



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 314 378**

(51) Int. Cl.:

C12Q 1/68 (2006.01)

(12)

TRADUCCIÓN DE PATENTE EUROPEA

T3

(96) Número de solicitud europea: **04712947 .3**

(96) Fecha de presentación : **19.02.2004**

(97) Número de publicación de la solicitud: **1597391**

(97) Fecha de publicación de la solicitud: **23.11.2005**

(54) Título: **Uso de ARN intrónico para medir la expresión génica.**

(30) Prioridad: **20.02.2003 US 448991 P**

(73) Titular/es: **Genomic Health, Inc.
301 Penobscot Drive
Redwood City, California 94063, US**

(45) Fecha de publicación de la mención BOPI:
16.03.2009

(72) Inventor/es: **Scott, Randal, W.;
Kiefer, Michael, C. y
Baker, Joffre, B.**

(45) Fecha de la publicación del folleto de la patente:
16.03.2009

(74) Agente: **Ungría López, Javier**

ES 2 314 378 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Uso de ARN intrónico para medir la expresión génica.

5 Antecedentes de la invención

Campo de la invención

Es bien sabido que la expresión génica dentro de las células y tejidos puede indicar el estado fisiológico y/o patológico de la célula, del tejido o del paciente. Durante varios decenios, la expresión génica, según lo medido por el análisis inmunohistoquímico de los marcadores de proteínas, se ha usado para efectuar decisiones de tratamiento. Por ejemplo, los niveles de los receptores de estrógenos y receptores de progesterona medidos de este modo ahora se usan habitualmente para seleccionar pacientes de cáncer de mama para el tratamiento con fármacos antiestrogénicos.

15 La bibliografía de investigación más reciente proporciona pruebas de que los niveles tisulares de ciertas especies de ARNm tienen valor de diagnóstico y de pronóstico. Esto es un cambio prometedor ya que las tecnologías para medir los niveles de ARN celular, según demuestran múltiples RT-PCR y plataformas de matrices de ADN, pueden ser muy sensibles, específicas y cuantitativas. La PCR-TR se considera generalmente más sensible que la tecnología de matrices de ADN. Sin embargo, el diseño y selección de sondas/cebadores de RT-PCR puede ser desafiante porque existen criterios múltiples para conseguir un rendimiento óptimo. Este desafío es particularmente grande cuando el ARN de la muestra a estudiar proviene de tejidos fijados, embebidos en cera, porque dicho ARN suele estar muy fragmentado (K. Specht *et al.*, *Am. J. Pathol.* 158: 419-29 [2001]; T. E. Godfrey *et al.*, *J. Mol. Diagnostics* 2: 84-91 [2000]).

25 Es una práctica aceptada medir la expresión de cualquier gen determinado ensayando el nivel de cualquiera de sus secuencias de ARNm maduro, procesado por corte y empalme, transcrita (secuencia de exón en lugar de secuencia de intrón). En teoría, un exón se define como cualquier segmento de un gen interrumpido representado en el producto de ARN maduro, y un intrón se define como un segmento de ADN que se transcribe pero se elimina del transcripto empalmando entre sí los exones de cada lado del mismo [B. Lewin. *Genes IV*, Cell Press, Cambridge Mass.1990]. La razón para la práctica aceptada del uso de secuencias de exones es teóricamente sencilla, porque los ARN maduros [ARNm] codifican proteínas, que definen fenotipos celulares, mientras que se considera que el ARN intrónico tiene comparativamente poca influencia en el fenotipo celular. Además, la opinión predominante es que los intrones se degradan rápidamente y son por tanto más difíciles de detectar que las secuencias de los exones {véanse las introducciones de 30 los siguientes artículos: Thomas *et al.*, *J. Virol.* 76:532-40 [2002]; Clement *et al.*, *J. Biol. Chem.* 276:16919-30 [2001]; 35 Sharp *et al.*, *Ann. Rev. Biochem.* 55:1119-1150 [1986]}.

40 Adicionalmente, Kapranov *et al.*, *Science* 296:916-919 [2002] describe matrices de oligonucleótidos que contienen sondas situadas en promedio cada 35 pares de bases a lo largo de los cromosomas humanos 21 y 22. Usando ARN citosólico poliadenilado obtenido a partir de 11 líneas celulares humanas, se presentan mapas que identifican áreas activas de la transcripción del ARN en estos cromosomas y los autores llegan a la conclusión de que se transcribe hasta un orden de magnitud más de la secuencia genómica que lo que sería justificable por los exones previstos y caracterizados previamente.

45 La presente invención se refiere al uso de ARN intrónico para medir la expresión génica. Se mostrará que las secuencias de ARN intrónico suelen ser fácilmente detectables mediante RT-PCR incluso cuando se usa ARN muy degradado procedente de tejidos fijados. Además, tienden a correlacionarse en su expresión con sus exones respectivos. El último punto es particularmente inesperado porque existen pocas o ninguna prueba de que las proporciones de las constantes de velocidad global para la síntesis y regeneración de las secuencias de intrones y exones transcritas son similares. De hecho, la bibliografía científica proporciona pruebas de la complejidad del pre-ARNm y la regeneración de intrones procesados por corte y empalme. Por ejemplo, puede existir pre-ARNm en múltiples 50 grupos cinéticos (Elliot and Rosbash, *Exp. Cell Res.* 229:181-8 [1996]), con subpoblaciones que contienen ARN intrónicos que no están eficazmente procesados por corte y empalme y se transportan al citoplasma como especies de ARNm “inmaduro”, donde pueden degradarse a diferente velocidad que las secuencias de ARN intrónicas nucleares (Wang *et al.*, *Proc. Natl. Acad. Sci. USA* 94:4360-5 [1997]). Además, parece ser que algunos ARN intrónicos procesados por corte y empalme entran en el citoplasma en estructura de lazo (Clement *et al.*, *RNA* 5:206-20 55 [1999]).

Sumario de la invención

60 La presente invención se basa en pruebas experimentales que demuestran que las secuencias de intrones transcritas, que por definición están presentes en el ARN nuclear heterogéneo pero típicamente no se incorporan en el ARNm, tienen utilidad de diagnóstico y de pronóstico. Esto es un descubrimiento importante por varias razones. Típicamente, las secuencias de los intrones son más largas que las secuencias de los exones, 20 veces o más. 65 De esta manera, los intrones, dado que tienen una longitud media mucho mayor, proporcionalmente proporcionan mayores oportunidades para un diseño óptimo de sondas de expresión génica, por ejemplo, en el caso de RT-PCR, la creación de conjuntos de sonda/cebador que poseen un mejor rendimiento técnico. Independientemente, como las secuencias intrónicas evolucionan más rápidamente que las secuencias de los exones, los ARN intróni-

ES 2 314 378 T3

cos son adecuados para controlar la expresión de distintos miembros estrechamente relacionados de una familia de genes.

En un aspecto, la invención se refiere a un método para controlar la expresión génica en una muestra biológica,
5 que comprende:

(a) proporcionar un polinucleótido complementario a una secuencia de ARN intrónica dentro de un gen diana, donde la expresión de tal secuencia de ARN intrónica se correlaciona con la expresión de una secuencia de ARNm exónica dentro del gen diana;

10 (b) hibridar el polinucleótido con la secuencia de ARN intrónica para formar un complejo polinucleótido-ARN intrónico; y

15 (c) detectar el complejo polinucleótido-ARN intrónico.

En un aspecto particular, la expresión del gen diana se mide mediante la RT-PCR, en cuyo caso en el proceso anterior puede usarse un conjunto de cebador/sonda basado en intrón.

20 En otro aspecto, la invención se refiere a métodos para usar secuencias basadas en intrón para diseñar y crear conjuntos de cebador-sonda para la RT-PCR. Dichos cebadores y sondas son particularmente adecuados para detectar y cuantificar niveles de ARN intrónico en muestras de tejidos fijados, embebidos en parafina (FPET), para el análisis de la expresión génica de alta sensibilidad. Por consiguiente, en otro aspecto, la invención se refiere al uso de conjuntos de cebador-sonda basados en intrón en ensayos de perfiles de la expresión génica, tales como el análisis de la expresión génica de muestras FPET para diagnosticar y/o predecir el pronóstico de diversos estados patológicos.

25 En particular, la invención se refiere a un método para preparar una molécula de oligonucleótido monocatenaria para la amplificación de un gen diana y medir el nivel de una especie de ARN intrónico, que comprende:

(a) identificar al menos un secuencia de intrón dentro del gen diana, donde la expresión de la secuencia de intrón se correlaciona con la expresión de una secuencia de exón dentro del gen diana;

30 (b) preparar una molécula de oligonucleótido monocatenaria que se corresponda con al menos una parte de la secuencia de intrón transcrita; y

35 (c) usar la molécula de oligonucleótido para medir la expresión génica.

Al igual que antes, la expresión génica puede medirse, por ejemplo, mediante la RT-PCR, en cuyo caso se usa un conjunto de cebador/sonda basado en intrón (que consta de dos cebadores y una sonda) para medir la expresión génica.

40 Si el oligonucleótido es un cebador directo, típicamente se diseña para comprender las secuencias 5' de una secuencia diana dentro de la secuencia intrónica transcrita. Si el oligonucleótido es un cebador inverso, típicamente se diseña para complementar las secuencias 5' de una secuencia diana aguas abajo del cebador directo dentro de la secuencia intrónica transcrita. Es importante identificar y usar una secuencia diana suficientemente larga para la amplificación por PCR. La secuencia diana debe ser generalmente al menos de aproximadamente 50 bases de nucleótidos de longitud, en particular al menos de 55 bases de nucleótidos de longitud, y en algunas realizaciones al menos de aproximadamente 60 bases de nucleótidos de longitud. Los cebadores y sondas de PCR se diseñan siguiendo principios bien conocidos. De este modo, el cebador de PCR es típicamente de 17-30 bases de nucleótidos de longitud y generalmente contiene aproximadamente del 20% al 80% de bases G + C. Es deseable diseñar los cebadores de PCR con una temperatura de fusión (Tm) entre aproximadamente 50°C y aproximadamente 70°C.

45 50 Cuando la molécula de oligonucleótido monocatenaria es una sonda de PCR, generalmente se diseña para comprender o complementar una parte interna de una secuencia diana dentro de la secuencia intrónica transcrita. Para la amplificación TaqMan®, la sonda de PCR se marca con un colorante fluorescente indicador y un resto de inactivación.

55 En otro aspecto, la invención se refiere a un método para medir la expresión de un gen mediante la amplificación de un gen diana por la reacción en cadena de la polimerasa (PCR), que comprende:

60 (a) identificar al menos una secuencia intrónica diana dentro del gen diana, donde la expresión de la secuencia intrónica se correlaciona con la expresión de una secuencia de exón correspondiente dentro del gen diana; y

(b) amplificar la secuencia intrónica diana transcrita usando un conjunto de cebador/sonda de PCR específico de intrón.

65 La secuencia de intrón diana es típicamente al menos aproximadamente de 50 bases de longitud, y el conjunto de cebador y sonda de PCR está diseñado para corresponder a secuencias únicas dentro de la secuencia intrónica diana transcrita.

ES 2 314 378 T3

En otro aspecto, la invención se refiere a un método para amplificar fragmentos de ARN en una muestra que representan al menos un gen de interés, que comprende las etapas de:

- (a) poner en contacto la muestra con al menos un conjunto de cebadores y sonda de PCR; y
- 5 (b) realizar la amplificación por PCR,

donde los cebadores y la sonda de PCR se diseñan basándose en una secuencia intrónica identificada dentro del gen de interés, y donde la expresión de la secuencia intrónica se correlaciona con la expresión de una secuencia de exón dentro del gen de interés.

10 En una realización particular, los cebadores y la sonda de PCR se diseñan típicamente basándose en una secuencia única dentro del intrón identificado. En otra realización, la muestra comprende ARN fragmentado que representa múltiples genes de interés y entra en contacto con un grupo de cebadores y sondas de PCR diseñados basándose en 15 secuencias únicas dentro de intrones presentes en los genes de interés.

En una realización preferida, la amplificación se realiza en una muestra de tejido fijado embebido en parafina (FPET), que puede proceder, por ejemplo, de una biopsia de tumor obtenida a partir de un paciente humano. El tumor puede ser cualquier tipo de tumor sólido tal como, por ejemplo, cáncer de mama, cáncer de pulmón o cáncer 20 colorrectal. El tejido tumoral puede recogerse mediante diversos métodos, incluyendo la biopsia de aguja fina, la biopsia de aguja gruesa o la resección.

25 En una realización particular, la invención se refiere a métodos que usan conjuntos de cebador-sonda de PCR basados en intrón en el análisis de la expresión génica para predecir la probabilidad de recurrencia en pacientes con cáncer de mama en sus primeros estadios.

30 En un aspecto adicional, la invención se refiere a una matriz que comprende una pluralidad de polinucleótidos que hibridan con los genes diana de interés, donde preferentemente al menos el 70% de los polinucleótidos comprenden secuencias intrónicas.

35 En otra aspecto más, la invención se refiere a secuencias de amplicón basadas en intrón y a su uso en el análisis de la expresión génica.

En una realización particular, la invención se refiere al análisis de la expresión génica de una muestra biológica 40 representativa de cáncer de mama invasivo, basado en la determinación de los niveles de expresión de transcritos de ARN o de los productos de expresión de un gen o conjunto de genes seleccionados del grupo consistente en:

- (a) Bcl2, cyclinG1, NFKBp65, NME1, EPHX1, TOP2B, DR5, TERC, Src, DIABLO;
- 45 (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;
- 50 (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;
- 55 (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;
- 60 (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;
- 65 (n) CA9, FOXM1, cyclinG1, XIAP, TS, Ki67, NFKBp65, CYP3A4, GSTM3, p27;
- (o) TFRC, XIAP, PRAME, p27, Contig51037, ErbB3, DPYD, TERC, NME1, VEGFC, y

ES 2 314 378 T3

(p) CEGP1, PRAME, hENT1, XIAP, Contig51037, ErbB3, DPYD, NFKBp65, ID 1, TS,

incluyendo el uso de secuencias basadas en intrones.

5 En otra realización, la invención se refiere al análisis de la expresión génica de una muestra biológica representativa de cáncer de mama ER-positivo basado en la determinación de los niveles de expresión de transcritos de ARN o de los productos de expresión de un gen o conjunto de genes seleccionados del grupo consistente en:

- (a) PRAME, p27, IGFBP2, HIF1A, TIMP2, ILT2, CYP3A4, ID 1, EstR1, DIABLO;
- (b) Contig51037, EPHX1, Ki67, TIMP2, cyclinG1, DPYD, CYP3A4, TP, AIB1, CYP2C8;
- (c) Bcl2, hENT1, FOXM1, Contig51037, cyclinG1, Contig46653, PTEN, CYP3A4, TIMP2, AREG;
- (d) HIF1A, PRAME, p27, IGFBP2, TIMP2, ILT2, CYP3A4, ID1, EstR1, DIABLO;
- (e) IGF1R, PRAME, EPHX1, Contig51037, cyclinG1, Bcl2, NME1, PTEN, TBP, TIMP2;
- (f) FOXM1, Contig51037, VEGFC, TBP, HIF1A, DPYD, RAD51C, DCR3, cyclinG1, BAG1;
- (g) EPHX1, Contig51037, Ki67, TIMP2, cyclinG1, DPYD, CYP3A4, TP, AIB1, CYP2C8;
- (h) Ki67, VEGFC, VDR, GSTM3, P27, upa, ITGA7, rhoC, TERC, Pin1;
- (i) CDC25B, Contig51037, hENT1, Bcl2, HLAG, TERC, NME1, upa, ID1, CYP;
- (j) VEGFC, Ki67, VDR, GSTM3, P27, upa, ITGA7, rhoC, TERC, Pin1;
- (k) CTSB, PRAME, p27, IGFBP2, EPHX1, CTSL, BAD, DR5, DCR3, XIAP;
- (l) DIABLO, Ki67, hENT1, TIMP2, ID1, p27, KRT19, IGFBP2, TS, PDGFB;
- (m) p27, PRAME, IGFBP2, HIF1A, TIMP2, ILT2, CYP3A4, ID1, EstR1, DIABLO;
- (n) CDH1; PRAME, VEGFC; HIF1A; DPYD; TIMP2, CYP3A4, EstR1, RBP4, p27;
- (o) IGFBP3, PRAME, p27, Bcl2, XIAP, EstR1, Ki67, TS, Src, VEGF;
- (p) GSTM3, PRAME, p27, IGFBP3, XIAP, FGF2, hENT1, PTEN, EstR1, APC;
- (q) hENT1, Bcl2, FOXM1, Contig51037, CyclinG1, Contig46653, PTEN, CYP3A4, TIMP2, AREG;
- (r) STK15, VEGFC, PRAME, p27, GCLC, hENT1, ID1, TIMP2, EstR1, MCP1;
- (s) NME1, PRAM, p27, IGFBP3, XIAP, PTEN, hENT1, Bcl2, CYP3A4, HLAG;
- (t) VDR, Bcl2, p27, hENT1, p53, PI3KC2A, EIF4E, TFRC, MCM3, ID1;
- (u) EIF4E, Contig51037, EPHX1, cyclinG1, Bcl2, DR5, TBP, PTEN, NME1, HER2;
- (v) CCNB1, PRAME, VEGFC, HIF1A, hENT1, GCLC, TIMP2, ID1, p27, upa;
- (w) ID1, PRAME, DIABLO, hENT1, p27, PDGFRa, NME1, BIN1, BRCA1, TP;
- (x) FBXO5, PRAME, IGFBP3, p27, GSTM3, hENT1, XIAP, FGF2, TS, PTEN;
- (y) GUS, HIA1A, VEGFC, GSTM3, DPYD, hENT1, EBX05, CA9, CYP, KRT18, y
- (z) Bclx, Bcl2, hENT1, Contig51037, HLAG, CD9, ID1, BRCA1, BIN1, HBEGF,

60 incluyendo el uso de secuencias basadas en intrones.

En otra realización, el cáncer es cáncer de mama y el gen o los genes analizados se seleccionan del grupo consistente en: FOXM1; PRAME; SKT15, Ki-67; CA9; NME1; SURV; TFRC; YB-1; RPS6KB1; Src; Chkl; CCNB1; Chk2;

65 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 y NFKBp65.

ES 2 314 378 T3

En otra una realización más, la invención se refiere al análisis de la expresión génica de una muestra biológica representativa de cáncer de mama invasivo, basado en la determinación de los niveles de expresión de los transcritos de ARN o productos de expresión de un gen o conjunto de genes seleccionados del grupo consistente en:

- 5 (a) p53BP2, Bcl2, BAD, EPHX1, PDGFR β , DIABLO, XIAP, YB1, CA9 y KRT8;
- (b) GRB7, CD68, TOP2A, Bcl2, DIABLO, CD3, ID1, PPM1D, MCM6 y WISP1;
- 10 (c) PR, p53BP2, PRAME, DIABLO, CTSL, IGFBP2, TIMP1, CA9, MMP9 y COX2;
- (d) CD68, GRB7, TOP2A, Bcl2, DIABLO, CD3, ID1, PPM1D, MCM6 y WISP1;
- 15 (e) Bcl2, p53BP2, BAD, EPHX1, PDGFR β , DIABLO, XIAP, YB1, CA9 y KRT8;
- (f) KRT14, KRT5, PRAME, p53BP2, GUS1, AIB1, MCM3, CCNE1, MCM6 e ID1;
- (g) PRAME, p53BP2, EstR1, DIABLO, CTSL, PPM1D, GRB7, DAPK1, BBC3 y VEGFB;
- 20 (h) CTSL2, GRB7, TOP2A, CCNB1, Bcl2, DIABLO, PRAME, EMS1, CA9 y EpCAM;
- (i) EstR1, p53BP2, PRAME, DIABLO, CTSL, PPM1D, GRB7, DAPK1, BBC3 y VEGFB;
- 25 (j) Chk1, PRAME, p53BP2, GRB7, CA9, CTSL, CCNB1, TOP2A, el tamaño del tumor e IGFBP2;
- (k) IGFBP2, GRB7, PRAME, DIABLO, CTSL, β -catenina, PPM1D, Chk1, WISP1 y LOT1;
- (l) HER2, p53BP2, Bcl2, DIABLO, TIMP1, EPHX1, TOP2A, TRAIL, CA9 y AREG;
- 30 (m) BAG1, p53BP2, PRAME, IL6, CCNB1, PAI1, AREG, el tamaño del tumor, CA9 y Ki67;
- (n) CEGP1, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, STK15, AKT2 y FGF18;
- (o) STK15, p53BP2, PRAME, IL6, CCNE1, AKT2, DIABLO, cMet, CCNE2 y COX2;
- 35 (p) KLK10, EstR1, p53BP2, PRAME, DIABLO, CTSL, PPM1D, GRB7, DAPK1 y BBC3;
- (q) AIB1, p53BP2, Bcl2, DIABLO, TIMP1, CD3, p53, CA9, GRB7 y EPHX1;
- 40 (r) BBC3, GRB7, CD68, PRAME, TOP2A, CCNB1, EPHX1, CTSL, GSTM1 y APC;
- (s) CD9, GRB7, CD68, TOP2A, Bcl2, CCNB1, CD3, DIABLO, ID1 y PPM1D;
- (t) EGFR, KRT14, GRB7, TOP2A, CCNB1, CTSL, Bcl2, TP, KLK10 y CA9;
- 45 (u) HIF1 α , PR, DIABLO, PRAME, Chk1, AKT2, GRB7, CCNE1, TOP2A y CCNB1;
- (v) MDM2, p53BP2, DIABLO, Bcl2, AIB1, TIMP1, CD3, p53, CA9 y HER2;
- 50 (w) MYBL2, p53BP2, PRAME, IL6, Bcl2, DIABLO, CCNE1, EPHX1, TIMP1 y CA9;
- (x) p27, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, STK15, AKT2 e ID1;
- (y) RAD51, GRB7, CD68, TOP2A, CIAP2, CCNB1, BAG1, IL6, FGFR1 y p53BP2;
- 55 (z) SURV, GRB7, TOP2A, PRAME, CTSL, GSTM1, CCNB1, VDR, CA9 y CCNE2;
- (aa) TOP2B, p53BP2, DIABLO, Bcl2, TIMP1, AIB1, CA9, p53, KRT8 y BAD;
- (ab) ZNF217, GRB7, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, APC4 y β -catenina.

60 En una realización diferente, la invención se refiere al análisis de la expresión génica de una muestra biológica, usando secuencias de polinucleótidos basadas en intrones que hibridan al menos con uno de los genes seleccionados del grupo consistente en: CD68; CTSL; FBXO5; SURV; CCNB1; MCM2; Chk1; MYBL2; HIF1A; cMET; EGFR; TS; STK15; IGFR1; BCI2; 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 y STK15.

ES 2 314 378 T3

El análisis de la expresión génica puede realizarse en un formato de matriz, y la matriz preferiblemente es una matriz de alta densidad que comprende al menos 100, más preferiblemente al menos 150, y aún más preferiblemente 200 secuencias en una sección de 5-10 μ . Finalmente, la presente invención se refiere al uso de ARN intrónico en pronósticos de cáncer como se define en las reivindicaciones.

5

Breve descripción de los dibujos

Las figuras 1A-M muestran secuencias intrónicas enmascaradas para los genes CEGP1, FOXM1, PRAME y STK15. Los amplicones usados para la RT-PCR se muestran en cursiva.

10

La Figura 2 muestra conjuntos de cebador/sonda para CEGP1, FOXM1, PRAME y STK15. Las secuencias de los cebadores directo e inverso se indican con "F" y "R", respectivamente. Las secuencias de los cebadores se designan con "P".

15

La Figura 3 muestra los coeficientes de correlación [R] para la co-expresión del ARN del exón de CEGP1 con otras 47 secuencias de ARN. Símbolos: rombo = exón de CEGP1 frente a sí mismo (= 1,0 por definición); cuadrados = intrones de CEGP1; triángulos = secuencias de otros genes.

20

La Figura 4 muestra los coeficientes de correlación [R] para la co-expresión del ARN del exón de PRAME con otras 47 secuencias de ARN. Símbolos: rombo = exón de PRAME frente a sí mismo (= 1,0 por definición); cuadrados = intrones de PRAME; triángulos = secuencias de otros genes.

25

La figura 5 muestra los coeficientes de correlación [R] para la co-expresión del ARN del exón de STK15 con otras 47 secuencias de ARN. Símbolos: rombo = exón de STK15 frente a sí mismo (= 1,0 por definición); cuadrados = intrones de STK15; triángulos = secuencias de otros genes.

La Figura 6 muestra un conjunto ejemplar de genes, cuya expresión puede analizarse por los métodos de la presente invención.

30

Descripción detallada de la realización preferida

A. Definiciones

35

A menos que se defina de otro modo, los términos técnicos y científicos usados en este documento tienen el mismo significado que el comúnmente entendido por un especialista habitual en la técnica a la que pertenece esta invención. Singleton *et al.*, *Dictionary of Microbiology and Molecular Biology 2nd ed.*, J. Wiley & Sons (New York; NY 1994), proporcionan al especialista en la técnica una guía general para muchos de los términos usados en la presente solicitud. Un especialista en la técnica reconocerá muchos métodos y materiales similares o equivalentes a los descritos en este documento, que podrían usarse en la práctica de la presente invención. De hecho, la presente invención no se limita en modo alguno a los métodos y materiales descritos. Para los fines de la presente invención, los siguientes términos se definen como se indica a continuación.

40

Los términos "corte y empalme" y "corte y empalme de ARN" se usan de manera indistinta y se refieren al procesamiento de ARN que elimina intrones y une exones para producir el ARNm maduro con secuencias codificantes continuas que se transportan al citoplasma de una célula eucariota.

45

En teoría, el término "exón" se refiere a cualquier segmento de un gen interrumpido que está representado en el producto de ARN maduro (B. Lewin. *Genes IV* Cell Press, Cambridge Mass. 1990). En teoría, el término "intrón" se refiere a cualquier segmento de ADN que se transcribe pero se elimina del interior del transcripto por corte y empalme de los exones que están a cada lado del mismo. Funcionalmente, en la secuencia del ARNm de un gen definido por los números de ID de Sec de Ref. aparecen secuencias de exones. Funcionalmente, las secuencias intrónicas son las secuencias intermedias dentro del ADN genómico de un gen, separadas por secuencias de exones y que tienen secuencias consenso de corte y empalme GT y AG en sus extremos 5' y 3'.

55

El término "micromatriz" se refiere a una disposición ordenada de elementos hibridables en una matriz, preferiblemente sondas polinucleotídicas sobre un sustrato.

