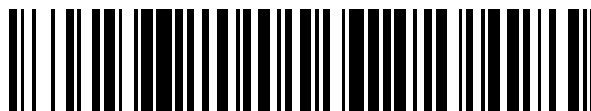


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 525 545**

51 Int. Cl.:

**G01N 33/574** (2006.01)

**G01N 33/48** (2006.01)

**C12Q 1/68** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **19.09.2006 E 11163930 (8)**

97 Fecha y número de publicación de la concesión europea: **10.09.2014 EP 2402758**

54 Título: **Métodos y usos para identificar el origen de un carcinoma de origen primario desconocido**

30 Prioridad:

**19.09.2005 US 718501 P**

**12.10.2005 US 725680 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

**26.12.2014**

73 Titular/es:

**JANSSEN DIAGNOSTICS, LLC (100.0%)**

**700 US Highway 202  
Raritan, NJ 08869 , US**

72 Inventor/es:

**WANG, YIXIN;  
MAZUMDER, ABHIJIT;  
TALANTOV, DMITRI;  
JATKOE, TIMOTHY y  
BADEN, JONATHAN**

74 Agente/Representante:

**IZQUIERDO FACES, José**

**ES 2 525 545 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).

**Métodos y usos para identificar el origen de un carcinoma de origen primario desconocido****Descripción**

## 5 CAMPO DE LA INVENCION

La invención proporciona métodos para identificar el origen de un carcinoma de origen primario desconocido.

## 10 ANTECEDENTES DE LA INVENCION

El carcinoma de origen primario desconocido (CUP) es un conjunto de tumores malignos confirmados por biopsia, heterogéneos, en donde la enfermedad metastásica presenta sin un sitio del tumor primario identificable o tejido de origen (toO). Este problema representa aproximadamente un 3-5% de todos los cánceres, haciendo de él el séptimo tumor maligno más común. Ghosh et al. (2005); y Mintzer et al. (2004). El pronóstico y régimen terapéutico de los pacientes son dependientes del origen del tumor primario, destacando la necesidad del sitio del tumor primario. Greco et al. (2004); Lembersky et al. (1996); y Schlag et al. (1994).

Se usan actualmente una variedad de métodos para resolver este problema. Varios métodos seguidos se esquematizan en las Figuras 1-2. Se pueden usar Marcadores tumorales séricos para diagnosis diferencial. Aunque carecen de especificidad adecuada, pueden ser usados en combinación con información patológica o clínica. Los métodos de Ghosh et al. (2005). Immunohistochemical (IHC) pueden ser usados para identificar el linaje tumoral pero pocos Marcadores IHC son 100% específicos. Por lo tanto, los patólogos usan a menudo un panel de Marcadores IHC. Varios estudios han demostrado precisiones del 66-88% usando de cuatro a 14 Marcadores IHC. Brown et al. (1997); DeYoung et al. (2000) y Dennis et al. (2005a). Estudios diagnósticos más caros incluyen métodos de obtención de imágenes como radiografía de tórax, tomografía computarizada (TC), y tomografías de emisión de positrones (PET). Cada uno de estos métodos puede identificar el primario en un 30 a 50% de los casos. Ghosh et al. (2005); y Pavlidis et al. (2003). A pesar de estas sofisticadas tecnologías, la capacidad de resolver casos de CUP es sólo del 20-30% ante mortem. Pavlidis et al. (2003); y Varadhachary et al. (2004).

Un nuevo enfoque prometedor reside en la capacidad de perfilado de expresión génica de todo el genoma para identificar el origen de los tumores. Ma et al. (2006); Dennis et al. (2005b); Su et al. (2001); Ramaswamy et al. (2001); Bloom et al. (2004); Giordano et al. (2001) y 20060094035. Estos estudios demostraron la factibilidad del tejido de identificación del origen en base al perfil de expresión génica. Para que estas tecnologías de perfilado de expresión sean útiles en el ámbito clínico, se deben superar dos obstáculos principales. Primero, como el perfilado de la expresión génica se realizó completamente en tejidos primarios, los candidatos a marcador de genes deben ser validados en tejidos metastásicos para confirmar que su expresión específica del tejido se conserva en la metástasis. Segundo, la tecnología de perfilado de expresión génica debe de ser capaz de usar tejido fijado en formalina, embebido en parafina (FFPE), ya que las muestras de tejido fijado son el material estándar en la práctica actual. Los fijación en formalina resulta en la degradación del ARN (Lewis et al. (2001) y Masuda et al. (1999)) por lo que los protocolos de micromatriz existentes no se realizarán con fiabilidad. Bibikova et al. (2004). Adicionalmente, la tecnología de perfilado debe ser robusta, reproducible y fácilmente accesible.

El RTPCR cuantitativo (qRTPCR) ha demostrado que genera resultados fiables de tejido FFPE. Abrahamsen et al. (2003); Specht et al. (2001); Godfrey et al. (2000) y Cronin et al. (2004). Por lo tanto, un enfoque más práctico sería usar un método de todo el genoma como una herramienta de descubrimiento y desarrollar un ensayo diagnóstico basado en una tecnología más robusta. Ramaswamy (2004). este paradigma, sin embargo, requiere un conjunto de genes más pequeño para ser desarrollado. Oien y sus compañeros usaron análisis en serie de la expresión génica (SAGE) para identificar 61 Marcadores tumorales de los que desarrollaron el método RTPCR en base a once genes para cinco tipos de tumores. Dennis et al. (2002). Otro estudio que acopla SAGE y qRTPCR desarrolló un panel de cinco genes para cuatro tipos de tumores y consiguió una precisión del 81%. Buckhaults et al. (2003). Un estudio más reciente acopló perfilado de micromatrices con qRTPCR, pero usó 79 Marcadores. Tohill et al. (2005).

