4683202, 4683195; 4965188; e 5075216.

2. Intrões e Excisão do ARN

A maioria dos genes em eucariotas superiores contém mais do que 100 000 pares de nucleótidos, alguns contendo mais de 2 milhões de pares de nucleótidos. Isto é significativamente mais longo do que a sequência de nucleótidos necessária para codificar uma proteína de tamanho médio (300-400 aminoácidos) que está na ordem dos cerca de 1000 nucleótidos. A maior parte do comprimento extra consiste em sequências não codificantes (intrão) que interrompem as sequências codificantes (exão) na sequência do gene. A maioria dos genes de eucariotas superiores que codificam para ARNm, ARNt e alguns que codificam para ARNr são interrompidos por sequências de intrão. Os genes para ARNm possuem tipicamente 0 a 60 intrões; enquanto que os genes para ARNt incluem tipicamente 0 ou 1 intrão.

Quando o ARNm é transcrito a partir de ADN, a princípio, tanto as sequências de exão como de intrão são transcritas para dentro do denominado ARN nuclear heterogêneo (hnARN) ou ARN imaturo ou pré-ARNm. Todavia, antes do ARN sair do núcleo, as sequências de intrão são frequentemente removidas do ARNm transcrito como um resultado de um processo conhecido como excisão do ARN. O processo da remoção de intrões envolve um processo de laços preciso, controlado por uma sequência de nucleótidos específica contígua os exões. Quase todos os intrões podem ser identificados através de sequências de consenso

específicas. As primeiras duas bases de um intrão são sempre GU, enquanto que as últimas duas bases são sempre AG, mas os sítios de excisão 5' e 3' possuem tipicamente sequências de consenso que se estendem para além dos motivos GU e AG. A excisão do ARNm ocorre numa partícula denominada spliceossoma, enquanto que o ARNt e o ARNr são excisados por mecanismos que não envolvem spliceossomas.

Os intrões são tipicamente muito mais compridos do que os exões (sequências que estão presentes no ARNm). Um exão de eucariota médio tem cerca de 150 nucleótidos de comprimento, enquanto que um único intrão humano pode ser tão comprido como próximo de 500000 nucleótidos, mas tipicamente são cerca de 2000-4000 nucleótidos. Em geral, um gene eucariótico contém muito mais sequências de intrões do que de exões, como ilustrado pela tabela seguinte (*Molecular Biology of the Cell*, Bruce Alberts et al., eds., 3ª edição, Garland Publishing Company, Nova Iorque, N.I., 1994, p. 340):

Tabela 1

Gene	Tamanho do gene (x 10 ³ nucleótidos)	Tamanho do ARNm (x 10 ³ nucleótidos)	Número de intrões
β-globina	1,5	0,6	2
Insulina	1,7	0,4	2
Proteinase C	11	1,4	7
Albumina	25	2,1	14
Catalase	34	1,6	12
Receptor de LDL	45	5,5	17
Factor VIII	186	9	25
Tiroglobina	300	8,7	36

Numa forma de realização particular da presente invenção, as sequências de intrão num gene de interesse são sujeitas a um processo de selecção para identificar a sequência ou sequências de ARN intrónico que co-expressa com as sequências de ARN do exão (*i. e.,* ARNm) do mesmo gene. Essas sequências de intrão seleccionadas, cuja expressão está correlacionada com a expressão das sequências de exão, possuem propriedades especialmente desejáveis como potenciais marcadores de diagnóstico: (1) devido ao seu desempenho técnico favorável (especificamente, otimizar a especificidade e a sensibilidade do ensaio); e, (2) qualquer que seja a importância biomédica ligada ao nível de ARNm do gene também está ligada aos níveis celulares das sequências intrónicas. Por exemplo, níveis elevados de uma espécie de ARNm que codifica um factor de crescimento potente estão provavelmente correlacionados com a elevada velocidade de crescimento de uma célula. Sequências intrónicas possuindo níveis celulares que estão correlacionados

com os níveis de ARNm deste mesmo gene possuem a mesma probabilidade de se correlacionar com velocidade de crescimento elevada de uma célula. Essas sequências intrónicas seleccionadas podem ser então utilizadas para rastrear espécimes de tecido valioso para pesquisar correlações clínicas e significativas para diagnóstico, previsão ou prognóstico.

Um processo exemplar para seleccionar sequências de intrão que co-expressa com o ARNm do mesmo gene é como se segue. Resumidamente, para qualquer gene de interesse, é ensaiado um conjunto de tecidos relevantes de uma população de doentes de interesse para medir os níveis de um conjunto de sequências intrónicas e de ARNm. As sequências intrónicas que se observaram possuir o coeficiente de correlação de Pearson mais elevado para co-expressão com as sequências de ARN do exão (ARNm) são depois seleccionados. O número de doentes estudados neste processo é, de um modo preferido, pelo menos, acima de 50 e, de um modo mais preferido, pelo menos, cerca de 100.

Num exemplo específico, o artigo biomédico de interesse refere-se a doentes com cancro da mama e o gene de interesse pode ser o marcador de crescimento tumoral Ki-67. Neste caso, são utilizados tumores de 50 ou mais doentes de cancro da mama para medir os níveis de ARNm de Ki-67 e os níveis de sequências de múltiplos intrões Ki-67 e os intrões que possuem o coeficiente de correlação de Pearson mais elevado co-expressando com ARN do exão são seleccionados.

Uma vantagem deste processo é que a selecção da sequência intrónica preferida pode ser realizada com espécimes de tecido que são relativamente fáceis de obter e abundantes (por exemplo, espécimes que não têm registos clínicos ligados valiosos).

Devido a esse tecido poder proporcionar grandes quantidades de ARN para rastreio, será possível detectar sinais de expressão genética a partir de sondas mesmo sub-óptimas. Os ensaios altamente sensíveis e específicos baseados nas sequências intrónicas seleccionadas podem ser depois utilizadas para rastrear espécimes de tecido valiosos, por exemplo, espécimes ligados a informação clínica importante, tais como recorrência de doença, morte, ou resposta a fármacos terapêuticos ou regimes de tratamento definidos.

3. *Perfis de Expressão Genética Utilizando Conjuntos de Sondas Iniciador/Sonda de PCR à base de Intrões*

Presentemente, iniciadores de PCR e sondas são concebidos com base na sequência de ARNm ou ADNc, sem considerar as sequências de intrão. Na verdade, os intrões são normalmente considerados como material de “empacotamento” que é removido durante a excisão e geralmente rapidamente degradado.

A presente invenção baseia-se na descoberta experimental não antecipada de que os ARN do intrão podem ser rapidamente detectados por RT-PCR, mesmo utilizando ARN altamente degradado a partir de espécimes de tecido fixo, embebido em parafina. Em particular, observou-se que em perfis de expressão genética para um dado gene, os sinais de RT-PCR de conjuntos de sonda/iniciador à base de intrão podem ser tão grandes como, ou maiores, do que os sinais de RT-PCR à base do exão. Embora esta descoberta seja suportada por algumas descobertas recentes com certas espécies de ARNm, não está de acordo com a opinião prevalente de que os intrões são degradados muito rapidamente após a excisão (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]).

Também inesperado, as constatações experimentais subjacentes à presente invenção, indicam que o ARN intrónico pode ser utilizado para perfis de expressão genética, porque as quantidades de tecido de sequências de intrão e exão expressas tendem a estar correlacionadas. Este resultado não é antecipado porque existe pouca ou nenhuma evidência de que a proporção das constantes da velocidade global para a síntese e processamento de sequências de intrão e exão transcritas são semelhantes. De facto, a literatura científica proporciona evidência sobre a complexidade do processamento de pré-ARNm e intrão excisado. Por exemplo, pode existir pré-ARNm em múltiplas misturas cinéticas (Elliott e Rosbash, *Exp. Cell Res.* 229:181-8 [1996]), com subpopulações contendo ARN de intrão que não são eficientemente removidas por excisão e são transportadas para o citoplasma em espécies de ARNm "imaturas", em que podem decair a velocidades diferentes das sequências de ARN de intrão nuclear (Wang *et al.*, *Proc. Natl. Acad. Sci. USA* 94:4360-5 [1997]). Existe evidência de que certos ARN de intrão excisados entram no citoplasma em estrutura de laço (Clement *et al.*, *RNA* 5:206-20 [1999]).

Finalmente, os resultados aqui apresentados indicam que as sequências de intrão podem servir como marcadores moleculares de diagnóstico ou prognóstico. O exame de quatro ARNm demonstraram previamente serem prognóstico no cancro, está demonstrado que as suas sequências de intrão correspondentes também são prognosticantes e nas mesmas direcções que as sequências de exão parentais transcritas (*i. e.*, prognóstico positiva ou negativo).

Em resumo, a abordagem da invenção foi demonstrada como se segue. O documento WO 0378662 descreve um conjunto de genes que prevê a probabilidade de recorrência do cancro da mama. Nesse estudo, os níveis de sequências de exão transcritas em espécimes de tecido de cancro da mama fixos embebidos em parafina de 146 doentes foram medidas por RT-PCR utilizando conjuntos de iniciador de PCR/sonda à base do exão. No estudo aqui descrito, os ensaios de RT-PCR foram criados para medir os níveis de sequência de intrão transcritos em quatro dos genes marcadores previamente identificados e depois utilizados para rastrear ARN de 60 espécimes de biopsia fixos embebidos em parafina (representando 60 doentes diferentes, um subconjunto dos doentes avaliados no estudo anterior). Os resultados apresentados nos exemplos abaixo demonstram que para cada gene os intrões e exões são co-expressos e que os intrões prevêm o risco de recorrência da doença, como previsto pelos resultados anteriores à base do exão.

3. Concepção de Iniciadores de PCR e Sondas à base de Intrões

De acordo com um aspecto da presente invenção, os iniciadores de PCR e sondas são concebidos com base nas sequências de intrão presentes no gene a ser amplificado. Consequentemente, o primeiro passo na concepção do iniciador/sonda é delinear as sequências de intrão nos genes. Isto pode ser realizado através de software disponível ao público, tal como o software DNA BLAT desenvolvido por Kent, W.J., *Genome Res.* 12(4):656-64 (2002), ou pelo software BLAST, incluindo as suas variações. Os passos subsequentes seguem

métodos bem estabelecidos de concepção de iniciador de PCR e sonda.

De modo a evitar sinais não específicos, é importante mascarar as sequências repetitivas nos intrões quando se concebem os iniciadores e sondas. Isto pode ser rapidamente realizado utilizando o programa Repeat Masker disponível online através de Baylor College of Medicine que rastreia as sequências de ADN contra uma biblioteca de elementos repetitivos e devolve uma sequência de análise, na qual os elementos repetitivos estão mascarados. As sequências de intrão mascaradas podem ser então utilizadas para conceber sequências de iniciador e sonda utilizando quaisquer programas de concepção de iniciador/sonda comercialmente ou de qualquer outra forma, disponíveis ao público, tais como Primer Express (Applied Biosystems); MGB assay-by-design (Applied Biosystems); Primer3 (Steve Rozen e Helen J. Skaletsky (2000) *Primer3 on the WWW for general users and for biologist programmers*. Em: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386).

Os factores mais importantes considerados na concepção de iniciadores para PCR, incluem o comprimento do iniciador, a temperatura de fusão (T_m) e o conteúdo em G/C, especificidade, complementaridade das sequências do iniciador e a sequência da extremidade 3'. Em geral, iniciadores de PCR óptimos têm geralmente 17-30 bases de comprimento e contêm cerca de 20-80%, tal como, por exemplo, cerca de 50-60% de bases G+C. As T_m 's entre 50 e 80 °C, e. g., cerca de 50 a 70 °C são tipicamente preferidas.

Para directrizes posteriores para concepção de iniciador de PCR e sonda ver, e. g. Dieffenbach, C.W. et al., "*General Concepts for PCR Primer Design*" em: *PCR Primer, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Nova Iorque, 1995, pp. 133-155; Innis e Gelfand, "*Optimization of PCRs*" em: *PCR Protocols, A Guide to Methods and Applications*, CRC Press, Londres, 1994, pp. 5-11; e Plasterer, T.N. *Primerselect: Primer and probe design. Methods Mol. Biol.* 70:520-527 (1997).

4. Aplicações

Os métodos da presente invenção e, especificamente, os iniciadores de PCR e sondas à base de intrões aqui apresentados, têm utilidade em todos os campos em que a amplificação de um ácido nucleico (incluindo ARN, ADN e, em geral, todos os oligo- e poli nucleótidos) representando um gene ou um fragmento de gene é necessária. Por isso, os iniciadores de PCR e sondas concebidos de acordo com a presente invenção podem ser utilizados para amplificar genes individuais, ou genes múltiplos presentes numa amostra biológica com o objectivo de determinar perfis de expressão genética através de qualquer metodologia incluindo, sem limitação, determinação de perfis de expressão genética baseada em PCR quantitativa (e. g. RT-PCR quantitativa) e análise de *microarray* e ensaios à base de esferas.

Por exemplo, numa forma de realização específica da técnica de *microarray*, são aplicadas inserções de clones de ADNc amplificados por PCR a um substrato num *array* denso. De um modo preferido, são aplicadas, pelo menos, 10000 sequências de nucleótidos ao substrato. Os genes dispostos em *microarrays*, imobilizados no microchip em, pelo menos, 10000 elementos cada,

são adequados para hibridação em condições de restrição. As sondas de ADNc marcadas com fluorescência podem ser criadas através da incorporação de nucleótidos fluorescentes por transcrição reversa de ARN extraído a partir dos tecidos de interesse. As sondas de ADNc marcadas aplicadas ao chip hibridam com especificidade a cada ponto de ADN no *array*. Após lavagem restrigente para remover sondas ligadas não especificamente, o chip é rastreado por microscopia de laser confocal ou por outro método de detecção, tal como uma câmara de CCD. A quantificação da hibridação de cada elemento disposto em *array* permite a avaliação da abundância do ARNm correspondente. Com fluorescência de cor dupla, sondas de ADNc marcadas separadamente criadas a partir de duas fontes de ARN são hibridadas par a par com o *array*. A abundância relativa dos transcritos a partir das duas fontes correspondentes a cada gene especificado, é assim determinada simultaneamente. A escala miniaturizada da hibridação permite uma avaliação conveniente e rápida de um padrão de expressão para números grandes de genes. Esses métodos demonstraram possuir a sensibilidade necessária para detectar transcritos raros que são expressos em algumas cópias por célula e detectar de forma reprodutível, pelo menos, diferenças de duas vezes nos níveis de expressão (Schena et al., *Proc. Natl. Acad. Sci. USA* 93(2):106-149 (1996)). A análise de *microarray* pode ser realizada através de equipamento disponível comercialmente, seguindo os protocolos do fabricante, tal como utilizando a tecnologia Affymetrix GenChip, ou a tecnologia de *microarray* da Agilent.

Um aspecto importante da presente invenção é utilizar a amplificação genética à base de intrões como parte de determinação de perfis de expressão genética para estar de acordo com doentes para os melhores fármacos ou combinações de

fármacos e para proporcionar informação de prognóstico. Por exemplo, a expressão de genes medida em tecido de cancro (e. g., tecido de biopsias de cancro da mama) pode ser utilizada para prever a probabilidade de sobrevivência a longo termo, sem doença, de doentes após cirurgia e/ou outra terapia de cancro, ou para prever a resposta do doente a uma abordagem terapêutica particular. Para este efeito, é tipicamente necessário corrigir (normalizar) ambas as diferenças na quantidade de ARN ensaiado e a variabilidade na qualidade do ARN utilizado. Por isso, os ensaios de uma invenção normalmente medem e incorporam a expressão de certos genes normalizantes, incluindo genes de referência bem conhecidos, tais como GAPDH e Cyp1. Alternativamente, a normalização pode ser baseada no sinal médio ou mediano (Ct) de todos os genes ensaiados ou um grande subconjunto destes (abordagem de normalização global). Numa base de gene-por-gene, a quantidade normalizada medida de um ARNm do tumor de um doente é comparada com a quantidade encontrada num cancro, e. g., conjunto de referência de tecido do cancro da mama. O número (N) de tecidos de cancro, e. g., cancro da mama, neste conjunto de referência deve ser suficientemente elevado para assegurar que conjuntos de referência diferentes (como um todo) se comportam essencialmente do mesmo modo. Se esta condição for satisfeita, a identidade dos tecidos de cancro da mama individual presentes num conjunto particular não terão impacto significativo nas quantidades relativas dos genes ensaiados. Normalmente, o conjunto de tecido de referência de cancro da mama consiste em, pelo menos, cerca de 30, de um modo preferido, pelo menos, cerca de 40 espécimes de tecido de cancro da mama diferentes fixos, embebidos em parafina (FPE). Salvo referência em contrário, os níveis de expressão normalizados para cada ARNm/tumor testado/doente serão expressos como uma percentagem de um nível de expressão medido no conjunto de

referência. Mais especificamente, o conjunto de referência de um número suficientemente elevado de (e. g., 40) tumores produz uma distribuição de níveis normalizados de cada espécie de ARNm. O nível medido numa amostra de tumor particular a ser analisada cai em algum percentil neste intervalo que pode ser determinado através de métodos bem conhecidos na técnica.