60

El término "polinucleótido", cuando se utiliza en singular o en plural, generalmente se refiere a cualquier polirribonucleótido o polidesoxirribonucleótido, que puede ser ARN o ADN no modificado o ARN o ADN modificado. De este modo, por ejemplo, los polinucleótidos como se definen en este documento incluyen, sin limitación, ADN mono- y bi-catenario, ADN que incluye regiones mono- y bi-catenarias, ARN mono- y bi-catenario, y ARN que incluye regiones mono- y bi-catenarias, moléculas hibridas que comprenden ADN y ARN que puede ser monocatenario o, más típicamente, bicatenario o incluyen regiones monocatenarias y bicatenarias. Además, el término "polinucleótido" como se usa en este documento se refiere a regiones tricatenarias que comprenden ARN o ADN o tanto ARN como ADN. Las cadenas en dichas regiones pueden ser de la misma molécula o de moléculas diferentes. Las regiones pueden incluir la totalidad de una o más de las moléculas, pero más típicamente incluyen sólo una región de algunas de las moléculas. Una de las moléculas de una región de triple-hélice a menudo es un oligonucleótido. El término "polinucleótido" específicamente incluye los ADNc. El término incluye ADN (incluyendo ADNc) y ARN que con-

tienen una o más bases modificadas. De este modo, los ADN o los ARN con estructuras modificadas para conseguir estabilidad o por otras razones son “polinucleótidos” como se usa este término en el presente documento. Además, se incluyen dentro del término “polinucleótidos” como se define en este documento ADN o ARN que comprenden bases inusuales, tales como inosina, o bases modificadas, tales como bases tritiadas. En general, el término “polinucleótido” abarca todas las formas modificadas química, enzimática y/o metabólicamente de polinucleótidos no modificados, así como las formas químicas de ADN y ARN características de virus y células, incluyendo células sencillas y complejas.

El término “oligonucleótido” se refiere a un polinucleótido relativamente corto que incluye, sin limitación, desoxirribonucleótidos monocatenarios, ribonucleótidos mono- o bi-catenarios, híbridos de ARN:ADN y ADN bicatenarios. Los oligonucleótidos, tales como los oligonucleótidos sonda de ADN monocatenarios a menudo se sintetizan por métodos químicos, por ejemplo, usando sintetizadores de oligonucleótidos automáticos que están disponibles comercialmente. Sin embargo, pueden obtenerse oligonucleótidos por una diversidad de métodos distintos, incluyendo técnicas mediadas por ADN recombinante *in vitro* y por la expresión de ADN en células y organismos.

Las expresiones “genes expresados diferencialmente”, “expresión génica diferencial” y sus sinónimos, que se usan indistintamente, se refieren a un gen cuya expresión es de un nivel superior o inferior en un paciente o sujeto de ensayo en relación a otro, por ejemplo, en un sujeto que padece una enfermedad, específicamente cáncer, tal como cáncer de mama, en relación a su expresión en un sujeto normal o de control. Las expresiones también incluyen genes cuya expresión está activada hacia un mayor o menor nivel en diferentes etapas de la misma enfermedad. También se entiende que un gen expresado diferencialmente puede estar activado o inhibido a nivel del ácido nucleico o a nivel de la proteína, o puede ser objeto de corte y empalme alternativo para dar como resultado un producto de polipéptido diferente. Dichas diferencias pueden demostrarse por un cambio en los niveles de ARNm, expresión en la superficie, secreción u otras divisiones de un polipéptido, por ejemplo. La expresión génica diferencial puede incluir una comparación de la expresión entre dos o más genes o sus productos génicos, o una comparación de las proporciones de la expresión entre dos o más genes o sus productos génicos, o incluso una comparación de dos productos procesados de manera diferente del mismo gen, que difieren entre sujetos normales y sujetos que padecen una enfermedad, específicamente cáncer, o entre diferentes estadíos de la misma enfermedad. La expresión diferencial incluye tanto diferencias cuantitativas como diferencias cualitativas en el patrón de expresión temporal o celular en un gen o sus productos de expresión entre, por ejemplo, células normales o enfermas, o entre células que han experimentado diferentes acontecimientos de enfermedad o estadíos de enfermedad. Para el fin de esta invención, se considera que está presente la “expresión génica diferencial” cuando hay al menos aproximadamente dos veces, preferiblemente al menos aproximadamente cuatro veces, más preferiblemente al menos aproximadamente seis veces, más preferiblemente al menos aproximadamente diez veces de diferencia entre la expresión de un gen determinado en sujetos normales y enfermos, o en los diversos estadíos del desarrollo de la enfermedad en un sujeto enfermo.

El término “sobre-expresión” con respecto a un transcripto de ARN se usa para referirse al nivel del transcripto determinado por normalización al nivel de los ARNm de referencia, que podría ser todos los transcritos medidos en la muestra o un conjunto de referencia particular de ARNm.

La frase “amplificación génica” se refiere a un proceso mediante el cual se forman múltiples copias de un gen o fragmento de gen en una célula o línea celular particular. La región duplicada (un tramo de ADN amplificado) se denomina a menudo “amplicón”. A menudo, la cantidad de ARN mensajero (ARNm) producido, es decir, el nivel de la expresión génica, también aumenta en la proporción del número de copias efectuadas del gen particular expresado.

El término “pronóstico” se usa en este documento para referirse a la predicción de la probabilidad de muerte o progresión atribuible al cáncer, incluyendo la recurrencia, la propagación de metástasis y la resistencia a fármacos, de una enfermedad neoplásica, tal como cáncer de mama. El término “predicción” se usa en este documento para referirse a la probabilidad de que un paciente responda de manera favorable o desfavorable a un fármaco o conjunto de fármacos, y también el alcance de esas respuestas, o de que el paciente sobreviva tras la extirpación quirúrgica del tumor primario y/o quimioterapia durante un cierto periodo de tiempo sin recurrencia de cáncer. Los métodos de predicción de la presente invención pueden usarse clínicamente para tomar decisiones de tratamiento escogiendo las modalidades de tratamiento más apropiadas para cualquier paciente particular. Los métodos de predicción de la presente invención son herramientas valiosas para predecir si es probable que un paciente responda favorablemente a un régimen de tratamiento, tal como una intervención quirúrgica, quimioterapia con un fármaco o combinación de fármacos determinada, y/o radioterapia, o si es probable la supervivencia a largo plazo del paciente, tras la cirugía y/o terminación de la quimioterapia u otras modalidades de tratamiento.

El término supervivencia a “largo plazo” se usa en este documento para referirse a la supervivencia durante al menos 3 años, más preferiblemente durante al menos 5 años, y mucho más preferiblemente durante al menos 10 años tras la cirugía u otro tratamiento.

El término “aumento de resistencia” a un fármaco u opción de tratamiento particular, cuando se usa de acuerdo con la presente invención, significa la disminución de respuesta a una dosis estándar del fármaco o a un protocolo de tratamiento estándar.

ES 2 314 378 T3

El término “disminución de la sensibilidad” a un fármaco u opción de tratamiento particular, cuando se usa de acuerdo con la presente invención, significa la disminución de respuesta a una dosis estándar del fármaco o a un protocolo de tratamiento estándar, cuando la disminución de la respuesta puede compensarse (al menos parcialmente) por el aumento de la dosis del fármaco o la intensidad del tratamiento.

5 La “respuesta del paciente” puede evaluarse usando cualquier punto final que indique un beneficio para el paciente, incluyendo, sin limitación, (1) la inhibición, en alguna medida, del crecimiento del tumor incluyendo la ralentización y detención completa del crecimiento; (2) la reducción del número de células tumorales; (3) la reducción del tamaño del tumor; (4) la inhibición (es decir, reducción, ralentización o detención total) de la infiltración de células tumorales 10 hacia tejidos y/u órganos periféricos adyacentes; (5) la inhibición (es decir, reducción, ralentización o detención total) de la metástasis; (6) la mejora de la respuesta inmune anti-tumoral, que puede, pero no tiene que, dar como resultado la regresión o el rechazo del tumor; (7) el alivio, en alguna medida, de uno más síntomas asociados con el tumor; (8) el aumento de la duración de la supervivencia después del tratamiento; y/o (9) el descenso de la mortalidad en un momento determinado después del tratamiento.

15 10 El término “tratamiento” se refiere tanto al tratamiento terapéutico como al tratamiento profiláctico o medidas preventivas, en el que el objetivo es prevenir o ralentizar (reducir) el estado patológico o trastorno diana. Los que necesitan tratamiento incluyen los que ya tienen el trastorno así como los propensos a tener el trastorno o aquellos en 20 los que se desea prevenir el trastorno. En el tratamiento de tumores (por ejemplo, cáncer), un agente terapéutico puede disminuir directamente la patología de las células tumorales, o hacer que las células tumorales sean más susceptibles al tratamiento por otros agentes terapeúticos, por ejemplo, radiación y/o quimioterapia.

25 El término “tumor”, como se usa en este documento, se refiere a todo el crecimiento celular neoplásico y proliferación, tanto maligna como benigna, y a todas las células y tejidos pre-cancerosos y cancerosos.

30 35 Los términos “cáncer” y “canceroso” se refieren a o describen el estado fisiológico en mamíferos que está caracterizado típicamente por un crecimiento celular no regulado. Los ejemplos de cáncer incluyen pero no se limitan a, cáncer de mama, cáncer de colon, cáncer de pulmón, cáncer de próstata, cáncer hepatocelular, cáncer gástrico, cáncer pancreático, cáncer cervical, cáncer de ovario, cáncer de hígado, cáncer de vejiga, cáncer del tracto urinario, cáncer de tiroides, cáncer renal, carcinoma, melanoma y cáncer cerebral.

35 40 La “patología” del cáncer incluye todos los fenómenos que comprometen el bienestar del paciente. Esto incluye, sin limitación, el crecimiento celular anormal o incontrolable, metástasis, la interferencia con el funcionamiento normal de las células vecinas, liberación de citoquinas u otros productos de secreción en niveles anormales, la supresión o el agravamiento de la respuesta inflamatoria o inmunológica, neoplasia, premalignidad, malignidad, invasión de tejidos u órganos circundantes o distantes, tales como ganglios linfáticos, etc.

45 50 La “rigurosidad” de las reacciones de hibridación es fácilmente determinable por un especialista normal en la técnica, y generalmente es un cálculo empírico, que depende de la longitud de la sonda, la temperatura de lavado y la concentración salina. En general, las sondas más largas requieren mayores temperaturas para una hibridación apropiada, mientras que las sondas más cortas necesitan temperaturas más bajas. La hibridación generalmente depende de la capacidad del ADN desnaturalizado para rehibridar cuando las cadenas complementarias están presentes en un medio por debajo de su temperatura de fusión. Cuanto más alto sea el grado de homología deseada entre la sonda y la secuencia hibridizable, mayor es la temperatura relativa que puede usarse. Como resultado, se deduce que el aumento de las temperaturas relativas tendería a hacer las condiciones de reacción más rigurosas, mientras que temperaturas más bajas menos. Si se desean detalles y explicaciones adicionales de la rigurosidad de las reacciones de hibridación, véase Ausubel *et al.*, *Current Protocols in Molecular Biology*, Wiley Interscience Publishers, (1995).

55 60 Las “condiciones rigurosas” o las “condiciones de alta rigurosidad”, como se definen en este documento, típicamente: (1) emplean baja fuerza iónica y alta temperatura de lavado, por ejemplo, cloruro sódico 0,015 M/citrato sódico 0,0015 M/dodecil sulfato sódico al 0,1% a 50°C; (2) emplean un agente desnaturalizante durante la hibridación, tal como formamida, por ejemplo, formamida al 50% (v/v) con albúmina de suero bovino al 0,1%/Ficoll al 0,1%/polivinilpirrolidona al 0,1%/tampón fosfato sódico 50 mM a pH 6,5 con cloruro sódico 750 mM, citrato sódico 75 mM a 42°C; ó (3) emplean formamida al 50%, SSC 5 x (NaCl 0,75 M, citrato sódico 0,075 M), fosfato sódico 50 mM (pH 6,8), pirofosfato sódico al 0,1%, solución de Denhardt 5 x, ADN de esperma de salmón sonificado (50 µg/ml), SDS al 0,1% y sulfato de dextrano al 10% a 42°C, con lavados a 42°C en SSC 0,2 x (cloruro sódico/citrato sódico) y formamida al 50% a 55°C, seguido de un lavado altamente riguroso que consiste en SSC 0,1 x que contiene EDTA a 55°C.

65 70 Las “condiciones moderadamente rigurosas” pueden identificarse como se describe por Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, New York: Cold Spring Harbor Press, 1989, e incluyen el uso de una solución de lavado y condiciones de hibridación (por ejemplo temperatura, fuerza iónica y % de SDS) menos rigurosas que las descritas anteriormente. Un ejemplo de condiciones moderadamente rigurosas es la incubación durante toda la noche a 37°C en una solución que comprende: formamida al 20%, SSC 5 x (NaCl 150 mM, citrato trisódico 15 mM), fosfato sódico 50 mM (pH 7,6), solución de Denhardt 5 x, sulfato de dextrano al 10% y 20 mg/ml de ADN de esperma de salmón cortado desnaturalizado, seguido por el lavado de los filtros en SSC 1 x a aproximadamente 37-50°C. El especialista en la técnica reconocerá cómo ajustar la temperatura, fuerza iónica, etc. cuando sea necesario para acomodar factores tales como la longitud de la sonda y similares.

ES 2 314 378 T3

En el contexto de la presente invención, la referencia a “al menos uno”, “al menos dos”, “al menos cinco” etc. de los genes enumerados en cualquier conjunto génico particular significa una cualquiera o todas y cada una de las combinaciones de los genes enumerados.

5

B. Descripción Detallada

La práctica de la presente invención empleará, a menos que se indique otra cosa, técnicas convencionales de biología molecular (incluyendo técnicas de recombinación), microbiología, biología celular y bioquímica, que están dentro de la experiencia en la técnica. Estas técnicas se explican a fondo en la bibliografía, tal como, “Molecular Cloning: A Laboratory Manual”, 2nd edition (Sambrook *et al.*, 1989); “Oligonucleotide Synthesis” (M.J. Gait, ed., 1984); “Animal Cell Culture” (R.I. Freshney, ed., 1987); “Methods in Enzymology” (Academic Press, Inc); “Handbook of Experimental Immunology” 4th edition (D.M. Weir & C.C. Blackwell, eds., Blackwell Science Inc, 1987); “Gene Transfer Vectors for Mammalian Cells” (J.M. Miller & M.P. Calos, eds., 1987); “Current Protocols in Molecular Biology” (F.M. Ausubel *et al.*, eds., 1987); y “PCR: The Polymerase Chain Reaction”, (Mullis *et al.*, eds., 1994).

1. Reacción en cadena de la polimerasa (PCR)

El fin de la reacción en cadena de la polimerasa (PCR) es hacer copias de un gen para proporcionar mayores cantidades de ácido nucleico para uso posterior. La PCR es un proceso basado en una enzima polimerasa especializada (por ejemplo la ADN polimerasa de Taq), que puede sintetizar una cadena complementaria a una cadena de ADN determinada en una mezcla que contiene los cuatro dNTP (dATP, dCTP, dGTP, dTTP) y dos cebadores oligonucleotídicos que flanquean la secuencia diana a amplificar. Los dos cebadores oligonucleotídicos se usan para generar un amplicón típico de una reacción de PCR. Un tercer nucleótido, o sonda, está diseñado para detectar la secuencia de nucleótidos localizada entre los dos cebadores de PCR. Aunque el diseño de la sonda puede diferir, en el método de PCR TaqMan® las señales de la sonda se controlan por la proximidad de un colorante fluorescente indicador y un colorante fluorescente de inactivación. Cualquier emisión inducida por láser desde el colorante indicador se inactiva por el colorante de inactivación cuando los dos colorantes se localizan próximos entre sí como están en la sonda. Durante la reacción de amplificación, la enzima polimerasa (por ejemplo la ADN polimerasa de Taq) escinde la sonda de una manera dependiente de la plantilla. Los fragmentos de la sonda resultante se disocian en solución, y la señal procedente del colorante indicador liberado está exenta del efecto de inactivación del segundo fluoróforo. Se libera una molécula de colorante indicador por cada nueva molécula sintetizada y la detección del colorante indicador no inactivado proporciona la base para la interpretación cuantitativa de los datos.

35

El material de partida para la PCR puede ser ADN, ADNc, ARNm o cualquier otro polinucleótido que necesite amplificarse. Como la PCR requiere ADN monocatenario como plantilla, si el material de partida es ADN bicatenario necesita desnaturizarse para producir ADN monocatenario.

Como el ARN no puede servir como plantilla para la PCR, si el material de partida es ARN, la primera etapa es la transcripción inversa de la plantilla de ARN en ADNc, seguido de su amplificación exponencial en una reacción de PCR. Esta versión de PCR se denomina generalmente PCR transcriptasa inversa (RT-PCR). Las dos transcriptas más comúnmente usadas son la transcriptasa inversa del virus de la mieloblastosis aviar (AMV-RT) y la transcriptasa inversa del virus de la leucemia murina de Moloney (MMLV-RT). La etapa de transcripción inversa se ceba típicamente usando cebadores específicos, hexámeros aleatorios o cebadores oligo-dT, dependiendo de las circunstancias. Por ejemplo, el ARN extraído de una muestra de tejido (por ejemplo FPET) se puede transcribir de forma inversa usando un kit de PCR de ARN GeneAmp (Perkin Elmer, CA, USA), siguiendo las instrucciones del fabricante. El ADNc obtenido puede usarse después como una plantilla en la siguiente reacción de PCR.

50

Aunque la etapa de PCR puede usar una variedad de ADN polimerasas dependientes de ADN termoestables, típicamente emplea la ADN polimerasa de Taq, que tiene una actividad nucleasa 5'-3' pero carece de una actividad endonucleasa correctora 3'-5'. De este modo, la PCR TaqMan® típicamente utiliza la actividad nucleasa 5' de la polimerasa de Taq o Tth para hidrolizar una sonda de hibridación unida a su amplicón diana, pero puede usarse cualquier enzima con actividad nucleasa 5' equivalente. En este caso, la sonda se diseña para ser no-extensible por la enzima ADN polimerasa de Taq. La RT-PCR TaqMan® puede realizarse usando un equipo comercialmente disponible tal como, por ejemplo, el ABI PRISM 7700TM Sequence Detection SystemTM (Perkin-Elmer-Applied Biosystems, Foster City, CA, USA), o Lightcycler (Roche Molecular Biochemicals, Mannheim, Germany). En una realización preferida, el procedimiento de la nucleasa 5' se realiza en un dispositivo de PCR cuantitativo en tiempo real tal como el ABI PRISM 7700TM Sequence Detection SystemTM. El sistema consiste en un termociclador, un láser, un dispositivo de carga acoplada (CCD), una cámara y un ordenador. El sistema amplifica muestras en un formato de 96 pocillos en un termociclador. Durante la amplificación, la señal fluorescente inducida por láser se recoge en tiempo real a través de cables de fibra óptica para los 96 pocillos y se detecta en el CCD. El sistema incluye un programa informático para ejecutar el instrumento y para el análisis de los datos.

65

Los datos de ensayo de la nucleasa 5' inicialmente se expresan como Ct, o el ciclo umbral. Como se ha señalado anteriormente, los valores de fluorescencia se registran durante cada ciclo y representan la cantidad de producto amplificado hasta ese momento en la reacción de amplificación. El momento en el que la señal fluorescente se registra por primera vez como estadísticamente significativa es el ciclo umbral (Ct).

ES 2 314 378 T3

Para minimizar errores y el efecto de la variación de una muestra a otra, la RT-PCR se realiza normalmente usando un patrón interno. El patrón interno ideal se expresa en un nivel constante entre diferentes tejidos, y no se ve afectado por el tratamiento experimental. Son ARN frecuentemente usados para normalizar los patrones de la expresión génica ARNm para los genes básicos administradores (“housekeeping genes”) gliceraldehído-3-fosfato-deshidrogenasa (GAPDG) y β -actina.

5 Para detalles adicionales de la PCR cuantitativa en tiempo real véase también Held *et al.*, *Genome Research* 6: 986-994 (1996). La PCR se describe en las Patentes de Estados Unidos Nº 4.683.202, 4.683.195, 4.965.188 y 5.075.216.

10

2. Intrones y Corte y Empalme de ARN

15

16

17

18

19

20

La mayoría de los genes en eucariotas superiores contienen más de 100.000 pares de nucleótidos, algunos contienen más de 2 millones de pares de nucleótidos. Esta longitud es significativamente mayor que la de la secuencia de nucleótidos requerida para codificar una proteína de tamaño medio (300-400 aminoácidos), que está en el orden de aproximadamente 1000 nucleótidos. La mayor parte de la longitud extra consta de secuencias no codificantes (intrones) que interrumpen las secuencias codificantes (exones) dentro de la secuencia génica. La mayoría de los genes de eucariotas superiores que codifican ARNm, ARNt y algunos que codifican ARNr se interrumpen por secuencias intrónicas. Los genes para el ARNm típicamente tienen de 0 a 60 intrones; mientras que los genes para el ARNt típicamente incluyen 0 ó 1 intrón.

25

26

27

28

29

30

Cuando se transcribe ARNm a partir de ADN, al principio se transcriben tanto las secuencias de los exones como las secuencias de los intrones en el denominado ARN nuclear heterogéneo (ARNhn) o ARN inmaduro o pre-ARNm. Sin embargo, antes de que el ARN salga del núcleo, las secuencias intrónicas a menudo se eliminan del ARNm transcrita como resultado de un proceso conocido como corte y empalme del ARN. El proceso de eliminación de intrones implica un proceso preciso de formación de bucles controlado por una secuencia específica de nucleótidos contigua a los exones. Casi todos los intrones pueden identificarse por secuencias consenso específicas. Las dos primeras bases de un intrón son siempre GU, mientras que las dos últimas bases son siempre AG, pero los sitios de corte y empalme 5' y 3' típicamente tienen secuencias consenso que se extienden más allá de los pares GU y AG. El corte y empalme del ARNm tiene lugar en una partícula llamada espliceosoma, mientras que el ARNt y el ARNr se cortan y empalan por mecanismos que no implican espliceosomas.

35

36

37

38

39

40

Los intrones son típicamente mucho más largos que los exones (secuencias que están presentes en el ARNm). Un exón eucariota medio es de aproximadamente 150 nucleótidos de longitud, mientras que un único intrón humano puede ser tan largo como de cerca de 500.000 nucleótidos, aunque típicamente es de aproximadamente 2000-4000 nucleótidos. En general, un gen eucariota contiene muchas más secuencias intrónicas que de exones, como se ilustra en la siguiente tabla (Molecular Biology of the Cell, Bruce Alberts *et al.*, eds., 3^a edición, Garland Publishing Company, New York, N.Y., 1994, p.340):

45

TABLA 1

Gen	Tamaño del gen (nucleótidos $\times 10^3$)	Tamaño del ARNm (nucleótidos $\times 10^3$)	Número de Intrones
β -globina	1,5	0,6	2
Insulina	1,7	0,4	2
Proteinasa C	11	1,4	7
Albúmina	25	2,1	14
Catalasa	34	1,6	12
Receptor LDL	45	5,5	17
Factor VIII	186	9	25
Tiroglobulina	300	8,7	36

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

En una realización particular de la presente invención, las secuencias intrónicas en un gen de interés están sometidas a un proceso de selección para identificar la secuencia o secuencias de ARN intrónicas que se coexpresan con secuencias de ARN de exones (es decir, ARNm) del mismo gen. Dichas secuencias intrónicas seleccionadas, cuya expresión se correlaciona con la expresión de secuencias de exones, tienen propiedades especialmente deseables como potenciales marcadores de diagnóstico: (1) a causa de su comportamiento técnico favorable (específicamente)

te, la optimización de la especificidad y sensibilidad de ensayo); y (2) cualquier importancia biomédica asociada al nivel de ARNm del gen también está asociada a los niveles celulares de secuencias intrónicas. Por ejemplo, es probable que altos niveles de una especie de ARNm que codifica un factor de crecimiento potente estén probablemente correlacionados con altos índices de crecimiento de una célula. Las secuencias intrónicas que tienen niveles celulares que correlacionan con niveles del ARNm de este mismo gen tienen la misma probabilidad de correlacionarse con altos índices de crecimiento de una célula. Dichas secuencias intrónicas seleccionadas pueden usarse para explorar muestras de tejido valiosas para buscar correlaciones clínicas y significado de diagnóstico, predicción o pronóstico.

Un proceso ejemplar para seleccionar secuencias intrónicas que se coexpresan con el ARNm del mismo gen es el siguiente. Brevemente, para cualquier gen de interés, se ensaya un conjunto de tejidos pertinentes de una población de pacientes de interés para medir los niveles de un conjunto de secuencias intrónicas y de ARNm. Después se seleccionan las secuencias intrónicas que se ha descubierto que tienen el mayor coeficiente de correlación de Pearson para la coexpresión con secuencias de ARN de exones (ARNm). El número de pacientes estudiados en este proceso es preferiblemente al menos superior a 50 y más preferiblemente de al menos aproximadamente 100.

En un ejemplo específico, el asunto biomédico de interés trata de pacientes con cáncer de mama y el gen de interés puede ser el marcador de crecimiento tumoral Ki-67. En este caso, se usan tumores de 50 o más pacientes de cáncer de mama para medir los niveles de ARNm de Ki-67 y los niveles de las secuencias de múltiples intrones de Ki-67 y se seleccionan los intrones que tienen el mayor coeficiente de correlación de Pearson para la co-expresión con ARN de exones.

Una ventaja de este proceso es que la selección de la secuencia intrónica preferida puede realizarse con muestras de tejido relativamente fáciles de obtener y abundantes (por ejemplo, muestras que carecen de registros clínicos asociados valiosos). Dado que dicho tejido puede proporcionar grandes cantidades de ARN para explorar, será posible detectar señales de expresión génica incluso a partir de sondas subóptimas. Después pueden usarse ensayos muy sensibles y específicos basados en las secuencias intrónicas seleccionadas para explorar muestras de tejido valiosas, por ejemplo, muestras asociadas a una información clínica importante, tal como la recurrencia de la enfermedad, muerte o la respuesta a fármacos terapéuticos o regímenes de tratamiento definidos.

30

3. Perfilado de la expresión génica usando intrones basados en conjuntos cebador/sonda de la PCR

En el momento actual, los cebadores y sondas de la PCR se han diseñado en base a la secuencia de ARNm o ADNc, sin considerar las secuencias intrónicas. De hecho, los intrones se consideran normalmente un material “de embalaje” que se elimina durante el corte y empalme y generalmente se degrada de forma rápida.

La presente invención está basada en el descubrimiento experimental inesperado de que los ARN de los intrones pueden detectarse fácilmente por RT-PCR, incluso usando ARN altamente degradado procedente de muestras de tejido fijado embebido en parafina. En particular, se ha descubierto que en el perfil de la expresión génica para un gen determinado, las señales de RT-PCR de conjuntos sonda/cebador basados en intrón pueden ser grandes, o más grandes, que las señales de RT-PCR basadas en exón. Aunque este descubrimiento se confirma por unos pocos descubrimientos recientes con ciertas especies de ARNm, no está de acuerdo con la opinión predominante de que los intrones se degradan muy rápidamente después del corte y empalme (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]).