## 55 RESUMEN DE LA INVENCION

La presente invención proporciona un método para identificar el origen de una metástasis de origen desconocido en una muestra que contiene células metastásicas que comprende los pasos de a) medir los Biomarcadores asociados con al menos seis carcinomas diferentes en donde los Biomarcadores son niveles de expresión de genes Marcadores en donde los genes Marcadores son SP-B, TTF1, DSG3, PSCA, F5, HPT1, PSA, MGB/MB, WT1 y PDEF, b) combinar los datos de los Biomarcadores en un algoritmo donde el algoritmo: normaliza los Biomarcadores frente a la referencia; e impone un punto de corte que optimiza la sensibilidad y especificidad de cada Biomarcador, pondera la prevalencia de carcinomas y selecciona un tejido de origen; c) determinar el origen en base a la probabilidad más alta determinada por el algoritmo o determinar que el carcinoma no se deriva de un conjunto particular de carcinomas; y d) medir opcionalmente Biomarcadores específicas para uno o más carcinomas

diferentes adicionales, y repetir los pasos c) y d) como sea necesario para Biomarcadores adicionales.

BREVE DESCRIPCION DE LOS DIBUJOS

5 Las Figuras 1-2 representan métodos del estado de la técnica para identificar el origen de una metástasis de origen desconocido.  
 La Figura 3 representa el algoritmo de diagnóstico CUP actual.  
 La Figura 4 representa datos de micromatrices que muestran intensidades de dos genes en un panel de tejidos. (A) Antígeno de células madre de próstata (PSCA). (B) Factor de coagulación V (F5). Los gráficos de barras muestran la intensidad en el eje y y el tejido en el eje x. Panc Ca, cáncer de páncreas; Panc N, páncreas normal.  
 10 La Figura 5 representa electroferogramas obtenidos de un Bioanalizador Agilent.  
 El ARN fue aislado de tejido FFPE usando una digestión de proteinasa K de tres horas (A) o dieciséis horas (B). La muestra C22 (roja) era un bloque de un año de edad mientras que la muestra C23 (azul) era un bloque de cinco años de edad. Se muestra una escala de tamaño en verde.  
 15 La Figura 6 representa una comparación de valores de Ct obtenidos de tres métodos qRT-PCR diferentes: cebado de hexámeros aleatorio en la transcripción inversa seguido por qPCR con el ADNc resultante (paso RH 2), cebado específico de genes (cebador inverso) en la transcripción inversa seguido por qPCR con el ADNc resultante (paso GSP 2), o cebado específico de genes y qRT-PCR en una reacción de una etapa (paso GSP 1). Se dividió ARN de once muestras en los tres métodos y se midieron los niveles de ARN para los tres genes;  $\beta$ -actina (A), HUMSPB (B), y TTF (C). El valor de Ct mediano obtenido con cada método se indica por la línea sólida.  
 20 La Figura 7 representa diagramas de placas de ensayo CUP.  
 La Figura 8 es una serie de gráficos que representan el rendimiento del ensayo sobre un intervalo de concentraciones ARN.  
 25 La Figura 9 es un diagrama de flujo de trabajo experimental; nominación y validación de candidatos a Marcadores (9A); y optimización de ensayos y construcción y prueba del algoritmo de predicción (9B).  
 La Figura 10 representa la expresión de 10 candidatos a Marcadores de genes específicos de tejidos seleccionados en carcinomas metastásicos de FFPE y adenocarcinoma primario de próstata. Para cada gráfico el eje X representa el valor de expresión del Marcador normalizado.  
 30 La Figura 11 representa la optimización del ensayo (a y B) Electroferogramas obtenidos de un Bioanalizador Agilent. El ARN se aisló de tejido FFPE usando una digestión de proteinasa K de tres horas (A) o dieciséis horas. La muestra C22 (roja) era un bloque de un año de edad mientras que la muestra C23 (azul) era un bloque de cinco años de edad. Se muestra una escala de tamaño en verde. (C y D). Comparación de valores de Ct obtenidos de tres métodos qRT-PCR diferentes: cebado de hexámeros aleatorio en la transcripción inversa seguido por qPCR con el ADNc resultante (paso RH 2), cebado específico de genes (cebador inverso) en la transcripción inversa seguido por qPCR con el ADNc resultante (paso GSP 2), o cebado específico de genes y qRT-PCR en una reacción de una etapa (paso GSP 1). Se dividió ARN de once muestras en los tres métodos y se midieron los niveles de ARN para los dos genes;  $\beta$ -actina (C) y HUMSPB (D). El valor de Ct mediano obtenido con cada método se indica por la línea sólida.  
 35 La Figura 12 es un mapa térmico que muestra los niveles de expresión relativos de los 10 paneles de Marcadores a través de 239 muestra. El rojo indica expresión más alta.  
 40

DESCRIPCION DETALLADA

45 El identificar el sitio primario en pacientes con carcinoma metastásico de origen primario desconocido (CUP) puede permitir la aplicación de regímenes terapéuticos específicos y puede prolongar la supervivencia. Los candidatos a marcador fueron entonces validados por reacción en cadena de la polimerasa de transcriptasa inversa (RT-PCR) en 205 carcinomas metastásicos FFPE originarios de estos seis tejidos así como metástasis originarias de otros tipos de cáncer para determinar la especificidad. Se seleccionó una firma de diez genes que predecía el tejido de origen de carcinomas metastásicos para estos seis tipos de cáncer. Después, se optimizaron los métodos de aislamiento de ARN y qRT-PCR para estos diez Marcadores, y se aplicó el ensayo qRT-PCR a un conjunto de 260 tumores metastásicos, generando una precisión general del 78%. Por último, se probó un conjunto independiente de 48 muestras metastásicas. Es importante destacar que treinta y siete muestras en este conjunto tenían o un primario conocido o se presentaban inicialmente como CUP pero se resolvieron posteriormente, y el ensayo demostró una precisión del 78%.