Num estudo de Fase II da expressão do gene em amostras de tecido fixo, embebido em parafina de carcinoma da câmara invasivo, a sobre expressão de qualquer dos genes seguintes no tecido de cancro da mama foi observado indicar uma probabilidade reduzida de sobrevivência sem recorrência de cancro após cirurgia: FOX1; PRAME; SKT15, Ki-67; CA9; NME1; SURV; TFRC; YB-1; RPS6KB1; Src; Chk1; CCNB1; Chk2; CDC25B; CYP3A4; EpCAM; VEGFC; hENT1; BRCA2; EGFR; TK1; VDR.

No mesmo estudo, a sobre expressão de qualquer um dos seguintes genes em cancro da mama indica um melhor prognóstico para sobrevivência sem recorrência de cancro após cirurgia: Blc12; CEGP1; GSTM1; PR; BBC3; GATA3; DPYD; GSTM3; ID1; EstR1; p27; XIAP; IGF1R; AK055699; P13KC2A; TGFB3; BAG1; pS2; WISP1; HNF3A; NFkBp65.

Neste mesmo estudo de Fase II da expressão genética em amostras de tecido fixo embebido em parafina de cancro da mama positivo para ER, a sobre expressão dos seguintes genes foi indicativa de uma probabilidade reduzida de sobrevivência sem recorrência de cancro após cirurgia: PRAME; FOXM1; EPHX1; HIF1A; VEGFC; Ki-67; VDR; NME1. Alguns destes genes (PRAME; FOXM1; VEGFC; Ki-67; VDR; e NME1) foram também identificados como indicadores de prognóstico fraco na análise prévia, não limitada a cancro da mama positivo para ER. A sobre expressão dos

restantes genes (EPHX1 e HIF1A) foi verificada ser um indicador negativo de sobrevivência sem doença apenas no cancro da mama positivo para ER. A sobre expressão dos seguintes genes em cancro positivo para ER foi verificado ser indicativo de um melhor prognóstico para sobrevivência sem recorrência de cancro após cirurgia: Bcl-2; DIABLO; IGF1R; GSTM3. Dos últimos genes, Bcl-2; IGF1R; e GSTM3 também foram identificados como indicadores de bom prognóstico na análise prévia, não limitada a cancro da mama positivo para ER. A sobre expressão de DIABLO pareceu ser um indicador positivo de sobrevivência sem doença apenas em cancro da mama positivo para ER. Para mais detalhes ver o documento WO 0378662.

Os estudos descritos acima foram realizados essencialmente como descrito no Exemplo 2 abaixo, excepto que a amplificação do gene foi estudada utilizando amplicões à base de exão. Para mais detalhes, ver o documento WO 0378662. Como atestado para os resultados apresentados no Exemplo 2, os resultados obtidos utilizando amplicões à base de intrões apresentam correlação excelente com os resultados anteriores e proporcionam tipicamente o benefício adicional de sensibilidade aumentada.

As constatações do estudo prévio de Fase II de carcinoma do ducto da mama invasivo foram submetidas a análise multivariada passo a passo, utilizando o Modelo de Riscos Proporcional de Cox utilizando a seguinte equação:

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

Nesta equação, os coeficientes para os genes que constituem uma previsão de resultados benéficos são números positivos e os

coeficientes para genes que são uma previsão de resultados desfavoráveis são números negativos. Os valores de “Ct” na equação são ΔC_t s, *i. e.*, reflectem a diferença entre o valor de Ct médio normalizado par uma população e o Ct normalizado medido para o doente em questão. A convenção utilizada na análise foi que os ΔC_t s abaixo e acima da média da população possuem sinais positivos e sinais negativos, respectivamente (reflectindo abundância de ARNm superior ou inferior). O risco relativo (RR) calculado resolvendo esta equação indica se o doente tem uma hipótese aumentada ou reduzida de sobrevivência a longo termo sem recorrência de cancro.

Numa análise multivariada, utilizando um conjunto de interrogação incluindo um número de genes reduzido, foram identificados os seguintes conjuntos de dez genes como possuindo valor de previsão particularmente forte de sobrevivência do doente sem recorrência de cancro após remoção cirúrgica do tumor primário.

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.

Numa análise multivariada, utilizando um conjunto de interrogação incluindo todos os genes identificados, foram identificados os seguintes conjuntos de dez genes como possuindo valor de previsão particularmente forte de sobrevivência do doente sem recorrência de cancro após remoção cirúrgica do tumor primário.

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, FOX1, 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.

Utilizando a mesma abordagem de análise multivariada para cancro da mama positivo para ER, foram identificados os seguintes conjuntos de dez genes como possuindo valor de previsão particularmente forte de sobrevivência de doentes sem recorrência de cancro após remoção cirúrgica do tumor primário.

1. PRAME, p27, IGFBP2, HIF1A, TIMP2, DLT2, 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.

À luz da excelente correlação entre os resultados de determinação de perfis de expressão genética à base de exões ou à base de intrões (ver Exemplo 2), é esperado que os mesmos

conjuntos de genes possuam um valor de prognóstico semelhante quando a determinação da expressão genética é baseada na quantificação de sinais de RT-PCR de conjuntos de iniciador/sonda à base de intrões.

Outros detalhes da invenção serão aparentes a partir dos seguintes exemplos não limitativos.

Exemplo 1

Concepção e Utilização de Conjuntos de Iniciador de PCR/Sonda Específicos de Intrões

Foi extraído ARN a partir de espécimes de biopsia de cancro da mama fixos em formalina, embebidos em parafina (FPET) (Clinomics Biosciences Inc., Pittsfield, MA) como se segue. Foram cortadas três secções de 10 µm e colocadas num tubo de 1,5 mL. A parafina foi removida por extracção com xileno (1 mL, 3 vezes) seguido por lavagem com etanol (1 mL, duas vezes). O ARN foi isolado de blocos de tecido seccionados utilizando o MasterPure™ Kit de Purificação (Epicentre, Madison, WI). O ARN foi quantificado pelo RiboGreen Fluorescence Method (Molecular Sondas). Foram então reunidas vinte amostras de ARN FPET e utilizadas como descrito abaixo.

A primeira cadeia do ADNc foi sintetizada utilizando Qiagen's Omniscript Reverse Transcriptase com iniciadores específicos para o gene reunidos (iniciadores reversos apresentados na Figura 2) hexâmeros aleatórios e Inibidor de ARNase, utilizando ARN FPET reunido (400 ng). Também foi realizada uma reacção sem transcriptase reversa (RT) com 150 ng

de ARN FPET reunido, ARN suficiente para realizar a amplificação Taqman a 5 ng/poço.

Tabela 2

Reagentes	RT	Sem RT	Conc. final
	Vol (µL)	Vol (µL)	
Tampão RT 10 X	4	2	1X
Mistura de dNTP, 5 mM de cada dNTP	4	2	500 µM cada
Hexâmero aleatório ABI 50 µM	1	0,5	1,23 µM
Mistura GSP, 1 µM	2	1	50 nM
Inibidor de ARNase ABI 20 U/µL	1	1	20 U/reacção
Omniscript RT, 4U/µL	2	0	8U ou 0 U/reacção
Água sem nuclease	10	5,5	
ARN FPET reunido (164 ng/µL)	16	8	65,6 ng/µL
Vol total	40	20	

Condições de reacção: 37 °C, 60 min,
93 °C, 5 min

Ensaio TaqMan

Ensaio TaqMan para o painel de 48 genes foram realizados em poços em triplicado com volume de reacção de 25 µL e incorporação de ARN de 5 ng por ensaio. Foi realizada uma

reação “sem RT” para cada gene num único poço como controle para verificar que o ARN mais do que os sinais de ADN estavam a ser medidos. Foi realizada quantificação em tempo real no ABI 7700 utilizando os seguintes parâmetros:

Condições dos ciclos: 95 °C, 10 min durante um ciclo, 95 °C, 20 seg seguido por 60 °C, 45 seg, 40 ciclos.

Volume de reação: 25 µL.

Condições de camada de tampão: FAM, (a referência passiva é ROX)

Resultados

Foram concebidos conjuntos de iniciador-sonda Taqman específicos com base em intrões mascarados os genes CEGP1, FOXM1, PRAME e STK15. Para delinear as sequências de intrão nos genes, a sequência de referência do NCBI para cada ARNm (NM_XXXXXX) foi alinhada com o genoma humano utilizando o programa de alinhamento de tipo BLAST (BLAT) disponível no sítio de recursos de genoma online da Universidade de Santa Cruz (<http://genome.ucsc.edu>). As sequências de intrão foram depois pesquisadas quando a sequências repetitivas utilizando o programa Repeat Masker disponível online através de Baylor College of Medicine (<http://searchlauncher.bcm.tmc.edu/seq-util/seq-util.html>). Sequências repetidas, tais como repetições de Alu, são identificadas por este programa e mascaradas. É importante excluir estas sequências antes de conceber iniciadores-sondas porque produzem sinais fortes, não específicos. As sequências de intrão mascaradas (Figuras 1A-M)

foram então utilizadas para conceber conjuntos de iniciador-sonda Taqman utilizando Primer Express (ABI). Outros programas adequados para conjuntos de iniciador-sonda incluem, por exemplo, o novo programa de concepção de sonda iniciadora para ensaios-por-desenho de MGB (ABI). Os amplicões para cada conjunto de iniciador-sonda são delineados em fonte a negrito na Figura 1. Cada conjunto iniciador-sonda específico é apresentado na Figura 2.

Os conjuntos de iniciador-sonda específicos para o intrão (genes de teste) foram utilizados em conjunto com o seu conjunto de iniciador-sonda específico para o exão correspondente (gene de referência) em experiências de perfil de expressão de genes Taqman convencionais utilizando ARN FPET reunido. A expressão normalizada foi calculada pela fórmula $2^{\Delta Ct}$ em que ΔCt é a diferença entre os Ct do conjunto de iniciador-sonda de gene de teste e os conjuntos de iniciador-sonda de gene de referência [Ct (referência)-Ct(teste)].

Exemplo 2

Um Estudo de Fase II de Expressão Genética em Tumores da Mama Pré-malignos e Malignos

Foi concebido um estudo de expressão genética e realizado com o objectivo principal de caracterizar molecularmente a expressão genética em amostras de tecido fixo embebido em parafina de carcinoma do ducto da mama invasivo e explorar a correlação entre esses perfis moleculares e sobrevivência sem doença.

Concepção do Estudo

Os ensaios moleculares foram realizados em tecidos de tumor da mama primário fixo com formalina, embebidos em parafina obtidos a partir de 60 doentes individuais diagnosticados com cancro da mama. Todos os doentes foram submetidos a cirurgia com diagnóstico de carcinoma invasivo da mama. Os doentes foram incluídos no estudo apenas se a avaliação histopatológica, realizada como descrito na secção de Materiais e Métodos, indicou quantidades adequadas de tecido tumoral e patologia homogénea.

Materiais e Métodos

Cada bloco de tumor representativo foi caracterizado por histopatologia convencional para o diagnóstico, avaliação semi-quantitativa de quantidade de tumor e grau tumoral. Foi preparado um total de 6 secções (10 micrones de espessura cada) e colocados em dois tubos de microcentrifuga da marca Costar (Polipropileno, tubos de 1,7 mL, transparentes; 3 secções em cada tubo). Se o tumor constituiu menos de 30% da área total do espécimen, a amostra pode ter sido dissecada de forma bruta pelo patologista, utilizando microdissecção grosseira, colocando o tecido de tumor directamente no tubo Costar.

Se foi obtido mais do que um bloco tumoral como parte do processo cirúrgico, todos os blocos tumorais foram submetidos à mesma caracterização, como descrito acima e o bloco mais representativo da patologia foi utilizado para análise.

Análise da Expressão Genética

O ARNm foi extraído e purificado a partir de amostras de tecido fixo, embebido em parafina e preparado para análise de expressão genética, como descrito acima.

Os ensaios moleculares de expressão genética quantitativa foram realizados por RT-PCR, utilizando o ABI PRISM 7900™ Sequence Detection System™ (Perkin-Elmer-Applied Biosystems, Foster City, CA, EUA). O ABI PRISM 7900™ consiste num termociclador, laser, dispositivo de carga acoplado (CCD), câmara e computador. O sistema amplifica amostras num formato de 384 poços num termociclador. Durante a amplificação, o sinal fluorescente induzido por laser é recolhido em tempo real através de cabos de fibra óptica para todos os 384 poços e detectados no CCD. O sistema inclui software para operar o instrumento e para analisar os resultados.

Análise e Resultados

O tecido tumoral foi analisado para expressão de 48 sequências de ARN diferentes representando produtos de 37 genes diferentes. Os valores de ciclo limiar (Ct) para cada doente foram normalizados com base na mediana de todos os genes para esse doente particular. Os resultados clínicos estavam disponíveis para todos os doentes a partir de uma revisão do registo de entrada e das fichas dos doentes seleccionados.

Os resultados foram classificados como:

0 morto devido a cancro da mama ou a causa desconhecida ou vivo com recorrência do cancro da mama;

1 vivo sem recorrência de cancro da mama ou morto devido a uma causa diferente de cancro da mama

A análise foi realizada por:

A análise da relação entre expressão genética normalizada e o tempo até ao resultado (0 ou 1, como definido acima), em que os doentes que estavam vivos sem recorrência de cancro da mama ou que morreram devido a uma causa diferente de cancro da mama foram censurados. Esta abordagem foi utilizada para avaliar o impacto de prognóstico de genes individuais e também conjuntos de múltiplos genes.

Para cada gene, foi definido um modelo de Riscos Proporcional de Cox (ver, e. g., Cox, D. R. e Oakes, D. (1984), *Analysis of Survival Data*, Chapman e Hall, Londres, Nova Iorque) com o tempo até à recorrência ou morte como a dependente variável e um nível de expressão no intervalo como a independente variável. Os genes que possuem um valor de $p < 0,05$ no modelo de Cox foram identificados. Para cada gene, o modelo de Cox proporciona o risco de recorrência relativo (RR) ou morte para uma alteração unitária na expressão do gene. Pode ser escolhida a divisão dos doentes em subgrupos em qualquer valor limiar da expressão medida (na escala de Ct), em que todos os doentes com valores de expressão acima do limiar possuem risco mais elevado e todos os doentes com valores de expressão abaixo

do limiar possuem risco inferior, ou vice versa, dependendo se o gene é um indicador de prognóstico fraco ($RR > 1,01$) ou bom ($RR < 1,01$). Assim, qualquer valor de limiar definirá subgrupos de doentes com risco respectivamente aumentado ou diminuído.

A Tabela 3, abaixo, apresenta correlação de expressão para a par (apresentada por coeficientes de correlação) entre os intrões e exões testados para os genes CEGP1, FOXM1, PRAME e STK15. Para dois dos quatro genes, CEGP1 e PRAME, verificou-se que os intrões produziam coeficientes de correlação [para co-expressão com os seus respectivos exões] acima de 0,90. No caso de STK15, um intrão correlacionado com a expressão de exões com um coeficiente de correlação de $\sim 0,80$. Para FOXM1, as correlações de expressão de intrão:exão foram significativamente mais baixas. Neste último caso, todavia, parece provável que a verdadeira expressão possa estar altamente correlacionada, mas não detectável por uma razão técnica. A expressão do exão FOXM1 em muitos doentes estava subjacente ao limiar de detecção do ensaio que potencialmente previne a detecção de correlações elevadas que pode existir. Se esta hipótese é correcta, os intrões de FOXM1 ainda iriam registar como marcadores de prognóstico clínico negativo como previamente demonstrado para FOXM1. Como apresentado mais tarde, este resultado ocorre.