También inesperados, los descubrimientos experimentales subyacentes a la presente invención indican que el ARN intrónico puede usarse para perfilar la expresión génica, porque las cantidades en tejido de las secuencias de intrones y exones expresadas tienden a correlacionarse. Este resultado es inesperado porque existen escasas o ninguna prueba de que la proporción de la constante de velocidad global para la síntesis y la renovación de las secuencias de intrones y exones transcritas son similares. De hecho, la bibliografía científica proporciona pruebas de la complejidad de la renovación de pre-ARNm y de intrones procesados (sometidos a cortes y empalme). Por ejemplo, puede existir pre-ARNm en múltiples grupos cinéticos (Elliot and Rosbash, *Exp. Cell Res.* 229:181-8 [1996]), con subpopulaciones que contienen ARN intrónicos que no están eficazmente empalmados y se transportan al citoplasma como especies de ARNm “inmaduros”, donde pueden degradarse a diferente velocidad que las secuencias de ARN intrónicas nucleares (Wang *et al.*, *Proc. Natl. Acad. Sci. USA* 94:4360-5 [1997]). Existen pruebas de que ciertos ARN intrónicos procesados por corte y empalme entran en el citoplasma en estructura de lazo (Clement *et al.*, *ARN* 5:206-20 [1999]).

Finalmente, los datos presentados en este documento indican que las secuencias intrónicas pueden servir como marcadores moleculares de diagnóstico o pronóstico. Examinando cuatro ARNm que previamente habían demostrado ser de carácter pronóstico en cánceres, se demuestra que sus secuencias intrónicas correspondientes son también de pronóstico, y en las mismas direcciones que las secuencias de exones transcritas parentales (es decir, de pronóstico tanto positivo como negativo).

65

En resumen, el enfoque de la invención se ha demostrado como sigue. El documento WO 0378662 describe un conjunto de genes que predicen la probabilidad de recurrencia de cáncer de mama. En este estudio, se midieron los niveles de las secuencias de exón transcritas en muestras de tejido de cáncer de mama fijado embebido

en parafina de 146 pacientes usando conjuntos de cebador/sonda de PCR basados en exón. En el estudio descrito en este documento, se crearon ensayos de RT-PCR para medir los niveles de las secuencias de intrón transcritas dentro de cuatro de los genes marcadores identificados anteriormente, y después se usaron para explorar el ARN de 60 muestras de biopsia fijadas embebidas en parafina (que representaban 60 pacientes diferentes, un sub-

5 conjunto de los pacientes evaluados en el estudio anterior). Los datos presentados en los ejemplos mostrados a continuación demuestran que para cada gen los intrones y los exones están co-expresados, y que los intrones predicen el riesgo de recurrencia de la enfermedad recurrente según lo predicho por los datos anteriores basados en exones.

10

3. Diseño de Cebadores y Sondas de PCR Basado en Intrones

De acuerdo con un aspecto de la presente invención, se diseñan cebadores y sondas de PCR en base a secuencias intrónicas presentes en el gen a amplificar. Por consiguiente, el primer paso en el diseño del cebador/sonda es la 15 delineación de las secuencias intrónicas dentro de los genes. Esto se puede hacer mediante el programa informático disponible públicamente, tal como el programa informático ADN BLAT desarrollado por Kent, W.J., *Genome Res.* 12 (4): 656-64 (2002), o por el programa informático BLAST incluyendo sus variaciones. Los pasos posteriores siguen métodos bien establecidos de diseño del cebadores y sondas de la PCR.

20 Para evitar señales no específicas, es importante enmascarar las secuencias repetitivas dentro de los intrones al diseñar los cebadores y las sondas. Esto puede lograrse fácilmente usando el programa Repeat Masker disponible online a través del Colegio Universitario de Medicina Baylor, que explora secuencias de ADN contra una biblioteca de elementos repetitivos y devuelve una secuencia problema en la cual los elementos repetitivos están enmascarados. Las secuencias intrónicas enmascaradas pueden usarse entonces para diseñar secuencias de cebador y sonda usando 25 cualquiera de los paquetes de diseño de cebador/sonda comerciales o disponibles al público de otra manera, tales como Primer Express (Applied Biosystems); MGB assay-by-desing (Applied Biosystems); Primer3 (Steve Rozen and Helen J. Skaletsky (2000); Primer3 en WWW para usuarios generales y para biólogos programadores. En: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386).

30

Los factores más importantes considerados en el diseño de cebadores de PCR incluyen la longitud del cebador, la temperatura de fusión (*T_m*), y el contenido de G/C, la especificidad, las secuencias de cebador complementarias y la secuencia 3' terminal. En general, los cebadores de PCR óptimos son generalmente de 17-30 bases de longitud, y contienen aproximadamente el 20-80%, tal como, por ejemplo, son aproximadamente el 50-35 60% de bases G+C. Las *T_m* típicamente preferidas están entre 50 y 80°C, por ejemplo, aproximadamente de 50 a 70°C.

40 Para más directrices para los diseños de cebador y sonda de la PCR véase, por ejemplo, Dieffenbach, C.W. et al., "General Concepts for PCR Primer Design" en: *PCR Primer, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, 1995, pp. 133-155; Innis y Gelfand, "Optimazation of PCRs" en: *PCR Protocols, A guide to Methods and Applications*, CRC Press, London, 1994, pp. 5-11; y Plasterer, T.N. Primerselect: Primer and probe desin. *Methods Mol. Biol.* 70: 520-527 (1997).

45

4. Aplicaciones

50 Los métodos de la presente invención y, específicamente, de los cebadores y sondas de PCR basados en intrones de este documento, tienen utilidad en todos los campos donde se requiere la amplificación de un ácido nucleico (incluyendo ARN, ADN y, en general, cualquier oligo- y polinucleótido) que representa un gen o un fragmento de gen. De esta manera, los cebadores y las sondas de PCR diseñados de acuerdo con la presente invención pueden 55 usarse para amplificar genes individuales, o genes múltiples presentes en una muestra biológica con el fin de perfilar la expresión génica por cualquier metodología incluyendo, sin limitación, el perfilado de la expresión génica basado en una PCR cuantitativa (por ejemplo, RT-PCR cuantitativa), y análisis de micromatrizes y ensayos basados en perlas.

55

Por ejemplo, en una realización específica de la técnica de micromatriz, se aplicaron insertos de clones de ADNc amplificados por PCR a un sustrato en una matriz densa. Preferiblemente se aplican al sustrato al menos 10.000 secuencias de nucleótidos. Los genes de la micromatriz, inmovilizados en el microchip en al menos 10.000 elementos cada uno, son adecuados para la hibridación en condiciones rigurosas. Pueden generarse sondas de ADNc marcadas con fluorescencia a través de la incorporación de nucleótidos fluorescentes por transcripción inversa del ARN extraídos de tejidos de interés. Las sondas de ADNc marcadas aplicadas al chip hibridan con especificidad en cada punto del ADN en la matriz. Después del lavado riguroso para eliminar las sondas no unidas específicamente, se explora el chip mediante microscopía láser confocal o mediante otro método de detección, tal como una cámara fotográfica CCD. La cuantificación de la hibridación de cada elemento de la matriz permite valorar la abundancia 60 del ARNm correspondiente. Con fluorescencia de color dual, se hibridan sondas de ADNc marcadas por separado generadas a partir de dos fuentes de ARN por pareja a la matriz. De este modo se determina de forma simultánea la abundancia relativa de los transcripto procedentes las dos fuentes que corresponden a cada gen específico. La escala miniaturizada de la hibridación proporciona una evaluación conveniente y rápida del patrón de expresión para una 65

gran cantidad de genes. Estos métodos han demostrado tener la sensibilidad requerida para detectar transcripciones poco frecuentes, que se expresan en algunas copias por célula, y para detectar de forma reproducible al menos aproximadamente dos veces más diferencias en los niveles de expresión (*Schena et al., Proc. Natl. Acad. Sci. USA* 93 (2): 106-149 (1996)). El análisis en micromatriz puede realizarse por equipos comercialmente disponibles, siguiendo 5 los protocolos del fabricante, tales como el uso de la tecnología Affymetrix GenChip, o la tecnología micromatriz de Agilent.

Un aspecto importante de la presente invención es usar la amplificación génica basada en intrones como parte del perfilado de la expresión génica para ajustar a los pacientes los mejores fármacos o combinaciones farmacológicas 10 y para proporcionar información de pronóstico. Por ejemplo, la expresión medida de genes en tejido canceroso (por ejemplo, tejido de biopsia de cáncer de mama) puede usarse para predecir la probabilidad de la supervivencia a largo plazo, sin la enfermedad de pacientes después de la cirugía y/u otra terapia para el cáncer, o predecir la respuesta del paciente a un enfoque terapéutico particular. Para este fin es típicamente necesario corregir (normalizando) tanto 15 las diferencias en la cantidad de ARN ensayado como la variabilidad en la calidad del ARN usado. Por lo tanto, los ensayos de la invención miden e incorporan generalmente la expresión de ciertos genes normalizadores, incluyendo 20 genes de referencia bien conocidos, tales como GAPDH y Cyp1. Como alternativa, la normalización puede basarse en la señal media o mediana (Ct) de todos los genes ensayados o de un subconjunto grande de los mismos (enfoque de normalización global). Sobre la base del gen-por-gen, la cantidad normalizada medida de un ARNm de un tumor de un paciente se compara con la cantidad encontrada en un cáncer, por ejemplo, el conjunto de referencia 25 de tejido de cáncer de mama. El número (N) de tejidos con cáncer, por ejemplo cáncer de mama, en este conjunto de referencia debe ser suficientemente alto para asegurar que diferentes conjuntos de referencia (en su conjunto) se comportan esencialmente de la misma manera. Si se cumple esta condición, la identidad de los tejidos con cáncer de mama individuales presentes en un conjunto particular no tendrá un impacto significativo en las cantidades relativas 30 de los genes ensayados. Generalmente, el conjunto de referencia de tejidos con cáncer de mama consiste en al menos aproximadamente 30, preferiblemente al menos aproximadamente 40 muestras diferentes de tejido con cáncer de mama fijado, embebido en parafina (FPE). A menos que se indique otra cosa, los niveles de expresión normalizados para cada ARNm/tumor/ensayado/paciente se expresarán como un porcentaje del nivel de expresión medido en el conjunto de referencia. Más específicamente, el conjunto de referencia de un número suficientemente alto (por ejemplo, 40) de tumores ofrece una distribución de niveles normalizados de cada especie de ARNm. El nivel medido en una muestra de tumor particular a analizar caerá en algún percentil dentro de este intervalo, que puede determinarse por métodos bien conocidos en la técnica.

En un estudio de la fase II de la expresión génica en muestras de tejido fijado embebido en parafina de carcinoma de mama invasivo, se descubrió que la sobreexpresión de cualquiera de los siguientes genes en el tejido de cáncer de mama indicaba una probabilidad reducida de supervivencia sin recurrencia del cáncer después de la cirugía: FOXM1; PRAME; SKT15, Ki-67; CA9; NME1; SURV; TFRC; YB-1; RPS6KB1; Src; Chk1; CCNB1; Chk2; 35 CDC25B; CYP3A4; EpCAM; VEGFC; hENT1; BRCA2; EGFR; TK1; VDR.

En el mismo estudio, la sobreexpresión de cualquiera de los siguientes genes en cáncer de mama indica un mejor 40 pronóstico para la supervivencia sin recurrencia del cáncer después de la cirugía: Blc12; CEGP1; GSTM1; PR; BBC3; GATA3; DPYD; GSTM3; ID1; EstR1; p27; XIAP; IGF1R; AK055699; P13KC2A; TGFB3; BAG1; pS2; WISP1; HNF3A; NFKBp65.

En este mismo estudio de la fase II de la expresión génica en muestras de tejido fijado embebido en parafina de 45 cáncer de mama ER positivo, la sobreexpresión de los siguientes genes fue indicativa de una probabilidad reducida de supervivencia sin recurrencia de cáncer después de la cirugía: PRAME; FOXM1; EPHX1; HIF1A; VEGFC; Ki-67; VDR; NME1. Algunos de estos genes (PRAME; FOXM1; VEGFC; Ki-67; VDR y NME1) también se identificaron como indicadores de un mal pronóstico en el análisis previo, sin limitarse al cáncer de mama ER-positivo. Se descubrió 50 que la sobreexpresión de los genes restantes (EPHX1 y HIF1A) era un indicador negativo de la supervivencia libre de la enfermedad sólo en cáncer de mama ER-positivo. Se descubrió que la sobreexpresión de los siguientes genes en cáncer ER-positivo era indicativa de un pronóstico mejor para la supervivencia sin recurrencia del cáncer después de la cirugía: Bcl-2; DIABLO; IGF1R; GSTM3. De los últimos genes, Bcl-2; IGFR1 y GSTM3 también se han identificado como indicadores de un buen pronóstico en el análisis previo, sin limitarse al cáncer de mama ER-positivo. La sobreexpresión de DIABLO parecía ser un indicador positivo de la supervivencia libre de enfermedad sólo en 55 cáncer de mama ER-positivo. Para detalles adicionales, véase el documento WO0378662.

Los estudios anteriormente descritos se realizaron esencialmente según lo descrito en el ejemplo 2 más abajo, excepto porque la amplificación génica se estudió usando amplicones basados en exones. Para detalles adicionales, véase el documento WO0378662. Según lo atestiguado por los datos que se muestran en el ejemplo 2, los datos obtenidos usando amplicones basados en intrones muestran una excelente correlación con los datos anteriores, y proporcionan típicamente la ventaja añadida del aumento de sensibilidad.

Los resultados del estudio de fase II anterior de carcinoma ductal de mama invasivo se sometieron a análisis multivariante en forma de etapas, usando el Modelo de Riesgo Proporcional de Cox usando la siguiente ecuación:

65

$$RR = \exp[\text{coef}(\text{genA}) \times Ct(\text{genA}) + \text{coef}(\text{genB}) \times Ct(\text{genB}) + \text{coef}(\text{genC}) \times Ct(\text{genC}) + \dots].$$

ES 2 314 378 T3

En esta ecuación, los coeficientes para los genes que son indicadores de resultado beneficioso son números positivos y los coeficientes para los genes que son indicadores de un resultado desfavorable son números negativos. Los valores de "Ct" en la ecuación son los ΔCt , es decir reflejan la diferencia entre el valor de Ct normalizado promedio para una población y el Ct normalizado medido para el paciente en cuestión. El convenio usado en el análisis ha sido que los ΔCt por debajo y por encima del promedio de la población tienen signos positivos y signos negativos, respectivamente (reflejando la mayor o menor abundancia de ARNm). El riesgo relativo (RR) calculado resolviendo esta ecuación indicaba si el paciente tiene una posibilidad aumentada o reducida de supervivencia a largo plazo sin recurrencia de cáncer.

En un análisis multivariante, usando un conjunto de interrogantes incluyendo un número reducido de genes, se han identificado los siguientes conjuntos de diez genes que tienen un valor de predicción particularmente grande de la supervivencia del paciente sin recurrencia del cáncer después de la extirpación quirúrgica del tumor primario.

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.

En un análisis multivariante, usando un conjunto de interrogantes incluyendo todos los genes identificados, se han identificado los siguientes conjuntos de diez genes que tienen un valor de predicción particularmente grande de la supervivencia del paciente sin recurrencia del cáncer después de la extirpación quirúrgica del tumor primario.

1. Bcl2, PRAME, cyclinG1, FOXM1, NFKBp65, TS, XIAP, Ki67, CYP3A4, p27;
2. FOXM1, cyclinG1, XIAP, Contig51037, PRAME, TS, Ki67, PDGFRa, p27, NFKBp65;
3. PRAME, FOXM1, cyclinG1, XIAP, Contig51037, TS, Ki6, PDGFRa, p27, NFKBp65;
4. Ki67, XIAP, PRAME, hENT1, contig51037, TS, CD9, p27, ErbB3, cyclinG1;
5. STK15, XIAP, PRAME, PLAUR, p27, CTSL, CD18, PREP, p53, RPS6KB1;
6. GSTM1, XIAP, PRAME, p27, Contig51037, ErbB3, GSTp, EREG, ID1, PLAUR;
7. PR, PRAME, NME1, XIAP, PLAUR, cyclinG1, Contig51037, TERC, EPHX1, DR5;
8. CA9, FOXM1, cyclinG1, XIAP, TS, Ki67, NFKBp65, CYP3A4, GSTM3, p27;
9. TFRC, XIAP, PRAME, p27, Contig51037, ErbB3, DPYD, TERC, NME1, VEGFC;
10. CEGP1, PRAME, hENT1, XIAP, Contig51037, ErbB3, DPYD, NFKBp65, ID1, TS.

Usando el mismo enfoque de análisis multivariante para el cáncer de mama ER-positivo, se han identificado los siguientes conjuntos de diez genes que tienen un valor de predicción particularmente grande de la supervivencia del paciente sin recurrencia del cáncer después de la extirpación quirúrgica del tumor primario.

1. PRAME, p27, IGFBP2, HIF1A, TIMP2, ILT2, CYP3A4, ID1, EstR1, DIABLO;
2. Contig51037, EPHX1, Ki67, TIMP2, cyclinG1, DPYD, CYP3A4, TP, AIB1, CYP2C8;
3. Bcl2, hENT1, FOXM1, Contig51037, cyclinG1, Contig46653, PTEN, CYP3A4, TIMP2, AREG;
4. HIF1A, PRAME, p27, IGFBP2, TIMP2, ILT2, CYP3A4, ID1, EstR1, DIABLO;
5. IGF1R, PRAME, EPHX1, Contig51037, cyclinG1, Bcl2, NME1, PTEN, TBP, TIMP2;
6. FOXM1, Contig51037, VEGFC, TBP, HIF1A, DPYD, RAD51C, DCR3, cyclinG1, BAG1;

ES 2 314 378 T3

7. EPHX1, Contig51037, Ki67, TIMP2, cyclinG1, DPYD, CYP3A4, TP, AIB1, CYP2C8;
8. Ki67, VEGFC, VDR, GSTM3, p27, upa, ITGA7, rhoC, TERC, Pin1;
- 5 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;
15. CDH1, PRAME, VEGFC, HIF1A, DPYD, TIMP2, CYP3A4, EstR1, RBP4, p27;
16. IGFBP3, PRAME, p27, Bcl2, XIAP, EstR1, Ki67, TS, Src, VEGF;
17. GSTM3, PRAME, p27, IGFBP3, XIAP, FGF2, hENT1, PTEN, EstR1, APC;
20. hENT1, Bcl2, FOXM1, Contig51037, CyclinG1, Contig46653, PTEN, CYP3A4, TIMP2, AREG;
18. STK15, VEGFC, PRAME, p27, GCLC, hENT1, ID1, TIMP2, EstR1, MCP1;
25. 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;
30. CCNB1, PRAME, VEGFC, HIF1A, hENT1, GCLC, TIMP2, ID1, p27, upa;
22. ID1, PRAME, DIABLO, hENT1, p27, PDGFRa, NME1, BIN1, BRCA1, TP;
23. FBXO5, PRAME, IGFBP3, p27, GSTM3, hENT1, XIAP, FGF2, TS, PTEN;
35. GUS, HIA1A, VEGFC, GSTM3, DPYD, hENT1, FBXO5, CA9, CYP, KRT18;
25. Bclx, Bcl2, hENT1, Contig51037, HLAG, CD9, ID1, BRCA1, BIN1, HBEGF.
- 40.

En vista de la excelente correlación entre los resultados del perfilado de la expresión génica basada en exones y basada en intrones (véase el ejemplo 2), se espera que los mismos conjuntos génicos tengan valor de pronóstico similar cuando el perfilado de la expresión génica se basa en la cuantificación de las señales de RT-PCR de conjuntos de cebador/sonda basados en intrones.

Serán evidentes detalles adicionales de la invención en los siguientes ejemplos no limitantes.

50 Ejemplo 1

Diseño y uso de conjuntos de cebador/sonda de PCR específicos de intrones

Se extrajo ARN de muestras de biopsia de cáncer de mama fijadas en formalina, embebidas en parafina (FPET) (Clionomics Biosciences Inc., Pittsfield, MA) como sigue. Se cortaron tres secciones de 10 µm y se colocaron en un tubo de 1,5 ml. Se retiró la parafina por extracción con xileno (1 ml, 3 veces) seguido de lavado con etanol (1 ml, dos veces). Se aisló el ARN de los bloques de tejido seccionados usando el kit de purificación MasterPure™ (Epicentre, Madison, WI). Se cuantificó el ARN por el método de fluorescencia de RiboGreen (Molecular Probes). Después se agruparon veinte muestras de ARN FPET y se usaron como se describe a continuación.

60 Se sintetizó ADNc de primera cadena usando la Transcriptasa Inversa Omniscript de Qiagen con los cebadores específicos de los genes agrupados (cebadores inversos mostrados en la figura 2), hexámeros aleatorios e inhibidor de la RNasa, usando el ARN FPET agrupado (400 ng). Se realizó también una reacción sin transcriptasa inversa (RT) con 150 ng de ARN FPET agrupado, suficiente ARN para realizar la amplificación Taqman en 5 ng/pocillo.

65

ES 2 314 378 T3

TABLA 2

Reactivos	RT	No RT	Con. Final
	Vol. (μ l)	Vol (μ l)	
Tampón RT 10X	4	2	1X
mezcla de dNTP, 5 mM cada dNTP	4	2	500 μ M cada uno
Hexámero aleatorio ABI, 50 μ M	1	0,5	1,25 μ M
Grupo GSP, 1 μ M	2	1	50 nM
Inhibidor de RNasa ABI, 20 U/ μ l	1	1	20 U/rxn
RT Omniscript 4 U/ μ l	2	0	8U ó 0 U/rxn
Agua sin nucleasa	10	5,5	
ARN FPET agrupado (164 ng/ μ l)	16	8	65,6 ng/ μ l
Vol. Total	40	20	
Condiciones de reacción	37° C, 60 min 93° C, 5 min		

Ensayo de TaqMan

Se realizaron ensayos de TaqMan para la lista de 48 genes en pocillos triplicados con un volumen de reacción de 25 μ l y una entrada de ARN de 5 ng por ensayo. Se realizó una reacción "no RT" para cada gen en un solo pocillo como control para verificar que se estaban midiendo las señales de ARN en lugar de ADN. La cuantificación en tiempo real se realizó en el ABI 7700 usando los siguientes parámetros:

Condiciones del ciclo: 95°C, 10 minutos para un ciclo, 95°C, 20 seg. seguido de 60°C, 45 seg, 40 ciclos.

Volumen de reacción: 25 μ l.

Ajuste de la capa de tinción: FAM, (la referencia pasiva es ROX).

55

Resultados

Los conjuntos de cebador-sonda de Taqman específicos de intrón se diseñaron en base a intrones enmascarados de los genes CEGP1, FOXM1, PRAME y STK15. Para delinear secuencias intrónicas dentro de los genes, la secuencia de referencia NCBI para cada ARNm (NM_XXXXXX) se alineó con el genoma humano usando el programa de herramienta de alineación de tipo BLAST (BLAT) disponible en el sitio de recursos de genoma on-line de la Universidad de Santa Cruz (<http://genome.ucsc.edu>). Después se buscaron las secuencias intrónicas para secuencias repetitivas usando el programa Repeat Masker disponible on-line a través del Colegio Universitario de Medicina Baylor (<http://searchlauncherbcm.tmc.edu/seq-util/seq-util.html>). Las secuencias repetidas, tales como repeticiones Alu, se identifican por este programa y enmascaran. Es importante excluir estas secuencias antes de diseñar los cebador-sondas porque producen señales fuertes, no específicas. Las secuencias intrónicas enmascaradas (figuras 1A-M) después se usaron para diseñar conjuntos de cebador-sonda usando Primer Express (ABI). Otros pro-

ES 2 314 378 T3

gramas adecuados para los conjuntos de cebador-sonda incluyen, por ejemplo, el programa de diseño de cebador-sonda más novedoso para ensayos MGB por diseño (ABI). Los amplicones para cada conjunto de cebador-sonda están delineados en negrita en la figura 1. Cada conjunto específico de cebador-sonda se muestra en la figura 2.

5 Los conjuntos de cebador-sonda específicos de intrón (genes de ensayo) se utilizaron junto con su correspondiente conjunto de cebador-sonda específico de exón (genes de referencia) en experimentos estándar del perfil de expresión génica de Taqman usando grupos de ARN FPET. La expresión normalizada se calculó mediante la fórmula $2^{\Delta Ct}$ donde ΔCt es la diferencia entre los Ct del conjunto de cebador-sonda del gen de ensayo y los conjuntos de cebador-sonda de los genes de referencia [Ct (referencia)- Ct (prueba)].

Ejemplo 2

15 *Un estudio de Fase II de la expresión génica en tumores de mama premalignos y malignos*

Se diseñó y se realizó un estudio de expresión génica con el objeto principal de caracterizar molecularmente la expresión génica en muestras de tejido fijado, embebido en parafina de carcinoma ductal de mama invasivo y para explorar la correlación entre dichos perfiles moleculares y la supervivencia libre de enfermedad.

20 *Diseño del estudio*

Se realizaron ensayos moleculares en tejidos tumorales primarios de mama, embebidos en parafina, fijados con formalina obtenidos de 60 pacientes individuales a los que se les había diagnosticado cáncer de mama. Todos los pacientes se sometieron a cirugía con diagnóstico de carcinoma invasivo de mama. Los pacientes se incluyeron en el estudio solamente si la valoración histopatológica, realizada según lo descrito en el apartado Materiales y Métodos, indicaba cantidades adecuadas de tejido tumoral y patología homogénea.

30 *Materiales y Métodos*

Cada bloque de tumor representativo se caracterizó por histopatológica estándar para el diagnóstico, valoración semi-cuantitativa de la cantidad de tumor, y grado del tumor. Se prepararon un total de 6 secciones (de 10 micrómetros de grosor cada una) y se colocaron en dos tubos de microcentrifuga Costar Brand (tubos de polipropileno de 1,7 mL, transparentes; 3 secciones en cada tubo). Si el tumor constituía menos del 30% del total del área de la muestra, haberse diseccionado de forma tosca por el patólogo, usando microdisección sin precisión, poniendo el tejido tumoral directamente dentro del tubo Costar.

40 Cuando se obtuvo más de un bloque del tumor como parte del procedimiento quirúrgico, todos los bloques del tumor se sometieron a la misma caracterización, según lo descrito anteriormente, y se usó para el análisis el bloque más representativo de la patología.

45 *Análisis de la expresión génica*

Se extrajo y purificó ARNm de muestras fijadas, embebidas en parafina y se preparó para el análisis de la expresión génica como se ha descrito anteriormente.

50 Los ensayos moleculares de la expresión génica cuantitativa se realizaron por la RT-PCR, usando el Sequence Detection System™ ABI PRISM 7900™ (Perkin-Elmer Biosistemas Applied, Foster City, CA, USA). ABI PRISM 7900™ consiste en un termociclador, un láser, un dispositivo de carga acoplada (CCD), una cámara y un ordenador. El sistema amplifica muestras en un formato de 384 pocillos en un termociclador. Durante la amplificación, la señal fluorescente inducida por láser se recoge en tiempo real a través de cables ópticos de fibra para los 384 pocillos, y se detecta en el CCD. El sistema incluye el programa informático para el funcionamiento del instrumento y para analizar los datos.

60 *Análisis y resultados*

65 Se analizó el tejido tumoral para la expresión de 48 secuencias de ARN diferentes que representaban los productos de 37 genes diferentes. Se normalizaron los valores del ciclo umbral (Ct) para cada paciente en base a la mediana de todos los genes para ese paciente particular. Los datos del resultado clínico estaban disponibles para todos los pacientes a partir de una revisión de datos registrados y de los gráficos de los pacientes seleccionados.