60 Un Biomarcador es cualquier indicio del nivel de expresión de un gen Marcador indicado. El indicio puede ser directo o indirecto y medir la sobre- o bajo-expresión del gen dados los parámetros fisiológicos y en comparación con un control interno, tejido normal u otro carcinoma. Los Biomarcadores incluyen, sin limitación, ácidos nucleicos (tanto sobre y -bajo-expresión y directos e indirectos). Usar ácidos nucleicos como Biomarcadores puede incluir cualquier método conocido en la técnica incluyendo, sin limitación, medir de la amplificación de ADN, ARN, micro ARN, pérdida de heterocigosidad (LOH), polimorfismos de nucleótido único (SNPs, Brookes (1999)), ADN de microsatélites, hipo- o hiper-metilación de ADN. Usar proteínas como Biomarcadores incluye cualquier método conocido en la técnica incluyendo, sin limitación, medir la cantidad, actividad, modificaciones como glicosilación,  
 65

fosforilación, ADP-ribosilación, ubiquitinación, etc., o inmunohistoquímica (IHC). Otros Biomarcadores incluyen obtención de imágenes, recuento celular y Marcadores de apoptosis.

5 Los genes indicados proporcionados en la presente son los asociados con un tumor o tipo de tejido particular. Un gen Marcador puede estar asociado con numerosos tipos de cáncer pero siempre que la expresión del gen esté lo suficientemente asociada con un tumor o tipo de tejido a ser identificado usando el algoritmo descrito en la presente para ser específico para un origen particular, el gen puede ser usado en la invención reivindicada para determinar el tejido de origen para un carcinoma de origen primario desconocido (CUP). Se conocen en la técnica numerosos genes asociados con uno o más cánceres. La presente invención proporciona genes Marcadores preferidos e incluso más preferido combinaciones de genes Marcadores. Estos se describen en la presente con detalle.

10 "Origen" como se refiere en "tejido de origen" significa o el tipo de tejido "pulmón, colon, etc.) o el tipo histológico (adenocarcinoma, carcinoma de células escamosas, etc.) dependiendo de las circunstancias médicas particulares y se entenderá por cualquiera experto en la técnica.

15 Un gen Marcador corresponde a la secuencia designada por una SEQ ID NO cuando contiene esa secuencia. Un segmento o fragmento de gen corresponde a la secuencia de dicho gen cuando contiene una porción de la secuencia referenciada o su complemento suficiente para distinguirlo como de la secuencia del gen. Un producto de la expresión del gen corresponde a dicha secuencia cuando su ARN, ARNm o ADNc hibrida con la composición que tiene dicha secuencia (por ejemplo, una sonda) o, en el caso de un péptido o proteína, está codificado por dicho ARNm. Un segmento o fragmento de un producto de la expresión del gen corresponde a la secuencia de dicho gen o producto de expresión del gen cuando contiene una porción del producto de expresión del gen referenciado o su complemento suficiente para distinguirlo como de la secuencia del gen o producto de expresión del gen.

20 Los métodos de la invención descritos y reivindicados en esta especificación incluyen genes Marcadores, como se definen en las reivindicaciones. "Marcador" o "gen Marcador" se usa a lo largo de esta especificación para referirse a genes y productos de la expresión de genes que corresponden con cualquier gen la sobre-o bajo-expresión del cual está asociada con un tumor o tipo de tejido. Los genes Marcadores se describen con más detalle en la Tabla 1.

Tabla 1

Panel CUP			
SEQ ID NO:	Nombre	Designación del Chip	secuencia
1	SP-B	209810_at	gaaaaaccagccactgctttacaggacagggggtgaagctgagccccgc ctcacaccacccccatgcactcaaagattggattttacagctacttgcaatt caaaattcagaagaataaaaaatgggaacatacagaactctaaaagataga catcagaaattgtaagttaagcttttcaaaaaatcagcaattcccagcgta gtcaagggtggacactgcacgctctggcatgatgggatggcgaccgggc aagctttctctcgagatgctctgctgcttgagagctattgctttgtaagatat aaaaagggtttcttttctttctgttaaggtggactccagatttgattgaaa gtcctaggggtattctatttctgctgtgattatctgctgaaagctcagctggg gttgtgcaagctagggaccattcctgtgtaatacaatgtctgcaccaatgct
2	TTF1	211024_s_at	gtgattcaaatgggtttccacgctagggcggggcacagattggagaggc tctgtgctgacatggctctggactctaaagaccaaactcactctgggcaca ctctgccagcaaaggactcgttgtaataaccaggattttttttttttgaa gggaggacgggagctggggagaggaaagagtcttaacataaccactt gtcactgacacaaaggaagtgtccccctccccggcacctctggcgccta ggctcagcggcgaccgccctccgcgaaaatagttgttaatgtgaactgt agctgtaaacgctgtcaaaagtggactaaatgcctagtttttagtaactgt acattttgtgtaaaaagaaaaccactcccagtcccagccctcacattttt atgggcattgacaaatctgtgtatattttggcagtttggtatttgcggcgtca gtcttttctgtgtaact