As Figuras 3, 4 e 5 demonstram que a correlação de expressão par a par dos ARN testados contra os ARN dos exões CEGP1, PRAME e STK15. Como apresentado, os intrões respectivos destes genes produziram as correlações mais elevadas. É de salientar que o painel de 48 genes incluiu genes que foram seleccionados, através de várias estratégias baseadas em bioinformática, como particularmente provável para correlacionar em expressão com CEGP1, PRAME, STK15 e FOXM1. As estratégias não baseadas em

intrões tiveram mais sucesso no caso de STK15, na medida em que vários genes candidatos tinham coeficientes de correlação de no intervalo de 0,6-0.7.

Tabela 3. Correlações entre a Expressão de Intrão e Exão para Quatro Genes

Coefficiente de Correlação de Expressão {R}

	<i>CEGP1</i> Intrão 1.1	<i>CEGP1</i> Intrão 3.1	<i>CEGP1</i> 4.1	<i>Intrão</i> 5.1	<i>CEGP1</i> Intrão	<i>CEGP1.2</i>
<i>CEGP1</i> Intrão 1.1	1,00					
<i>CEGP1</i> Intrão 3.1	0,89	1,00				
<i>CEGP1</i> Intrão 4.1	0,97	0,82	1,00			
<i>CEGP1</i> Intrão 5.1	0,91	0,87	0,88	1,00		
<i>CEGP1.2</i>	0,91	0,80	0,90	0,87	1,00	
	<i>FOXM1</i> intrão3.3	<i>FOXM1</i> intrão5.1	<i>FOXM1</i> intrão7.	<i>FOXM1.</i>		
			1	1		
<i>FOXM1</i> intrão3.3	1,00					
<i>FOXM1</i> intrão5.1	0,48	1,00				
<i>FOXM1</i> intrão7.1	0,54	0,73	1,00			
<i>FOXM1.1</i>	0,44	0,33	0,36	1,00		
	<i>STK15</i> intrão1.1	<i>STK15</i> intrão2.1	<i>STK15</i> intrão4.	<i>STK15.</i>		
			1	2		
<i>STK15</i> intrão1.1	1,00					
<i>STK15</i> intrão2.1	0,78	1,00				
<i>STK15</i> intrão4.1	0,69	0,74	1,00			
<i>STK15.2</i>	0,63	0,70	0,78	1,00		
	<i>PRAME</i> intrão2.1	<i>PRAME.3</i>				
<i>PRAME</i> intrão2.1	1,00					
<i>PRAME.3</i>	0,97	1,00				

A Tabela 4 abaixo, apresenta o impacto após sobrevivência do doente de expressão de CEGPl, FOXM1, PRAME e STK15, exões e intrões. Os exões parentais tinham todos impacto estatisticamente significativo em risco relativo [RR], como foi previamente determinado pelos autores, excepto no caso de FOXM1. Dado que o presente estudo avaliou 60 doentes do grupo original de 146 doentes, o marcador FOXM1 pode ter caído na sua significância devido ao risco estatístico de examinar um conjunto de resultados reduzido. Muito notoriamente, para todos os quatro genes testados, a expressão do intrão teve impacto significativo em RR e na mesma direcção que os exões parentais.

Tabela 4. Resultados do Modelo de Cox para 60 Doentes com Cancro da Mama

Correlações de Prognóstico

Gene	Coef	RR=exp(coef)	se(coef)	z	p
CEGP1.2	-0,202	0,817	0,050	-4,024	0,00006
CEGP1intrão1.1	-0,329	0,720	0,087	-3,771	0,00016
CEGP1intrão3.1	-0,261	0,770	0,078	-3,335	0,00085
CEGP1intrão4.1	-0,275	0,760	0,073	-3,774	0,00016
CEGP1intrão5.1	-0,312	0,732	0,082	-3,817	0,00014
FOXM1.1	0,175	1,192	0,136	1,289	0,19700
FOXM1intrão3.3	0,304	1,355	0,120	2,523	0,01160
FOXM1intrão5.1	0,514	1,673	0,195	2,639	0,00832
FOXM1intrão7.1	0,546	1,726	0,182	2,993	0,00276
PRAME.3	0,125	1,133	0,054	2,294	0,02180
PRAMEintrão2.1	0,125	1,133	0,052	2,397	0,01650
STK15.2	0,692	1,998	0,201	3,450	0,00056

(continuação)

Gene	Coef	RR=1xp(coef)	se(coef)	z	p
STK15intrão1.1	0,357	1,429	0,149	2,400	0,01640
STK15intrão2.1	0,391	1,479	0,154	2,536	0,01120
STK15intrão4.1	0,410	1,506	0,133	3,084	0,00204

Existe uma percepção comum de que os níveis de estado estável de sequências de exão transcritas excedem grandemente as sequência de intrão transcritas (Sharp *et al.* *Ann. Rev. Biochem.* 55: 1119-50 [1986]). Todavia, o nosso exame da expressão de exões e intrões de CEGP1, FoxM1, PRAME e STK15, utilizando TaqMan[TM] RT-PCR para avaliar ARN de tecido de cancro da mama fixo embebido em parafina, demonstraram que as intensidades de sinal de intrão e exão estavam no mesmo intervalo e em todos os casos no intervalo de detecção útil do ensaio [resultados não apresentados]. A detecção de ARN intrónico neste estudo é ainda mais relevante porque o tecido utilizado foi fixo em formalina que degrada ARN e desse modo limita substancialmente a capacidade para detectar ARN (T.E. Godfrey *et al.* *J. Mol. Diag.* 2: 84-91 [2000]). No caso de CEGP1 três dos intrões testados produziram sinais inferiores e um deles um sinal superior ao exão. No caso de FOXM1, cinco de nove intrões testados produziram sinais superiores aos do exão. No caso de intensidade de sinal de PRAME a partir do intrão e exão testados foram quase idênticos. Finalmente, para STK15 todos os intrões tinham intensidades de sinal que foram 1/4 a 1/20 os do exão, mas ainda assim no intervalo útil do ensaio. Assim, estes resultados indicam que os níveis de estado estável dos intrões expressos são adequados para utilização de ARN de intrão como marcadores moleculares.

Embora a presente invenção tenha sido descrita com referência ao que é considerado como sendo formas de realização

específicas, deve ser entendido que a revelação não está limitada. Por exemplo, embora a revelação inclua vários genes e conjuntos de genes associados ao cancro da mama, estão contemplados genes semelhantes e conjuntos de genes e métodos que se referem a outros tipos de cancro.

LISTAGEM DE SEQUÊNCIAS

<110> GENOMIC HEALTH, INC.

SCOTT, Randal W.

BAKER, Joffre B.

KIEFER, Michael C.

<120> UTILIZAÇÃO DE ARN INTRÓNICO PARA MEDIR A EXPRESSÃO
GENÉTICA

<130> 39740-0007 PCT

<140> Não Atribuído

<141> 2004-02-19

<150> documento US 60/448,991

<151> 2003-02-20

<160> 67

<170> FastSEQ para Windows Versão 4.0

<210> 1

<211> 1566

<212> ADN

<213> *Homo sapiens*

<220>

<221> característica miscelânea

<222> (802)...(1027)

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

<221> característica miscelânea

<222> (1126)...(1518)

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

<221> intrão

<222> (0)...(0)

<223> MGB-CEGP1 int 1.1

<400> 1

```
gtgagtgtcc ggccgcgggg ggcacactgg cacagcaggc agggccagga agagtgttta 60
ggccccggc ggagtccaga gccgggcgcg cggggctcgg ggctggcggc tgcagctccg 120
cgggggcctc tgctcccccc gggacctcac ccgccggccg ggccaaggcg ccacgaccgc 180
tgggggcctg agtccttcgg cccggcctcg gacccggagc tgctgacggg tcccggccccg 240
gtccggatgc ctccagagcg cctgctagtc agaccgtcgc cggcgagcag gcaggagggt 300
gcggaacctg gccttggggg cccgcgcctc agcgtaggcg gggaaactga gggccggggc 360
gggcacatcc gcgaggcggt ggcagctttg ccgtttcttt ctttgggggc cggcaagtgc 420
tgctgatggc ttcgggggtg gctccagaga cttttctgtc agcggaaacag cgcctgttcc 480
gatctgggaa ttaccctgaa gcagcaacaa gcctagggtt tcagcagaga actttgggtt 540
ccagagagga ctctggacgt gctgtgctta ctggacttgc aatactttca aaatgctttt 600
gtttttaatt aatatcctgg agtagtgtca acccaggaaa tacttctgcc aaggcgggtt 660
tccaggttga gaggatgggc aggggtggga gtgcaggggg ccggccatgg ggacaccatc 720
cccgttcgc 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 catagcctaa gaagcctatac caagatctca cttatctctt 1566

ttgcag
```

<210> 2

<211> 4985

<212> ADN

<213> *Homo sapiens*

<220>

<221> característica miscelânea

<222> (83)...(253)

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

<221> característica miscelânea

<222> (1975)...(2296)

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

<221> característica miscelânea

<222> (2445)...(2468)

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

<221> característica miscelânea

<222> (2499)...(2551)

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

<221> característica miscelânea

<222> (4295)...(4352)

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

<221> intrão

<222> (0)...(0)

<223> MGB-CEGP1 int 3.1

<400> 2

```
.....-
gtacctctgc ccagctgtgg atggggggcag agccacatct gagaccctct cccttgcacg 60
cgcacacaca cactgactct agnnnnnnnnn 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 aaaagggtgct gaaaaccaca gcaccagtat 360
gagcctgctc cctgctctgg gtgggtaggg aggaggctgg atccttcca tgcagacttt 420
caatgaagtg ccctgttttc agccccaagc tagatccggc ctttccatgt 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 actcttcctt gctctttctt aagctaaaga gtcacatttt ggtggctgtg 780
ttcctcttgg agttgcatct gcctattttt aggggaagtg ccctaaatac tagcctatta 840
accccttttg ccatgtgctg cttattcttt cccattactt aagaatgagg tcattttaat 900
ttcttctact atttaatcac aaatttatag attgttttaa tcctgggtct ggtaactttt 960
caagggtttc ttcattggaag atgatttttg tctcattttc caaggatggc agctcacacc 1020
ttatacttaa ctagaatacc tgtttgggta ccaagaaaaa ttgtcagagg aacccccagg 1080
ggccaatggg tttgatggct atcatcacc agagcctgct cattctcagc gtttggggcg 1140
gggaagtcac acatactggc tttgatcagg cagatttcct atcttgtgcc aggtgtggcc 1200
cttgataaag tagcagttgg gtttcatttt cctgccaggt tctctggggg cattggtgtg 1260
ccctgcactc ttgtccaatg taggccaaat tcgagatggg aatgaattag gaggccagtg 1320
gcacagagtg atccgaatct cagggcatct ctcttttga ttgctcaaag ctgcttcctg 1380
ggaagtcact ttggcttcct ctgcaggtgg ctggggaggg atgtgggaac tgcaggttaa 1440
agccatcgct tgagccctca cggctctgggt cccaccaggt tacaaagcag ctggtagcga 1500
----
```

ttaagatcac	ctcttatccc	tgtacttcca	gagccctggc	tcagcccccac	tctcccctcc	1560
tgcaagcccc	cggactgatt	agagacacag	gctcctcata	ccagaagcaa	atacaaatgc	1620
agttcctttc	tgcaaaactgt	gttttctaaa	ttttctacaa	ttcagacatt	cttgatccc	1680
ctaaagagta	tttgaagtga	acatttttgt	ctggaactaa	aaccaaaatc	taagaatttg	1740
cgttggtggtc	tggaagtgtc	ctctgtgatt	ttctgttggtg	tttcaacctg	attgcttggc	1800
aaattcatgg	gagtgtcagc	caacagatta	tagcaattgg	taacggagaa	cctttgcac	1860
ctagggtttt	gattcttcaa	atagaacagc	ctgtaaaaag	ttttcttcta	ggatttcctc	1920
tctgatatgc	acattaaact	ctatgaaact	gtaggcttaa	aaaccacag	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	tggtgttttt	aataattatt	taccaaatgt	tctttaatat	2400
agggcatcat	aatcattgac	tctgagggaa	agctcaagat	actgnnnnnn	nnnnnnnnnn	2460
nnnnnnnncc	ttagagactc	caaagctgtg	ataaagagnn	nnnnnnnnnn	nnnnnnnnnn	2520
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	ncccgccccg	cctttgttcc	tattcatggg	2580
tgctcaggct	ctcagaatga	gcactcctct	tttgttttgt	gtgttctgag	aatattttaga	2640
tggtgtactg	atgccttttc	agggcaacag	ggaagggtgc	aggggtggcaa	agtggaggct	2700
gtgctttcag	caggacctgt	taccggtttt	atgtcatgtt	ttcctcccaa	ttcacaaggc	2760
atatttttgt	ttgggtttcca	gaaataatct	tcagtggagc	cctgatcttg	gggtgcacca	2820
gaatggggga	tttccaatgt	ttctgagctg	tttcccttct	ggtgaacgaa	ccatcctgga	2880
cgtgacaacc	agaccaattt	tggaaagagc	tagggccatt	tgctgggctg	cctagtttgg	2940
aacagattaa	tctgtccacc	ccagcagtg	tcttgcatta	agtcagagt	tcacaaaggc	3000
tttgaggtca	cttcttgaaa	agctgtcagc	gtttccagag	ccatttaagt	ctctattatg	3060
tcttggtaac	ttcagggtga	gcttgatgtg	gtaggacatt	aggtggtagg	ttctctgtgt	3120
atcacaatgg	catctggcat	acaggcattc	ttacgaaata	tttcttgtgt	aggtgaatta	3180
ctctgaggca	gtaaagggtca	ctttgcaa	gtcttaacag	tcttgtaa	agagtga	3240
agcagcagca	gctggcctgt	ttgggaggtg	actttccagg	tgctcctgcc	cccatttctt	3300
gggcagttat	atatttacc	ccgagcacta	gttacttccc	atgctcggct	gacccaagga	3360
caaacacaa	gctttctggg	ccttctcaga	caggacactg	cttctagagg	cactgtgcac	3420
ctcccgcgcc	atctcagtac	tgggggtgcaa	atcacatctt	cgggaattacc	agccagagca	3480
agagaaagct	ttccaccaat	ccagtgcag	tctctttctg	tgtaattga	cagccaccct	3540
tggcatggat	gaatgaatcc	cagcaaccag	cagactgagt	gctggagtgc	aggcagctca	3600
taactgtcag	gcaaaagagc	aagaggggtt	taagagagac	tccagaaagt	atgggatata	3660
ttaacccttg	cactgtcttc	tggaaatagga	atgacatctg	tttgatttaa	aacaattgtt	3720
ccgtttaagc	acagtttgac	agctctggag	tgggagctgg	agagagaact	ttgacttcac	3780
tagaacctgt	tggctaagg	tttaggggca	caatatagaa	gggtgttgga	ttctagagaa	3840
gtgaaagcaa	cctttttgta	ctcgtgttga	aaacagtgcc	ctactagtat	tagagtgtct	3900
cattgataga	gagccaatga	caaccaagtc	cctactctca	gagatgtttt	agagttacat	3960
tgcacgaatg	caaagaagca	acataggaac	aggtaatata	taataaagta	taaactgagc	4020
agatgtcttg	aaagtattct	agggtatgaa	aagaattcct	tcaggatgct	ggtaggcagc	4080
aggatctcaa	agaattagtt	ttgagatgag	gcagaatgct	ggtaaaccac	acgggcagtt	4140
accttgctgt	gccccctcat	ttagatgtgt	gccgagccct	gcaagaacag	aagcagctgt	4200
tccccctccc	accatcatat	tacaagggtta	agcctaata	gaatttactg	tataacctca	4260
aaaaattgta	cagcagctac	cacacacgag	cacannnnnn	nnnnnnnnnn	nnnnnnnnnn	4320
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnngttctgt	tgctgttgct	gtgtgatgct	4380
gtcagaggct	tatgccctga	gggagggatc	aagggagtg	ctgagggtgg	tcacagaaga	4440
cagattccgg	ggcatgtggc	ccgtacgagg	atgccaaaat	gccacagtca	cactcacctc	4500
agaaggggtg	gattgggtgg	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	gccaccctt	agccctttgc	4920
cgctctccca	gggctcagga	ttctgaccac	agcctagtca	cctgtcgcac	actgctgttt	4980
ttcag						4985

<210> 3

<211> 2556

<212> ADN

<213> *Homo sapiens*

<220>

<221> característica miscelânea

<222> (507)...(743)