ES 2 314 378 T3

Los resultados se clasificaron como:

- 0 muerte debida al cáncer de mama o por causa desconocida o el paciente sigue vivo con recurrencia de cáncer de mama;
5
1 vivo sin recurrencia de cáncer de mama o muerto debido a una causa distinta al cáncer de mama.

El análisis se realizó mediante:

10 El análisis de la relación entre la expresión génica normalizada y el tiempo transcurrido hasta el resultado (0 ó 1 según se ha definido anteriormente) donde se censuraron los pacientes que estaban vivos sin recurrencia del cáncer de mama o que murieron debido a una causa distinta del cáncer de mama. Este enfoque se usó para evaluar el impacto del pronóstico de genes individuales y también conjuntos de genes múltiples.

15 Para cada gen se definió un Modelo de Riesgo Proporcional de Cox (véase, por ejemplo, Cox, D.R. y Oakes, D. (1984), *Analysis of Survival Data*, Chapman and Hall, London, new York) con el tiempo transcurrido hasta la recurrencia o hasta la muerte como variable dependiente, y el nivel de expresión del gen como variable independiente. Se identificaron los genes que tienen un valor-p < 0,05 en el modelo de Cox. Para cada gen, el modelo de Cox proporciona el riesgo relativo (RR) de recurrencia o muerte para un cambio unitario en la expresión del gen. Se puede elegir repartir a los pacientes en subgrupos a cualquier valor umbral de la expresión medida (en la escala Ct), donde todos los pacientes con valores de expresión por encima del umbral tienen mayor riesgo, y todos los pacientes con valores de expresión por debajo del umbral tienen menor riesgo, o viceversa, dependiendo de si el gen es un indicador de pronóstico malo (RR>1,01) o bueno (RR<1,01). De esta manera, cualquier valor umbral definirá subgrupos de pacientes con un riesgo aumentado o reducido respectivamente.

25 La tabla 3, a continuación, muestra la correlación por parejas de la expresión (presentada por coeficientes de correlación) entre los intrones y exones ensayados para los genes CEGP1, FOXM1, PRAME y STK15. Para dos de los cuatro genes, CEGP1 y PRAME, se encontró intrones que producían coeficientes de correlación [para la co-expresión con sus exones respectivos] por encima de 0,90. En el caso de STK15, un intrón se correlacionó con la expresión del exón con un coeficiente de correlación ~0,80. Para FOXM1, las correlaciones de la expresión intrón:exón eran significativamente más bajas. En este último caso, sin embargo, parece probable que la expresión real pueda estar altamente correlacionada pero no sea detectable por una razón técnica. La expresión del exón FOXM1 en muchos pacientes estaba por debajo del umbral de detección del ensayo, que potencialmente previene la detección de las altas correlaciones que puedan existir. Si esta hipótesis es correcta, los intrones FOXM1 se registrarían aún como marcadores de pronóstico clínico negativo según lo demostrado previamente para FOXM1. Como se demuestra más adelante, este resultado ocurre.

30 Las figuras 3, 4 y 5 muestran la correlación por parejas de la expresión de los ARN ensayados contra los ARN de los exones CEGP1, PRAME y STK15. Como se muestra, los intrones respectivos de estos genes ofrecieron las 35 correlaciones más altas. Es notable que la lista de 48 genes incluyera genes seleccionados, mediante varias estrategias basadas en bioinformática, genes con una probabilidad particular de correlacionarse en expresión CEGP1, PRAME, STK15 y FOXM1. Estas estrategias no basadas en intrones fueron las más satisfactorias en el caso de STK15, ya que varios genes candidatos tenían coeficientes de correlación de expresión en el intervalo de 0,6-0,7.

45

(Tabla pasa a página siguiente)

50

55

60

65

ES 2 314 378 T3

TABLA 3

Correlaciones entre la expresión de intrones, exones para cuatro genes

5

		Coeficiente de correlación de la expresión {R}				
		CEGP1intrón1.1	CEGP1intrón3.1	CEGP1intrón4.1	CEGP1intrón5.1	CEGP1.2
10	CEGP1intrón1.1	1,00				
	CEGP1intrón3.1	0,89	1,00			
15	CEGP1intrón4.1	0,97	0,82	1,00		
	CEGP1intrón5.1	0,91	0,87	0,88	1,00	
	CEGP1.2	0,91	0,80	0,90	0,87	1,00
20		FOXM1intrón3.3	FOXM1intrón5.1	FOXM1intrón7.1	FOXM1.1	
	FOXM1intrón3.3	1,00				
	FOXM1intrón5.1	0,48	1,00			
25	FOXM1intrón7.1	0,54	0,73	1,00		
	FOXM1.1	0,44	0,33	0,38	1,00	
		Intrón1,1STK15	Intrón2,1STK15	Intrón4,1STK15	STK15.2	
30	STK15intrón1.1	1,00				
	STK15intrón2.1	0,78	1,00			
	STK15intrón4.1	0,69	0,74	1,00		
	STK15.2	0,63	0,70	0,78	1,00	
40		PRAMEintrón2.1	PRAME.3			
	PRAMEintrón2.1	1,00				
	PRAME.3	0,97	1,00			
45						

La tabla 4, a continuación, muestra el impacto sobre la supervivencia de los pacientes de la expresión de los exones e intrones de CEGP1, FOXM1, PRAME, y STK15. Todos los exones parentales tenían un impacto significativo en el riesgo relativo [RR], como se ha determinado previamente, excepto en el caso de FOXM1. Como el presente estudio evaluó a 60 pacientes del grupo original de 146 pacientes, el marcador FOXM1 puede haber perdido significado debido al riesgo estadístico de examinar un conjunto de datos reducido. Muy notablemente, para los cuatro genes ensayados, la expresión de los intrones afectaba significativamente al RR, y en la misma dirección que los exones parentales.

55

60

65

ES 2 314 378 T3

TABLA 4

Resultados del modelo de Cox para 60 pacientes con Cáncer de mama

5

Correlaciones de Pronóstico

	Gen	Coef	RR=exp(coef)	se (coef)	z	p
10	CEGP1.2	-0,202	0,817	0,050	-4,024	0,00006
	CEGP1intrón1.1	-0,329	0,720		-3,771	0,00016
15	CEGP1intrón3.1	-0,261	0,770	0,078	-3,335	0,00085
	CEGP1intrón4.1	-0,275	0,760		-3,774	0,00016
20	CEGP1intrón5.1	-0,312	0,732	0,082	-3,817	0,00014
	FOXM1.1	0,175	1,192		1,289	0,19700
	FOXM1intrón3.3	0,304	1,355		2,523	0,01160
	FOXM1 intrón5.1	0,514	1,673		2,639	0,00832
25	FOXM1 intrón7.1	0,546	1,726	0,182	2,993	0,00276
	PRAME.3	0,125	1,133		2,294	0,02180
	PRAMEintrón2.1	0,125	1,133		2,397	0,01650
	STK15.2	0,692	1,998	0,201	3,450	0,00056
35	STK15intrón1.1	0,357	1,429		2,400	0,01640
	STK15intrón2.1	0,391	1,479		2,536	0,01120
	STK15intrón4.1	0,410	1,506		3,084	0,00204

- 45 Existe una percepción común de que los niveles constantes de las secuencias de exones transcritas exceden enormemente de las secuencias de intrones transcritas (Sharp *et al.* *Ann Rev. Biochem.* 55: 1119-50 [1986]). Sin embargo, el examen de la expresión de exones e intrones de CEGP1, FoxM1, PRAME y STK15, usando TaqMan [TM] RT-PCR para ensayar el ARN de tejido de cáncer de mama fijado y embebido en parafina, ha demostrado que las intensidades de la señal de los intrones y de los exones estaban en el mismo intervalo, y en todos los casos en el intervalo de detección útil del ensayo [datos no mostrados]. La detección del ARN intrónico en este estudio es aún más notable porque el tejido usado se fijó en formalina, que degrada el ARN, y por tanto limita sustancialmente la capacidad para detectar el ARN (T.E. Godfrey *et al.* *J. Mol. Diag.* 2: 84-91 [2000]). En el caso de CEGP1 tres de los intrones ensayados produjeron señales más bajas y uno una señal más alta que el exón. En el caso de FOXM1, cinco de nueve intrones ensayados produjeron señales más altas que el exón. En el caso de PRAME, las intensidades de la señal del intrón y el exón ensayados eran casi idénticas. Finalmente, para STK15 todos los intrones tenían intensidades de señal que eran de 1/4 a 1/20 las del exón, pero todavía estaban en el intervalo útil del ensayo. De esta manera, estos resultados indican que los niveles de estado estacionario de los intrones expresados son adecuados para el uso de los ARN, de los intrones como marcadores moleculares.
- 50
- 55
- 60 Aunque la presente invención se ha descrito con referencia a lo que se considera que son realizaciones específicas, se entiende que la descripción no está limitada a éstas. Por ejemplo, aunque la descripción incluye diversos genes y conjuntos de genes asociados al cáncer de mama, se contemplan genes y conjuntos de genes similares y métodos que conciernen a otros tipos de cáncer.

65

REIVINDICACIONES

1. Un método para controlar la expresión génica en una muestra biológica, que comprende:

5 (a) proporcionar un polinucleótido complementario a una secuencia de ARN intrónica dentro de un gen diana, donde la expresión de dicha secuencia de ARN intrónica se correlaciona con la expresión de una secuencia de ARNm exónica dentro de dicho gen;

10 (b) hibridar dicho polinucleótido con dicha secuencia de ARN intrónica para formar un complejo de polinucleótido-ARN intrónico; y

15 (c) detectar el complejo de polinucleótido-ARN intrónico.

2. El método de la reivindicación 1 en el que dicha secuencia de ARN intrónica se selecciona identificando secuencias intrónicas que se co-expresan con el ARNm de dicho gen diana, y seleccionando una secuencia de ARN intrónica que tiene el coeficiente de correlación más alto para dicha co-expresión.

3. El método de la reivindicación 1 en el que dicha secuencia de ARN intrónica es de al menos 50 bases de nucleótidos de longitud.

4. El método de la reivindicación 1 en el que dicha secuencia de ARN intrónica es de al menos 55 bases de nucleótidos de longitud.

5. El método de la reivindicación 1 en el que dicha secuencia de ARN intrónica es de al menos 60 bases de nucleótidos de longitud.

6. El método de la reivindicación 1 en el que dicha muestra biológica es una muestra de tejido.

30 7. El método de la reivindicación 6 en el que dicho tejido es un tejido tumoral.

8. El método de la reivindicación 7 en el que dicho tumor es cáncer.

9. El método de la reivindicación 8 en el que dicho cáncer se selecciona del grupo consistente en cáncer de mama, cáncer de colon, cáncer de pulmón, cáncer de próstata, cáncer hepatocelular, cáncer gástrico, cáncer pancreático, cáncer cervical, cáncer ovárico, cáncer de hígado, cáncer de vejiga, cáncer del tracto urinario, cáncer de tiroides, cáncer renal, carcinoma, melanoma y cáncer cerebral.

40 10. El método de la reivindicación 6 en el que dicha muestra dicha de tejido es una muestra de tejido fijada, embebida en cera.

11. El método de la reivindicación 10 en el que dicha muestra de tejido es una muestra de tejido fijada en formalina, embebida en parafina.

45 12. El método de la reivindicación 10 en el que dicho ARN exónico está fragmentado.

13. El método de la reivindicación 1 en el que dicha muestra biológica es un fluido biológico.

14. El método de la reivindicación 1 en el que dicha hibridación se realiza en condiciones rigurosas.

50 15. El método de la reivindicación 1 que comprende adicionalmente la etapa de cuantificar la expresión de dicho ARN intrónico.

16. El método de la reivindicación 1 en el que dicho polinucleótido es un oligonucleótido monocatenario.

55 17. El método de la reivindicación 16 en el que dicho oligonucleótido monocatenario es una sonda de PCR.

18. El método de la reivindicación 17 en el que dicha sonda de PCR se diseña en base a una secuencia única dentro de dicha secuencia de ARN intrónico.

60 19. El método de la reivindicación 16 en el que dicho oligonucleótido monocatenario es un cebador de PCR.

20. El método de la reivindicación 19 en el que dicho cebador de PCR se diseña en base a una secuencia única dentro de dicha secuencia de ARN intrónico.

65 21. El método de la reivindicación 1 en el que se controla la expresión de más de un gen diana.

22. El método de la reivindicación 21 que comprende controlar simultáneamente al menos 50 genes diana.

ES 2 314 378 T3

23. El método de la reivindicación 21 que comprende controlar simultáneamente al menos 100 genes diana.
24. El método de la reivindicación 21 que comprende controlar simultáneamente al menos 500 genes diana.
- 5 25. El método de la reivindicación 21 que comprende controlar simultáneamente al menos 10.000 genes diana.
26. El método de la reivindicación 25 en el que las secuencias de ARN intrónico correspondientes a una pluralidad de dichos genes diana se disponen como una matriz inmovilizada sobre una superficie sólida.
- 10 27. El método de la reivindicación 1 en el que dicho gen diana se selecciona de los genes enumerados en la Figura 6.
28. Un método para preparar una molécula de oligonucleótido monocatenaria para la amplificación de un gen diana que comprende:
- 15 (a) identificar al menos una secuencia de intrón dentro de dicho gen diana, cuya expresión correlaciona con la expresión de una secuencia de ARNm exónico dentro de dicho gen diana; y
- 20 (b) preparar una molécula de oligonucleótido monocatenaria que corresponde al menos a una parte de la secuencia de intrón transcrita.
29. El método de la reivindicación 28 que comprende identificar secuencias de repetición dentro de dicha secuencia de intrón antes de preparar dicha molécula de oligonucleótido monocatenaria.
- 25 30. El método de la reivindicación 29 en el que dichas secuencias de repetición se enmascaran antes de preparar dicha molécula de oligonucleótido.
31. El método de la reivindicación 28 en el que dicha molécula de oligonucleótido monocatenaria es un cebador de PCR.
- 30 32. El método de la reivindicación 31 en el que dicho cebador de PCR es un cebador directo diseñado para comprender secuencias 5' de una secuencia diana dentro de dicha secuencia de intrón transcrita.
33. El método de la reivindicación 31 en el que dicho cebador de PCR es un cebador inverso diseñado para complementar secuencias 5' de una secuencia diana aguas abajo del cebador directo dentro de dicha secuencia de intrón transcrita.
34. El método de la reivindicación 32 ó 33 en el que dicha secuencia diana es de al menos 50 bases de nucleótidos de longitud.
- 40 35. El método de la reivindicación 32 ó 33 en el que dicha secuencia diana es de al menos 55 bases de nucleótidos de longitud.
36. El método de la reivindicación 32 ó 33 en el que dicha secuencia diana es de al menos 60 bases de nucleótidos de longitud.
- 45 37. El método de la reivindicación 31 en el que dicho cebador de PCR es de 17-30 bases de nucleótidos de longitud.
38. El método de la reivindicación 31 en el que dicho cebador de PCR contiene aproximadamente del 20% al 80% de bases G+C.
- 55 39. El método de la reivindicación 31 en el que la temperatura de fusión (Tm) de dicho cebador de PCR está entre aproximadamente 50°C y aproximadamente 70°C.
40. El método de la reivindicación 28 en el que dicha molécula de oligonucleótido monocatenaria es una sonda de PCR.
41. El método de la reivindicación 40 en el que dicha sonda de PCR está diseñada para comprender o complementar una parte interna de una secuencia diana dentro de la secuencia de intrón transcrita.
42. El método de la reivindicación 41 en el que dicha sonda de PCR se marca con un tinte fluorescente indicador y un tinte fluorescente de inactivación.
- 65 43. El método de la reivindicación 28 en el que dicho gen diana se selecciona de los genes enumerados en la Figura 6.

ES 2 314 378 T3

44. Un método para amplificar ARN intrónico en una muestra de tejido fijada embebida en parafina que representa al menos un gen de interés, que comprende las etapas de:

5 (a) poner en contacto el ADN obtenido por transcripción inversa del ARN intrónico, cuya expresión correlaciona con la expresión de un ARN exónico correspondiente, con al menos un conjunto de cebadores y sonda de PCR que corresponden a dicho ARN intrónico; y

(b) realizar la amplificación por PCR.

10 45. El método de la reivindicación 44 en el que dichos cebadores y sonda de PCR se diseñan en base a una secuencia única dentro de dicho ARN intrónico.

15 46. El método de la reivindicación 44 en el que dicha muestra comprende ARN fragmentado que representa múltiples genes de interés.

15 47. El método de la reivindicación 46 en el que dicha muestra se pone en contacto con un grupo de cebadores y sondas de PCR diseñados en base a secuencias únicas dentro de los intrones, cuya expresión se correlaciona con la expresión de los exones correspondientes, presentes en dichos genes de interés.

20 48. El método de la reivindicación 47 en el que dicho grupo comprende al menos uno de los conjuntos de cebador/sonda basados en intrón mostrados en la Figura 2.

25 49. El método de la reivindicación 47 en el que dicho grupo comprende al menos un conjunto de cebador directo o inverso o sonda mostrado en la Figura 2.

50 50. El método de la reivindicación 46 en el que dicha muestra de tejido es de una biopsia tumoral.

51. El método de reivindicación 50 en el que dicha biopsia tumoral se obtiene de un paciente humano.

30 52. El método de la reivindicación 51 en el que dicho tumor es cáncer de mama.

53. El método de la reivindicación 51 en el que dicho tumor es cáncer pulmón.

35 54. El método de la reivindicación 51 en el que dicho tumor es cáncer colorectal.

55 55. El método de la reivindicación 51 que comprende adicionalmente la etapa de determinar los niveles de expresión de los transcritos de ARN de dichos genes de interés o sus productos de expresión.

40 56. El método de la reivindicación 55 en el que la expresión diferencial de dichas transcripciones de ARN o sus productos se correlaciona con una respuesta del paciente pronosticada al tratamiento o con la supervivencia del paciente.

45 57. El método de la reivindicación 44 en el que dicho gen de interés se selecciona de los genes enumerados en la Figura 6.

58. El método de la reivindicación 52 en el que el tumor es cáncer de mama invasivo, y el método comprende:

50 (1) determinar los niveles de expresión de los transcritos del ARN o los productos de expresión de un gen o conjunto de genes seleccionado de grupo consistente en:

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

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

55 (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;

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

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

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

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

ES 2 314 378 T3

- (j) Ki67, XIAP, PRAME, hENT1, contig51037, TS, CD9, p27, ErbB3, cyclinG1;
- (k) STK15, XIAP, PRAME, PLAUR, p27, CTSL, CD18, PREP, p53, RPS6KB1;
- 5 (l) GSTM1, XIAP, PRAME, p27, Contig51037, ErbB3, GSTp, EREG, ID1, PLAUR;
- (m) PR, PRAME, NME1, XIAP, PLAUR, cyclinG1, Contig51037, TERC, EPHX1, DR5;
- 10 (n) CA9, FOXM1, cyclinG1, XIAP, TS, Ki67, NFKBp65, CYP3A4, GSTM3, p27;
- (o) TFRC, XIAP, PRAME, p27, Contig51037, ErbB3, DPYD, TERC, NME1, VEGFC; y
- (p) CEGP1, PRAME, hENT1, XIAP, Contig51037, ErbB3, DPYD, NFKBp65, ID1, TS

15 en dicha muestra;

(2) someter los datos obtenidos en la etapa (a) a análisis estadístico; y

20 (3) determinar si la probabilidad de supervivencia a largo plazo de dicho paciente, sin cáncer de mama recurrente ha aumentado o disminuido.

25 59. El método de reivindicación 58 en el que los niveles de expresión de dichas transcripciones de ARN o de sus productos de expresión se normalizan contra los niveles de expresión de todas las transcripciones del ARN o de sus productos de expresión en dicha muestra de tejido de cáncer de mama, o de un conjunto de referencia de transcripciones del ARN o de sus productos.

60. El método de la reivindicación 52 en el que el tumor es cáncer de mama invasivo receptor estrógeno (ER)-positivo, y el método comprende

30 (1) determinar los niveles de expresión de las transcripciones del ARN o de los productos de expresión de un gen o conjunto de genes seleccionados del grupo consistente en:

- (a) PRAME, p27, IGFBP2, HIF1A, TIMP2, ILT2, CYP3A4, ID1, EstR1, DIABLO;
- 35 (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;
- 40 (e) IGF1R, PRAME, EPHX1, Contig51037, cyclinG1, Bcl2, NME1, PTEN, TBP, TIMP2;
- (f) FOXM1, Contig51037, VEGFC, TBP, HIF1A, DPYD, RAD51C, DCR3, cyclinG1, BAG1;
- 45 (g) EPHX1, Contig51037, Ki67, TIMP2, cyclinG1, DPYD, CYP3A4, TP, AIB1, CYP2C8;
- (h) Ki67, VEGFC, VDR, GSTM3, p27, upa, ITGA7, rhoC, TERC, Pin1;
- 50 (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;
- 55 (l) DIABLO, Ki67, hENT1, TIMP2, ID1, p27, KRT19, IGFBP2, TS, PDGFB;
- (m) p27, PRAME, IGFBP2, HIF1A, TIMP2, ILT2, CYP3A4, ID1, EstR1, DIABLO;
- 60 (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;
- 65 (q) hENT1, Bcl2, FOXM1, Contig51037, CyclinG1, Contig46653, PTEN, CYP3A4, TIMP2, AREG;
- (r) STK15, VEGFC, PRAME, p27, GCLC, hENT1, ID1, TIMP2, EstR1, MCP1;

ES 2 314 378 T3

- (s) NME1, PRAM, p27, IGFBP3, XIAP, PTEN, hENT1, Bcl2, CYP3A4, HLAG;
- (t) VDR, Bcl2, p27, hENT1, p53, PI3KC2A, EIF4E, TFRC, MCM3, ID1;
- 5 (u) EIF4E, Contig51037, EPHX1, cyclinG1, Bcl2, DR5, TBP, PTEN, NME1, HER2;
- (v) CCNB1, PRAME, VEGFC, HIF1A, hENT1, GCLC, TIMP2, ID1, p27, upa;
- 10 (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; y
- 15 (z) Bclx, Bcl2, hENT1, Contig51037, HLAG, CD9, ID1, BRCA1, BIN1, HBEGF;

(2) someter los datos obtenidos en la etapa (1) al análisis estadístico; y

20 (3) determinar si la probabilidad de la supervivencia a largo plazo de dicho paciente, sin cáncer de mama recurrente ha aumentado o disminuido.

25 61. El método de la reivindicación 60 en el que los niveles de expresión de dichas transcripciones del ARN o de sus productos de expresión se normalizan contra los niveles de expresión de todas las transcripciones del ARN o de sus productos de expresión en dicha muestra del tejido de cáncer de mama, o de un conjunto de referencia de transcripciones del ARN o de sus productos.

62. El método de la reivindicación 52 en el que el tumor es cáncer de mama, y el método comprende

30 (1) determinar los niveles de expresión de las transcripciones del ARN o de los productos de expresión de un gen o conjunto de genes seleccionados del grupo consistente en: 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; y NFKBp65, normalizados contra los niveles de expresión de todas las transcripciones del ARN o de sus productos en dicha muestra, o de un conjunto de referencia de transcripciones del ARN o de sus productos de expresión;

35 (2) someter los datos obtenidos en la etapa (a) al análisis estadístico; y

40 (3) determinar si la probabilidad de supervivencia a largo plazo de dicho paciente, sin cáncer de mama recurrente ha aumentado o disminuido.

45 63. El método de la reivindicación 62 en el que los niveles de expresión de dichas transcripciones del ARN o de sus productos de expresión se normalizan contra los niveles de expresión de todas las transcripciones del ARN o de sus productos de expresión en dicha muestra de tejido de cáncer de mama, o de un conjunto de referencia de transcripciones del ARN o de sus productos.

50 64. El método de la reivindicación 52 en el que el tumor es cáncer de mama invasivo, y el método comprende determinar los niveles de expresión de las transcripciones del ARN o de los productos de expresión de un gen o conjunto de genes del grupo consistente en:

- (a) p53BP2, Bcl2, BAD, EPHX1, PDGFR β , DIABLO, XIAP, YB1, CA9, y KRT8;
- (b) GRB7, CD68, TOP2A, Bcl2, DIABLO, CD3, ID1, PPM1D, MCM6, y WISP1;
- 55 (c) PR, p53BP2, PRAME, DIABLO, CTSI, IGFBP2, TIMP1, CA9, MMP9, y COX2;
- (d) CD68, GRB7, TOP2A, Bcl2, DIABLO, CD3, ID1, PPM1D, MCM6, y WISP1;
- (e) Bcl2, p53BP2, BAD, EPHX1, PDGFR β , DIABLO, XIAP, YB1, CA9, y KRT8;
- 60 (f) KRT14, KRT5, PRAME, p53BP2, GUS1, AIB1, MCM3, CCNE1, MCM6, e ID1;
- (g) PRAME, p53BP2, EstR1, DIABLO, CTSI, PPM1D, GRB7, DAPK1, BBC3, y VEGFB;
- (h) CTSI2, GRB7, TOP2A, CCNB1, Bcl2, DIABLO, PRAME, EMS1, CA9, y EpCAM;
- (i) EstR1, p53BP2, PRAME, DIABLO, CTSI, PPM1D, GRB7, DAPK1, BBC3, y VEGFB;

ES 2 314 378 T3

- (j) Chkl, PRAME, p53BP2, GRB7, CA9, CTSL, CCNB1, TOP2A, tamaño del tumor, e IGFBP2;
- (k) IGFBP2, GRB7, PRAME, DIABLO, CTSL, β -Catenina, PPM1D, Chkl, WISP1, y LOT1;
- 5 (l) HER2, p53BP2, Bcl2, DIABLO, TIMP1, EPHX1, TOP2A, TRAIL, CA9, y AREG;
- (m) BAG1, p53BP2, PRAME, IL6, CCNB1, PAI1, AREG, tamaño del tumor, CA9, y Ki67;
- 10 (n) CEGP1, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, STK15, y AKT2, y FGF18;
- (o) STK15, p53BP2, PRAME, IL6, CCNE1, AKT2, DIABLO, cMet, CCNE2, y COX2;
- (p) KLK10, EstR1, p53BP2, PRAME, DIABLO, CTSL, PPM1D, GRB7, DAPK1, y BBC3;
- 15 (q) AIB1, p53BP2, Bcl2, DIABLO, TIMP1, CD3, p53, CA9, GRB7, y EPHX1
- (r) BBC3, GRB7, CD68, PRAME, TOP2A, CCNB1, EPHX1, CTSL, GSTM1, y APC;
- (s) CD9, GRB7, CD68, TOP2A, Bcl2, CCNB1, CD3, DIABLO, ID1, y PPM1D;
- 20 (t) EGFR, KRT14, GRB7, TOP2A, CCNB1, CTSL, Bcl2, TP, KLK10, y CA9;
- (u) HIF1 α , PR, DIABLO, PRAME, Chkl, AKT2, GRB7, CCNE1, TOP2A, y CCNB1;
- (v) MDM2, p53BP2, DIABLO, Bcl2, AIB1, TIMP1, CD3, p53, CA9, y HER2;
- 25 (w) MYBL2, p53BP2, PRAME, IL6, Bcl2, DIABLO, CCNE1, EPHX1, TIMP1, y CA9;
- (x) p27, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, STK15, AKT2, e ID1;
- 30 (y) RAD51, GRB7, CD68, TOP2A, CIAP2, CCNB1, BAG1, IL6, FGFR1, y p53BP2;
- (z) SURV, GRB7, TOP2A, PRAME, CTSL, GSTM1, CCNB1, VDR, CA9; y CCNE2;
- 35 (aa) TOP2B, p53BP2, DIABLO, Bcl2, TIMP1, AIB1, CA9, p53, KRT8, y BAD;
- (ab) ZNF217, GRB7, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, APC4, y β -catenina,

en una muestra de tejido de cáncer de mama obtenida del dicho paciente,

40 (2) someter los datos obtenidos en la etapa (a) a análisis estadístico; y

(3) determinar si la probabilidad de dicha supervivencia a largo plazo ha aumentado o disminuido.