(continuada)

Panel CUP			
SEQ ID NO:	Nombre	Designación del Chip	secuencia
3	DSG3	205595_at	ccatcccatagaagtccagcagacaggatttgtaagtccagactttgtcaggaaagcaaggagcttctgcttctgcccctctgggtctgtccagccagctgttccatccctgaccctctgcagcatggaactatttagtaacggagacttactcggcttctggttccctcgtgcaaccttccactgcaggctttgatccacttctcacaaaaatgtgatagtgacagaaaagggtgatctgtccatttccagtgttctcggcaacctagctggccaacgcagctacgagggtcacatactatgctctgtaacagaggatcctgtcccgtctaataatgaccagaatgagctggaataaccacactgaccaaatctggatctttggactaaagtattcaaaaatagcatagcaaaagctcactgtattgggctaataatttggcacttattagcttctctcataaactgatcacgattataaattaatgttgggttcatacccaaaagcaaatatgttgcactcctaatctcaagtac
4	HPT1	209847_at	ctgcaccacactactagatatttcatgtgctatagacattagagagattttcattttccatgacatttttctctctgcaaaggcttagctacttgtgttttccctttggggcaagacagactcattaaatattctgtacatttttctttatcaaggagatatcagtggtgtctcatagaactgcctggattccattatgtttttctgattccatcctgtgtccccttcatccttgcactccttgggtatttcaactgaatttcaaacatttgc
5	PSCA	205319_at	ttcctgaggcacatcctaacgcaagtttgacctgatgtttgcaccccttttccnaacctgaccttccatgggcctttccaggattcnaaccnggcagatcagtttttagtanacanatccgcntgcagatggcccctcaaccntttntgtgntgtttccatggcccagcattttccaccttaacctgtgttcaggcactnttccccaggaagccttccctgccaccccattatgaattgagccaggttggfcccgtggtgtccccgcaccagcaggggacaggcaatcaggagggcccagtaaaggctgagatgaagtggactgagtagaactggaggacaagagttgacgtgagttcctgggagttccagagatg
6	F5	204713_s_at	atcctctacagccagatgtcacagggafactgtactttcacttgggtgctgga gaattcanaagtaagaacatgctaagcntaagggaccaaggtagaaag agatcaagcagcaaagcacaggttctcctggatgaaactactagcacataa agttgggagacacctaagccaagacactggttctcctccggaatgaggcc ctgggaggaccttctagccaagacactggttctcctccagaatgaggcc ctggaaggacctcctagtgatctgttactctfaaacaaagtaactcatctaa gatttgggtgggagatggcatttggcttctgagaaaggtagctatgaataat ccaagatactgatgaagacacagctgttaacaattggctgatcagcccca gaatgcctcacgtgcttggggagaaagcaccctcttccaacaagcctgg aaag
7	MGB1	206378_at	gcagcagcctcaccatgaagttgctgatggcttcatgctggcggccctctc ccagcactgctacgcaggctctggctgcccctattggagaatgtgattcca agacaatcaatccacaagtgtctaagactgaatacaagaacttctcaaga gttcatagacgacaatgccactacaaatgccatagatgaattgaaggatgt tttcttaaccaaaccggatgaaactctgagcaatgttgagggtttatgcaatta atafatgacagcagctttgtgatttatttaacttctgcaagaccttggctca cagaactgcagggtatggtgagaaccaactacggattgctgcaaaccac accttcttcttctatgtcttttact

(continuada)

Panel CUP			
SEQ ID NO:	Nombre	Designación del Chip	secuencia
8	PDEF	220192_x_at	gagtgggggcccttaaactggattcaaaaaatgctctaacaataggaatggtt gaagaggctctgcagcttcagatgaaactaatctctagaagaggcacia gaatggctaaagcaattcatccaagggccaccggaagtaattagagctttg aaaaaatctgittgtcaggcagagagctatattggaggaagcattacagaa cgaaagagatcttttaggaacagttgggggtggccctgcaaattagaggct attgctaagaaaggaaaatttaataaataattggttttcgtgtggatgtactcc aagtaaagctccagtgactaatatgtataaatgtaaatgatattaaatgaa catcagttaaaaaaaaattctttaaggctactattaatgcagactacttta atcattgaaatctgaactcattactcatttctgccaattactcccttgggtat ttactgcgta
9	PSA	204582_s_at	tggtgtaattttgctcctctctgtgtcctggggaataactggccatgctggagac atatcactcaatttctctgaggacacagataggatgggggtgtctgtgtattttg ggggtacagagatgaaagaggggtgggaccacactgagagagtgagg agtgcacatgtgtctggacactgtccatgaagcactgagcagaagctggagg cacaacgcaccagacactcacagcaaggatggagctgaaaacataacc actctgtcc
10	WT1	206067_s_at	atagatgtacatacctccttgcacaaatggaggggaattcatttcatcactgg gagtgctcttagtgataaaaaccatgctggatagggctcaagttgtaaaaa tgaaagtactttaaagaaaataggggatggccaggatctccactgataa gactgttttaagtaactaaggaccttgggtctacaagtatatgtgaaaaaa atgagacttactgggtgaggaaatccattgttaagatggctgtgtgtgtgt gtgtgtgtgtgtgtgtgtgtgtgtgtgttttttaaggagggaatttattatt taccgttcttgaactactgtgataatatagtctgataatgatttctctttgac aactaaaataggactgtataagtactagatgcatcactgggtgtgatcttac aagat