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

<221> característica miscelânea

<222> (1584)...(1678)

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

<221> característica miscelânea

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

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

<221> intrão

<222> (0)...(0)

<223> MGB-CEGP1 int 4.1

<400> 3

```
gtaagtatgg gccagtgcac acctgccatg ggaaccgtcg tattccacag gctgccttct 60
gtggcccagc tcagaagcac cacctcatgg cacggctgca gcagcagggg aggcagttag 120
cacgggatac cgacctctac caagtacttg ttcactgcag aaggggtggc 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 aaatccttg atgcctccag 420
gcctttccca agccctcctg ccttcctggc gtggtggact ctactcaac 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 catcttgtag tttggctccc acagcacttt cctcagctgt atagctctgg 840
gttgacttgt gtgttgatgt gtctgtctcc ccaggatga gccccctcca agtcagggac 900
cttgctcat ttttctctc agtctctccc tggtagctgc tatgggatat gctcagtaca 960
cttgtgttta atgagtgggt aaatgggtgg cctacaccat cgggccgcag ctctgcacc 1020
acgattgtag taacaaaact ccacctggga acaggaaacc actggcaatt catggtgttc 1080
ctaaaccacg atttatgcca ggggaagcac tgaggagtgc ccttaggaa cttcccaaa 1140
gccatggaca gaagaccct gccatttggg ggggatggg gtttatggg agtaggagat 1200
gaggggacag tttactggg gagggacttc tctccattgt ctccctcaca aagcagactg 1260
ccaccccaaa gctgtccaag ccaaggctgg tgccaccatc acactcaagc aacaggttct 1320
gacatgctct tagggcccct cgaagtcagg ctgtccctga gggcttccag tgagctagca 1380
gagtggagac cattttccca cctccagatc ttcggaagga agaccagac cctccaagac 1440
tcacctgagg ggcgagaccc tcaacatttc atagtcttc agggaaacagt tgctgaaggg 1500
ggcggggggg tgggcacctg taagcttggt tttaaagatt ttaaagtgt ttaagatatc 1560
actgctcaaa taatattggt ctgnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1620
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnncg 1680
tttgaggaat taacaaagaa aaaaactaag acctagaatc tcaccacata accagctggt 1740
tcaatttttc catattccta tttagttggt gttcatatgc atacacaatt tttacatagc 1800
tataatcaca ggacaacaca aatatgtaat tagttctttt gaattagaaa aattacaaag 1860
ggcctatgta aaatgcaaac actccaaagc atataaagaa aacatgcagt ttccgcctc 1920
ccgtttccct tgccagaggt aaccacgggt agcagtttga tgaatagata gttttgtagt 1980
gttactgttc tctttttggc ctatcatcaa tacattcata tatagtcttg ataattacca 2040
cttagtgaa tctgatgaga gaaattctga catatatgta caaattaaat attgtctgtt 2160
ttattccagc ataaagtgt atagcatttc ccaaagcccc agtacagctg tattaatagg 2220
taaacttctc tagatagaac aaagcagtag tctagaatct cttggtataa tttcccttat 2280
ataataaaag tctctcccc aactctccca tctccctctt cctgtatgac tttgtttaaa 2340
cccatgtttc agcatttcta caatttgat tgtaactatc tgcatacaca gacaccacag 2400
ggctgacttt ggagttatgt ctttcgtnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2460
nnnnnnnnnn nnnaatcag aatttctctg gagcaaacac agccctgtgt ttgtggaaat 2520
ctcagtgcct tatgtattga ttcattttgc tgcag 2556
```

<210> 4

<211> 716

<212> ADN

<213> *Homo sapiens*

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

<400> 4

gtgagtggca	acccaacac	tgagtgaggg	tctgcaccag	cctgcctgtc	cctacccta	60
ccccttaatg	gtgttttagca	cagatgcagg	ctgtttcctg	tgcatttgcc	ccccagcag	120
gccctgtgct	gcttcgcatg	ctacagtggg	agtgggtctag	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	catttttagag	gtttccaggc	420
ccaaggctct	ttttctctc	gactcagggg	actgaagctt	gcattcccta	gtgtctcttt	480
ggtcagtgca	atatacctcc	aaaatctttt	ccatgtttta	tgtttgctaa	ggatctgtgg	540
ccctttaacg	ggctgtgtct	cccacagagc	ctcattacaa	cacattttta	ttgcgtgaac	600
agagtcacat	atctttcatt	cctcttatgt	ctgggatttc	agcaaacaca	gttgtagggg	660
gatgagcaat	ctaactcatt	cagtctgaga	accgtgctct	tttgcttctc	ttgtag	716

<210> 5
 <211> 2041
 <212> ADN
 <213> *Homo sapiens*

<220>
 <221> característica miscelânea
 <222> (559)...(869)
 <223> n= a, t, c ou g

<221> característica miscelânea
 <222> (888)...(1052)
 <223> n= a, t, c ou g

<221> característica miscelânea
 <222> (1102)...(1380)
 <223> n= a, t, c ou g

<221> característica miscelânea

<222> (1393)...(1692)

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

<221> intrão

<222> (1841)...(1907)

<223> FOXM1 int 3.2

<221> intrão

<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 acttcagtgt ttactaccta aagtcctatgg agggctcttca 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 gttggtcggg 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 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 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
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 nntaattccc tagtttctta atttctctga gccacctttc ttgctattga 1740
tactacctc acagccttac tctgcttttc tagccctga cagctatcta ggtcttttct 1800
ttatcacaat ctaagggttg catcagtcct tattcccgta gaatagatgg gtttatggct 1860
gaagggtgac gctctgcggt gtggagtgtc aggagagttg ccaagagggc tgcaaagaca 1920
ccagacgaag cctgtgctga gcacagtggg aggggcctga ggctgggttc cccatgtgtt 1980
tgaaggggtga tgtttctgaa tctaaagtag ctgataacca gttgtcttgc tcttcttcca 2040
g 2041
```

<210> 6
<211> 993
<212> ADN
<213> *Homo sapiens*

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

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

<221> característica miscelânea
<222> (249)...(397)
<223> n= a, t, c ou g

<221> característica miscelânea
<222> (442)...(749)
<223> n= a, t, c ou g

<400> 6

```
gtgaatgccc tgctttcctc taaatagggc ctaagttgga gggtgtcata gccatctcaa 60
aaggaaacaa gttctgctag tgatgctttc atttgatcag gggagagtta gaagccagcc 120
acccaattag tgacttgcac aaaaccagc 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 nnnnnnnntac 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 nnnnnnnnnng cctgccttag gctggagacc agaagctgag 780
ctaccagaac gtcttttcag aaagaagtta ttttggtttt tcagagtgcc cataaggctg 840
ctggtagctg taaccattct cctgggaggg gcagtgtgtc ggggtgtctt ttgtcatcag 900
tcaggaataa gtgtttttcc caatccggtc aaattgacca cgttggtggt aacttcatct 960
catttctctc ccacaatgcc tggccgccac cag 993
```

<210> 7

<211> 602

<212> ADN

<213> *Homo sapiens*

<220>

<221> intrão

<222> (0)...(0)

<223> FOXM1 int 5.1

<400> 7

```
gtaaggttct ttccctctgg ctcggggctt ggccttggtt tcctttcact gctcagcatg 60
gcttttagtg acagagacaa gatgtgatgt ggggaagggt ccctatggcc atgttttgtc 120
taggtgccag ccctagacac agaacaccct gaggtgcagg cacacacca cticcctccc 180
cttccatggg catcacaagg gcacactgag cagagcaggg cacagcaggg gagcatgctg 240
cagcagccac aagcgcagtg caccagcctc aggggaggca gttcgttcgc tcaactttgt 300
gcctagcttt tctttgccac gcatatagct acctgctctg gcatcccca ggggtgttga 360
ggacacgtgg gtgaagcggg agtgccactc tgccatcatg tgtctgtagg ccaccacct 420
gccactcat cacagttttg gagactgctc gcctacgtcc atcccctcag gttggcctcc 480
tctctctggg ctgtcattaa ctcaagcaca caccaccaga gcagctggtg gggttttgcc 540
atcccctctt taccttattg tggttaacata ggtttctttc tctccccatc tgccacaagc 600
ag 602
```

<210> 8

<211> 4656

<212> ADN

<213> *Homo sapiens*

<220>

<221> intrão

<222> (0)...(0)

<223> FOXM1 int 7.1

<221> característica miscelânea

<222> (316)...(615)

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

<221> característica miscelânea

<222> (626)...(930)

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

<221> característica miscelânea

<222> (1044)...(1259)

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

<221> característica miscelânea

<222> (1452)...(1802)

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

<221> característica miscelânea

<222> (2085)...(3116)

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

<221> característica miscelânea

<222> (3138)...(3433)

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

<221> característica miscelânea

<222> (3442)...(3728)

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

<221> característica miscelânea

<222> (3814)...(4081)

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

<221> característica miscelânea

<222> (4289)...(4371)

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

<400> 8

gtgggtgtcc	tattttcctc	tgaagagaga	ttctggccaa	ttaagaatgt	tggaccttca	60
gcttgcaaag	cactctgata	agtgttcctt	gagagcttat	aaatctagtt	gggtagaaaa	120
ggcataaaaa	cataggggaag	tgtaatagca	ttagaagagc	taaaaaggta	tttggattac	180
aatgtaagt	gtgtcagaag	gcccataaat	acctgatgag	cttgtaagaa	ttcagacaaa	240
agtgattgtg	atagatgggc	taggattatt	aaggaagata	cacaagggag	gcaggcctta	300
gaaagagatg	gatttntnnn	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	aatgggagct	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	nnnnnnnnnn	1260
tctagcttga	ggggaaggtg	agaagggtaa	attcagagcc	aacttggatc	agccatcaga	1320
tctgcactta	acactgttaa	agggttctgt	gagtagcggc	tgacatgtaa	ccaaagtga	1380
aagcttcccc	catccccctt	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	tttctccctt	ttcctttttg	1920
gctccctctc	caccctacca	tcaccacat	caccctact	cccaagataa	ctggttgata	1980
atztatgatg	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	nnnnnnnnnn	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	ttttaagagt	4140
tattgtagtc	aaatgtctct	cttctctttt	acagtttctg	ggtttccagt	cttggttaag	4200
aagggtcacc	gcaccctcag	attgtatatg	tagtctccta	gattctcttt	caggatttgt	4260
atgattttta	ggttttcatt	ttttttttan	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	4320
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	ntaaggttct	4380
tattttcatg	cattaaatct	ttgtatacag	tgtaagacaa	gcatgcaatt	ttatttcctc	4440
tcggatgaat	gctattataa	ttatgccact	acatactaca	taccgcgcatc	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> ADN
 <213> *Homo sapiens*

<220>
 <221> intrão
 <222> (0)...(0)

<223> MGB-PRAME int 2.1

<400> 9

gtaagttcga	gccctgattc	ctccgcttcc	ccgcaggggtg	accttgggct	tgtgcccccg	60
gcaccacccc	tgtcccgggt	ccctgttttc	tctctggaaa	tgggttgaag	accaaagaaa	120
ataatgtgcg	ccacttgggt	caccccgggc	cgcttggccc	ggaaaattgg	ccccagttga	180
ggagttgtgg	ctgtaaggat	gccttgaacc	gaggcggcgg	tgctcgtggt	tggagctctc	240
caggggtgggt	gcgcatttgt	aatgcggtgg	atgctctggg	actcggcccc	tctgaagggtg	300
ctggggggttg	gggacggccc	aggcagtggc	gtaggcgctc	taggaaggcg	ggagcagagg	360
cagaaatgtc	gctgcaagac	cgtagtcagg	gtccttgacc	acaggggtca	cttgtgacca	420
accacatggt	ctgttggttc	tcctgcccc	tggttcagcc	caggaaacac	tgggtgctcag	480
gtttggagcc	agagatttgc	actgaaagg	cgggattgag	tcgccagttg	tcagtttcct	540
cagcagtatt	tgcggagggt	ttcacaggag	gccgttgctt	cgtaaataatt	atacatgtat	600
tcttcttttt	ggag					614

<210> 10
 <211> 432
 <212> ADN
 <213> *Homo sapiens*

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

<400> 10

```
gtaaggggtga cctagcagct tgggtgtgggg ccctgggaac ctgagcagga tgcagctggg 60
gtcaggggagc atggagcgcc taaggctggg ccagaggctc tgatgggtgc cagcaaggaa 120
gttcaggggag gccttggggc tactgcaggg gtcactcttg gaatgggctt ctggacatgg 180
ggcactgatt aaaatgcaga ggtgtctgaa ggaacatgca cctgcttcct cctgggtggg 240
tgggaattgg ggaccaggaa ggatcccagg atcctagtgg gaaagggagc agctgatgcc 300
tgaagtacga agtaaaagtg cagatctaag gtggatgtct gtttggttct tacctacatt 360
atgagactca tgggtcttatt ttgagttgat cttaaagcat catctcagct aattacctgt 420
ttttccccac ag 432
```

<210> 11

<211> 3740

<212> ADN

<213> *Homo sapiens*

<220>

<221> intrão

<222> (0)...(0)

<223> MGB-STK15 int 1.1

<221> característica miscelânea

<222> (84)...(151)

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

<221> característica miscelânea

<222> (321)...(491)

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

<221> característica miscelânea

<222> (537)...(615)

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

<221> característica miscelânea

<222> (853)...(894)

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

<221> característica miscelânea

<222> (1259)...(1309)

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

<221> característica miscelânea

<222> (1337)...(2516)

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

<221> característica miscelânea

<222> (2521)...(3567)

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

<400> 11

gtacaagggg	tttgttgagt	ggtgttgaca	tgcgcgggag	gggtgggtgg	gcttcagatt	60
ggattttgtc	ctccgagatc	accnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	120
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nggtaagcgt	acggagaact	tgagctggg	180
gtgggtgtta	cagaggaaaa	gcaggagtgc	ggtttaacgg	gggccgcttt	agatagaata	240
gcctaagaag	gcccttgtcc	tggctggatg	agtgggtgaa	ttgatgaatg	agaacctcct	300
tcagagggcc	ttcccgggtcc	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	360
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	420
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	480
nnnnnnnnnn	ngggatgcag	accggtgcat	acaaatcgtc	tggggacggt	aaaatgnnnn	540
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	600
nnnnnnnnnn	nnnnncactg	tccttaactc	tcgtaatgtc	tcttcctctt	ccgtaacctt	660
ccttgctccct	tgaattaaac	gtttttcagc	aacctactca	gttcgctcct	cccttcattt	720
ctgcagacat	gcacaggtct	gagggaggaa	ggaataaacc	gtataaacct	cctgcgctat	780
tagcctaaca	gctttttctat	tcaaaatagt	aggacttctg	gtttgaactg	aatggatcct	840
gtgaaagtca	tcnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnagtctc	900
cttggcgttg	tctccagaat	tctggattag	aatcttattc	cattctgctt	gttattcaat	960
ttccctagaa	agaaaggtag	aataaattgg	agcaaattgc	tgtagcttct	gtcagaagaa	1020
tgttgaataa	atgttggttag	gcctatgtga	tctcattaga	ctgctactta	gaattgtaag	1080
ggaagttaaag	cattagagca	tgtgtgaaat	taaatatttg	attaacacaa	gtgtgcattt	1140
ccttgttgct	gtttatcaac	ttttacttac	ccactgtttt	tttataaggg	ctgcagcctg	1200
tagtctgggc	ctggcttcat	catggaatta	tttgcttaat	tgtaaaatgg	taatcttaan	1260
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	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

tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	1980
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2040
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2100
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2160
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2220
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2280
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2340
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2400
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2460
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2520
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2580
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2640
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2700
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2760
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2820
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2880
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2940
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	3000
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	3060
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	3120
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	3180
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	3240
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	3300
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	3360
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	3420
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	3480
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	3540
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	3600
agcttcatttt	ggacaaaaca	tgtaaaatcc	gggtgtgttgt	ggaggccttt	tgattgggga	3660
actgtaacgc	tgcctatcga	gcaacagcac	tttaagcagg	tggcttttgt	caaattaaag	3720
gttctttcttt	ttctttttcag					3740

<210> 12

<211> 1622

<212> ADN

<213> *Homo sapiens*

<220>

<221> intrão

<222> (0)...(0)

<223> MGB-STK15 int 2.1

<221> característica miscelânea

<222> (705)...(989)

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

<400> 12