45 65. El método de la reivindicación 64 en el que los niveles de expresión de dichas transcripciones del ARN o de sus productos de expresión se normalizan contra los niveles de expresión de todas las transcripciones del ARN o sus productos de expresión en dicha muestra de tejido de cáncer de mama, o de un conjunto de referencia de las transcripciones del ARN o de sus productos.

50 66. Un método de pronóstico de enfermedad a largo plazo - supervivencia libre de un paciente de cáncer usando ARN intrónico para medir la expresión génica usando una serie que comprende una pluralidad de polinucleótidos, que hibridan con genes diana de interés, inmovilizados en una superficie sólida, en la que al menos uno de dichos polinucleótidos comprende una secuencia basada en intrón cuya expresión correlaciona con la expresión de una secuencia de exón correspondiente.

55 67. El método de la reivindicación 66 en el que todos dichos polinucleótidos comprenden secuencias intrónicas.

68. El método de la reivindicación 66 que comprende al menos uno de los amplicones mostrados en las Figuras 1 A-M, o el complemento del mismo.

60 69. El método de reivindicación 66 que comprende dos o más de los amplicones mostrados en las Figuras 1 A-M, o el complemento del mismo.

70. El método de la reivindicación 66 que comprende todos los amplicones mostrados en las Figuras 1 A-M, o el complemento del mismo.

ES 2 314 378 T3

71. El método de la reivindicación 66 que comprende el uso de secuencias de polinucleótidos basadas en intrón que hibridan al menos un gen de interés del grupo consistente en: FOXM1, PRAME, Bcl2, STK15, CEGP1, Ki-67, GSTM1, PR, BBC3, NME1, SURV, GATA3, TFRC, YB-1, DPYD, CA9, Contig51037, RPS6K1 y Her2, en el que al menos el 80% de las secuencias de dicha serie está basada en intrón.

5 72. El método de la reivindicación 71 que comprende el uso de secuencias de polinucleótidos basadas en intrón que hibridan con al menos 5 de dichos genes.

10 73. El método de la reivindicación 71 que comprende el uso de secuencias de polinucleótidos basadas en intrón que hibridan con al menos 10 de dichos genes.

15 74. El método de la reivindicación 71 que comprende el uso de secuencias de polinucleótidos basadas en intrón que hibridan con todos dichos genes.

20 75. El método de la reivindicación 66 que comprende el uso de secuencias de polinucleótidos basadas en intrón que hibridan con al menos un gen de interés seleccionado del grupo consistente en: FOXM1, PRAME, Bcl2, STK15, CEGP1, Ki-67, GSTM1, CA9, PR, BBC3, NME1, SURV, GATA3, TFRC, YB-1, DPYD, GSTM3, RPS6KB1, Src, Chkl, 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 y DIABLO.

25 76. El método de la reivindicación 75 que comprende el uso de secuencias de polinucleótidos basadas en intrón que hibridan con al menos 5 de dichos genes.

30 77. El método de la reivindicación 75 que comprende el uso de secuencias de polinucleótidos basadas en intrón que hibridan con al menos 10 de dichos genes.

35 78. El método de la reivindicación 75 que comprende el uso de secuencias de polinucleótidos basadas en intrón que hibridan con todos dichos genes.

40 79. El método de la reivindicación 66 que comprende el uso de secuencias de polinucleótidos basadas en intrón que hibridan con al menos un conjunto de genes seleccionado del grupo consistente en:

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

65

ES 2 314 378 T3

80. El método de la reivindicación 66 que comprende el uso de secuencias de polinucleótidos basadas en intrón que hibridan con al menos un conjunto de genes seleccionados del grupo consistente en:

- (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, SENTI, 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; y
- (z) Bclx, Bcl2, hENT1, Contig51037, HLAG, CD9, ID1, BRCA1, BIN1, HBEGF.

81. El método de la reivindicación 66 que comprende el uso de secuencias de polinucleótidos basadas en intrón que hibridan con al menos un conjunto de genes seleccionados del grupo consistente en:

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

ES 2 314 378 T3

- (f) KRT14, KRT5, PRAME, p53BP2, GUS1, AIB1, MCM3, CCNE1, MCM6, e ID1;
- (g) PRAME, p53BP2, EstR1, DIABLO, CTSL, PPM1D, GRB7, DAPK1, BBC3, y VEGFB;
- 5 (h) CTSL2, GRB7, TOP2A, CCNB1, Bcl2, DIABLO, PRAME, EMS1, CA9, y EpCAM;
- (i) EstR1, p53BP2, PRAME, DIABLO, CTSL, PPM1D, GRB7, DAPK1, BBC3, y VEGFB;
- 10 (j) Chkl, PRAME, p53BP2, GRB7, CA9, CTSL, CCNB1, TOP2A, tamaño del tumor, e IGFBP2;
- (k) IGFBP2, GRB7, PRAME, DIABLO, CTSL, β -catenina, PPM1D, Chkl, WISP1, y LOT1;
- (l) HER2, p53BP2, Bcl2, DIABLO, TIMP1, EPHX1, TOP2A, TRAIL, CA9, y AREG;
- 15 (m) BAG1, p53BP2, PRAME, IL6, CCNB1, PAI1, AREG, tamaño del tumor, CA9, y Ki67;
- (n) CEGP1, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, STK15, y AKT2, y FGF18;
- 20 (o) STK15, p53BP2, PRAME, IL6, CCNE1, AKT2, DIABLO, cMet, CCNE2, y COX2;
- (p) KLK10, EstR1, p53BP2, PRAME, DIABLO, CTSL, PPM1D, GRB7, DAPK1, y BBC3;
- (q) AIB1, p53BP2, Bcl2, DIABLO, TIMP1, CD3, p53, CA9, GRB7, y EPHX1;
- 25 (r) BBC3, GRB7, CD68, PRAME, TOP2A, CCNB1, EPHX1, CTSL GSTM1, y APC;
- (s) CD9, GRB7, CD68, TOP2A, Bcl2, CCNB1, CD3, DIABLO, ID1, y PPM1D;
- (t) EGFR, KRT14, GRB7, TOP2A, CCNB1, CTSL, Bcl2, TP, KLK10, y CA9;
- 30 (u) HIF1 α , PR, DIABLO, PRAME, Chkl, AKT2, GRB7, CCNE1, TOP2A, y CCNB1;
- (v) MDM2, p53BP2, DIABLO, Bcl2, AIB1, TIMP1, CD3, p53, CA9, y HER2;
- (w) MYBL2, p53BP2, PRAME, IL6, Bcl2, DIABLO, CCNE1, EPHX1, TIMP1, y CA9;
- 35 (x) p27, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, STK15, AKT2, e ID1;
- (y) RAD5 1, GRB7, CD68, TOP2A, CIAP2, CCNB1, BAG1, IL6, FGFR1, y p53BP2;
- 40 (z) SURV, GRB7, TOP2A, PRAME, CTSL, GSTM1, CCNB1, VDR, CA9; y CCNE2;
- (aa) TOP2B, p53BP2, DIABLO, Bcl2, TIMP1, AIB1, CA9, p53, KRT8, y BAD; y
- 45 (ab) ZNF217, GRB7, p53BP2, PRAME, DIABLO, Bcl2, COX2, CCNE1, APC4, y β -catenina.

82. El método de la reivindicación 66 que comprende secuencias de polinucleótidos basadas en intrón que hibridan con al menos uno de los genes seleccionados del grupo consistente en: CD68; CTSL; FBXO5; SURV; CCNB1; MCM2; Chk1; MYBL2; HIF1A; cMET; EGFR; TS; STK15, IGRF1; 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; Chkl; MYBL2; HIF1A; cMET; EGFR; TS; y STK15.

83. El método de la reivindicación 66 que comprende el uso secuencias de polinucleótidos basadas en intrón correspondientes a al menos un gen seleccionado de los genes enumerados en la Figura 6.

55 84. El método de la reivindicación 66 que comprende el uso secuencias de polinucleótidos basadas en intrón correspondientes a una pluralidad de genes seleccionados de los genes enumerados en la Figura 6.

85. El método de la reivindicación 66 que comprende el uso de ambas secuencias de polinucleótidos basadas en 60 intrón y basadas en exón.

86. El método de la reivindicación 66 que comprende el uso de ambas secuencias de polinucleótidos basadas en intrón y basadas en exón que hibridan el mismo gen diana de interés.

65 87. El método de la reivindicación 66 en el que dicha serie comprende al menos 100 genes.

ES 2 314 378 T3

88. El método de la reivindicación 87 en el que dicha serie comprende al menos 100 genes en una sección de 100 μ .

5 89. El método de la reivindicación 87 en el que dicha serie comprende al menos 150 genes en una sección de 100 μ .

10 90. El uso de al menos un polinucleótido que comprende una secuencia basada en intrón en el que la expresión del intrón correlaciona con la expresión de una secuencia de exón correspondiente, para medir la expresión génica usando ARN intrónico en el pronóstico de cáncer.

15

20

25

30

35

40

45

50

55

60

65

<CEGP1 intrón 1, 1566 bases

MGB-CEGP1 int1.1

FIG.1A

>CEGP1 intrón 4 4985 bases

GTACCTCTGCCAGCTGTGGATGGGGCAGAGCCACATCTGAGACCCCTCT
CCCTTGACCGCGCACACACACACTGACTCTAGNNNNNNNNNNNNNNNNNN
NN
NN
NN
NN
NNNNATCTTACATAGAACATTTCAACATGACTAGATGTCAGGAGC
AATATACTGGATGATCTGCCAAGTTTCAAAAAGGTGCTGAAACCACA
GCACCAAGTATGAGCCTGCTCCCTGCTCTGGGTGGTAGGGAGGAGGCTGG
ATCCTTCCCATGCAAGACTTTCAATGAAGTGCCTGTTTCAGCCCCAAGC
TAGATCCGGCCCTCCATGTTTGCACTTTGAGCTCGAGGGGAGAAG
GGCTCCCTCCCTGGACTTCCGTGCTGGTTCTCGCTACGTCAAC
ATTTATCATCCCTGTAAATTGCCGAAACTCTCTCTGATGTC
TTCTCTCATCTCTTGCTTGAGTTATACCTTTCTGATCCCTG
TACTTAGTAGATTCTGAGAGGAAGGGCATTAGTACATGTCAGGCAATC
AGTTATTTTAACTGAATGTCATCCCTTAACCTCTCCCTGCTTTCTT

FIG. 1B

GAATGAATCCAGCAACCAGCAGACTGAGTGTGGAGTCAGGCAGCTCA
TAACCTGTAGGCAAAGAGCAAGAGGGTTTAAGAGAGACTCCAGAAAGT
ATGGGATATATAACCCCTTGCACTGTCTCTGGAAATAGGAATGACATCTG
TTTGTATTAACAAATGTCGTTAAGCAGTTGACAGCTCTGGAG
TGGGAGCTGGAGAGAGAACCTTGACTTCACAGAACCTGTTGGCTAAGGT
TTTAGGGCACAATATAGAAGGGTGTGGATTCTAGAGAAAGTGAACCAA
CCTTTTGTACTCGTGTGAAAACAGTGCCTACTAGTATTAGAGTGTCT
CATTGATAGAGAGCCAATGACAACCAAGTCCCTACTCTCAGAGATGTTT
AGAGTTACATTGCAAGAATGCAAAAGAACATAGGAACAGGTAATTAA
TAATAAAGTATAAAACTGAGCAGATGCTTGAAAGTATTCTAGGGTATGAA
AAGAATTCTTCAGGATGCTGGTAGGCAGCAGGATCTCAAAGAAATTAGTT
TTGAGATGAGGAGAGGAGGATGCTGGTAAACCACACGGCGAGTTACCTTGCT
GCCCTCTCATTTAGATGTGTGCCAGGCCCTGCAAGAACAGAACAGCTGT
TCCCCCTTCCCACCATCATACTACAAGGGTAAGCCTAATCAGAATTACTG
TATAACCTCAAAAAAAATTGTACAGCAGCTACCCACACCGAGCACANNNNN
NN
NNNGTCTGTTGCTGTTGCTGTGATGCTGTCAGAGGTTATGCCCTGA
GGGAGGGATCAAGGGAGTGGCTGAGGGTGTGCAAGAACAGAACAGATTCCGG
GGCATGTGGCCCGTACGAGGATGCCAAATGCCACAGTCACACTCACCTC
AGAAGGGTGGGATTGGTGGGGCAGAGAGGGGGCGTTGAATGTTTGAAGA
ATTATCTCAAGAGTATGTGAAAAAAATTGAGAATCTGATCATTCTATCT
GAACATTCTTAGGAGGATTCTCTTCTCTTACATTCTGATCAGC
TCTGGGTAAGACATGGCAGAGATAAGAGCGTGTGATCACCAGTTCCTGGG
GTCAGCAGGCTCTGATCCTGCATGCAATAGAGAGCTCAGTGTATTGGG
AGGCTCCAACCTGTTAGGAGAGCTGAGAACATCGTATCTCTGGGTGACA
GAATAAAATTCTCATGTCATTAATTGGCCTAGGGTGTGACTTAAATGACAT
ATACTTTCAAAATGTGGGGCTGATGGAGACCTAAGCAGAACAGATCTGTGG
GCCACCCCTTAGCCCTTGCCGCTCTCCAGGGCTCAGGATTCTGACCCAC
AGCCTAGTCACCTGTGCGCACACTGCTGTTTTCAG

MGB-CEGP1 int 3.1

>CEGP1 intrón 5, 2556 bases

MGB-CEGP1 int 4.1

FIG. 1C

GAGGGGACAGTTCACTGGTGAGGGACTTCTCCATTGTCCTCCCTCACA
 AAGCAGACTGCCACCCAAAGCTGTCAAGCCAAGGCTGGTGCCACCATC
 ACACCTCAAGCAACAGGTTCTGACATGCTTAGGGCCCTCGAAGTCAGG
 CTGTCCTGAGGGCTTCCAGTGAGCTAGCAGAGTGGAGACCAATTTC
 CCTCCAGATCTCGGAAGGAAGACCCAGACCTCCAAGACTCACCTGCGG
 GGCAGACCTCAACATTCTAGTCTTCAAGGAACAGTTGCTGAAGGG
 GGCGGGGGGGGGGCCTGTAAAGCTGTTTTAAAGATTTAAATGCT
 TTAAGATATCACTGCTCAAATAATATTGTTCTGNNNNNNNNNNNN
 NNN
 NNN
 AAAAACAAGACCTAGAACATTCACCAATAACAGCTGTTCAATTTC
 CATATTCTATTAGTTGTGTTCATATGCATACACAATTTCACATAGC
 TATAATCACAGGACAACACAAATATGTAATTAGTTCTTGAATTAGAAA
 AATTACAAGGGCTATGTAATGCAAACACTCCAAGCATATAAGAA
 AACATGCAGTTCCCGCCTCCGTTCCCTTGCCAGAGGTAACCACGGT
 AGCAGTTGATGAATAGATAGTTGTAGTTGGCTTTTCTTTTGGC
 CTATCATCAATACATTATAGTCAGATAATTACAGTTACTGTCA
 CGTTAATTGTGTCAGAACATCCTGTGATTATCTTCTCTAACTAAT
 CTAGATTGAATCTGTGAGAGAAATTCTGACATATATGTAACAAATTAAAT
 ATTGTCGTTTATTCCAGCATAAAGTGTATAGCATTCCAAAGCCCC
 AGTACAGCTGTTAAATAGGTAAACCTCTAGATAGAACAAAGCAGTAG
 TCTAGAACATCTTGGTATAATTCCCTTATATAATAAAAGTCTCTCCCC
 AACTCTCCATCTCCCTCTGTATGACTTTGTTAACCCATGTTTC
 AGCATTCTACAATTGTATTGTAACTATCTGCATACACAGACACCAAG
 GGTCTGACTTGGAGTTATGCTTCTGNNNNNNNNNNNNNNNNNN
 NNNNNNNNNNNNNNNNNNNNNNAATCAGAACATTCTGAGCAACAC
 AGCCCTGTGTTGTGAAATCTCAGTGTATTGATTGATTCATTGTC
 TGTCAG

>CEGP1 intrón 6, 716 bases

GTGAGTGGCAACCCCAACACTGAGTGAGGTCTGCACCAAGCCTGCCCTGTC
 CCTACCCCTACCCCTTAATGGTCTTAGCACAGATGCAGGCTGTTCTG
 TGCATTGCCCCCCCAGCAGGCCCTGTGCTGCTCGCATGCTACAGTGG
 AGTGGCTAGGCCCTGTGGGAAGGCCCTCTCTCCCTGTGACCTTGG
 AAGCCCTTCCCTCTCTGGACTAGGCTGCTCTAACGCTGGTATTCCA
 GAGACTGGCACAAACACCTCCAGGAGGCCAGGGCAGCACGAAGTTAGAGC
 TGTGTTATAATGATGCGGACTTCTGGCCAGCAGGAGGCCAGGGCCGTATAT
 TTCTGGGGATGCTGCCCTGCCCTCACGGTGTGCTTCACTAGCTC
 CATTAGAGGTTCCAGGCCAAGGCTCTTTCTCCCTGACTCAGGG
 ACTGAAGCTTGCAATTCCCTAGTGTCTTTGGTCAGTGCAATATAACCTCC
 AAAATCTTTCCATGTTAATGTTGTAAGGATCTGTGGCCCTTTAACG
 GGCTGTGCTCCCACAGAGCCTCATTACAACACATTTTATGCGTGAAAC
 AGAGTCACATATCTTCAATTCTCTTATGTCGGGATTTCAGCAACACA
 GTTGTATGGGATGAGCAACTAACATTCACTGAGAACCGTGCTCT
 TTTGCTCTCTGTAG

CEGP1 int 5.1

FIG. 1D

FOXM1 intrón 3, 2041 bases

GTAATGTGTCACAGCAACCAAAATCAAGGTCAAGCCAGCCTGACAGTC
 TCTCCAGTGCTGACTGCACTTGTATCTGGACAGCAGTTAAGTGC
 AAAGACACTAGAACAAATGTAATTAGATTGTGACTCAATCTT
 ATTGAATCCAGGCCAAATCATTAAAGAACAGCTCTTAACCTACCTCATGTG
 TTACTACCTAAAGTCATGGAGGGCTTCAATGTCAGCACTCAAGCCC
 ACTTTCTGCTACACTCAACAGCCGTCTAGATGCCAGCAGCTAGAGTGGCTA
 AGTAGTTTATGAAAAATGCTTGTATTAAAAAAATGCTGCTGTGAG

FOXM1 int 3.2

CCTCATGACCCAAAGATGTCATCTCTGTAGCGTCACATAGCATTCTAGT
 GGCAGGGGTTTCCTTCACTTCATTCACTGAAAGACCGAGATGCCGTGT
 GAGTCAACATAGCTCACGCAGTGGTCGGTGTCAAGAGCCACAAATGAGGT
 CTTCTGACGGGTGCTCAATTCCAAGTCAAGTGTGCTTTGTTTCTCATG
 GTAGAACTCNCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
 NNN
 NNN
 NNN
 NNN
 NNCATGGTAGAATTAACTCCCTCCATCAGCTTACTTTCTA
 GNNN
 NNN
 TAGTTCTTAATTCTCTGAGCCACCTTCTGCTATTGATCACTACCTC
 ACAGCCTACTCTGCTTTCTAGCCCCCTGACAGCTATCTAGGTCTTTCT
 TTATCACAACTAAGGTTGGCATTAGCTTTATTCCCGTAGAATAGATGG
GTTTATGGCTGAAGGTGACGGCTCTGGGTGTGGAGTGTCAGGAGAGTTG
CCLAGAGGGCTGCAAAAGACACCAGACGAAGCCCTGTCTGAGCACAGTGGG
 AGGGCCTGAGGCTGGTTCCCATGTTGAAGGGTATGTTCTGAA
 TCTAAAGTAGCTGATAACCAGTTGCTTGTCTCTCCAG

>FOXM1 intrón 4, 993 bases

GTGAATGCCCTGCTTCTCTAAATAGGGCTAACAGTTGGAGGTGTCATA
GCCATCTCAAAGGAAACAAGTTCTGCTAGTGTATGCTTCAATTGATCAG
GGGAGAGTAGAACGCCACCAATTAGTGAATTGACACATTTGTGCCAGACC
GAATTAAAGTACACTTGACAAATTACCAATTGACACATTTGTGCCAGACC
 AGAGCAAGGAGAAGGTGTTCTGACCCAACAGAAAGGGCTCCCCAGGGNN
 NNN
 NNN
 NNN
 NNN
 AAGAAAATTCTGGAAATGCTTGTCTAAAGGCTTCCCTNNNNNNNNNNNN
 NNN
 NNN
 NNN
 NNN
 NNN
 CCTGCCCTTGGCTGGAGACAGAACGCTGAGCTACAGAACGTCCTTTCAG
 AAAGAAGTTATTGTTGTTTCAAGAGTGGCCATAAGGCTGCTGGTAGCTG
 TAACCATTCTCCTGGAGGGGAGTTGTCTGGGTGTCTTGTACATCAG
 TCAGGAATAAGTGTGTTTCCCAATCCGGTCAAATTGACCACGTTGGTGGT

FIG. 1E

MGB-FOXM1 int 3.2
 (subrayado)
FOXM1 int 3.1 (negrita)

FOXM1 int 4.1 (subrayado)
FOXM1 int 4.2 (negrita)

FIG. 1F

AACTTCATCTCATTCTCTCCCACAATGCCCTGGCCGCCACCAAG

ES 2 314 378 T3

>FOXM1 intrón 5, 602 bases

GTAAGGTCTTCCCTCTGGCTCGGGCTGGCTTGCCCTGTTCCCTCACT
GCTCAGCATGGCTTAGTGACAGAGACAAGATGTATGTGGGGAAAGGGT
CCCTATGGCCATGTTTGCTAGGTGCCAGCCCTAGACACAGAACACCC
GAGGGTCAGGCACACACCCACTTCCCTCCCTTCCATGGCATCAAGG
GCACACTGAGCAGAGCAGGGCACAGCAGGGGAGCATGTCAGCAGGCCAC
AAGCGCATGGCACCGCCTCAGGGGGGGCAGTTCGTTGCTCACTTTGT
GCCTAGCTTCTTGCCACGCATATACTACCTGCTCTGGCATCCCCCA
GGGGTGTGAGGACACGTGGGTGAAGCGGTAGTGCACACTGCCCACATG
TGTCTGTAGGCCACCCACCTGCCCACTCATCACAGTTGGAGACTGCTC
GCCTACGTCCATCCCTCAGGTTGGCCTCTCTCTGGCTGTCATTAA
CTCAAGCACACACCACAGAGCAGCTGGGGGTTTCCATCCCCCTCT
TACCTTATTGTGTTAACATAGGTTCTTCTCCCCATCTGCCACAAGC
AG

MGB-FOXM1 int 5.1

>FOXM1 intrón 7, 4656 bases

MGB-FOXM1 int 7.1

FIG. 1H

MGB-FOXM1 int 7.2

TTTTAAGACTTCAAAACQACACAAGAGGATCAGAACCGTATGTGATAT
TTTTGTGCGTCCTGTCTGGTACCGTTGGTTACCTTATCTCTGTTCCC
TTTCAG

MGB-FOXM1 int 7.3

>PRAME intrón 2, 614 bases

GTAAGTTCGAGCCCTGATTCCCTCCGCTTCCCCGCAGGGTGACCTTGGGCT
TGTGCCCGGGCACCAACCCCTGCTCCGGTCCCTGTTCTCTCGAAA
TGGGTGAAGACCAAAAGAAAAATAATGTGCGCCACTTGGGTACCCCGGGC
CGCTGCCCCGGAAAAATTGGCCCCAGTTGAGGAGTTGTTGGCTGTAAGGAT
GCCTTGAAACGAGGGCGGTGCTCGTGGTTGGAGCTCTCCAGGGTGGGT
GCGCATTTGTAATGGGTGGATGCTCTGGGACTCGGCCCCCTGTAAGGTG
CTGGGGTTGGGACGGCCCAAGGCAGTGGCTAGGCCTCTAGGAAGGCG
GGAGCAGAGGCAGAAATGTCGCTGCAAGACCGTAGTCAGGGTCTTGACC
ACAGGGGTCACTTGTGACCAACCAATGGTCTGTTCTCTGCCCCC
TGGTTCAAGCCCAGGAAACACTGCTGCTCAGGTTGGAGCCAGAGATTGCG
ACTGAAAGGGCGGATTGACTGCCAGTTGTCAGTTCTCAGCAGTATT
TGCAGGAGGTTTACAGGAGGCCGTTGCTCGTAAATATTACATGTAT
TCCTCTTTGGAG

MGB-PRAME int 2.1

FIG. 1I

>PRAME intrón 4, 432 bases

GTAAGGGTACCTAGCAGCTGGTGTGGGCCCTGGAACCTGAGCAGGA
TGCAGCTGGGTCAAGGGAGCATGGAGGCCCTAAGGCTGGGCCAGAGGCTC
TGATGGTTGCCAGCAAGGAAGTTCAAGGGAGGCCCTGGGCTACTGCAGGG
GTCACTCTGGAAATGGGCTTCTGGACATGGGCACTGATTAAATGCGAGA
GGTGTCTGAAAGGAAATGCACTGCCCTGCTTCTCCTGGTGGGTGGAAATTGG
GGACCAAGGAAGGATCCCAGGATCTAGTGGGAAAGGGAGCAGCTGATGCC
TGAAGTACGAAGTAAAGTGCAGATCTAAGGTGGATGCTGTTGGTCT
TACCTACATTATGAGACTCATGGCTTATTTGAGTTGATCTAAAGCAT
CATCTCAGCTAATTACCTGTTTCCCCACAG

MGB-PRAME int 4.1

FIG. 1J

ES 2 314 378 T3

>STK15 intrón 1, 3740 bases

FIG. 1K

MGB-STK15 int 1.1

NN
 NNN
 NNNNNNNNNNNNNNNNNNGAGGGAAAGCACATGCCTGCCCCCATT
 AGCTTCATTGACAAAATGTAAATCCGGTGTGTTGGAGGCCTT
 TGATTGGGAACGTAAACGCTGCCTATGAGCAACAGCACTTAAGCAGG
 TGGCTTGTCAAATTAAAGGTTCTCTTTCTTTCTAG