La presente invención proporciona un método para identificar el origen de una metástasis de origen desconocido midiendo Biomarcadores asociados con al menos seis carcinomas diferentes en una muestra que contiene células metastásicas en donde los Biomarcadores son niveles de expresión de genes y donde los genes Marcadores son SP-8, JTF1, DSG3, PSCA, F5, HPT1, PSA, MGB/MG, WT1 y PDEF; combinando los datos de los Biomarcadores en un algoritmo donde el algoritmo: normaliza los Biomarcadores frente a la referencia; e impone un punto de corte que optimiza la sensibilidad y especificidad de cada Biomarcador, pondera la prevalencia de carcinomas y selecciona un tejido de origen; determina el origen en base a la probabilidad más alta determinada por el algoritmo o determina que el carcinoma no se deriva de un conjunto particular de carcinomas; y mide opcionalmente Biomarcadores específicos para uno o más carcinomas diferentes adicionales, y repite los pasos como sea necesario para Biomarcadores adicionales.

La presente invención proporciona un método para identificar el origen de una metástasis de origen desconocido en una muestra que contiene células metastásicas como se define en las reivindicaciones, que comprende: medir los Biomarcadores asociados con al menos seis carcinomas diferentes; combinar los datos de los Biomarcadores en un algoritmo donde el algoritmo i) normaliza los Biomarcadores frente a la referencia; y ii) impone un punto de corte que optimiza la sensibilidad y especificidad de cada Biomarcador, pondera la prevalencia de carcinomas y selecciona un tejido de origen; determinar el origen en base a la probabilidad más alta determinada por el algoritmo o determinar que el carcinoma no se deriva de un conjunto particular de carcinomas; y medir opcionalmente Biomarcadores específicos para uno o más carcinomas diferentes adicionales, y repetir los pasos c) y d) para Biomarcadores adicionales.

Los genes Marcadores son i) SP-B, TTF1, DSG3, PSCA, F5, HPT1, PSA, MCB/MC, WT1 y PDEF. Los genes Marcadores pueden incluir además KRT6F, p73H, y/o SFTPC.

Preferiblemente, los genes Marcadores pueden además incluir ITGB6, KLK10, CLDN18, TR10 y/o FKB 10.

En otra realización, los genes Marcadores pueden además incluir CDX1 y/o FABP1.

5 En una realización, la expresión génica se mide usando al menos una de las SEQ ID NOs: 11-58.

La presente divulgación también proporciona métodos que miden la expresión génica obteniendo y midiendo la formación de al menos uno de los amplicones SEQ ID NOs: 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54 y/o 58.

10 En una realización, los genes Marcadores pueden comprender además un Marcador específico del género seleccionado de al menos uno de: i) en el caso de un paciente masculino KLK2, NGEP o NPY; o ii) en el caso de un paciente femenino PIP, B305D, B726 o GABA-Pi; y/o PAX8, STAR o EMX2. Preferiblemente los genes Marcadores comprenden además KLK2. En esta realización, los genes marcadores pueden comprender además NGEP y/o NPY.  
15 En una realización, en el caso de un paciente femenino, los genes Marcadores comprenden además PIP, B305D, B726 o GABA-Pi. En una realización, en el caso de un paciente femenino los genes Marcadores comprenden además, PAX8, STAR o EMX2.

20 La presente divulgación proporciona métodos para obtener información clínica adicional incluyendo el sitio de metástasis para determinar el origen del carcinoma; obtener conjuntos de biomarcadores óptimos para carcinomas que comprende los pasos de usar metástasis de origen conocido, determinar Biomarcadores para ello y comparar los Biomarcadores con Biomarcadores de metástasis de origen desconocido; proporcionar dirección de terapia determinando el origen de una metástasis de origen desconocido e identificar el tratamiento apropiado para ello; y proporcionar un pronóstico determinando el origen de una metástasis de origen desconocido e identificar el pronóstico correspondiente para ello.  
25

La presente divulgación proporciona además métodos de encontrar Biomarcadores determinando el nivel de expresión de un gen Marcador en una metástasis particular, medir un Biomarcador para el gen Marcador para determinar la expresión del mismo, analizar la expresión del gen Marcador de acuerdo con cualquiera de los métodos proporcionados en la presente o conocidos en la técnica y determinar si el gen Marcador es efectivamente específico para el tumor de origen.  
30

La presente divulgación proporciona además una composición que contiene al menos una secuencia aislada seleccionada de las SEQ ID NOs: 11-58. La presente invención proporciona además el uso de kits para realizar un ensayo de acuerdo con los métodos proporcionados en la presente, en donde el kit comprende ARN o ADNc que hibrida con los genes marcadores de la invención, una micromatriz o un chip génico.  
35

La presente invención proporciona además el uso de micromatrices o chips génicos para realizar los métodos descritos en la presente.  
40

La presente divulgación proporciona además carpetas de diagnóstico/pronóstico que contienen secuencias de ácidos nucleicos aislados, sus complementos, o porciones de los mismos de una combinación de genes como se describe en la presente donde al combinación es suficiente para medir o caracterizar la expresión génica en una muestra biológica que tiene células metastásicas con respecto a células de carcinomas o tejido normal diferentes.  
45

Cualquier método descrito en la presente invención puede además incluir medir la expresión de al menos un gen expresado constitutivamente en la muestra.  
50