```
gtaaattgaa taatctgtaa tctcattcac atttataaac ccacatggag gttggctctg 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 gtgtgctaatt ttcaggagta actgactttg tatttggaag 480
tcttaacacc ctctctttgt agagcactca taccgttgag ctggggatgg actttgaggc 540
tttcatttct agcacttgtc cctcacttac aatgagctgt tgaagctgaa ggaaatctca 600
tccctcctac cccttttagt ttgattagct gaggtgttta 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 nnnnnnnnnnt aataataata atacatttta gtagtaactt 1020
tgtgaagtgt ctacatttgt ttcctctttg tcagtttttt gctcaattcc attttgtcaa 1080
tacttggaata atgaaacatt ggttaatcaa tagtacagta ataagcttat tgtggaaaat 1140
cttcgatata tgaaaactta gactcttcta aaacttcata 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 tcagtgcacat tcattatttt gctatactag agaacaaact ttgtgaaatt 1560
ctcaatatat tcatcttttg ctttcatgaa tgccagaaag tttattttct cttccattct 1620
ag
```

<210> 13

<211> 1093

<212> ADN

<213> *Homo sapiens*

<220>

<221> intrão

<222> (0)...(0)

<223> MGB-STK15 int 4.1

<400> 13

```
gtaagctttc ttatttaca agttctgtac tgttctacta gaatatatta tttcgttgca 60
aatttcgttg tgggaactct ggggaaaaaa atgaggcctt tatttgcatt tagaggatat 120
aatgtttcc agatttccaa tcttaaaaaa aatggaattt tgtgtaatga ggtattttac 180
```

taggaactca	agtgccttaa	aaaatggctt	tcaaatttag	aaaaagcttg	tatgaatctt	240
ttatagaaat	gtgtggaagt	tcctctctgt	ccttagaaat	aaccactaca	tatggtttat	300
gcgtctgtac	ttttttattg	tacaaaagtg	caagttttta	aaaaatagaa	tatgttgcag	360
aactatatac	tcatatatga	ctgaggggtt	tgacagtatt	atagtttttag	ttctttattg	420
taaaggtttg	ctgtaatgtc	ttccccaggg	cttttctaaa	agcctcctct	cagtctctga	480
actatctgga	ctctagaatg	taccgggagg	agcgaggaat	gaaccacacag	actcttttgc	540
ttttagcggt	ctaacagagg	ctaagagtct	aaatccactg	gttctcatgc	cccagctagc	600
ctgtgggctc	catcccgtt	ccattagtaa	cagtggctct	gtctccacca	ccagagtggg	660
tctccacca	gagagaatta	gcacctctgg	gactggaggg	agcagctggg	gttagtttga	720
aacatgcccc	cagatgggtc	ggaagcattc	ctccctctct	ggtcacttat	cctttttgtg	780
gtcttcagcg	ttgtcatggc	cctgttcctc	tgagcatagt	acgggcttgg	gacatttccc	840
atagagtgtc	tcaggtctaa	aacccgagac	tgctccttgt	cactgactct	cacacctgac	900
ggcagctagg	gacgtcaggg	tttcatgtcg	tggcagctct	ttgatatgtg	ttattgcctt	960
ggttcttggc	gaggatgcat	attgagtga	gttggaatac	gaaattattt	gtagaatgtg	1020
tctgctactc	attgaaaatt	tgtagaaaa	gctttgtttt	cttcacattc	taaagtgttc	1080
aaattcctcc	tag					1093

<210> 14

<211> 18

<212> ADN

<213> Sequência Artificial

<220>

<223> iniciador de sentido directo

<400> 14

agcgcctgtt ccgatctg

18

<210> 15

<211> 24

<212> ADN

<213> Sequência Artificial

<220>

<223> iniciador de sentido reverso

<400> 15

aaccaaagtt ctctgctgaa aacc

24

<210> 16
<211> 15
<212> ADN
<213> Sequência Artificial

<220>
<223> sonda

<400> 16

ccctgaagca gcaac

15

<210> 17
<211> 22
<212> ADN
<213> Sequência Artificial
<220>
<223> iniciador de sentido directo

<400> 17

ctgttgctgt gtgatgctgt ca

22

<210> 18
<211> 20
<212> ADN
<213> Sequência Artificial

<220>
<223> iniciador de sentido reverso

<400> 18
cctcagccac tcccttgatc 20

<210> 19
<211> 15
<212> ADN
<213> Sequência Artificial

<220>
<223> sonda

<400> 19
tcagggcata agcct 15

<210> 20
<211> 19
<212> ADN
<213> Sequência Artificial

<220>
<223> iniciador de sentido directo

<400> 20
tccccttgcc tttggagaa 19

<210> 21
<211> 19
<212> ADN
<213> Sequência Artificial

<220>

<223> iniciador de sentido reverso

<400> 21

aaaggcctgg aggcattaa

19

<210> 22

<211> 13

<212> ADN

<213> Sequência Artificial

<220>

<223> sonda

<400> 22

cagcccaaat cct

13

<210> 23

<211> 25

<212> ADN

<213> Sequência Artificial

<220>

<223> iniciador de sentido directo

<400> 23

cttaatggtg ttttagcacag atgca

25

<210> 24

<211> 20

<212> ADN

<213> Sequência Artificial

<220>

<223> iniciador de sentido reverso

<400> 24

ccactgtagc atgcgaagca

20

<210> 25

<211> 16

<212> ADN

<213> Sequência Artificial

<220>

<223> sonda

<400> 25

caaatgcaca ggaaac

16

<210> 26

<211> 19

<212> ADN

<213> Sequência Artificial

<220>

<223> iniciador de sentido directo

<400> 26

gctctgcggt gtggagtgt

19

<210> 27
<211> 21
<212> ADN
<213> Sequência Artificial

<220>
<223> iniciador de sentido reverso

<400> 27
cacaggcttc gtctggtgtc t

21

<210> 28
<211> 24
<212> ADN
<213> Sequência Artificial

<220>
<223> sonda

<400> 28
tgcagccctc ttggcaactc tcct

24

<210> 29
<211> 23
<212> ADN
<213> Sequência Artificial

<220>
<223> iniciador de sentido directo

<400> 29
aaaatgctgt ctgtgagcct cat

23

<210> 30
<211> 21
<212> ADN
<213> Sequência Artificial

<220>
<223> iniciador de sentido reverso

<400> 30
aaccccctgcc cactagaaat g

21

<210> 31
<211> 30
<212> ADN
<213> Sequência Artificial

<220>
<223> sonda

<400> 31
acccaagatg tcatctcctg tagcgtcaca

30

<210> 32
<211> 25
<212> ADN
<213> Sequência Artificial

<220>
<223> iniciador de sentido directo

<400> 32
aatagatggg tttatggctg aaggt

25

<210> 33
<211> 23
<212> ADN
<213> Sequência Artificial

<220>
<223> iniciador de sentido reverso

<400> 33
ctcttggcaa ctctcctgac act

23

<210> 34
<211> 13
<212> ADN
<213> Sequência Artificial

<220>
<223> sonda

<400> 34
ccgcagagcc gtc

13

<210> 35
<211> 25
<212> ADN
<213> Sequência Artificial

<220>
<223> iniciador de sentido directo

<400> 35
ccatctcaaa aggaaacaag ttctg

25

<210> 36
<211> 21
<212> ADN
<213> Sequência Artificial

<220>
<223> iniciador de sentido reverso

<400> 36
gggtggctgg cttctaactc t

21

<210> 37
<211> 24
<212> ADN
<213> Sequência Artificial

<220>
<223> sonda

<400> 37
ccctgatcaa atgaaagcat cact

24

<210> 38
<211> 20
<212> ADN
<213> Sequência Artificial

<220>
<223> iniciador de sentido directo

<400> 38
agaagccagc cacccaatta

20

<210> 39
<211> 26
<212> ADN
<213> Sequência Artificial

<220>
<223> iniciador de sentido reverso

<400> 39
tgtgtcattt ggtatttgtc aagtgt

26

<210> 40
<211> 26
<212> ADN
<213> Sequência Artificial

<220>
<223> sonda

<400> 40
tgacttgcac aaaaccagt gaatta

26

<211> 24
<212> ADN
<213> Sequência Artificial
<220>
<223> iniciador de sentido directo

<400> 41
tggacagaga caagatgtga tgtg

24

<210> 42
<211> 22
<212> ADN
<213> Sequência Artificial

<220>
<223> iniciador de sentido reverso

<400>.42
gctggcacct agacaaaaca tg 22

<211> 15
<212> ADN
<213> Sequência Artificial

<220>
<223> sonda

<400> 43
ccatagggac ccttc 15

<210> 44
<211> 25
<212> ADN
<213> Sequência Artificial

<220>
<223> iniciador de sentido directo

<400> 44
ggtgtcctat tttcctctga agaga 25

<210> 45
 <211> 21
 <212> ADN
 <213> Sequência Artificial

 <220>
 <223> iniciador de sentido reverso

<400> 45
tgcaagctga aggtccaaca t 21

<210> 46
 <211> 15
 <212> ADN
 <213> Sequência Artificial

<220>
 <223> sonda

<400> 46
ttctggccaa ttaag 15

<210> 47
 <211> 21
 <212> ADN
 <213> Sequência Artificial

<220>
 <223> iniciador de sentido directo

<400> 47
tcattcatgt ggccgtagca t 21

<210> 48
<211> 19
<212> ADN
<213> Sequência Artificial

<220>
<223> iniciador de sentido reverso

<400> 48
ggtggagagg gagccaaaa 19

<210> 49
<211> 16
<212> ADN
<213> Sequência Artificial

<220>
<223> sonda

<400> 49
cctgtttggg ttttca 16

<210> 50
<211> 23
<212> ADN
<213> Sequência Artificial

<220>
<223> iniciador de sentido directo

<400> 50

agaggatcag aacccgtatg tga

23

<210> 51

<211> 24

<212> ADN

<213> Sequência Artificial

<220>

<223> iniciador de sentido reverso

<400> 51

gggaaacaga gataaggtga acca

24

<210> 52

<211> 15

<212> ADN

<213> Sequência Artificial

<220>

<223> sonda

<400> 52

tgtgcgtcct gtctg

15

<210> 53

<211> 19

<212> ADN

<213> Sequência Artificial

<220>

<223> iniciador de sentido directo

<400> 53

gggtgacctt gggcttgtg

19

<210> 54

<211> 23

<212> ADN

<213> Sequência Artificial

<220>

<223> iniciador de sentido reverso

<400> 54

cttcaaccca tttccagaga gaa

23

<210> 55

<211> 14

<212> ADN

<213> Sequência Artificial

<220>

<223> sonda

<400> 55

cccgggtccc tggtt

14

<210> 56
<211> 21
<212> ADN
<213> Sequência Artificial

<220>
<223> iniciador de sentido directo

<400> 56
agggtgacct agcagcttgg t 21

<210> 57
<211> 18
<212> ADN
<213> Sequência Artificial

<220>
<223> iniciador de sentido reverso

<400> 57
gcctctggcc cagcctta 18

<210> 58
<211> 16
<212> ADN
<213> Sequência Artificial

<220>
<223> sonda

<400> 58

tccctgaccc cagctg

16

<210> 59

<211> 25

<212> ADN

<213> Sequência Artificial

<220>

<223> iniciador de sentido directo

<400> 59

cgtaatgtct cttcctcttc cgtaa

25

<210> 60

<211> 24

<212> ADN

<213> Sequência Artificial

<220>

<223> iniciador de sentido reverso

<400> 60

acgaactgag taggttgctg aaaa

24

<210> 61

<211> 16

<212> ADN

<213> Sequência Artificial

<220>

<223> sonda

<400> 61

tcaagggaca aggaag

16

<210> 62

<211> 26

<212> ADN

<213> Sequência Artificial

<220>

<223> iniciador de sentido directo

<400> 62

cattcacatt tataaaccca catgga

26

<210> 63

<211> 23

<212> ADN

<213> Sequência Artificial

<220>

<223> iniciador de sentido reverso

<400> 63

aatccaaagt aaaggcggaa aga

23

<210> 64

<211> 16

<212> ADN

<213> Sequência Artificial

<220>

<223> sonda

<400> 64

tggtcttgtc gggaat

16

<210> 65

<211> 20

<212> ADN

<213> Sequência Artificial

<220>

<223> iniciador de sentido directo

<400> 65

gcgaggaatg aaccacaga

20

<210> 66

<211> 25

<212> ADN

<213> Sequência Artificial

<220>

<223> iniciador de sentido reverso

<400> 66

gcatgagaac cagtggattt agact

25

<210> 67

<211> 16

<212> ADN

<213> Sequência Artificial

<220>

<223> sonda

<400> 67

cgctaaaagc aaaaga

16

Lisboa, 5 de Dezembro de 2008

REIVINDICAÇÕES

1. Método para monitorizar a expressão genética numa amostra biológica, compreendendo:
 - (a) proporcionar um polinucleótido complementar a uma sequência de ARN intrónico num gene alvo, em que a expressão da referida sequência de ARN intrónico correlaciona-se com a expressão de uma sequência de ARN exónico no referido gene;
 - (b) hibridar o referido polinucleótido com a referida sequência de ARN intrónico para formar um complexo polinucleótido-ARN intrónico; e
 - (c) detectar o complexo polinucleótido-ARN intrónico.
2. Método da reivindicação 1, em que a referida sequência de ARN intrónico é seleccionada por identificação de sequências intrónicas que são co-expressas com o ARNm do referido gene alvo e seleccionando uma sequência de ARN intrónico possuindo o maior coeficiente de correlação para a referida co-expressão.
3. Método da reivindicação 1, em que a referida sequência de ARN intrónico tem, pelo menos, 50 bases de nucleótido de comprimento.
4. Método da reivindicação 1, em que a referida sequência de ARN intrónico tem, pelo menos, 55 bases de nucleótido de comprimento.

5. Método da reivindicação 1, em que a referida sequência de ARN intrónico tem, pelo menos, 60 bases de nucleótido de comprimento.
6. Método da reivindicação 1, em que a referida amostra biológica é uma amostra de tecido.
7. Método da reivindicação 6, em que o referido tecido é um tecido de tumor.
8. Método da reivindicação 7, em que o referido tumor é cancro.
9. Método da reivindicação 8, em que o referido cancro é seleccionado do grupo consistindo de cancro da mama, cancro do cólon, cancro do pulmão, cancro da próstata, cancro hepatocelular, cancro gástrico, cancro pancreático, cancro do colo do útero, cancro ovárico, cancro do fígado, cancro da bexiga, cancro do tracto urinário, cancro da tiróide, cancro renal, carcinoma, melanoma e cancro do cérebro.
10. Método da reivindicação 6, em que a referida amostra de tecido é uma amostra de tecido fixa, embebida em cera.
11. Método da reivindicação 10, em que a referida amostra de tecido é uma amostra de tecido fixa em formalina, embebida em parafina.
12. Método da reivindicação 10, em que o referido ARN exónico é fragmentado.

13. Método da reivindicação 1, em que a referida amostra biológica é um fluido biológico.
14. Método da reivindicação 1, em que a referida hibridação é efectuada sob condições de restringência.
15. Método da reivindicação 1 compreendendo ainda o passo de quantificar a expressão do referido ARN intrónico.
16. Método da reivindicação 1, em que o referido polinucleótido é um oligonucleótido de cadeia simples.
17. Método da reivindicação 16, em que o referido oligonucleótido de cadeia simples é uma sonda de PCR.
18. Método da reivindicação 17, em que a referida sonda de PCR é concebida com base numa sequência única na referida sequência de ARN intrónico.
19. Método da reivindicação 16, em que o referido oligonucleótido de cadeia simples é um iniciador de PCR.
20. Método da reivindicação 19, em que o referido iniciador de PCR é concebido com base na sequência única na referida sequência de ARN intrónico.
21. Método da reivindicação 1, em que é monitorizada a expressão de mais de um gene alvo.
22. Método da reivindicação 21 compreendendo a monitorização simultânea de, pelo menos, 50 genes alvo.

23. Método da reivindicação 21 compreendendo a monitorização simultânea de, pelo menos, 100 genes alvo.
24. Método da reivindicação 21 compreendendo a monitorização simultânea de, pelo menos, 500 genes alvo.
25. Método da reivindicação 21 compreendendo a monitorização simultânea de, pelo menos, 10000 genes alvo.
26. Método da reivindicação 25, em que as sequências de ARN intrónico correspondentes a uma pluralidade dos referidos genes alvo são apresentadas como uma matriz imobilizada numa superfície sólida.
27. Método da reivindicação 1, em que o referido gene alvo é seleccionado dos genes listados na Figura 6.
28. Método de preparação de uma molécula de oligonucleótido de cadeia simples para amplificação de um gene alvo compreendendo:
 - (a) a identificação de, pelo menos, uma sequência de intrão num referido gene alvo, cuja expressão correlaciona com a expressão de uma sequência de ARN exónico no referido gene alvo; e
 - (b) a preparação de uma molécula de oligonucleótido de cadeia simples que corresponde a, pelo menos, uma porção da sequência de intrão transcrita.

29. Método da reivindicação 28 compreendendo a identificação de sequências repetidas numa referida sequência de intrão antes da preparação da referida molécula de oligonucleótido de cadeia simples.
30. Método da reivindicação 29, em que as referidas sequências repetidas são mascaradas antes da preparação da referida molécula de oligonucleótido.
31. Método da reivindicação 28, em que a referida molécula de oligonucleótido de cadeia simples é um iniciador de PCR.
32. Método da reivindicação 31, em que o referido iniciador de PCR é um iniciador de sentido directo concebido para compreender sequências 5' de uma sequência alvo na referida sequência de intrão transcrita.
33. Método da reivindicação 31, em que o referido iniciador de PCR é um iniciador reverso concebido para complementar sequências 5' de uma sequência alvo a montante do iniciador de sentido directo na referida sequência de intrão transcrita.
34. Método da reivindicação 32 ou 33, em que a referida sequência alvo tem, pelo menos, 50 bases de nucleótido de comprimento.
35. Método da reivindicação 32 ou 33, em que a referida sequência alvo tem, pelo menos, 55 bases de nucleótido de comprimento.

36. Método da reivindicação 32 ou 33, em que a referida sequência alvo tem, pelo menos, 60 bases de nucleótido de comprimento.
37. Método da reivindicação 31, em que o referido iniciador de PCR tem 17-30 bases de nucleótido em comprimento.
38. Método da reivindicação 31, em que o referido iniciador de PCR contém cerca de 20% a 80% de bases G+C.
39. Método da reivindicação 31, em que a temperatura de fusão (T_m) do referido iniciador de PCR está entre cerca de 50 °C e cerca de 70 °C.
40. Método da reivindicação 28, em que a referida molécula de oligonucleótido de cadeia simples é uma sonda de PCR.
41. Método da reivindicação 40, em que a referida sonda de PCR é concebida para compreender ou complementar uma porção interna de uma sequência alvo na sequência de intrão transcrita.
42. Método da reivindicação 41, em que a referida sonda de PCR é marcada com um corante fluorescente repórter e um corante fluorescente extintor.
43. Método da reivindicação 28, em que o referido gene alvo é seleccionado dos genes listados na Figura 6.
44. Método para amplificar ARN intrónico numa amostra de tecido fixa, embebida em parafina, representando, pelo menos, um gene de interesse, compreendendo os passos de:

- (a) colocar em contacto o ADN obtido por transcrição reversa do ARN intrónico, cuja expressão correlaciona com a expressão de um correspondente RNA exónico com, pelo menos, um conjunto de iniciadores de PCR e sondas correspondentes ao referido ARN intrónico; e
 - (b) efectuar amplificação por PCR.
45. Método da reivindicação 44, em que os referidos iniciadores de PCR e sonda são concebidos com base na sequência única no referido ARN intrónico.
46. Método da reivindicação 44, em que a referida amostra compreende o ARN fragmentado que representa múltiplos genes de interesse.
47. Método da reivindicação 46, em que a referida amostra é colocada em contacto com uma mistura de iniciadores de PCR e sondas concebidas com base em sequências únicas nos intrões, cuja expressão correlaciona com a expressão de correspondentes exões, presentes nos referidos genes de interesse.
48. Método da reivindicação 47, em que a referida mistura compreende, pelo menos, um dos conjuntos iniciador com base em intrão/sonda apresentados na Figura 2.
49. Método da reivindicação 47, em que a referida mistura compreende, pelo menos, um iniciador de sentido directo ou reverso ou sonda apresentados na Figura 2.

50. Método da reivindicação 46, em que a referida amostra de tecido é de uma biopsia de tumor.
51. Método da reivindicação 50, em que a referida biopsia de tumor é obtida de um doente humano.
52. Método da reivindicação 51, em que o referido tumor é cancro da mama.
53. Método da reivindicação 51, em que o referido tumor é cancro do pulmão.
54. Método da reivindicação 51, em que o referido tumor é cancro colorrectal.
55. Método da reivindicação 51 compreendendo ainda o passo de determinar os níveis de expressão dos transcritos de ARN dos referidos genes de interesse ou dos seus produtos de expressão.
56. Método da reivindicação 55 em que expressão diferencial dos referidos transcritos de ARN ou dos seus produtos é correlacionada com a resposta prevista do doente ao tratamento ou sobrevivência do doente.
57. Método da reivindicação 44, em que o referido gene de interesse é seleccionado dos genes listados na Figura 6.
58. Método da reivindicação 52, em que o tumor é cancro da mama invasivo e o método compreende

(1) determinar os níveis de expressão dos transcritos de ARN ou dos produtos de expressão de um gene ou conjunto de genes seleccionado do grupo consistindo em:

- (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; e
- (p) CEGP1, PRAME, hENT1, XIAP, Contig51037, ErbB3, DPYD, NFkBp65, ID1, TS

na referida amostra;

- (2) submeter os dados obtidos no passo (a) a análise estatística; e
- (3) determinar se a probabilidade da sobrevivência a longo prazo do referido doente, sem recaída de cancro de mama, aumentou ou diminuiu.

59. Método da reivindicação 58, em que os níveis de expressão dos referidos transcritos de ARN dos ou seus produtos de expressão são normalizados contra os níveis de expressão de todos os transcritos de ARN ou dos seus produtos de expressão na referida amostra de tecido de cancro da mama ou de um conjunto de referência de transcritos de ARN ou dos seus produtos.

60. Método da reivindicação 52, em que o tumor é o cancro da mama invasivo positivo para o receptor de estrogénio (ER) e o método compreende

- (1) determinar os níveis de expressão dos transcritos de ARN ou dos produtos de expressão de um gene ou conjunto de genes seleccionados do grupo consistindo em:
 - (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) FBX05, PRAME, IGFBP3, p27, GSTM3, hENT1, XIAP, FGF2, TS, PTEN;
- (y) GUS, HIA1A, VEGFC, GSTM3, DPYD, hENT1, EBX05, CA9, CYP, KRT18; e
- (z) Bclx, Bcl2, CENT1, Contig51037, HLAG, CD9, ID1, BRCA1, BIN1, HBEGF;

(2) submeter os dados obtidos no passo (1) a análise estatística; e

(3) determinar se a probabilidade de sobrevivência a longo prazo do referido doente, sem a recaída de cancro da mama aumentou ou diminuiu.

61. Método da reivindicação 60, em que os níveis de expressão dos referidos transcritos de ARN ou dos seus produtos de expressão são normalizados contra os níveis de expressão de todos os transcritos de ARN ou dos seus produtos de expressão na referida amostra de tecido de cancro da mama

ou de um conjunto de referência de transcritos de ARN ou dos seus produtos.

62. Método da reivindicação 52, em que o tumor é cancro da mama e o método compreende

(1) determinar os níveis de expressão dos transcritos de ARN ou dos produtos de expressão de um gene ou conjunto de genes seleccionados do grupo consistindo em:

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; BAG1; pS2; WISP1; HNF3A; e NFkBp65, normalizados contra os níveis de expressão de todos os transcritos de ARN ou dos seus produtos na referida amostra, ou de um conjunto de referência de transcritos de ARN ou dos seus produtos de expressão;

(2) submeter os dados obtidos no passo (a) a análise estatística; e

(3) determinar se a probabilidade de sobrevivência a longo prazo do referido doente, sem a recaída de cancro da mama aumentou ou diminuiu.

63. Método da reivindicação 62, em que os níveis de expressão dos referidos transcritos de ARN ou dos seus produtos de expressão são normalizados contra os níveis de expressão de todos os transcritos de ARN ou dos seus produtos de expressão na referida amostra de tecido de cancro da mama

ou de um conjunto de referência de transcritos de ARN ou dos seus produtos.

64. Método da reivindicação 52, em que o tumor é cancro da mama invasivo e o método compreende determinar os níveis de expressão de transcritos de ARN ou dos produtos de expressão de um gene ou conjunto de genes do grupo consistindo em:

- (a) p53BP2, Bcl2, BAD, EPHX1, PDGFR , DIABLO, XIAP, YB1, CA9 e KRT8;
- (b) GRB7, CD68, TOP2A, Bcl2, DIABLO, CD3, ID1, PPM1D, MCM6 e WISP1;
- (c) PR, p53BP2, PRAME, DIABLO, CTSL, IGFBP2, TIMP1, CA9, MMP9 e COX2;
- (d) CD68, GRB7, TOP2A, Bcl2, DIABLO, CD3, ID1, PPM1D, MCM6 e WISP1;
- (e) Bcl2, p53BP2, BAD, EPHX1, PDGFR , DIABLO, XIAP, YB1, CA9 e KRT8;
- (f) KRT14, KRT5, PRAME, p53BP2, GUS1, AIB1, MCM3, CONE1, MCM6 e ID1;
- (g) PRAME, p53BP2, EstR1, DIABLO, CTSL, PPM1D, GRB7, DAPK1, BBC3 e VEGFB;
- (h) CTSL2, GRB7, TOP2A, CCNB1, Bcl2, DIABLO, PRAME, EMS1, CA9 e EpCAM;
- (i) EstR1, p53BP2, PRAME, DIABLO, CTSL, PPM1D, GRB7, DAPK1, BBC3 e VEGFB;
- (j) Chk1, PRAME, p53BP2, GRB7, CA9, CTSL, CCNB1, TOP2A, tamanho de tumor e IGFBP2;
- (k) IGFBP2, GRB7, PRAME, DIABLO, CTSL, β -Catenina, PPM1D, Chk1, WISP1 e LOT1;

- (l) HER2, p53BP2, Bcl2, DIABLO, TIMP1, EPHX1, TOP2A, TRAIL, CA9 e AREG;
- (m) BAG1, p53BP2, PRAME, IL6, CCNB1, PAI1, AREG, tamanho de tumor, CA9 e Ki67;
- (n) CEGP1, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, STK15 e AKT2 e FGF 18;
- (o) STK15, p53BP2, PRAME, IL6, CCNE1, AKT2, DIABLO, cMet, CCNE2 e COX2;
- (p) KLK10, EstR1, p53BP2, PRAME, DIABLO, CTSL, PPM1D, GRB7, DAPK1 e BBC3;
- (q) AIB1, p53BP2, Bcl2, DIABLO, TIMP1, CD3, p53, CA9, GRB7 e EPHX1
- (r) BBC3, GRB7, CD68, PRAME, TOP2A, CCNB1, EPHX1, CTSL GSTM1 e APC;
- (s) CD9, GRB7, CD68, TOP2A, Bcl2, CCNB1, CD3, DIABLO, ID1 e PPM1D;
- (t) EGFR, KRT14, GRB7, TOP2A, CCNB1, CTSL, Bcl2, TP, KLK10 e CA9;
- (u) HIF1 α , PR, DIABLO, PRAME, Chk1, AKT2, GRB7, CCNE1, TOP2A e CCNB1;
- (v) MDM2, p53BP2, DIABLO, Bcl2, AIB1, TIMP1, CD3, p53, CA9 e HER2;
- (w) MYBL2, p53BP2, PRAME, IL6, Bcl2, DIABLO, CCNE1, EPHX1, TIMP1 e CA9;
- (x) p27, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, STK15, AKT2 e ID1;
- (y) RAD51, GRB7, CD68, TOP2A, CIAP2, CCNB1, BAG1, IL6, FGFR1 e p53BP2;
- (z) SURV, GRB7, TOP2A, PRAME, CTSL, GSTM1, CCNB1, VDR, CA9; e CCNE2;
- (aa) TOP2B, p53BP2, DIABLO, Bcl2, TIMP1, AIB1, CA9, p53, KRT8 e BAD;

(ab) ZNF217, GRB7, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, APC4 e β -Catenina,

numa amostra de tecido de cancro da mama obtido do referido doente,

(2) submeter os dados obtidos no passo (a) a análise estatística; e

(3) determinar se a probabilidade da referida sobrevivência a longo prazo aumentou ou diminuiu.

65. Método da reivindicação 64, em que os níveis de expressão dos referidos transcritos de ARN ou dos seus produtos de expressão são normalizados contra os níveis de expressão de todos os transcritos de ARN ou dos seus produtos de expressão na referida amostra de tecido de cancro da mama ou de um conjunto de referência de transcritos de ARN ou dos seus produtos.

66. Método de prognóstico de sobrevivência sem doença a longo prazo de um doente com cancro utilizando ARN intrónico para medir a expressão genética utilizando uma matriz compreendendo uma pluralidade de polinucleótidos, hibridando com genes alvo de interesse, imobilizados numa superfície sólida, em que, pelo menos, um dos referidos polinucleótidos compreende uma sequência com base em intrão cuja expressão correlaciona com a expressão de uma sequência de exão correspondente.

67. Método da reivindicação 66, em que todos os referidos polinucleótidos compreendem sequências de intrão.

68. Método da reivindicação 66 compreendendo, pelo menos, um dos amplicões apresentados nas Figuras 1A-M, ou o seu complementar.
69. Método da reivindicação 66 compreendendo dois ou mais dos amplicões apresentados nas Figuras 1A-M ou o seu complementar.
70. Método da reivindicação 66 compreendendo todos os amplicões apresentados nas Figuras 1A-M, ou o seu complementar.
71. Método da reivindicação 66 compreendendo a utilização de sequências de polinucleótidos com base em intrão que hibridam com, pelo menos, um gene de interesse seleccionado do grupo consistindo em: FOXM1, PRAME, Bcl2, STK15, CEGP1, Ki-67, GSTM1, PR, BBC3, NME1, SURV, GATA3, TFRC, YB-1, DPYD, CA9, Contig51037, RPS6K1 e Her2, em que pelo menos, 80% das sequências na referida matriz são com base em intrão.
72. Método da reivindicação 71 compreendendo a utilização de sequências de polinucleótidos com base em intrão que hibridam com, pelo menos, 5 dos referidos genes.
73. Método da reivindicação 71 compreendendo a utilização de sequências de polinucleótidos com base em intrão que hibridam com, pelo menos, 10 dos referidos genes.
74. Método da reivindicação 71 compreendendo a utilização de sequências de polinucleótidos com base em intrão que hibridam com todos os referidos genes.

75. Método da reivindicação 66 compreendendo a utilização de sequências de polinucleótidos com base em intrão que hibridam com, pelo menos, um gene de interesse seleccionado do grupo consistindo em: 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, BAG1, CYP3A4, EpCAM, VEGFC, pS2, hENT1, WISP1, HNF3A, NFkBp65, BRCA2, EGFR, TK1, VDR, Contig51037, pENT1, EPHX1, IF1A, CDH1, HIF1 α , IGFBP3, CTSB, Her2 e DIABLO.
76. Método da reivindicação 75 compreendendo a utilização de sequências de polinucleótidos com base em intrão que hibridam com, pelo menos, 5 dos referidos genes.
77. Método da reivindicação 75 compreendendo a utilização de sequências de polinucleótidos com base em intrão que hibridam com, pelo menos, 10 dos referidos genes.
78. Método de matriz da reivindicação 75 compreendendo a utilização de sequências de polinucleótidos com base em intrão que hibridam com todos os referidos genes.
79. Método da reivindicação 66 compreendendo a utilização de sequências de polinucleótidos com base em intrão que hibridam com, pelo menos, um conjunto de genes seleccionado do grupo consistindo em:
- (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; e
- (p) CEGP1, PRAME, hENT1, XIAP, Contig51037, ErbB3, DPYD, NFKBp65, ID1, TS.

80. Método da reivindicação 66 compreendendo a utilização de sequências de polinucleótidos com base em intrão que

hibridam com, pelo menos, um conjunto de genes seleccionado do grupo consistindo em:

- (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, CTSB, BAD, DR5, DCR3, XIAP;
- (l) DIABLO, Ki67, CENT1, 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, HIAIA, VEGFC, GSTM3, DPYD, hENT1, EBXO5, CA9, CYP, KRT18; e
- (z) Bclx, Bcl2, hENT1, Contig51037, HLAG, CD9, ID1, BRCA1, BIN1, HBEGF.

81. Método da reivindicação 66 compreendendo a utilização de sequências de polinucleótidos com base em intrão que hibridam com, pelo menos, um conjunto de genes seleccionado do grupo consistindo em:

- (a) p53BP2, Bcl2, BAD, EPHX1, PDGFR β , DIABLO, XIAP, YB1, CA9 e KRT8;

- (b) GRB7, CD68, TOP2A, Bcl2, DIABLO, CD3, ID1, PPM1D, MCM6 e WISP1;
- (c) PR, p53BP2, PRAME, DIABLO, CTSL, IGFBP2, TIMP1, CA9, MMP9 e COX2;
- (d) CD68, GRB7, TOP2A, Bcl2, DIABLO, CD3, ID1, PPM1D, MCM6 e WISP1;
- (e) Bcl2, p53BP2, BAD, EPHX1, PDGFR β , DIABLO, XIAP, YB1, CA9 e KRT8;
- (f) KRT14, KRT5, PRAME, p53BP2, GUS1, AIB1, MCM3, CCNE1, MCM6 e ID1;
- (g) PRAME, p53BP2, EstR1, DIABLO, CTSL, PPM1D, GRB7, DAPK1, BBC3 e VEGFB;
- (h) CTSL2, GRB7, TOP2A, CCNB1, Bcl2, DIABLO, PRAME, EMS1, CA9 e EpCAM;
- (i) EstR1, p53BP2, PRAME, DIABLO, CTSL, PPM1D, GRB7, DAPK1, BBC3 e VEGFB;
- (j) Chk1, PRAME, p53BP2, GRB7, CA9, CTSL, CCNB1, TOP2A, tamanho de tumor e IGFBP2;
- (k) IGFBP2, GRB7, PRAME, DIABLO, CTSL, β -Catenina, PPM1D, Chk1, WISP1 e LOT1;
- (l) HER2, p53BP2, Bcl2, DIABLO, TIMP1, EPHX1, TOP2A, TRAIL, CA9 e AREG;
- (m) BAG1, p53BP2, PRAME, IL6, CCNB1, PAI1, AREG, tamanho de tumor, CA9 e Ki67;
- (n) CEGP1, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, STK15 e AKT2 e FGF18;
- (o) STK15, p53BP2, PRAME, IL6, CCNE1, AKT2, DIABLO, cMet, CCNE2 e COX2;
- (p) KLK10, EstR1, p53BP2, PRAME, DIABLO, CTSL, PPM1D, GRB7, DAPK1 e BBC3;
- (q) AIB1, p53BP2, Bcl2, DIABLO, TIMP1, CD3, p53, CA9, GRB7 e EPHX1

- (r) BBC3, GRB7, CD68, PRAME, TOP2A, CCNB1, EPHX1, CTSL, GSTM1 e APC;
- (s) CD9, GRB7, CD68, TOP2A, Bcl2, CCNB1, CD3, DIABLO, ID1 e PPM1D;
- (t) EGFR, KRT14, GRB7, TOP2A, CCNB1, CTSL, Bcl2, TP, KLK10 e CA9;
- (u) HIF1 α , PR, DIABLO, PRAME, Chk1, AKT2, GRB7, CCNE1, TOP2A e CCNB1;
- (v) MDM2, p53BP2, DIABLO, Bcl2, AIB1, TIMP1, CD3, p53, CA9 e HER2;
- (w) MYBL2, p53BP2, PRAME, IL6, Bcl2, DIABLO, CCNE1, EPHX1, TIMP1 e CA9;
- (x) p27, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, STK15, AKT2 e ID1;
- (y) RAD51, GRB7, CD68, TOP2A, CIAP2, CCNB1, BAG1, IL6, FGFR1 e p53BP2;
- (z) SURV, GRB7, TOP2A, PRAME, CTSL, GSTM1, CCNB1, VDR, CA9; e CCNE2;
- (aa) TOP2B, p53BP2, DIABLO, Bcl2, TIMP1, AIB1, CA9, p53, KRT8 e BAD; e
- (ab) ZNF217, GRB7, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, APC4 e β -Catenina.

82. Método da reivindicação 66 compreendendo sequências de polinucleótidos com base em intrão que hibridam com, pelo menos, um gene seleccionado do grupo consistindo em: CD68; CTSL; FBXO5; SURV; CCNB1; MCM2; Chk1; MYBL2; HIF1A; cMET; EGFR; TS; STK15, IGFR1; BC12; 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; e STK15.

83. Método da reivindicação 66 compreendendo a utilização de sequências de polinucleótidos com base em intrão correspondentes a, pelo menos, um gene seleccionado dos genes listados na Figura 6.
84. Método da reivindicação 66 compreendendo a utilização de sequências de polinucleótidos com base em intrão correspondentes a uma pluralidade de genes seleccionados dos genes listados na Figura 6.
85. Método da reivindicação 66 compreendendo a utilização de ambas as sequências de polinucleótido com base em intrão e com base em exão.
86. Método da reivindicação 66 compreendendo a utilização de ambas as sequências de polinucleótido com base em intrão e com base em exão que hibridam com o mesmo gene alvo de interesse.
87. Método da reivindicação 66, em que a referida matriz compreende, pelo menos, 100 genes.
88. Método da reivindicação 87, em que a referida matriz compreende, pelo menos, 100 genes numa secção de 100 μ .
89. Método da reivindicação 87, em que a referida matriz compreende, pelo menos, 150 genes numa secção de 100 μ .
90. Utilização de, pelo menos, um polinucleótido compreendendo uma sequência com base em intrão, em que a expressão do intrão correlaciona com a expressão de uma sequência de

exão correspondente, para medir a expressão genética utilizando ARN intrónico no prognóstico de cancro.

Lisboa, 5 de Dezembro de 2008

MGB-CEGP1 int1.1

FIG. 1B

[illegible]