>STK15 intrón 2, 1622 bases

GTAAATTGAATAATCTGTAATCTCATTCACATTATAAACCCACATGGAG
 GTGGTCTTGTGCGGAATTCTTCCGCCCTTACTTGGATTTAATTTAG
 ATCCCTTACTGTGATCCTGGATATGAATTAGTCACTTTCTCGTGTTCAG
 TAACATTTTGCTGCTCTTAGAGTAGCTTTTGTCTGCTTTGTCTTAT
 AATCGCTGCTTAAGTTCTATATCCCTCACTGTATGCAAGGATAATAGT
 AATAATGCACTGGCAGGAGTCAAAACTTTAAATTGGCCATAAATAT
 AAAATAATTAGAAAAGGCTACCTTGAATTACTGTATTTGATTCTAAGTT
 CCTATGATAACGCCATTAAAAATTGCTCTATATTAAARTGTTCTT
 TTATTTGCTTGTCTGAATGCCTGCTGCGTTGAGCAGTGTGCTAAT
 TTCAGGAGTAACGTACTTTGTTATTGGAAAGTCCTAACACCCCTCTCTTGT
 AGAGCACTCATACCGTTGAGCTGGGATGGACTTTGAGGTTTCATTCT
 AGCACTTGTCCCTCACTTACAATGAGCTTTGAAGCTGAAGGAAATCTCA
 TCCCTCTACCCCTTTAGTTGATTAGCTGAGGGTGTAGAGTTAACCT
 AACAAATTAAAGTTGTAATACAGTACTTACAGGCGTATAAAATAACATT
 TCAANNN
 NNN
 NNN
 NNN
 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTAATAATA
 ATACATTTAGTAGTAACTTGTGAAGTGTCTACATTGTCTCTTIG
 TCAGTTTTGTCTCAATTCCATTGTCAATACTTGGAAAATGAAACATT
 GGTTAATCAATAGTACAGTAATAAGCTATTGTGAGGAAATCTCGATATA
 TGAAAACCTAGACTCTTCAAAACTTCATGAAGATAATACCACTGTTGAA
 CGTTTGACGTATTGTTGGTCTTTCTTAAACGTATATTATCAAA
 GAAATTCAATGGAACGTGAGATTGGCATAAAGTTTGTATCATAGCT
 TTTGCCAATAGCAATGTAGTGTCTATTCCAATTATTGAGAAATT
 AGAAAGTGTCTCTTCATTAAATGGATATTGTAAATAAGCATGATT
 AGGGGTGAGGAATTGGAGGGATAGAAGGTATCATTCAGGTATTCTAGC
 CACATACTAACTATCCTCTGGAGGTACTGATTAAACACCTTTCACCTT
 CCATCTCTTACAGTGAACATTCAATTGTCTACTAGAGAACAAACT
 TTGTGAAATTCTCAATATATTCACTTTGCTTCTGATGAATGCCAGAAAG
 TTTATTTCTCTTCCATTCTAG

MGB-STK15 int 2.1

FIG. 1L

>STK15 intrón 4, 1093 bases

GTAAAGCTTCTTATTACAAAGTTCTGTAATGTTCTACTAGAAATATATTA
TTTCGTTGCAAATTCTGTTGTGGGAACCTCTGGGGAAAAAAATGAGGCCTT
TATTTGCAATTAGAGGGATAAAATGTTCCAGATTTCAAATCTTAAAAAA
AATGGAATTTGTGTAATGAGGTATTACTAGGAACCTCAAGTGCTTTAA
AAAATGGCTTCAAATTAGAAAAAGCTGTATGAATCTTTATAGAAAT
GTGTTGAAAGTTCTCTCTGCTTCTAGAAATAACCACTACATATGGTTAT
GCGTCTGTACTTTTTATTGTACAAAAGTGCAGTTTTAAAAAATAGAA
TATGTTGCAAGAACATATATACTCATATATGACTGAGGGTTTGACAGTATT
ATAGTTTAGTTCTTATTGTAAAGGTTGGCTGTAATGTCCTCCCCAGGG
CTTTCTAAAAGCCTCCTCTCAGTCTCTGAACATCTGGACTCTAGAATG
TACCGGGAGGAGCAGGAATGAAACCACAGACTCTTGTCTTTAGCGGT
CTAACAGAGGCTAAGAGTCTAAATCCACTGGTTCTCATGCCAGCTAGC
CTGTTGGCTCCATCCCCTTCATTAGTAACAGTGGCTCTGTCCTCACCA
CCAGAGTGGTTCTCCACCCAGAGAGAAATTAGCACCTCTGGACTGGAGGG
AGCAGCTGGGGTTAGTTGAAACATGCCAGATGGTCTGGAAGCATTC
CTCCCTCTCTGGTCACTTATCCCTTTTGTTGCTTCAGCGTTCTCATGCC
CTGTTCCCTCTGAGCATAGTACGGGCTTGGAACATTCCCATAGAGTGCT
TCAGGTCTAAAACCCGAGACTGCTCTGTCACTGACTCTCACACCTGAC
GGCAGCTAGGGACCTCAGGGTTCTGTCGTGGCAGCTTTGATAGTGG
TTATTGCTTGGTTCTGCTGAGGATGCAATTGAGTGAAGTTGGAATAC
GAAATTATTGTAGAATGTTGCTACTCATTGAAAATTGTTAGAAAA
GCTTTGTTCTTCACATTCTAAAGTGTCAAATTCCCTCTAG

FIG. 1M

MGB-STK15 int 4.1

ES 2 314 378 T3

Nombre	Secuencia SEC ID N°:
MGB-CEGP1 Int 1.F1	AGCGCCTGTCGGATCTG 14
MGB-CEGP1 Int 1.R1	AACCAAAGTCTCTGCTGAAAACC 15
MGB-CEGP1 Int 1.P1	CCCTGAAGCAGCAAC 16
MGB-CEGP1 Int 3.F1	CTGTTGCTGTGATGCTGTC 17
MGB-CEGP1 Int 3.R1	CCTCAGCCACTCCCTGATC 18
MGB-CEGP1 Int 3.P1	TCAGGGCATAAAGCCT 19
MGB-CEGP1 Int 4.F1	TCCCCCTGCCCTGGAGA 20
MGB-CEGP1 Int 4.R1	AAAGGCCTGGAGGCATCAA 21
MGB-CEGP1 Int 4.P1	CAGCCCCAAATCCT 22
MGB-CEGP1 Int 5.F1	CTTAATGGTGTAGCACAGATGCA 23
MGB-CEGP1 Int 5.R1	CCACTGTAGCATGGGAAGCA 24
MGB-CEGP1 Int 5.P1	CAAATGCACAGGAAAC 25
FOXM1 Int 3.F1	GCTCTGCCTGTGGAGTGT 26
FOXM1 Int 3.R1	CACAGGCTTCGCTCTGGTGTCT 27
FOXM1 Int 3.P1	TGCAGCCCTCTGGCAACTCTCT 28
FOXM1 Int 3.F2	AAAATGCTGTCTGTGAGCCTCAT 29
FOXM1 Int 3.R2	AACCCCTGCCCACTAGAAATG 30
FOXM1 Int 3.P2	ACCCAAGATGTCATCTCTGTAGCGTCACA 31
MGB-FOXM1 Int 3.F2	AATAGATGGGTTATGGCTGAAGGT 32
MGB-FOXM1 Int 3.R2	CTCTTGGAACCTCTCCGTACACT 33
MGB-FOXM1 Int 3.P2	CCGCAGAGCCGTC 34
FOXM1 Int 4.F1	CCATCTAAAAGGAAACAAGTTCTG 35
FOXM1 Int 4.R1	GGGTGGCTGGCTTAACTCT 36
FOXM1 Int 4.P1	CCCTGATCAAATGAAAGCATCACT 37
FOXM1 Int 4.F2	AGAAGCCAGGCCACCCATT 38
FOXM1 Int 4.R2	TGTGTCATTTGGTATTTGTCAAGTGT 39
FOXM1 Int 4.P2	TGACTTGACAAAACCCAGTGAATTA 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	GGTGTCCATTTCCTCTGAAGAGA 44
MGB-FOXM1 Int 7.R1	TGCAAGCTGAAGGTCCAACAT 45
MGB-FOXM1 Int 7.P1	TTCTGGCCAATTAAAG 46
MGB-FOXM1 Int 7.F2	TCATTGATGTGGCCGTAGCAT 47
MGB-FOXM1 Int 7.R2	GGTGGAGAGGGAGCCAAAA 48
MGB-FOXM1 Int 7.P2	CTGTGTTGGGTTTCA 49
MGB-FOXM1 Int 7.F3	AGAGGATCAGAACCGTAGTGTGA 50
MGB-FOXM1 Int 7.R3	GGGAAACAGAGATAAGGTGAACCA 51
MGB-FOXM1 Int 7.P3	TGTGCGTCTGTCTG 52
MGB-PRAME Int 2.F1	GGGTGACCTTGGGCTTGTG 53
MGB-PRAME Int 2.R1	CTTCACCCATTTCAGAGAGAA 54
MGB-PRAME Int 2.P1	CCCGGGTCCCTGTT 55
MGB-PRAME Int 4.F1	AGGGTAGCCTAGCAGCTGGT 56
MGB-PRAME Int 4.R1	GCCTCTGGCCCAGCCTTA 57
MGB-PRAME Int 4.P1	TCCCTGACCCCCAGCTG 58
MGB-STK15 Int 1.F1	CGTAATGTCCTCTCTTCCGTAA 59
MGB-STK15 Int 1.R1	ACGAACGTGAGTAGGTTGCTGAAAAA 60
MGB-STK15 Int 1.P1	TCAAGGGACAAGGAAG 61
MGB-STK15 Int 2.F1	CATTCACATTATAAACCCACATGGA 62
MGB-STK15 Int 2.R1	AATCCAAAGTAAGGGCGGAAGA 63
MGB-STK15 Int 2.P1	TGGCTTGTCGGGAAT 64
MGB-STK15 Int 4.F1	GCGAGGAATGAACCCACAGA 65
MGB-STK15 Int 4.R1	GCATGAGAACCGAGTGGATTAGACT 66
MGB-STK15 Int 4.P1	CGCTAAAAGCAAAAGA 67

FIG. 2

Expresión normalizada de CEGP1: Coeficientes de Correlación

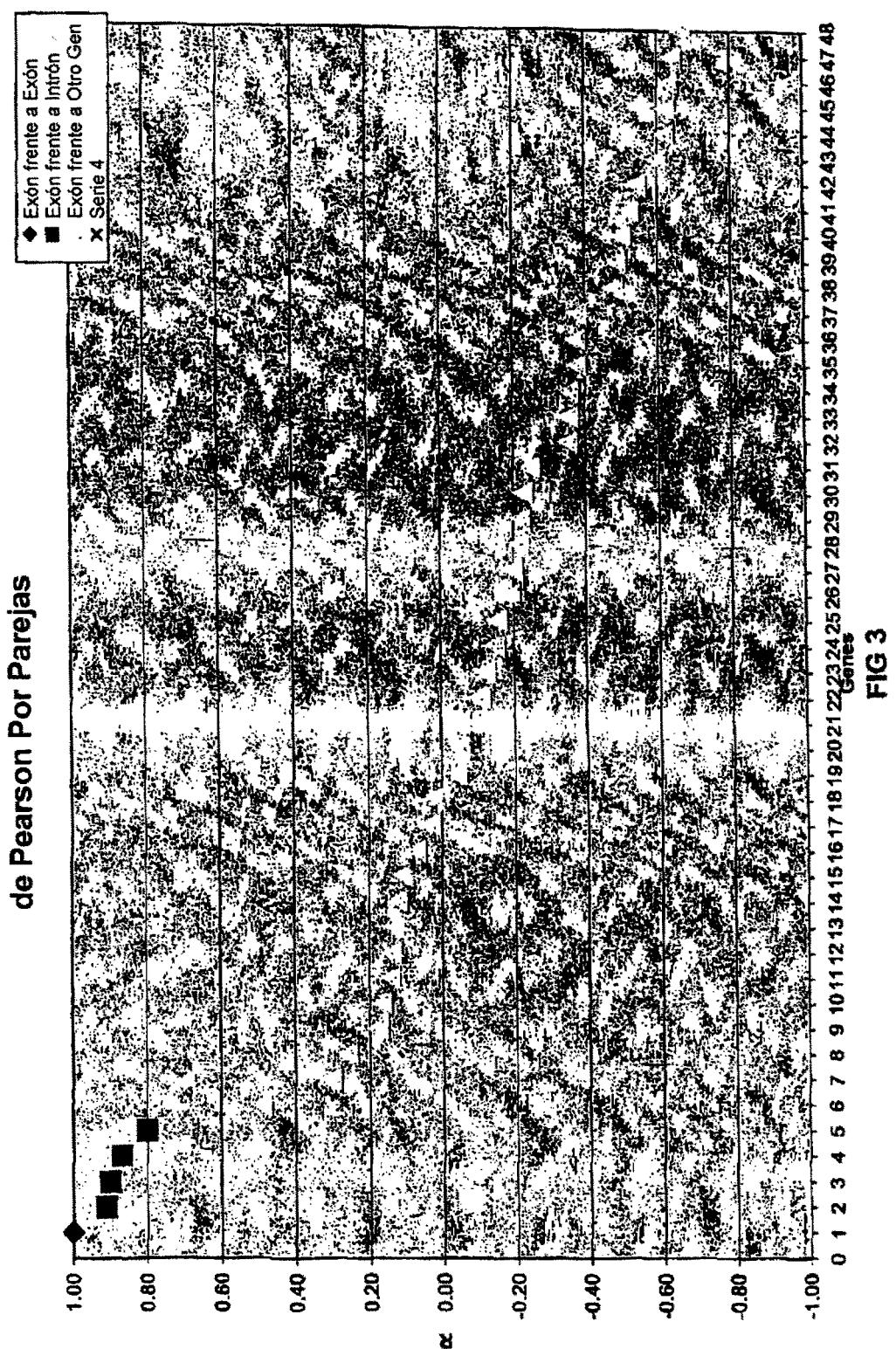


FIG 3

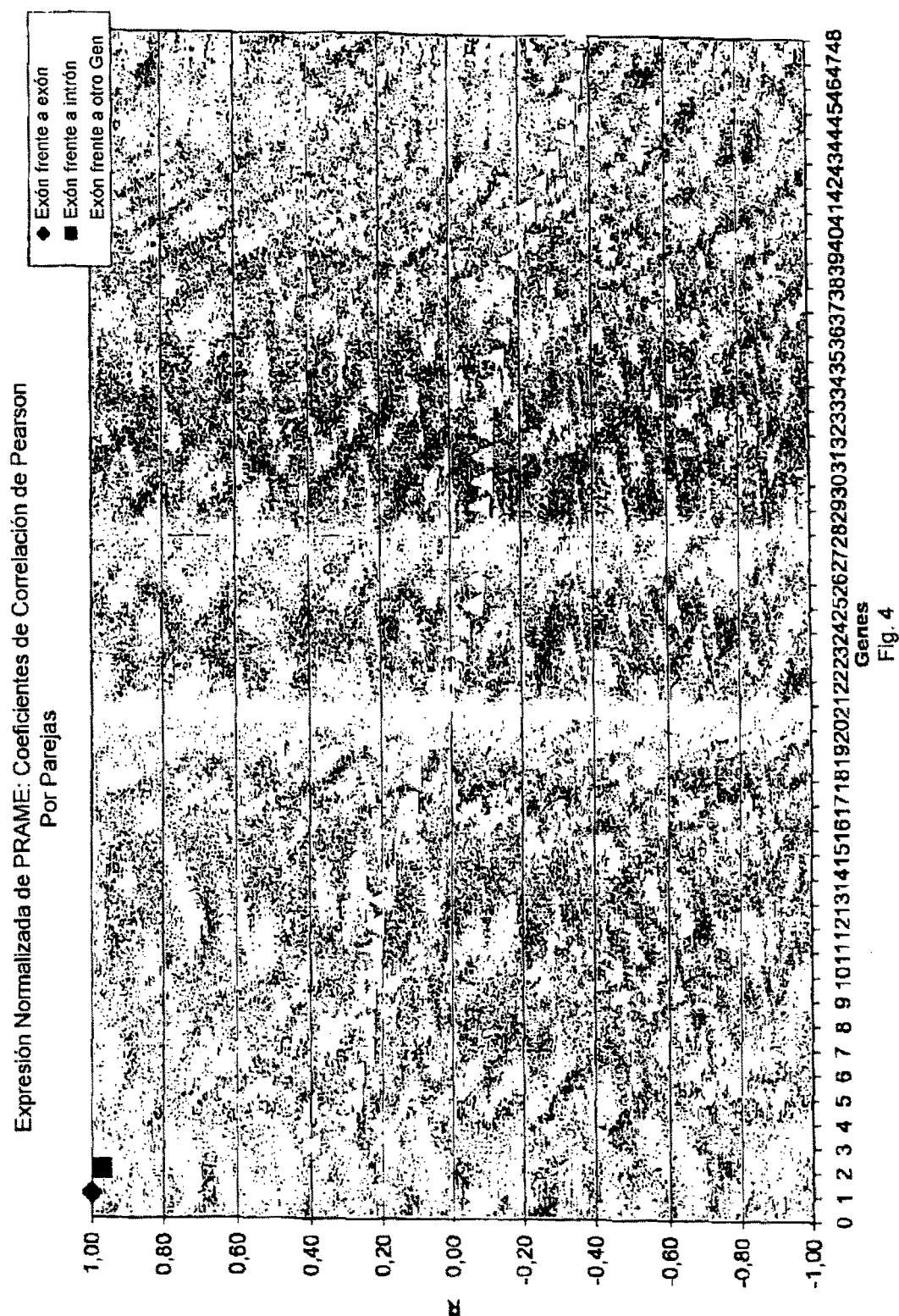


Fig. 4

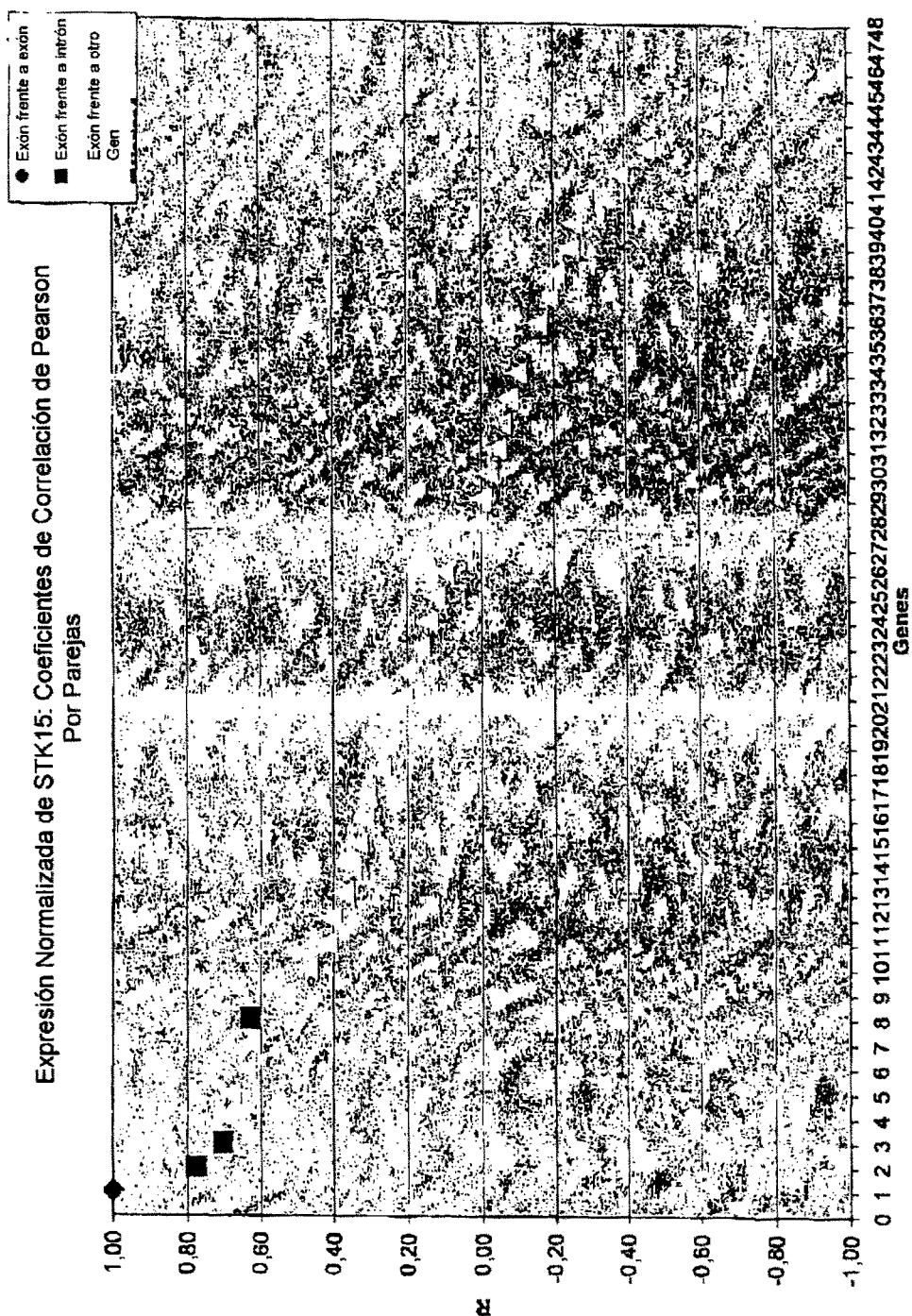


Fig. 5

ES 2 314 378 T3

Nombre del Gen	Secuencia ID	Procedencia Secuencia	Nº de acceso de GHallgenes2-03.txt
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

ES 2 314 378 T3

Nombre del Gen	Secuencia ID	Procedencia Secuencia	Nº de acceso de GHallgenes2-03.txt
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
relacionado con BCL2	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

ES 2 314 378 T3

Nombre del Gen	Secuencia ID	Procedencia Secuencia	Nº de acceso de GHialligenes2-03.txt
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 2	NM_004702	NCBI	NM_004702
CD10S	NM_000118	NCBI	NM_000118
CD134	NM_003327	NCBI	NM_003327
CD18	NM_000211	NCBI	NM_000211
CD31	NM_000442	NCBI	NM_000442
CD3z	NM_000734	NCBI	NM_000734
CD40	NM_000074	NCBI	NM_000074
CD44E	X55150	NCBI	X55150
CD44s	M59040	NCBI	M59040
CD44v3	AJ251595v3	NCBI	AJ251595v3
CD44v6	AJ251595v6	NCBI	AJ251595v6
CD68	NM_001251	NCBI	NM_001251
CD82	NM_002231	NCBI	NM_002231
CD9	NM_001769	NCBI	NK-001769
CDC20	NM_001255	NCBI	NM_001255
cdc25A	NM_001789	NCBI	NM_001789
CDC25B	NM_021874	NCBI	NM_021874
CDH1	NM_004360	NCBI	NM_004360
CDX2	NM_001265	NCBI	NM_001265
CEACAM6	NM_002483	NCBI	NM_002483
CEGP1	NM_020974	NCBI	NM_020974
CEGP1 intron 1	NM_020974int1	NCBI	NM_020974int1
CEGP1 intron 2	NM_020974int2	NCBI	NM_020974int2
CEGP1 intron 3	NM_020974int3	NCBI	NM_020974int3

FIG. 6

ES 2 314 378 T3

Nombre del Gen	Secuencia ID	Procedencia Secuencia	Nº de acceso de GHallgenes2-03.txt
CEGP1 intron 4	NM_020974int4	NCBI	NM_020974int4
CEGP1 intron 5	NM_020974int5	NCBI	NM_020974int5
CEGP1 intron 7	NM_020974int7	NCBI	NM_020974int7
CGA	NM_001275	NCBI	NM_001275
CHAF1B	NM_005441	NCBI	NM_005441
chk1	NM_001274	NCBI	NM_001274
chk2	NM_007194	NCBI	NM_007194
CIAP1	NM_001166	NCBI	NM_001166
CIAP2	NM_001165	NCBI	NM_001165
CKAP4	NM_006825	NCBI	NM_006825
Claudin	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	AJ744123	NCBI	AJ744123
Contig44799	Contig44799	Rosetta Contigs	Contig44799
Contig46653	Contig46653	Rosetta Contigs	Contig46653
Contig47405	Contig47405	Rosetta Contigs	Contig47405
COX2	NM_000963	NCBI	NM_000963
CRBP	NM_002899	NCBI	NM_002899
cripto	NM_003212	NCBI	NM_003212
CRK7	NM_016507	NCBI	NM_016507
CRMP1	NM_001313	NCBI	NM_001313
CSF1	NM_000757	NCBI	NM_000757
CSF1R	NM_005211	NCBI	NM_005211
CSF3	NM_000759	NCBI	NM_000759
CSNK1D	NM_001893	NCBI	NM_001893
CSTF1	NM_001324	NCBI	NM_001324
CTSB	NM_001908	NCBI	NM_001908
CTSD	NM_001909	NCBI	NM_001909
CTSH	NM_004390	NCBI	NM_004390
CTSL	NM_001912	NCBI	NM_001912
CTSL2	NM_001333	NCBI	NM_001333
Cyclin C	NM_005190	NCBI	NM_005190
Cyclin G1	NM_004060	NCBI	NM_004060

FIG. 6

ES 2 314 378 T3

Nombre del Gen	Secuencia ID	Procedencia Secuencia	Nº de acceso de GHallgenes2-03.txt
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_exones18-23	X76132_18-23	NCBI	X76132_18-23
DCC_exones6-7	X76132_6-7	NCBI	X76132_6-7
DCK	NM_000788	NCBI	NM_000788
DCR3	NM_016434	NCBI	NM_016434

FIG. 6

Nombre del Gen	Secuencia ID	Procedencia	Nº de acceso de GHalgenes2-03.txt
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 Habitual	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
EmR1	NM_000125	NCBI	NM_000125
F2	NM_000506	NCBI	NM_000506
F2R	NM_001992	NCBI	NM_001992
F2RL2	NM_004101	NCBI	NM_004101
fas	NM_000043	NCBI	NM_000043
fasl	NM_000639	NCBI	NM_000639
FBXO5	NM_012177	NCBI	NM_012177
FGF 19	NM_005117	NCBI	NM_005117
FGF1	NM_000800	NCBI	NM_000800
FGF18	NM_003862	NCBI	NM_003862

FIG. 6

ES 2 314 378 T3

Nombre del Gen	Secuencia ID	Procedencia Secuencia	Nº de acceso de GHallgenes2-03.txt
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
FLJ310713	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 intrón 3	NM_021953 int3	NCBI	NM_021953 int3
FOXM1 intrón 4	NM_021953 int4	NCBI	NM_021953 int4
FOXM1 intrón 5	NM_021953 int5	NCBI	NM_021953 int5
FOXM1 intrón 7	NM_021953 int7	NCBI	NM_021953 int7
FPPS	NM_002004	NCBI	NM_002004
FRP1	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-catinina	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
gp130	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