Los Marcadores para el cáncer de páncreas son factor de coagulación V (F5), antígeno de células madre de próstata (PSCA), integrina,  $\beta 6$  (ITGB6), calicreína 10 (KLK10), claudina 18 (CLDN18), isoforma trío (TR10), y proteína hipotética FLJ22041 similar a las proteínas de unión a FK506 (FKBP10). Los Biomarcadores para F5 y PSCA se miden juntos. Los Biomarcadores para ITGB6, KLK10, CLDN18, TR10, y FKBP10 se pueden medir además de para F5 y PSCA. El F5 se describe por ejemplo por 20040076955; 20040005563 y WO2004031412. El PSCA se describe por ejemplo por la WO1998040403; 20030232350 y WO2004063355. La ITGB6 se describe por ejemplo por la WO2004018999 y 6339148. La KLK 10 se describe por ejemplo por la WO20040770 y 20030235820. El CLDN18 se describe por ejemplo por la WO2004063355 y WO2005005601. El TR10 se describe por ejemplo por la 20020055627. El FKBP10 se describe por ejemplo por la WO2000055320.  
55

Los genes Marcadores para el cáncer de colon son el transportador asociado a péptidos intestinal HPT-1 (CDF17), el factor de transcripción de la caja homeótica de tipo caudal 1 (CDX 1) y la proteína de enlace de ácidos grasos 1 (FABP1). Los Biomarcadores para CDX1 y FABP1 se pueden medir en adición a un Biomarcador para CDH 17. El CDH17 se describe por ejemplo por Takamura et al. (2004) y la WO2004063355. El CDX1 se describe por ejemplo por Pillozzi et al. (2004); 20050059008y la 20010029020. El FABP11 se describe por ejemplo por Borchers et al. (1997); Chan et al. (1985); Chen et al. (1986); y Lowe et al. (1985).  
60  
65

5 Los genes Marcadores para el cáncer de pulmón son la proteína surfactante B (SP-B), el factor de transcripción tiroideo (TTF), desmogleína 3 (DSG3), isoforma 6F de la queratina 6 (KRT6F), gen relacionado con el p53 (p73H), y la proteína surfactante C (SFTPC). Los Biomarcadores para SP-B, TTF y DSG3 se miden juntos. Los Biomarcadores para KRT6F, p73H y SFTPC se pueden medir en adición a cualquiera de los Biomarcadores para SP-B, TTF y/o DSG3. La SP-B se describe por ejemplo por Pilot-Mathias et al. (1989); 20030219760 y 20030232350. El TTF se describe por ejemplo por Jones et al. (2005); US20040219575; WO1998056953; WO2002073204; 20030138793 y WO2004063355. La DSG3 se describe por ejemplo por Wan et al. (2003); 20030232350; aWO2004030615; y WO2002101357. La KRT6F se describe por ejemplo por Takahashi et al. (1995); 20040146862 y 20040219572. El p73H se describe por ejemplo por Senoo et al. (1998) y 20030138793. La SFTPC se describe por ejemplo por Glasser et al. (1988).

10 Los genes Marcadores pueden ser seleccionados además de un Marcador específico del género como, en el caso de un paciente masculino KLK2, NGEP o NPY; o en el caso de un paciente femenino PIP, B305D, B726 o GABA-Pi; y/o PAX8, STAR o EMX2.

15 Los genes Marcadores para el cáncer de mama son el factor epitelial derivado de la próstata (PDEF), mamaglobina (MG), proteína prolactininducible (PIP), B305D, B726, y GABA-π. Los biomarcadores para PDEF y MG se miden juntos. Los Biomarcadores para PIP, B305D, B726 y GABA-Pi se pueden medir en adición a los Biomarcadores para PDEF y MG. El PDES se describe por ejemplo por la WO2004030615; WO2000006589; WO2001073032; Wallace et al. (2005); Feldman et al. (2003); y Oettgen et al. (2000). La MG se describe por ejemplo por la WO2004030615; 20030124128; Fleming et al (2000); Watson et al. (1996 and 1998) y 5668267. La PIP se describe pro ejemplo por Autiero et al. (2002); Clark et al. (1999); Myal et al. (1991) y Murphy et al. (1987). B305D, B726 y GABA-Pi se describen por Reinholz et al. (2005). El NGEP se describe por ejemplo por Bera et al. (2004).

20 Los Marcadores para el cáncer de ovarios son el tumor de Wilm I (WT1), PAX8, proteína reguladora aguda esteroideogénica (STAR) y EMX2. Preferiblemente, se miden los Biomarcadores para WT1. Los Biomarcadores para STAR y EMX2 pueden ser medidos en adición a los Biomarcadores para WT1. El WT1 se describe por ejemplo por 350840; 6232073; 6225051; 20040005563 y Bentov et al. (2003). El PAX8 se describe por ejemplo por 20050037010; Poleev et al. (1992); Di Palma et al. (2003); Marques et al. (2002); Cheung et al. (2003); Goldstein et al. (2002); Oji et al. (2003); Rauscher et al. (1993); Zapata-Benavides et al. (2002) y Dwight et al. (2003). La STAR se describe por ejemplo por Gradi et al. (1995) y Kim et al. (2003). El EMX2 se describe por ejemplo por Noonan et al. (2001).

25 Los Marcadores para el cáncer de próstata son KLK3, KLK22, NGEP y NPY. Se miden los Biomarcadores para KLK3. Los Biomarcadores para KLK2, NGEP y NPY pueden ser medidos en adición a KLK3. El KLK2 y KLK3 se describen por ejemplo por Magklara et al. (2002). El KLK2 se describe por ejemplo por US 20030215835 y US 5786148, El KLK3 se describe por ejemplo por US 6261766.