```
>CEGP1 intrão 5, 2556 bases
```

MGB-CEGP1 int 4.1

FIG. 1C

GTGAGTGGCAACCCCAACACTGAGTGAGGGTCTGCACCAGCCTGCCTGTCT
CCTACCCCTACCCCTTAATGGTGTTTAGCACAGATGCAGGCTGTTTCCTG
TGCATTTGCCCCCCCAGCAGGCCCTGTGCTGCTTCGCATGCTACAGTGGG
AGTGGTCTAGGCCTGTGGGGAAGGCCCTCTCTCCCTGTGTGACCTTGGG
AAGCCCTTCCTCCTCTCCTGGACTAGGCTGCTCCTAACGCTGGTATTCCA
GAGACTGGCACAACACCTCCCAGGAGGCCAGGGCAGCACGAAGTTAGAGC
TGTTTATAATGATGCGGCACCTTCTGGCCAGCAGGAGCCAGGGCCGTATAT
TTCTGGCGGGATGCCTGCCCTTGCCCTTCACGGTGTGTCTTCACTAGCTC
CATTTTAGAGGTTTTCCAGGCCCAAGGCTCTTTTTCTCCTCGACTCAGGGG
ACTGAAGCTTGCATTCCCTAGTGTCTCTTTGGTCAGTGCAATATACCTCC
AAAAATCTTTTCCATGTTTAATGTTTGCTAAGGATCTGTGGCCCTTTAACG
GGCTGTGTCTCCACAGAGCCTCATTACAACACATTTTTATTGCGTGAAC
AGAGTCACATATCTTTTCATTCCTCTTATGTCTGGGATTTACAGCAAACACA
GTTGTATGGGATGAGCAATCTAACTCATTACGTCTGAGAACCGTGCTCT
TTTGTCTTCTCTTGTAG

FIG. 1D

GTAATGTGTCCCACAGCAACCAAAATCAAGGTCAGCCAGCCTGACAGTC
TCTCCAGTGCTGTACTGCAACTTGTATCTGGGACAGCAGTTAAGTGCAA
GGACACTAGAATGATAAACAAATGTATCTTTTAGATTGTGACTCAATCTT
ATTGAATCCAGGCAAAATCATTAAGAAGAGCTCCTTAAC TACTTCATGTG
TTACTACCTAAAGTCCATGGAGGGTCTTCAATGTAGCAGCTCAAGCCCAC
TTTCTGCTAACACTCAACAGCCGTCCTAGATGCCAGCAGCTAGAGTGGCTA
AGTAGTTTTATGAAAATGTCTTTAGATAAAAAAAAAAAATGCTGCTGTGAG

FOXM1 int 3.2

[illegible]

```
>FOXM1 intrão 4, 993 bases
```

[illegible]

FIG. 1E

MGB-FOXM1 int 3.2
(sublinhado)
FOXM1 int 3.1(negrito)

FOXM1 int 4.1(sublinhado)
FOXM1 int 4.2(negrito)

FIG. 1F

FIG. 1G

FIG. 1H

MGB-FOXM1 int 7.1

FIG. 1H

[illegible]

[illegible]

TTTTAAGACTTCAAAACGACACAAAGAGGATCAGAACCCGTATGTGATAT
 TTTTGTGCGTCCTGTCTGGTGACCGTTGGTTCACCTTATCTCTGTTTCCC
 TTTCAG

MGB-FOXMI int 7.3

>PRAME intrão 2, 614 bases

GTAAGTTCGAGCCCTGATTCCTCCGCTTCCCCGCAGGGTGACCTTGGGCT
 TGTGCCCCCGGCACCACCCCTGTCCCGGGTCCCTGTTTTCTCTCTGGAAA
 TGGGTGAAGACCAAAGAAAAATAATGTGCGCCACTTGGGTACCCCCGGC
 CGCCTGCCCGGAAAATTGGCCCCAGTTGAGGAGTTGTGGCTGTAAGGAT
 GCCTTGAACCGAGGCGGCGGTGCTCGTGGTTGGAGCTCTCCAGGGTGGGT
 GCGCATTTGTAAATGCGGTGGATGCTCTGGGACTCGGCCCTCTGAAGGTG
 CTGGGGGTGGGGACGGCCAGGCAGTGGCGTAGGCGTCTAGGAAGGCG
 GGAGCAGAGGCAGAAATGTCGCTGCAAGACCGTAGTCAGGGTCCTTGACC
 ACAGGGGTCACTTGTGACCAACCACATGGTCTGTTGTTCCCTCCTGCCCCC
 TGGTTCAGCCCAGGAAACACTGGTGCTCAGGTTTGGAGCCAGAGATTTGC
 ACTGAAAGGGCGGGATTGAGTCGCCAGTTGTGAGTTTCTCAGCAGTATT
 TGCGGAGGTTTTACAGGAGGCCGTTGCTTCGTAAATATTATACATGTAT
 TCTTCTTTTTGGAG

MGB-PRAME int 2.1

FIG. 1I

>PRAME intrão 4, 432 bases

GTAAGGGTGACCTAGCAGCTTGGTGTGGGGCCCTGGGAACCTGAGCAGGA
 TGCAGCTGGGGTCAGGGAGCATGGAGCGCCTAAGGCTGGGCCAGAGGCTC
 TGATGGTTGCCAGCAAGGAAGTTGAGGAGGCCCTTGGGGCTACTGCAGGG
 GTCACCTCTTGAATGGGCTTCTGGACATGGGGCACTGATTAAAATGCAGA
 GGTGTCTGAAGGAACATGCACCTGCTTCCTCCTGGTGGGGTGGGAATTGG
 GGACCAGGAAGGATCCCAGGATCCTAGTGGGAAAGGAGCAGCTGATGCC
 TGAAGTACGAAGTAAAAGTGCAGATCTAAGGTGGATGTCTGTTTGGTTCT
 TACCTACATTATGAGACTCATGGTCTTATTTTGGAGTTGATCTTAAAGCAT
 CATCTCAGCTAATTACCTGTTTTTCCCCACAG

MGB-PRAME int 4.1

FIG. 1J

FIG. 1K

[illegible]

MGB-STK15 int 1.1

```
>STK15 intrão 2, 1622 bases
```

FIG. 1L

```

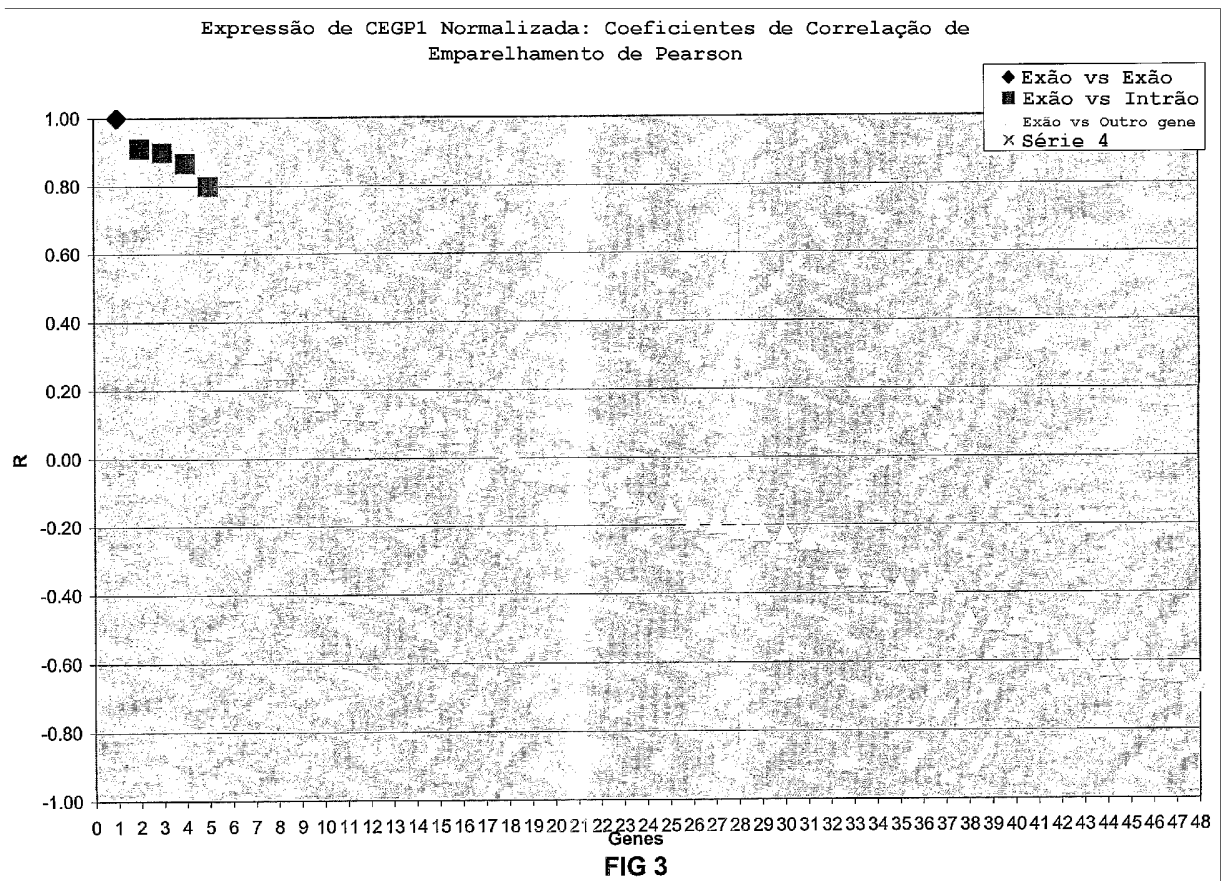
>STK15 intrão 4, 1093 bases
GTAAGCTTTCTTATTTACAAAAGTTCTGTACTGTTCTACTAGAATATATTA
TTTCGTTGCAAATTTTCGTTGTGGGAACTCTGGGGAAAAAAATGAGGCCTT
TATTTGCATTTAGAGGATATAAATGTTTCCAGATTTCCAATCTTAAAAAA
AATGGAATTTTGTGTAATGAGGTATTTTACTAGGAACTCAAGTGCTTTAA
AAAATGGCTTTCAAATTTAGAAAAAGCTTGTATGAATCTTTTATAGAAAT
GTGTGGAAGTTCCCTCTCTGTCTTAGAAAATAACCACTACATATGGTTTAT
GCGTCTGTACTTTTTTTATTTGTACAAAAGTGCAAGTTTTTAAAAAATAGAA
TATGTTGCAGAACTATATACTCATATATGACTGAGGGTTTTGACAGTATT
ATAGTTTTTAGTTCTTTATTTGTAAAGGTTGGCTGTAATGTCTTCCCCAGGG
CTTTTCTAAAAGCCCTCCTCTCAGTCTCTGAACTATCTGGACTCTAGAATG
TACCGGGAGGAGCGAGGAATGAACCCACAGACTCTTTTGCTTTTAGCGGT
CTAACAGAGGCTAAGAGTCTAAATCCACTGGTTCTCATGCCCCAGCTAGC
CTGTGGGCTCCATCCCGCTTCCATTAGTAACAGTGGCTCTGTCTCCACCA
CCAGAGTGGTTCTCCACCCAGAGAGAATTAGCACCTCTGGGACTGGAGGG
AGCAGCTGGGGTTAGTTTGAAACATGCCCCCAGATGGTCTGGAAGCATTC
CTCCCTCTCTGGTCACTTATCCTTTTTTGTGGTCTTCAGCGTTGTCATGGC
CCTGTTCCCTCTGAGCATAGTACGGGCTTGGGACATTTCCCATAGAGTGCT
TCAGGTCTAAAACCCGAGACTGCTCCTTGTCACTGACTCTCACACCTGAC
GGCAGCTAGGGACGTCAGGGTTTCATGTCGTGGCAGCTCTTTGATAGTGG
TTATTGCCCTTGGTTCTTGCTGAGGATGCATATTGAGTGAAGTTGGAATAC
GAAATTATTTGTAGAATGTGTCTGCTACTCATTGAAAAATTTGTTAGAAAA
GCTTTGTTTTCTTCACATTCTAAAGTGTTCAAATTCTCCTAG