ES 2 314 378 T3

Nombre del Gen	Secuencia ID	Procedencia	Nº de acceso de GHallgenes2-03.txt
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-BGF	NM_001945	NCBI	NM_001945
hCRA	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	XM_004688	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

ES 2 314 378 T3

Nombre del Gen	Secuencia ID	Procedencia Secuencia	Nº de acceso de GHallgenes2-03.txt
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
Kitng	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_000224
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

ES 2 314 378 T3

Nombre del Gen	Secuencia ID	Procedencia Secuencia	Nº de acceso de GH\algenes2-03.txt
Maspina	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

Nombre del Gen	Secuencia ID	Procedencia Secuencia	Nº de acceso de GHallgenes2-03.txt
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
Osetonectina	NM_003118	NCBI	NM_003118
P14ARF	S78535	NCBI	S78535
p14ARF	NM_000077P14	NCBI	NM_000077
p16-INK4	L27211	NCBI	L27211
P16INK4	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
PAJ1	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
PDGFR α	NM_006206	NCBI	NM_006206
PDGFR β	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

ES 2 314 378 T3

Nombre del Gen	Secuencia ID	Procedencia Secuencia	Nº de acceso de GHallgenes2-03.txt
P13Kc2A	NM_002645	NCBI	NM_002645
PIK3RI	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 intrón 1	NM_006115int1	NCBI	NM_006115int1
PRAME intrón 2	NM_006115int2	NCBI	NM_006115int2
PRAME intrón 3	NM_006115int3	NCBI	NM_006115int3
PRAME intrón 4	NM_006115int4	NCBI	NM_006115int4
PRAME intrón 5	NM_006115int5	NCBI	NM_006115int5
PREP	NM_002726	NCBI	NM_002726
PRKCD	NM_006254	NCBI	NM_006254
PRKCG	NM_002739	NCBI	NM_002739
PRO2000	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
RBMS	NM_005778	NCBI	NM_005778

FIG. 6

ES 2 314 378 T3

Nombre del Gen	Secuencia ID	Procedencia Secuencia	Nº de acceso de GHallgenes2-03.txt
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	RNasaP	GHI Habitual	RNasaP
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_0031-52
STAT5B	NM_012448	NCBI	NM_012448
STC1	NM_003155	NCBI	NM_003155
STC2	AK027663	NCBI	AK027663
STC2	NM_003714	NCBI	NM_003714
STK11	NM_000455	NCBI	NM_000455
STK15	NM_003600	NCBI	NM_003600
STK15 intrón 1	NM_003600int1	NCBI	NM_003600int1
STK15 intrón 2	NM_003600int2	NCBI	NM_003600int2
STK15 intrón 3	NM_003600int3	NCBI	NM_003600int3
STK15 intrón 4	NM_003600int4	NCBI	NM_003600int4
STK15 intrón 6	NM_003600int6	NCBI	NM_003600int6
STMY3	NM_005940	NCBI	NM_005940

FIG. 6

ES 2 314 378 T3

Nombre del Gen	Secuencia ID	Procedencia	Nº de acceso del GHallgenes2-03.txt
SUHW1	NM_080740	NCBI	NM_080740
SURF6	NM_006753	NCBI	NM_006753
Surfact A1	NM_005411	NCBI	NM_005411
SURV	NM_001168	NCBI	NM_001168
SXR	NM_003889	NCBI	NM_003889
SYK	NM_003177	NCBI	NM_003177
TAGLN	NM_003186	NCBI	NM_003186
TBD	NM_016261	NCBI	NM_016261
TBP	NM_003194	NCBI	NM_003194
TBX2	NM_005994	NCBI	NM_005994
TEK	NM_000459	NCBI	NM_000459
TERC	U86046	NCBI	U86046
TERT	NM_003219	NCBI	NM_003219
TFF3	NM_003226	NCBI	NM_003226
TFRC	NM_003234	NCBI	NM_003234
TGFA	NM_003236	NCBI	NM_003236
TGFb1	NM_000660	NCBI	NM_000660
TGFB3	NM_003239	NCBI	NM_003239
TGFBR2	NM_003242	NCBI	NM_003242
Trombospondina 1	NM_003246	NCBI	NM_003246
Tinosina 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
TTTF1	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

ES 2 314 378 T3

Nombre del Gen	Secuencia ID	Procedencia	Nº de acceso de GHallgenes2-03.txt
	Secuencia		
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_003317
VEGFC	NM_005429	NCBI	NM_005429
vim	NM_003380	NCBI	NM_003380
WISP1	NM_003882	NCBI	NM_003882
Wnt-5a	NM_003392	NCBI	NM_003392
Wnt-5b	NM_032642	NCBI	NM_032642
wwox	NM_016373	NCBI	NM_016373
XIAP	NM_001167	NCBI	NM_001167
XIST	M97168	NCBI	M97168
YB-1	NM_004559	NCBI	NM_004559
ZNF217	NM_006526	NCBI	NM_006526

FIG. 6

ES 2 314 378 T3

LISTA DE SECUENCIAS

<110> GENOMIC HEALTH, INC.
 SCOTT, Randal W.
 5 BAKER, Joffre B.
 KIEFER, Michael C.
 <120> USO DE ARN INTRÓNICO PARA MEDIR LA EXPRESIÓN GÉNICA
 <130> 39740-0007 PCT
 10 <140> No asignado
 <141> 2004-02-19
 <150> US 60/448.991
 15 <151> 2003-02-20
 <160> 67
 <170> FastSEQ para Windows Versión 4.0
 <210> 1
 20 <211> 1566
 <212> ADN
 <213> *Homo sapiens*
 25 <220>
 <221> misc_feature
 <222> (802)...(1027)
 <223> n= a, t, c, o g
 30 <221> misc_feature
 <222> (1126)...(1518)
 <223> n= a, t, c, o g
 35 <221> intrón
 <222> (0)...(0)
 <223> MGB-CEP1 int 1.1
 40 <400> 1

	gtgagtgtcc	ggccgcgggg	gcgcacctgg	cacagcaggc	agggccagga	agagtgttta	60
45	ggtccccggc	ggagtcaga	gccgggcgcg	cggggctcg	ggctggcggc	tgcagctccg	120
	cggggggcctc	tgcctccccc	gggacacctac	ccgcccggccg	ggccaaggcg	ccacgaccgc	180
	tggggccctg	agtccttcgg	cccgccctcg	gaccggagac	tgctgacggt	tcccggcccg	240
	gtccggatgc	ctccagagcg	cctgctagtc	agaccgtcgc	cggcgacgag	gcaggagggt	300
50	gcggacacctg	gccttgggtt	cccgcgcctc	agcgtaggcg	ggaaactga	gggcccgggccc	360
	gggcacatcc	gcgaggcggt	ggcagcttg	ccgtttcttt	ctttggggc	cggcaagtgc	420
	tgctgatggc	ttcgggggtgg	gctccagaga	cttttctgtc	agcggAACAG	cgcctgttcc	480
	gatctggaa	ttaccctgaa	gcagcaacaa	gcctagggtt	ttagcagaga	actttggttt	540
55	ccagagagga	ctctggacgt	gctgtgctta	ctggacttgc	aataacttca	aaatgcittt	600
	gttttaatt	aatatcctgg	agtagtgtca	acccaggaaa	tacttctgcc	aaggcggggtt	660
	tccaggttga	gaggatgggc	aggggtggga	gtgcaggggg	ccggccatgg	ggacaccatc	720
	cccgctcgc	agcatctgag	agccctggat	gacatctgct	ccgatcccgg	gacagacttc	780
60	ccataaaatac	tctaaaccag	cnnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnnn	840
	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	900
	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	960
	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	1020
65	NNNNNNNNTC	tgagctccga	gaaagctgac	agacagctgc	ttggtgttca	gagcttgct	1080
	gtccgtttgg	tccttcctc	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	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	1560

ES 2 314 378 T3

<210> 2
<211> 4985
<212> ADN
5 <213> *Homo sapiens*
<220>
<221> misc_feature
10 <222> (83)...(253)
<223> n= a, t, c, o g
<221> misc_feature
<222> (1975)...(2296)
15 <223> n= a, t, c, o g
<221> misc_feature
<222> (2445)...(2468)
20 <223> n= a, t, c, o g
<221> misc_feature
<222> (2499)...(2551)
<223> n= a, t, c, o g
25 <221> misc_feature
<222> (4295)...(4352)
<223> n= a, t, c, o g
<221> intrón
30 <222> (0)...(0)
<223> MGB-CEGP1 int 3.1

35

40

45

50

55

60

65

ES 2 314 378 T3

<400> 2

5	gtacctctgc ccagctgtgg atggggcag agccacatct gagaccctct ccctgcacg 60 cgcacacaca cactgactct agnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnnn 120 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnnn 180 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnnn 240 nnnnnnnnnn nnnatcttta catagaatac atttcaaaca tgactagatg tctcaggagc 300 aatatagtgg atgatctgcc aagttttca aaaagggtgc gaaaaccaca gcaccagtat 360 10 gaggctgctc cctgctctgg gtgggttaggg aggaggctgg atcctccca tgcagactt 420 caatgaagtg ccctgtttc agccccaaagc tagatccggc cttccatgt tttgcattt 480 tgagctccga gggcagaag ggctccccc ctggactttc cgtgcgtgg ttcccttcgc 540 ctacgtcacc attatattttt cctctgtaaa ttgcggaa actctctct tctgatgtcc 600 ttctcttcat tctctttgtt ttgagttat acctttttt attcccttgt tacttagtag 660 attcttgaga ggaaggggca ttaagtatcat gtggccaatc agttttttt aactgaatgt 720 15 catctttta actcttccc gctcttctt aagctaaaga gtcacatttt ggtggctgtg 780 ttcctcttgg agttgcattt gcctattttt agggaaagtg ccctaataac tagcctatta 840 acccctttgg ccatgtctg cttattttt cccattactt aagaatgagg tcattttaa 900 20 ttcttctact attaatcac aaattttatag attttttaa tcctggctt ggttaactttt 960 caagggtttc ttcatggaaatg atgatttttt tctcatttt caaggatggc agctcacacc 1020 ttatacttaa ctagaataacc ttgtggta ccaagaaaaa ttgtcagagg aacccccagg 1080 ggccaatggg ttgtatggct atcatcaccc agagctgtt catttcagc gtttggggcg 1140 gggaagtacac acataactggc ttgtatcagg cagatttctt atcttgcctt aggtgtggcc 1200 25 ctgtataaaatg tagcagttgg gtttcatttt cctgcaggat tctctgggtt caattgggtg 1260 ccctgcactc ttgtccaaat taggc当地atggg aatgaattttt gaggccagg 1320 gcacagatgt atccgaatctt cagggcatctt ctcccttttta ttgctcaaag ctgcttcctg 1380 ggaagtactt ttggcttctt ctgcaggatgg ctggggaggg atgtggaaac tgcaggtaa 1440 agccatcgct tgagccctca cggctctgggtt cccacccagt tacaaagcag ctggtagcga 1500 30 ttaagatcac ctcttataccc ttgtacttcca gagccctggc tcagccccac tctccctcc 1560 tgcaagcccc cgactgtt agagacacag gctcctcata ccagaagcaa atacaaatgc 1620 agttcccttc tgc当地aaactgt gttttctaaa ttttctacaa ttgc当地acattt cttggatccc 1680 ctaaagatgtt ttgtatgtt acatttttgtt ctggactt aacccaaatc taagaatttg 1740 cggtgtggc tggaaatgtt ctctgttattt ttctgttgc ttcaacctt attgtttggc 1800 35 aaattcatgg gagggtcaggc caacagatgg tagcaattgg taacggagaa ctttgcac 1860 ctagggtttt gattcttcaa atagaacagc ctgtaaaatgg ttttcttca ggatttccctc 1920 tctgtatgc acattaaactt ctatgaaactt gtaggctttaa aaacccacag tggtnnnnnnn 1980 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2040
---	---

40

45

50

55

60

65

ES 2 314 378 T3

nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	2100
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	2160
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	2220
nnnnnnnnnn	nnnnnnccac	agtggatacc	ttcaaagtga	ttcaaagt	ttaaaagaag	gtaacacagg	ttaaaagaag	2280
aagcttagtat	tttctattgc	tgttgaaaa	aataattatt	taccaaatgt	tctttaatat	actggnnnnn	nnnnnnnnnn	2340
agggcatcat	aatcattgac	tctgagggaa	agctcaagat	actgnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	2400
nnnnnnnnnc	ttagagactc	caaagctgt	ataaaagagnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	2460
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	ncccggcccg	ccttgttcc	taticatggg	ttccctccaa	ttcacaaggc	2520
tgctcaggct	ctcagaatga	gcactcctct	tttgggggt	gtttctgag	aatatttaga	gggtggccaa	agtggaggct	2580
tggtgtactg	atgcctttc	agggcaacag	ggaagggtgc	agggtggca	aggtggaggct	ttccctccaa	ttcacaaggc	2640
gtgcitttcag	cagacacctgt	taccctttt	atgtcatgtt	tttgggggt	tttgggggt	tttgggggt	tttgggggt	2700
atatttttgt	ttggtttcca	gaaataatct	tcagtgggagc	tttgggggt	tttgggggt	tttgggggt	tttgggggt	2760
gaatggggga	tttccaatgt	ttctgagctg	tttcccttct	tttgggggt	tttgggggt	tttgggggt	tttgggggt	2820
cgtgacaacc	agaccaattt	tggaaagagc	tagggccatt	tttgggggt	tttgggggt	tttgggggt	tttgggggt	2880
aacagattaa	tctgctcacc	ccagcagttg	tcttgcatta	tttgggggt	tttgggggt	tttgggggt	tttgggggt	2940
tttgggggtca	tttcttgaaa	agctgtcagc	gtttccagag	tttgggggt	tttgggggt	tttgggggt	tttgggggt	3000
tcttggtaac	ttcaggtgt	gtttgatgt	gttggacatt	tttgggggt	tttgggggt	tttgggggt	tttgggggt	3060
atcacaatgg	catctggcat	acaggcattc	ttacgaaaata	tttgggggt	tttgggggt	tttgggggt	tttgggggt	3120
ctctgaggca	gtaaagggtca	cttggggat	gtctttaacag	tttgggggt	tttgggggt	tttgggggt	tttgggggt	3180
agcagcagca	gctggcctgt	tttggggatgt	actttccagg	tttgggggt	tttgggggt	tttgggggt	tttgggggt	3240
gggcagttt	atatttacc	ccggactact	gttacttccc	tttgggggt	tttgggggt	tttgggggt	tttgggggt	3300
caaacacaac	gctttctggg	cttcttcaga	caggacactg	tttgggggt	tttgggggt	tttgggggt	tttgggggt	3360
ctcccgcc	attcagttac	tggggtgcaa	atcacatctt	tttgggggt	tttgggggt	tttgggggt	tttgggggt	3420
agagaaagct	ttccaccaat	ccagtgcaag	tctttttctg	tttgggggt	tttgggggt	tttgggggt	tttgggggt	3480
ttgcattggat	gaatgaatcc	cagcaaccag	cagactgagt	tttgggggt	tttgggggt	tttgggggt	tttgggggt	3540
taactgtcag	gaaaaagagc	aagagggttt	taagagagac	tttgggggt	tttgggggt	tttgggggt	tttgggggt	3600
ttaacccttg	caactgttcc	tggaaatagga	atgacatctg	tttgggggt	tttgggggt	tttgggggt	tttgggggt	3660
ccgtttaagc	acagtttgac	agctctggag	tggggactgg	tttgggggt	tttgggggt	tttgggggt	tttgggggt	3720
tagaacctgt	ttggtaaggt	tttaggggca	caatataaaaa	tttgggggt	tttgggggt	tttgggggt	tttgggggt	3780
gtgaaagcaa	ccttttgt	ctcggttga	aaacagtgcc	tttgggggt	tttgggggt	tttgggggt	tttgggggt	3840
cattgataga	gagccaatga	caaccaagtc	cctactctca	tttgggggt	tttgggggt	tttgggggt	tttgggggt	3900
tgcacgaatg	caaagaagca	acataggaac	aggttaattaa	tttgggggt	tttgggggt	tttgggggt	tttgggggt	3960
agatgtttg	aaagtattct	agggatgtaa	aagaattccct	tttgggggt	tttgggggt	tttgggggt	tttgggggt	4020
aggatctcaa	agaatttagtt	ttgagatgag	gcagaatgt	tttgggggt	tttgggggt	tttgggggt	tttgggggt	4080
accctgttgt	gccccctcat	ttagatgtgt	gccgagccct	tttgggggt	tttgggggt	tttgggggt	tttgggggt	4140
tccccttccc	accatcatac	tacaaggta	agcttaatca	tttgggggt	tttgggggt	tttgggggt	tttgggggt	4200
aaaaatgt	cagcagctac	cacacacgag	cacannnnnnn	tttgggggt	tttgggggt	tttgggggt	tttgggggt	4260
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	tttgggggt	tttgggggt	tttgggggt	tttgggggt	4320
gtcagaggct	tatgcccgt	gggagggtac	aaggggatgg	tttgggggt	tttgggggt	tttgggggt	tttgggggt	4380
cagattccgg	ggcatgtggc	ccgtacagg	atgcccataat	tttgggggt	tttgggggt	tttgggggt	tttgggggt	4440
agaagggtgg	gatgtgtggg	ggcagagagg	ggcgttggaaa	tttgggggt	tttgggggt	tttgggggt	tttgggggt	4500
agagtatgt	aaaaaatgt	gaatctgtat	cattctatct	tttgggggt	tttgggggt	tttgggggt	tttgggggt	4560
tctcctttc	tctttagatt	cttgatcagc	tcttgggtaa	tttgggggt	tttgggggt	tttgggggt	tttgggggt	4620
cgtgagttacc	agttcctgggg	gtcagcaggc	tctgatctgt	tttgggggt	tttgggggt	tttgggggt	tttgggggt	4680
tgtatggga	aggctcccaa	ctcgtagga	gagttgagac	tttgggggt	tttgggggt	tttgggggt	tttgggggt	4740
gaataaattt	ttcatgtcta	ttaattggcc	taggttact	tttgggggt	tttgggggt	tttgggggt	tttgggggt	4800
aatgtggggc	tgtatggagac	ctaagcagac	agatctgtgg	tttgggggt	tttgggggt	tttgggggt	tttgggggt	4860
cgctcccca	gggctcagga	tctgaccac	agcctagtca	tttgggggt	tttgggggt	tttgggggt	tttgggggt	4920
ttcag				tttgggggt	tttgggggt	tttgggggt	tttgggggt	4980

<210> 3
<211> 2556
<212> ADN
<213> *Homo sapiens*
<220>
<221> misc_feature
<222> (507)...(743)
<223> n= a, t, c, o g
<221> misc_feature
<222> (1584)...(1678)
<223> n= a, t, c, o g
<222> (2428)...(2474)

ES 2 314 378 T3

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

<221> intrón

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

5 <223> MGB-CEGP1 int 4.1

<400> 3

10

	gtaagtatgg	gccagtgcac	acctgccatg	ggaaccgtcg	tattccacag	gctgccttct	60
	gtggcccagc	tcagaagcac	cacctcatgg	cacggctgca	gcagcaggga	aggcagttag	120
	cacggatac	cgacccctac	caagttacttg	ttcactgcag	aagggtggtc	tcccttaggg	180
	aaggaaatg	atattttaaa	aaggaactca	tcaggaggaa	atgaaattca	ggagtaagga	240
15	gtgtgaatgt	tggggggcag	ttctccctgt	tcccacagaa	taaaaccaa	tgtccatc	300
	tggcaatcac	agctctttgc	caccagggtcc	tgcttccccct	ataaaaccta	tctgcctcct	360
	ttccgcagac	actactcccc	ttgcctttgg	agaacagccc	aaatcccttg	atgcctccag	420
	gcctttccca	agccctcccg	ctttcctggc	gtggtgact	ctcaactaac	cttcaatatt	480
20	ctgttttaact	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	nnncatcct	ctgttatcag	aagcctctg	gtgcttcag	780
	acagggcagc	catttgcac	tttggctccc	acagcacttt	cctcagctgt	atagctctgg	840
25	gttgacttgt	gtttgtatgt	gtctgtctcc	ccaggtatga	gccccctcca	agtcaggac	900
	cttgcctcat	tttcctctc	agtcctcccc	tggtacctgc	tatggatat	gctcagtaca	960
	cttgtgttta	atgagtgggt	aatatgggtgg	cctacaccat	cgggcccgcag	ctcctgcacc	1020
	acgattgttag	taacaaaact	ccacctggga	acaggaaacc	actggcaatt	catgggtttc	1080
	ctaaaccacg	atttatgcca	gggaaagcac	tgaggagttc	ccttttaggaa	ccttcccaa	1140
30	gcccatttgt	ggggatgggt	gcccatttgt	ggggatgggt	gtttatggtg	agtaggagat	1200
	gaggggacag	tttcaacttgt	gagggacttc	tctccattgt	ctccctcaca	aagcagactg	1260
	ccaccccaa	gtgttccaag	ccaaggctgg	tgccaccatc	acactcaagc	aacaggttct	1320
	gacatgtct	tagggcccc	cgaagtctgg	ctgtccctga	gggcttccag	tgagctagca	1380
	gagttggagac	cattttccca	cctccagatc	ttcggaaagga	agacccagac	cctccaagac	1440
35	tcacctgcgg	ggcgagaccc	tcaacatttc	atagtcttc	agggaaacagt	tgctgaaggg	1500
	ggcggggggg	tggcaccctg	taagttgtt	tttaaagatt	ttaaatgtct	ttaagatatc	1560
	actgtctaaa	taatatttgtt	ctgnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	1620
	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnncg	1680
	tttggagaat	taacaaagaa	aaaaactaag	acctagaatc	tcaccacata	accagctgtt	1740
	tcaatttttc	catatccctt	tttagttgtt	gttcatatgc	atacacaatt	tttacatagc	1800
40	tataatcaca	ggacaacaca	aatatgtat	tagttcttt	gaatttagaaa	aattacaag	1860
	ggcctatgtt	aaatgcacaa	actccaaagc	atataaagaa	aacatgcagt	ttccgcctc	1920
	ccgtttccct	tgccagaggt	aaccacgtt	agcagtttgc	tgaatagata	gttttgttagt	1980
	tggctttttt	tcttttggc	ctatcatcaa	tacattcata	tatagtcttg	ataattacca	2040
	gttactgtca	cgttaattgt	gtgcagaatc	atccctgtat	tatccctctt	tctaactaat	2100
	ctagattgaa	tctgtatgaga	gaaattctga	catatatgtt	caaattaaat	attgtctgtt	2160
45	ttatccctcgc	ataaaagtgct	atagcatttc	ccaaaggcccc	agtacagctg	tattaatagg	2220
	taaacttctc	tagatagaac	aaagcagtag	tctagaatct	tttggatataa	tttcccttat	2280
	ataataaaag	tctctcccc	aactctcca	tctccctctt	cctgtatgac	tttggtaaaa	2340
	cccatgtttc	agcatttcta	caattttgtat	tgtaactatc	tgcatacaca	gacaccacag	2400
	ggtctgactt	ggagttatgt	ctttcgtnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	2460
50	nnnnnnnnnn	nnnnnaatcag	aatttctctg	gagcaaacac	agccctgtgt	ttgtggaaat	2520
	ctcagtgttt	tatgtatttga	ttcatttttgc	tgtcag			2556

55 <210> 4

<211> 716

<212> ADN

<213> *Homo sapiens*

60 <220>

<221> intrón

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

<223> MGB-CEGP 1 int 5.1

65

ES 2 314 378 T3

<400> 4

5 **gtgagtgcc** **accccaaac** **tgagtgg** **tctgcacc** **cctgcctg** **cctacc** **ta** **60**
 cccccta **atg** **gttttag** **ca** **gatgcagg** **ctgttcc** **tc** **cat** **tttgcc** **ccccca** **gcag** **120**
 gccctgtg **ct** **gttgc** **cat** **g** **ta** **ca** **gtgg** **gg** **ag** **gtgg** **ctag** **gc** **ctgtgg** **aagg** **ccc** **tc** **180**
 tctccctgt **tg** **acctgg** **aa** **gccc** **ttcc** **tc** **ctc** **ctg** **gactagg** **ctg** **ctc** **ccta** **acgc** **240**
 tggtattcca **gagactg** **ca** **acac** **cctcc** **caggagg** **cca** **ggcagc** **acg** **aagttag** **agc** **300**

10 **tgttataat** **gatgcgg** **cac** **ttctggcc** **ag** **caggagg** **ccag** **ggccgtat** **at** **ttctggcg** **gg** **360**
 atgcctgc **c** **tgccct** **ticac** **ggtgtgc** **cct** **tcactag** **ctc** **cattt** **tagag** **gtttcc** **cagg** **420**
 ccaaggct **c** **ttttctcc** **c** **gactcagg** **gg** **actgaag** **ctt** **gcattcc** **cta** **gtgtct** **ctt** **480**
 ggtcagtgc **ca** **atatac** **ctcc** **aaaatctt** **tt** **ccatgtt** **aa** **tgttgc** **taa** **ggatctgt** **gg** **540**
 ccctttaac **g** **gcgtgt** **ct** **cccacag** **agc** **ctcattaca** **aa** **cacat** **tttta** **ttgcgt** **gaac** **600**
 agagtca **cat** **atcttcatt** **cctttatgt** **ctgggatt** **tc** **agcaa** **acaca** **gttgtat** **ggg** **660**
 gatgagca **at** **taactcatt** **cagtctgaga** **accgtgc** **tct** **tttgctt** **c** **ttgt** **tag** **716**

20 <210> 5

<211> 2041

<212> ADN

25 <213> *Homo sapiens*

<220>

<221> misc_feature

<222> (559)...(869)

30 <223> n= a, t, c, o g

<221> misc_feature

<222> (888)...(1052)

35 <223> n= a, t, c, o g

<221> misc_feature

<222> (1102)...(1380)

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

40 <221> misc_feature

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

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

<221> intrón

45 <222> (1841)...(1907)

<223> FOXM1 int 3.2

<221> intrón

50 <222> (1871)...(1936)

<223> FOXM1 int 3.1

55

60

65

ES 2 314 378 T3

<400> 5

5	gtaatgtgtc ccacagcaac caaaaatcaag gtcagcccg cctgacagtc ttcgcgtgc 60 tgtactgcaa cttgttatctg ggacagcagt taagtgc当地 ggacactaga atgataaaca 120 aatgtatctt ttagattgtg actcaatctt attgaatcca ggcaaaatca ttaagaagag 180 ctccttaact acttcatgtg ttactaccta aagtcctatgg agggtctca atgttagcact 240 caagcccact tttctgtac actcaacagg cgctccatagat gccagcagct agagtggcta 300 agtagttta taaaaatgtc ttgattaaaa aaaaaaatgc tgtctgttag cctcatgacc 360 caagatgtca ttcctgttag cgtcacatag catttctatgg gggcagggtt tttcttca 420 cttcattcat gaaaaagaccg agatgcctgt gagtcacat agctcacgca gttggtcgg 480 gtcagagcca caaatgaggt cttctgacgg gtgctcaatt ccaagtcaag tgtcttg 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 ttactccct 1080 tccatcagct tactttccta gnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1140 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1200 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1260 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1320 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1380 ttaatttcct agnnnnnnnnn 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
35	nnnnnnnnnn nntaattccc tagtttctta atttctctga gccaccttcc ttgctattga 1740 tcactacctc acagccttac tctgttttc tagccccctga cagctatcta ggtctttct 1800 ttatcacaat ctaaggttgg catcaatctt tattcccctga gaatagatgg gtttatggct 1860 gaaggtgacg gctctgcggg gtggagtgcc aggagagttg ccaagaggcgc tgcaaaagaca 1920 ccagacgaag cctgtgtctga gcacagtggg agggggcctga ggctggtttc cccatgtgtt 1980 tgaagggtga tgtttctgaa tctaaagtag ctgataacca gttgtcttgc tcttcttcca 2040 g 2041