30 El método puede también incluir obtener información clínica adicional que incluya el sitio de metástasis para determinar el origen del carcinoma. En la Figura 3 se proporciona un diagrama de flujo.

35 La divulgación proporciona además un método para obtener conjuntos de Biomarcadores óptimos para carcinomas usando metástasis de origen conocido, determinando Biomarcadores para ello y comparando los Biomarcadores con Biomarcadores de metástasis de origen desconocido.

40 La divulgación proporciona además un método para proporcionar dirección de terapia determinando el origen de una metástasis de origen desconocido de acuerdo con los métodos descritos en la presente e identificando el tratamiento apropiado para ello.

45 La divulgación proporciona además un método para proporcionar un pronóstico determinando el origen de una metástasis de origen desconocido de acuerdo con los métodos descritos en la presente e identificando el pronóstico correspondiente para ello.

50 La divulgación proporciona además un método para encontrar Biomarcadores que comprende determinar el nivel de expresión de un gen Marcador en una metástasis particular, medir un Biomarcador para el gen Marcador para determinar la expresión del mismo, analizar la expresión del gen Marcador de acuerdo con los métodos descritos en la presente y determinar si el gen Marcador es efectivamente específico para el tumor de origen.

55 La divulgación proporciona además composiciones que comprenden al menos una secuencia aislada seleccionada de las SEQ ID NOs: 11-58.

60 La invención proporciona además el uso de kits, micromatrices o chips génicos para llevar a cabo los ensayos descritos en la presente, como se define en las reivindicaciones.

65

Se ha encontrado sólo en raras ocasiones que la mera presencia o ausencia de secuencias de ácidos nucleicos particulares en una muestra de tejido tenga valor diagnóstico o de pronóstico. La información sobre la expresión de varias proteínas, péptidos o ARNm, por otro lado, se considera cada vez más como importante. La mera presencia de secuencias de ácidos nucleicos que tienen el potencial de expresar proteínas, péptidos o ARNm (tales secuencias referidas como "genes") dentro del genoma por sí mismo no es determinante de si una proteína, péptido a ARNm se expresa en una célula dada. Si un gen dado capaz de expresar proteínas, péptidos o ARNm lo hace y a que extensión tiene lugar tal expresión, en todo caso, se determina por una variedad de factores complejos. Independientemente de las dificultades de entender y evaluar estos factores, el ensayar la expresión génica puede proporcionar información útil sobre la ocurrencia de eventos importantes como tumorigénesis, metástasis, apoptosis, y otros fenómenos clínicamente relevantes. Las indicaciones relativas del grado a los que los genes son activos o inactivos puede encontrarse en los perfiles de expresión génica. Los perfiles de expresión génica obtenidos de los métodos de esta invención se usan para proporcionar diagnóstico y tratar pacientes de CUP.

La preparación de la muestra requiere la recogida de muestras de pacientes. Las muestras de pacientes usadas en el método de la invención son las que se sospecha que contienen células enfermas como las células tomadas de un nódulo en un aspirado de aguja fina (FNA) del tejido. La preparación del tejido bruto obtenido de una biopsia o espécimen quirúrgico y microdissección por captura láser también son adecuadas para el uso. La tecnología por microdissección por captura láser (LCM) es una manera de seleccionar las células a ser estudiadas, minimizando la variabilidad causada por la heterogeneidad del tipo celular. Consecuentemente, los cambios moderados o pequeños en la expresión del gen Marcador entre células normales o benignas y cancerosas pueden ser detectados fácilmente. Las muestras pueden también comprender células epiteliales circulantes extraídas de sangre periférica. Estas se pueden obtener de acuerdo con un número de métodos pero el método más preferido es la técnica de separación magnética descrita en la 6136182. Una vez que se ha obtenido la muestra que contiene las células de interés, se obtiene un perfil de expresión génica usando un Biomarcador, para los genes en las carpetas apropiadas.

Los métodos preferidos para establecer los perfiles de expresión génica incluyen determinar la cantidad de ADN que se produce por un gen que puede codificar para una proteína o péptido. Esto se consigue por PCR con transcriptasa inversa (TR-PCR), RT-PCR competitiva, RT-PCR en tiempo real, RT-PCR con visualización diferencial, análisis Northern Blot y otras pruebas relacionadas. Aunque es posible llevar a cabo estas técnicas usando reacciones de PCR individuales, es mejor amplificar el ADN complementario (ADNc) o el ARN complementario (ARNc) producido por el ARNm y analizarlo a través de micromatrices. Se conocen una variedad de configuraciones de matrices y métodos para su producción diferentes por los expertos en la técnica y se describen e por ejemplo, 5445934; 5532128; 5556752; 5242974; 5384261; 5405783; 5412087; 5424186; 5429807; 5436327; 5472672; 5527681; 5529756; 5545531; 5554501; 5561071; 5571639; 5593839; 5599695; 5624711; 5658734 y 5700637.