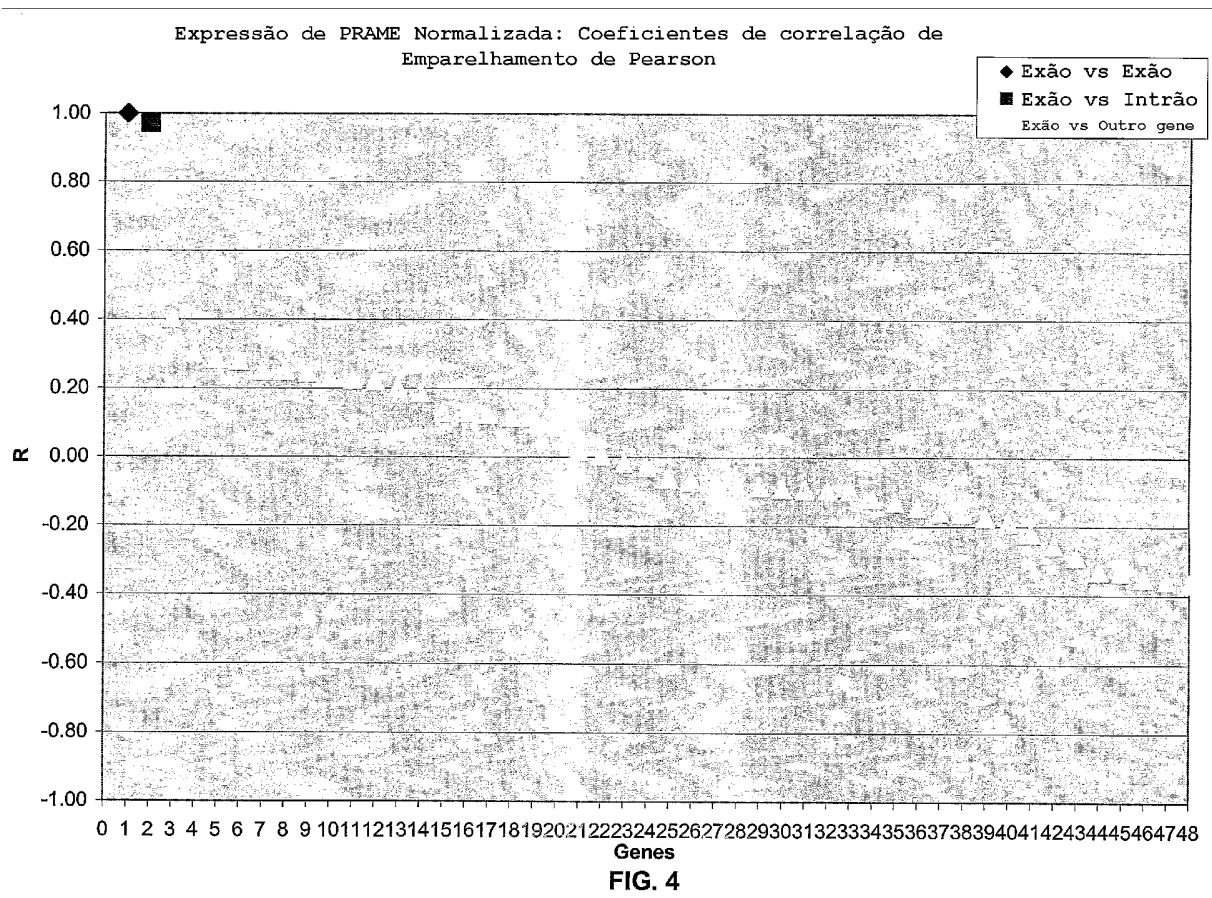
```

FIG. 1M

MGB-STK15 int 4.1

Nome	Sequência SEQ ID N°:	
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	CTGTTGCTGTGTATGCTGTCA	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	CCACTGTAGCATGCCAAGCA	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	CTCTTGGAACCTCTCCTGACACT	33
MGB-FOXM1 int 3.P2	CCGCAGAGCCGTC	34
FOXM1 int 4.F1	CCATCTCAAAGGAAACAAGTTCTG	35
FOXM1 int 4.R1	GGGTGGCTGGCTTCTAACTCT	36
FOXM1 int 4.P1	CCCTGATCAAATGAAAGCATCACT	37
FOXM1 int 4.F2	AGAAGCCAGCCACCAATTA	38
FOXM1 int 4.R2	TGTGTCATTTGGTATTTGTCAAGTGT	39
FOXM1 int 4.P2	TGACTTGACACAAAACCCAGTGAATTA	40
MGB-FOXM1 int 5.F1	TGGACAGAGACAAGATGTGATGTG	41
MGB-FOXM1 int 5.R1	GCTGGCACCTAGACAAAACATG	42
MGB-FOXM1 int 5.P1	CCATAGGGACCCCTTC	43
MGB-FOXM1 int 7.F1	GGTGTCCCTATTTTCTCTGAAGAGA	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	CCTGTTTGGGTTTCA	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	CTTCAACCCATTTCAGAGAGAA	54
MGB-PRAME int 2.P1	CCCGGGTCCCTGTT	55
MGB-PRAME int 4.F1	AGGGTGACCTAGCAGCTTGTT	56
MGB-PRAME int 4.R1	GCCTCTGGCCCAGCCTTA	57
MGB-PRAME int 4.P1	TCCCTGACCCCAGCTG	58
MGB-STK15 int 1.F1	CGTAATGTCTCTTCTCTCCGTAA	59
MGB-STK15 int 1.R1	ACGAACTGAGTAGGTTGCTGAAAA	60
MGB-STK15 int 1.P1	TCAAGGGACAAGGAAG	61
MGB-STK15 int 2.F1	CATTACATTTATAAACCCACATGGA	62
MGB-STK15 int 2.R1	AATCCAAAGTAAAGGCGGAAAGA	63
MGB-STK15 int 2.P1	TGGTCTTGTGCGGAAT	64
MGB-STK15 int 4.F1	GCGAGGAATGAACCCACAGA	65
MGB-STK15 int 4.R1	GCATGAGAACCAGTGGAATTTAGACT	66
MGB-STK15 int 4.P1	CGCTAAAAGCAAAAGA	67





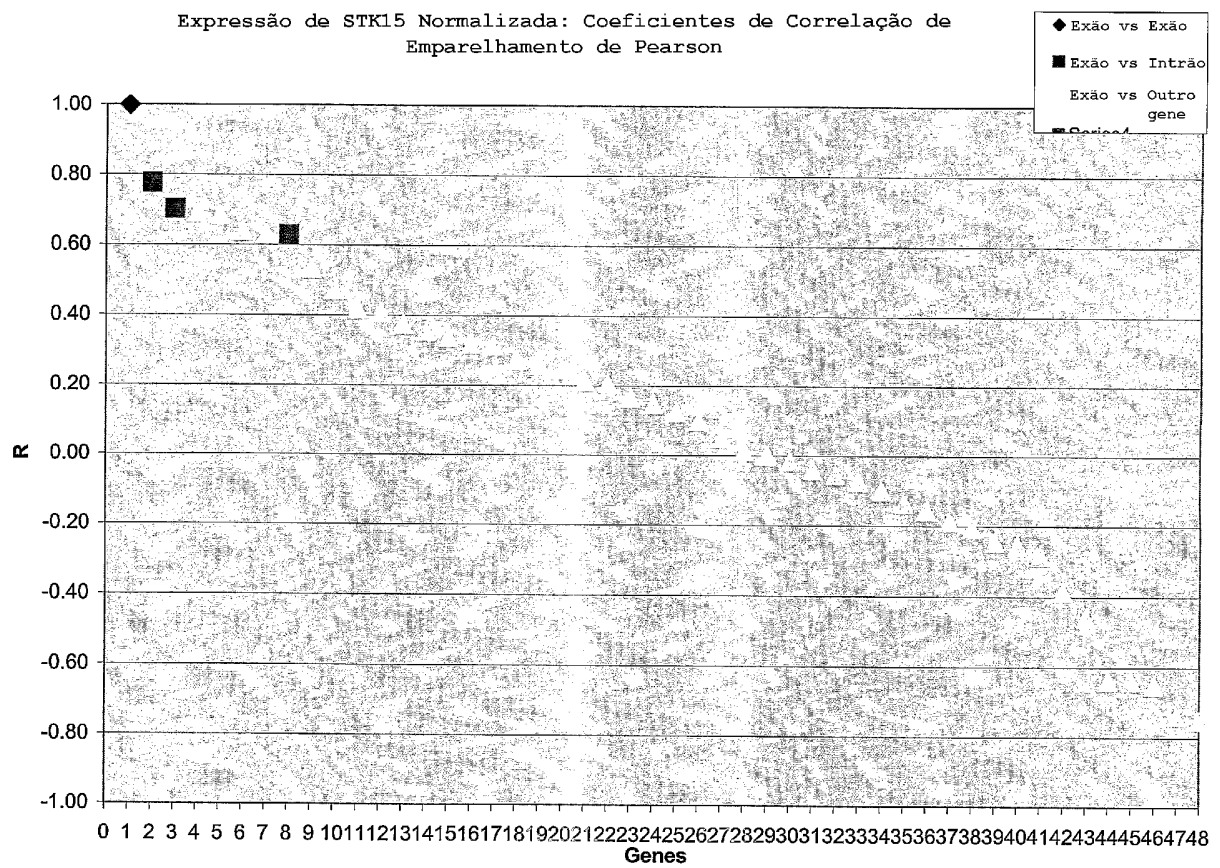


FIG. 5

			GHIallgenes2-03.txt
Nome do Gene	Sequência ID	Origem da Sequência	Número de acesso
18srRNA	M10098	NCBI	M10098
28srRNA	M11167	NCBI	M11167
A-Catenina	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
anexina I	NM_000700	NCBI	NM_000700
anexina 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-actina	NM_001101	NCBI	NM_001101
B-Catenina	NM_001904	NCBI	NM_001904

FIG. 6

Nome do Gene	Sequência ID	Origem de Sequência	GHIallgenes2-03.txt Número de acesso
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-relacionado	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

Nome do Gene	Sequência ID	Origem da Sequência	GHIallgenes2-03.txt Número de acesso
c-myb	NM_005375	NCBI	NM_005375
C20 orf1	NM_012112	NCBI	NM_012112
C20orf103	NK-012261	NCBI	NM_012261
c20orf108	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 variante 1	NM_057749var1	NCBI	NM_057749
CCNE2 variante 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 intrão 1	NM_020974int1	NCBI	NM_020974int1
CEGP1 intrão 2	NM_020974int2	NCBI	NM_020974int2
CEGP1 intrão 3	NM_020974int3	NCBI	NM_020974int3

FIG. 6

Nome do Gene	Sequência ID	Origem da Sequência	GHIallgenes2-03.txt Número de acesso
CEGP1 intrão 4	NM_020974int4	NCBI	NM_020974int4
CEGP1 intrão 5	NM_020974int5	NCBI	NM_020974int5
CEGP1 intrão 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
Claudina	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
Ciclina C	NM_005190	NCBI	NM_005190
Ciclina G1	NM_004060	NCBI	NM_004060

FIG. 6

Nome do Gene	Sequência ID	Origem da Sequência	GHIallgenes2-03.txt Número de acesso
Ciclina G2	NM_004354	NCBI	NM_004354
Ciclina 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_exões18-23	X76132-18-23	NCBI	X76132_18-23
DCC_exões7	X76132_6-7	NCBI	X76132_6-7
DCK	NM_000788	NCBI	NM_000788
DCR3	NM_016434	NCBI	NM_016434

FIG. 6

			GHIallgenes2-03.txt
Nome do Gene	Sequência ID	Origem da Sequência	Número de acesso
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 endotelina	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
fas1	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

Nome do Gene	Sequência ID	Origem da Sequência	GHIallgenes2-03.txt Número de acesso
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 intrão 3	NM_021953 int3	NCBI	NM_021953 int3
FOXM1 intrão 4	NM_021953 int4	NCBI	NM_021953 int4
FOXM1 intrão 5	NM_021953 int5	NCBI	NM_021953 int5
FOXM1 intrão 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- Catenina	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

			GHIallgenes2-03.txt
Nome do Gene	Sequência ID	Origem da Sequência	Número de acesso
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
Hepsina	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

Nome do Gene	Sequência ID	Origem da Sequência	GHIallgenes2-03.txt Número de acesso
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
Kitlmg	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 variante 1	NM_002656	NCBI	NM_002656
Lot1 variante 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

			GHIallgenes2-03.txt
Nome do Gene	Sequência ID	Origem da Sequência	Número de acesso
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

Nome do Gene	Sequência ID	Origem da Sequência	GHIallgenes2-03.txt Número de acesso
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, osteopontina	NM_000582	NCBI	NM_000582
ORC3	NM_012381	NCBI	NM_012381
Osteonectina	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

			GHIallgenes2-03.txt
Nome do Gene	Sequência ID	Origem da Sequência	Número de acesso
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 intrão1	NM_006115int1	NCBI	NM_006115int1
PRAME intrão2	NM_006115int2	NCBI	NM_006115int2
PRAME intrão3	NM_006115int3	NCBI	NM_006115int3
PRAME intrão4	NM_006115int4	NCBI	NM_006115int4
PRAME intrão5	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

			GHIallgenes2-03.txt
Nome do Gene	Sequência ID	Origem da Sequência	Número de acesso
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	NCBI	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_003152
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 intrão1	NM_003600int1	NCBI	NM_003600int1
STK15 intrão2	NM_003600int2	NCBI	NM_003600int2
STK15 intrão3	NM_003600int3	NCBI	NM_003600int3
STK15 intrão4	NM_003600int4	NCBI	NM_003600int4
STK15 intrão6	NM_003600int6	NCBI	NM_003600int6
STMY3	NM_005940	NCBI	NM_005940

FIG. 6

			GHIallgenes2-03.txt
Nome do Gene	Sequência ID	Origem da Sequência	Número de acesso
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
TFR3	NM_003234	NCBI	NM_003234
TGFA	NM_003236	NCBI	NM_003236
TGFb1	NM_000660	NCBI	NM_000660
TGFB3	NM_003239	NCBI	NM_003239
TGFB2	NM_003242	NCBI	NM_003242
Trombospondina 1	NM_003246	NCBI	NM_003246
Timosina 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

			GHIallgenes2-03.txt
Nome do Gene	Sequência ID	Origem da Sequência	Número de acesso
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

RESUMO

"UTILIZAÇÃO DE ARN INTRÓNICO PARA MEDIR A EXPRESSÃO GENÉTICA"

A invenção baseia-se na utilização de ARN intrónico na monitorização de expressão genética. Consequentemente, a presente invenção refere-se a métodos de determinação de perfis de expressão genética utilizando ARN intrónico, cuja expressão está correlacionada com a expressão do ARN exónico correspondente e métodos de diagnóstico e prognóstico baseados nos resultados desses estudos de expressão genética.