45 <210> 6
 <211> 993
 <212> ADN
 <213> *Homo sapiens*
 50 <220>
 <221> intrón
 <222> (52)...(124)
 55 <223> FOXM1 int 4.1
 <221> intrón
 <222> (110)...(185)
 <223> FOXM1 int 4.2
 60 <221> misc_feature
 <222> (249)...(397)
 <223> n= a, t, c, o g
 65 <221> misc_feature
 <222> (442)...(749)
 <223> n= a, t, c, o g

ES 2 314 378 T3

<400> 6

5	gtgaatgccc	tgcttcctc	taaataggc	ctaagttgga	ggtgtcata	gccatctcaa	60
	aaggaaacaa	gttctgctag	tgatgcttc	atttgcac	gggagagtt	gaagccagcc	120
	acccaattag	tgacttgac	aaaaccagt	gaattaagta	cacctgacaa	ataccaaatg	180
	acacattttt	tgcccaagacc	agagcaagga	gaaggctgtt	ctgacccaac	agaaaggct	240
	ccccagggnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	300
10	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	360
	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnntac	aagaattct	ggaaatgctt	420
	gctctaaaaa	aagcccttcc	ttnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	480
	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	540
	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	600
15	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	660
	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	720
	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	cctgccttag	gctggagacc	agaagctgag	780
	ctaccagaac	gtcttttcag	aaagaagtt	ttttgggttt	tcagagtgcc	cataaggctg	840
	ctggtagctg	taaccatct	cctgggaggg	gcagttgtct	gggggtgtctt	ttgtcatcag	900
20	tcaggaataa	gtgttttcc	caatccggtc	aaattgacca	cgttggtggt	aacttcatct	960
	catttcctc	ccacaatgcc	tggccggcac	cag			993

25	<210> 7
	<211> 602
	<212> ADN
	<213> <i>Homo sapiens</i>
30	<220>
	<221> intrón
	<222> (0)...(0)
	<223> FOXM1 int 5.1
35	<400> 7

40	gtaagttct	ttccctctgg	ctcggggctt	ggccttgttt	tcctttact	gctcagcatg	60
	gctttagtgg	acagagacaa	gatgtgtgt	ggggaaagggt	ccctatggcc	atgtttgtc	120
	taggtgccag	ccctagacac	agaacacccct	gagggtcagg	cacacaccca	cttccctccc	180
	cttccatggg	catcacaagg	gcacactgag	cagagcaggg	cacagcaggg	gagcatgctg	240
45	cagcagccac	aagcgcattgg	caccaggcctc	aggggcggca	gttcgttcgc	tcactttgt	300
	gcctagctt	tcttgcac	gcataatagct	acctgtctg	gcataccccca	gggggtgttga	360
	ggacacgtgg	gtgaagcggt	agtgccactc	tgccatcatg	tgtctgttagg	ccacccacac	420
	gcccactcat	cacagtttg	gagactgctc	gcctacgtcc	atcccctcag	gttggccctcc	480
50	tctctctggg	ctgtcattaa	ctcaagcaca	caccaccaga	gcagctggtg	gggttttgcc	540
	atcccctctt	taccttattg	tgttaacata	ggtttcttc	tctcccatc	tgccacaagc	600
	ag					602	

55	<210> 8
	<211> 4656
	<212> ADN
	<213> <i>Homo sapiens</i>
60	<220>
	<221> intrón
	<222> (0)...(0)
65	<223> FOXM1 int 7.1
	<221> misc_feature
	<222> (316)...(615)

ES 2 314 378 T3

<223> n= a, t, c, o g
 <221> misc_feature
 <222> (626)...(930)
 5 <223> n= a, t, c, o g
 <221> misc_feature
 <222> (1044)...(1259)
 10 <223> n= a, t, c, o g
 <221> misc_feature
 <222> (1452)...(1802)
 <223> n= a, t, c, o g
 15 <221> misc_feature
 <222> (2085)...(3116)
 <223> n= a, t, c, o g
 <221> misc_feature
 20 <222> (3138)...(3433)
 <223> n= a, t, c, o g
 <221> misc_feature
 25 <222> (3442)...(3728)
 <223> n= a, t, c, o g
 <221> misc_feature
 <222> (3814)...(4081)
 30 <223> n= a, t, c, o g
 <221> misc_feature
 <222> (4289)...(4371)
 35 <223> n= a, t, c, o g

 <400> 8

 40 **gtgggtgtcc tattttcctc tgaagagaga ttctggccaa ttaagaatgt tggaccttca 60**
 gcttgcaaag cactctgata agtgtccctt gagagcttat aaatcttagtt gggttagaaaaa 120
 ggcataaaaaa catagggaaag tctaatacgca ttagaagagc taaaaggtt tttggattac 180
 aatgttaagtg gtgtcagaag gcccataaaat acctgtatgag cttgttaagaa ttcatagacaaa 240
 45 agtgattgtg atagatgggc taggattatt aaggaagata cacaaggag gcaggcctta 300
 gaaagagatg gattnnnnnn 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
 50 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
 nnnnnnnnnn nnnnnngtggaa tttgarnnnn 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
 55 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 900

60

65

ES 2 314 378 T3

	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	gtagattgg	gtaagcaaac	aggtagttag	960
5	agagcatgct	aatgggcagt	gccatggagg	Cggaaatgc	agttcgtaacc	tggcagtagt	1020
	aaagtgactg	ggtcagacta	actnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	1080
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	1140
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	1200
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	1260
10	tctagcttga	gggaaagggt	agaagggtaa	attcagagcc	aacttggatc	agccatcaga	1320
	tctgactta	acactgttaa	agggttctgt	gagtacggc	tgacatgtaa	ccaaagtgaa	1380
	aagctcccc	catcccccttc	agagagatga	aaatagcata	gagtctggag	tttagagcga	1440
	cttgggtttg	cnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	1500
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	1560
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	1620
15	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	1680
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	1740
	nncatattac	acacaaaatt	ataccacaca	tacataattt	agcgtaaaatt	cattcatgtg	1800
	gccgttagcat	gtgccctgtt	tgggtttca	tgcagtgggt	tttctccct	ttccttttg	1860
	gctccctctc	caccctacca	tcacccacat	caccctact	cccaagataa	ctggttgata	1920
20	atttatgtatg	ctttcttgca	tattttatca	atgctcttag	tataactata	catgtatagc	1980
	gatagccatt	ttatatgtac	acatacaaca	cacagaacat	tgatnnnnnn	nnnnnnnnnnn	2040
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	2100
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	2160
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	2220
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	2280
25	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	2340
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	2400
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	2460
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	2520
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	2580
30	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	2640
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	2700
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	2760
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	2820
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	2880
35	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	2940
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	3000
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	3060
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	3120
	tcaatttgtg	agagctcnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	3180
40	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	3240
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	3300
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	3360
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	3420
	nnnnnnnnnnn	nnntgtgaga	gnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	3480
45	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	3540
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	3600
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	3660
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	3720
	nnnnnnnnnnncc	ctgtgagagc	tcttttgtgt	taaaataatc	ttctttcttt	tatgtcaag	3780
50	atattttct	acttctattg	ttatctctt	tacnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	3840
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	3900
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	3960
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	4020
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	4080
	ntgttttgtt	tttcttaact	ttgtttatgg	tatcttgc	cacagtaaaa	ttttaaagtt	4140
55	ttatgtatgc	aaatgtctct	cttctctttt	acagttctg	ggtttccagt	cttggtaag	4200
	aagtcaccc	gcacccctag	attgtatatg	tagtctcta	gattctcttt	caggattgt	4260
	atgatttaa	ggtttctatt	ttttttttan	nnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	4320
	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	nnnnnnnnnnn	4380
	tattttcatg	cattaaatct	ttgtatatacg	tgtaagacaa	gcatgcaatt	ttatccctc	4440
	tcggatgaat	gttattataa	ttatgccact	acatactaca	tacccgcac	ttttacccc	4500
60	agaattgaac	taccaacttc	aacatacatc	gtattctcat	atthaataga	ttttaagact	4560
	tcaaaaacgac	acaaagagga	tcagaacccg	tatgtatata	ttttgtgcgt	cctgtctgg	4620
	gaccgttggt	tcaccttatac	tctgtttcccc	tttcag			4656

ES 2 314 378 T3

<210> 9
 <211> 614
 <212> ADN
 5 <213> *Homo sapiens*
 <220>
 <221> intrón
 10 <222> (0)...(0)
 <223> MGB-PRAME int 2.1

 <400> 9

 15

 gtaagttcga gccctgattc ctccgcgttcc ccgcagggtg accttggct tggcccccg 60
 gcaccacccc tgtccgggtt ccctgtttc tcgtggaaa tgggttgaag accaaagaaa 120
 ataatgtgcg ccacttgggt caccggggc cgccgtcccc ggaaaaattgg ccccaagtgt 180
 20 ggagttgtgg ctgttaaggat gccttgaacc gaggcgccgg tgctcggtt tgagactctc 240
 cagggtgggt ggcgcattgt aatgcgggtt atgcgttgg actcgcccccc tctgaagggt 300
 ctgggggtt gggacggccc aggcagtggc gtaggcgtcc taggaaggcg ggagcagagg 360
 cagaaatgtc gctgcagac cgttagtcagg gtcccttgacc acagggtca ctgtgacca 420
 25 accacatggc ctgttgttcc tcctgcccccc tgttcagcc cagggaaacac tggtgctcag 480
 gtttggagcc agagatttgc actgaaaggg cgggattttag tcgcccatttgc tca 540
 cagcgttatt tgcggagggtt ttcacaggag gccgttgctt cgtaaatatt atacatgtat 600
 tcttctttt ggag 614

30 <210> 10
 <211> 432
 <212> ADN
 35 <213> *Homo sapiens*
 <220>
 <221> intrón
 <222> (0)...(0)
 40 <223> MGB-PRAME int 4.1

 <400> 10

 45 gtaagggtga cctagcagct tgggtgtgggg ccctgggaac ctgagcagga tgcagctggg 60
 gtcaggggagc atggagcgcc taaggctggg ccaggaggctc tggatgggtgc cagcaaggaa 120
 gttcaggggag gccttggggc tactgcagggt gtcactcttgc gaatgggctt ctggacatgg 180
 50 ggcactgtatt aaaatgcaga ggtgtctgaa ggaacatgca cctgcttcct cctgggtgggg 240
 tgggaatttgg ggaccaggaa ggatcccagg atccttagtgg gaaaggggagc agctgtgccc 300
 tgaagtacga agtaaaagtg cagatctaag gtggatgtct gtttgggtct tacctacatt 360
 atgagactca tggtcttatt ttgagttgat cttaaagcat catctcagct aattacatgt 420
 ttttccccac ag 432

55 <210> 11
 <211> 3740
 60 <212> ADN
 <213> *Homo sapiens*
 <220>
 <221> intrón
 65 <222> (0)...(0)

ES 2 314 378 T3

<223> MGB-STK15 int 1.1
<221> misc_feature
<222> (84)...(151)
5 <223> n= a, t, c, o g
<221> misc_feature
<222> (321)...(491)
10 <223> n= a, t, c, o g
<221> misc_feature
<222> (537)...(615)
<223> n= a, t, c, o g
15 <221> misc_feature
<222> (853)...(894)
<223> n= a, t, c, o g
<221> misc_feature
20 <222> (1259)...(1309)
<223> n= a, t, c, o g
<221> misc_feature
25 <222> (1337)...(2516)
<223> n= a, t, c, o g
<221> misc_feature
<222> (2521)...(3567)
30 <223> n= a, t, c, o g

35

40

45

50

55

60

65

ES 2 314 378 T3

<400> 11

ES 2 314 378 T3

<210> 12
 <211> 1622
 <212> ADN
 5 <213> *Homo sapiens*
 <220>
 <221> intrón
 10 <222> (0)...(0)
 <223> MGB-STK15 int 2.1
 <221> misc_feature
 <222> (705)...(989)
 15 <223> n= a, t, c, o g

 <400> 12

 20
 gtaaaattgaa taatctgtaa ttcattcac atttataaac ccacatggag gttggcttg 60
 tcgggaattc ttccgcctt tactttggat ttaaaatttag atcccttact gtgatcctgg 120
 atatgaatta gtcactttc tcgtgttcag taacattttg ctgccttcta gagtagcttt 180
 tttgttctgc ttgtcttat aatcggtcgc ttaagtttct atatcccctcc actgtatgca 240
 ggataaatgt aataatgcat ctggcaggag ttcaaaaactt ttaaaaatttgg ccataaatat 300
 aaaataatTA gaaaaaggct accttgaatt actgtatTTG attctaaagtT cctatgataa 360
 cggccattTA aaaaatttgc ctatattttaa aatgtttCTC tttatTTGTC tttgtctgaa 420
 tgccTGTcGc gttgtggaca gtgtgctaat ttcaaggatTA actgactttg tatttggaaag 480
 tcttaacacc ctctcttgc agagcactca taccgttgag ctggggatgg actttggaggc 540
 tttcatttct agcacttgc cctcaTTac aatgagctgt tgaagctgaa ggaaatctca 600
 tccCTCCTAC cccttttagt ttgatttagct gagggtgtta gagttAACTT aacaatttaa 660
 ggttgtataA cagttactac aggctataaa ataatacatt tcaannnnnn nnnnnnnnnnnn 720
 nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn 780
 nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn 840
 35 nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn 900
 nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn 960
 nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn 1020
 tgtgaagtgt ctacatttg ttccttttgc tcaTTTTT gctcaattcc attttgtcaa 1080
 tacttggaaa atgaaacatt ggttaatcaa tagtacagta ataagcttata tttggaaaat 1140
 40 cttcgatata tgaaaactta gactcttcta aaacttcatg aagataatac cactgttgaa 1200
 cgTTTgacG tattttttt tggctttttt cttaaacgtt tattatcaa gaaatttcaa 1260
 tggaaactgag attttggcat aaagtttttg tatcatagct ttitgccaAA tagcaatgta 1320
 gtgtctattt ccaaattttt gagaattttt agaaagtgtc tccttcattt atggatattt 1380
 gttataaaag catgattttt aggggtgagg aattggaggg gatagaaggt atcattcagg 1440
 45 tatttttagc cacatactaa ctatccctcg gaggtaactg ttttttttcc ttttcacctt 1500
 ccattcttta tcagtgacat tcattttttt gctatactag agaacaact ttgtgaaatt 1560
 ctcaatatat tcatttttg ctttcatgaa tgccagaaag ttatTTCT cttccattct 1620
 ag 1622

 50
 <210> 13
 <211> 1093
 <212> ADN
 55 <213> *Homo sapiens*
 <220>
 <221> intrón
 60 <222> (0)...(0)
 <223> MGB-STK15 int 4.1

ES 2 314 378 T3

<400> 13

5	gtaagcttt ttatttacaa agttctgtac tggttctacta gaatataatta tttcggttgc	60
	aatttcgttg tggttgcatt tagaggatataatgtttcc agatttccaa tcttaaaaaaaaatggaaatttgtgtaatga ggtatttac	120
	taggaactca agtgctttaa aaaatggctt tcaaatttag aaaaagcttg tatgaatctt	240
10	ttatagaaat gtgttggaaat tcctctctgt ccttagaaat aaccactaca tatggtttat	300
	gcgtctgtac ttttttattt tacaaaatgtt caagtttttta aaaaatagaa tatgtttgcag	360
	aactatatac tcataatatac ctgagggttt tgacagtatt atagtttag ttctttatgt	420
	taaagggttgg ctgttaatgtc tcccccaggg cttttctaaa agcctctct cagtctctga	480
	actatcttggaa ctctttagaaatg taccggggagg agcgaggaaat gaaccacag actcttttgc	540
15	tttttagcggt ctaacagagg ctaagagtctt aaatccactg gttctcatgc cccagctagc	600
	ctgtgggctc catcccgtt ccatttagaa cagtggctt ctgtccacca ccagagtgt	660
	tctccaccca gagagaatttgcaccccttgg gactggggagg agcagctggg gtttagtttga	720
	aacatccccc cagatggctt ggaagcatttgcctt ggtcaacttgcctt gacatttccc	780
20	gtcttcagcg ttgtcatggc cctgttccctc tgagcatatg acgggcttgg gacatttccc	840
	atagagtctt tcaggtctaa aacccgagac tgctccctgtt cactgactct cacacctgac	900
	ggcagctagg gacgtcagggtttcatgtcg tggcagcttctt tgatagttgg ttattgcctt	960
	ggttcttgctt gaggatgcatttgatgaa gttggaaatac gaaatttattt gtagaatgtg	1020
	tctgtactc attggaaaatt tgtttagaaaaa gctttgtttt cttcacattc taaagtgttc	1080
	aaatttcctcc tag	1093

25

<210> 14

<211> 18

30 <212> ADN

<213> Secuencia Artificial

<220>

35 <223> Cebador directo

<400> 14

40	agcgccctgtt ccgatctg
----	-----------------------------

<210> 15

<211> 24

45 <212> ADN

<213> Secuencia Artificial

<220>

<223> Cebador inverso

50 <400> 15

aacccaaatgtt ctctgctgaa aacc

55 <210> 16

<211> 15

<212> ADN

60 <213> Secuencia Artificial

<220>

<223> Sonda

65 <400> 16

ccctgaagca gcaac

ES 2 314 378 T3

<210> 17
<211> 22
<212> ADN
5 <213> Secuencia Artificial
<220>
<223> Cebador directo

10 <400> 17

ctgttgctgt gtgatgctgt ca

15 <210> 18
<211> 20
<212> ADN
20 <213> Secuencia artificial
<220>
<223> Cebador inverso

25 <400> 18

cctcagccac tcccttgate

30 <210> 19
<211> 15
<212> ADN
<213> Secuencia Artificial
35 <220>
<223> sonda

<400> 19
40 tcaggcata agcct

<210> 20
45 <211> 19
<212> ADN
<213> Secuencia Artificial
<220>
50 <223> Cebador directo

<400> 20

55 tccccttgcc ttggagaa

<210> 21
<211> 19
60 <212> ADN
<213> Secuencia Artificial
<220>
65 <223> Cebador inverso

ES 2 314 378 T3

1 <400> 21

aaaggcctgg aggcataa

5 <210> 22

<211> 13

<212> ADN

10 <213> Secuencia Artificial

<220>

<223> Sonda

15 <400> 22

cagcccaaat cct

20 <210> 23

<211> 25

<212> ADN

25 <213> Secuencia Artificial

<220>

<223> Cebador directo

30 <400> 23

cttaatggtg tttagcacag atgca

35 <210> 24

<211> 20

<212> ADN

<213> Secuencia Artificial

40 <220>

<223> Cebador inverso

<400> 24

45 ccactgtac atgcgaagca

<210> 25

50 <211> 16

<212> ADN

<213> Secuencia Artificial

<220>

55 <223> Sonda

<400> 25

60 caaatgcaca ggaaac

<210> 26

<211> 19

65 <212> ADN

<213> Secuencia Artificial

ES 2 314 378 T3

<220>
<223> Cebador directo

5 <400> 26
gctctcggt gtggagtg

10 <210> 27
<211> 21
<212> ADN
<213> Secuencia Artificial

15 <220>
<223> Cebador inverso
<400> 27

20 cacaggcttc gtctgggtc t

25 <210> 28
<211> 24
<212> ADN
<213> Secuencia Artificial

30 <220>
<223> Sonda
<400> 28

35 tgcagccctc ttggcaactc tcct
<210> 29
<211> 23

40 <212> ADN
<213> Secuencia Artificial
<220>

45 <223> Cebador directo
<400> 29

50 aaaatgctgt ctgtgagcct cat
<210> 30
<211> 21

55 <212> ADN
<213> Secuencia Artificial
<220>
<223> Cebador directo

60 <400> 30
aacccctgcc cactagaaat g

65 <210> 31
<211> 30

ES 2 314 378 T3

<212> ADN
<213> Secuencia Artificial
<220>
5 <223> Sonda

<400> 31

10 acccaagatg tcatctcctg tagcgtcaca

<210> 32
<211> 25
15 <212> ADN
<213> Secuencia Artificial
<220>
20 <223> Cebador directo

<400> 32

25 aatagatggg tttatggctg aaggt

<210> 33
<211> 23
30 <212> ADN
<213> Secuencia Artificial
<220>
<223> Cebador inverso

35 <400> 33

ctttggcaa ctctcctgac act

40 <210> 34
<211> 13
<212> ADN
45 <213> Secuencia Artificial
<220>
<223> Sonda

50 <400> 34

ccgcagagcc gtc

55 <210> 35
<211> 25
<212> ADN
<213> Secuencia Artificial
60 <220>
<223> Cebador directo

65 <400> 35

ccatctcaa aggaaacaag ttctg

ES 2 314 378 T3

<210> 36
<211> 21
<212> ADN
5 <213> Secuencia Artificial
<220>
<223> Cebador inverso

10 <400> 36

gggtggctgg cttctaactc t

15 <210> 37
<211> 24
<212> ADN
20 <213> Secuencia Artificial
<220>
<223> Sonda

25 <400> 37

ccctgatcaa atgaaagcat cact

30 <210> 38
<211> 20
<212> ADN
<213> Secuencia Artificial
35 <220>
<223> Cebador directo

<400> 38
40
agaaggccagc cacccaaata

45 <210> 39
<211> 26
<212> ADN
<213> Secuencia Artificial
<220>
50 <223> Cebador inverso

<400> 39

55 tggttcattt ggtatttgc aagtgt

60 <210> 40
<211> 26
<212> ADN
<213> Secuencia Artificial
<220>
65 <223> Sonda

ES 2 314 378 T3

<400> 40

tgacttgcac aaaacccagt gaatta

5 <210> 41
<211> 24

<212> ADN

10 <213> Secuencia Artificial
<220>
<223> Cebador directo

15 <400> 41

tggacagaga caagatgtga tgtg

20 <210> 42
<211> 22
<212> ADN
<213> Secuencia Artificial
25 <220>
<223> Cebador inverso

30 <400> 42

gctggcacct agacaaaaca tg

35 <210> 43
<211> 15
<212> ADN
<213> Secuencia Artificial
40 <220>
<223> Sonda

<400> 43

45 ccatagggac ccttc

<210> 44
50 <211> 25
<212> ADN
<213> Secuencia Artificial
<220>
55 <223> Cebador directo

<400> 44

60 ggtgtcctat tttcctctga agaga

<210> 45
<211> 21
65 <212> ADN
<213> Secuencia Artificial

ES 2 314 378 T3

<220>
<223> Cebador inverso

5 <400> 45
tgcaagctga aggtccaaca t

10 <210> 46
<211> 15
<212> ADN
<213> Secuencia Artificial

15 <220>
<223> Sonda

<400> 46

20 ttctggccaa ttaag

<210> 47

25 <211> 21
<212> ADN
<213> Secuencia Artificial

30 <220>
<223> Cebador directo

<400> 47

35 tcattcatgt ggccgttagca t

<210> 48

40 <211> 19
<212> ADN
<213> Secuencia Artificial

45 <220>
<223> Cebador inverso

<400> 48

50 ggtggagagg gagccaaaa

<210> 49

55 <211> 16
<212> ADN
<213> Secuencia Artificial

60 <220>
<223> Sonda

<400> 49

cctgtttggg ttttca

65 <210> 50
<211> 23

ES 2 314 378 T3

<212> ADN
<213> Secuencia Artificial
<220>
5 <223> Cebador directo

<400> 50

10 agaggatcag aaccgtatg tga

<210> 51
<211> 24
15 <212> ADN
<213> Secuencia Artificial
<220>
20 <223> Cebador inverso

<400> 51

25 gggaaacaga gataaggta acca

<210> 52
<211> 15
30 <212> ADN
<213> Secuencia Artificial
<220>
<223> Sonda

35 <400> 52

tgcgttcct gtctg

40 <210> 53
<211> 19
<212> ADN
45 <213> Secuencia Artificial
<220>
<223> Cebador directo

50 <400> 53

gggtgaccc ttccatgt

55 <210> 54
<211> 23
<212> ADN
<213> Secuencia Artificial
60 <220>
<223> Cebador inverso

65 <400> 54

ttcaaccc ttccatgt gaa

ES 2 314 378 T3

<210> 55
<211> 14
<212> ADN
5 <213> Secuencia Artificial
<220>
<223> Sonda

10 <400> 55

cccggtccc tgtt

15 <210> 56
<211> 21
<212> ADN
20 <213> Secuencia Artificial
<220>
<223> Cebador directo

25 <400> 56

agggtgacct agcagcttgg t

30 <210> 57
<211> 18
<212> ADN
<213> Secuencia Artificial
35 <220>
<223> Cebador inverso

<400> 57

40 gcctctggcc cagccta

<210> 58
45 <211> 16
<212> ADN
<213> Secuencia Artificial
<220>
50 <223> Sonda

<400> 58

55 tccctgaccc cagctg

<210> 59
<211> 25
60 <212> ADN
<213> Secuencia Artificial
<220>
65 <223> Cebador directo

ES 2 314 378 T3

<400> 59

cgtaatgtct cttectcttc cgtaa

5 <210> 60
<211> 24

<212> ADN
10 <213> Secuencia Artificial
<220>
<223> Cebador inverso

15 <400> 60

acgaactgag tagtttgctg aaaa

20 <210> 61
<211> 16
<212> ADN
25 <213> Secuencia Artificial
<220>
<223> Sonda

30 <400> 61

tcaagggaca aggaag

35 <210> 62
<211> 26
<212> ADN
<213> Secuencia Artificial
40 <220>
<223> Cebador directo

<400> 62

45 cattcacatt tataaccca catgga

<210> 63
50 <211> 23
<212> ADN
<213> Secuencia Artificial
<220>
55 <223> Cebador inverso

<400> 63

60 aatccaaagt aaaggcgaa aga

<210> 64
<211> 16
65 <212> ADN
<213> Secuencia Artificial

ES 2 314 378 T3

<220>

<223> Sonda

5 <400> 64

tggtcttgtc ggaaat

10 <210> 65

<211> 20

<212> ADN

<213> Secuencia Artificial

15 <220>

<223> Cebador directo

<400> 65

20

gcgaggaatg aaccacaga

<210> 66

25 <211> 25

<212> ADN

<213> Secuencia Artificial

30 <220>

<223> Cebador inverso

<400> 66

35

gcatgagaac cagtggattt agact

<210> 67

40 <211> 16

<212> ADN

<213> Secuencia Artificial

<220>

45 <223> Sonda

<400> 67

50

cgctaaaagc aaaaga

55

60

65