La tecnología de micromatrices permite la medición del nivel de ARNm en estado estacionario de miles de genes proporcionando simultáneamente una herramienta poderosa para identificar efectos como el inicio, detención o modulación de la proliferación celular no controlada. Hay actualmente en uso generalizado dos tecnologías de micromatrices, ADNc y matrices de oligonucleótidos. Aunque existen diferencias en la construcción de estos chips, esencialmente todo el análisis de datos secuencia abajo y salida son lo mismo. El producto de estos análisis son típicamente mediciones de la intensidad de la señal recibida de una sonda etiquetada usada para detectar una secuencia de ADNc de la muestra que hibrida con una secuencia de ácidos nucleicos en una localización conocida en la micromatriz. Típicamente, la intensidad de la señal es proporcional a la cantidad de ADNc, y por lo tanto de ARNm, expresada en las células de la muestra. Un gran número de dichas técnicas están disponibles y son útiles. Los métodos preferidos para determinar la expresión génica se pueden encontrar en 6271002; 6218122; 6218114 y 6004755.

El análisis de los niveles de expresión se lleva a cabo comparando dichas intensidades de señal. Esto se hace mejor generando una matriz de relación de las intensidades de la expresión de los genes en una muestra de prueba frente a las de las muestras de control. Por ejemplo, las intensidades de la expresión génica de un tejido enfermo se pueden comparar con las intensidades de la expresión generadas de tejido benigno o normal del mismo tipo. Una relación de estas intensidades de expresión indica las veces de cambio en la expresión génica entre las muestras de prueba y control.

La selección se puede basar en pruebas estadísticas que producen listas clasificadas relacionadas con la evidencia de significancia de cada expresión diferencial del gen entre factores relacionados con el sitio de origen original del tumor. Ejemplos de tales pruebas incluyen ANOVA y Kruskal-Wallis. Las clasificaciones pueden ser usadas como ponderaciones en un modelo diseñado para interpretar la suma de dichas ponderaciones, hasta un punto de corte, como la preponderancia de la evidencia a favor de una clase sobre otra. La evidencia anterior como se describe en la bibliografía también se puede usar para ajustar las ponderaciones.

En la presente invención, se eligieron 10 marcadores que mostraron evidencia significativa de expresión diferencial entre 6 tipos de tumores. El proceso de selección incluía una recogida ad-hoc de pruebas estadísticas, optimización de varianza media y conocimiento experto. En una realización alternativa los métodos de extracción de características podrían ser automatizados para seleccionar y probar marcadores a través de enfoques de

aprendizaje supervisados. A medida que crece la base de datos, la selección de marcadores se puede repetir para producir la precisión de diagnóstico más alta posible en cualquier estado dado de la base de datos.

5 Una realización preferida es normalizar cada medición identificando un conjunto de control estable y modificando a escala este conjunto a varianza cero en todas las muestras. Este conjunto de control se define como cualquier transcripción endógena individual o conjunto de transcripciones endógenas afectadas por el error sistemático en el ensayo, y que no se conoce que cambie independientemente de este error. Todos los marcadores son ajustados por el factor específico de la muestra que genera la varianza cero para cualquier estadística descriptiva del conjunto de control, como la media o mediana, o para una medición directa. Alternativamente, si la premisa de variación de los controles relacionados sólo con el error sistemático no es cierta, todavía el error de clasificación resultante es menor que cuando se realiza la normalización, el conjunto de control todavía se usará como se ha expuesto. Los controles pico no endógenos también podrían ser útiles, pero no se prefieren.

15 Después de la selección de marcadores, estas variables seleccionadas se usan en un clasificador diseñado para producir una precisión de clasificación tan alta como sea posible. Se puede usar un algoritmo de aprendizaje supervisado diseñado para relacionar un conjunto de mediciones de entrada con un conjunto de salidas de predictores a fin de construir un modelo de las 10 entradas para predecir el tejido de origen. El problema se puede plantear como: dado los datos de aprendizaje  $\{(x_1, y), \dots, (x_n, y)\}$  producir un clasificador  $h: X \rightarrow Y$  que mapea una muestra  $x \in X$  de su tejido de etiqueta de origen  $y \in Y$ . Las predicciones se basan en casos anteriormente resueltos que están contenidos en la base de datos y por lo tanto componen el conjunto de aprendizaje.

25 El algoritmo de aprendizaje supervisado debería encontrar parámetros en base a las relaciones de las variables de entrada con las salidas conocidas que minimizarán el error de clasificación esperados. Estos parámetros pueden ser entonces usados para predecir el tejido de origen de una nueva entrada de la muestra. Ejemplos de estos algoritmos incluyen modelos de clasificación lineal, clasificadores cuadráticos, métodos basados en árbol, redes neuronales y métodos prototipo como un clasificador de elemento afín k o algoritmos de cuantificación del vector de aprendizaje.

30 Una realización específica al modelo de 10 marcadores normalizados es el método LDA, usando parámetros por defecto, como se describe en Venables y Ripley (2002). Este método se basa en el Análisis discriminante lineal de Fisher, donde las medias dadas  $\mu_{y=0}, \mu_{y=1}$  y las covarianzas  $\Sigma_{y=0}, \Sigma_{y=1}$  para las etiquetas clases y 0 y 1, buscamos una combinación lineal de  $w \cdot x$  que tendría medias  $w \cdot \mu_{y=i}$  y varianzas  $w^T \Sigma_{y=i} w$  que maximizarán la proporción de la varianza entre las clases con la varianza dentro de las clases.

35

$$S = \frac{\sigma^2_{entre}}{\sigma^2_{dentro}} = \frac{(\bar{w} \cdot \bar{\mu}_{y=1} - \bar{w} \cdot \bar{\mu}_{y=0})^2}{\bar{w}^T \Sigma_{y=1} \bar{w} + \bar{w}^T \Sigma_{y=0} \bar{w}} = \frac{(\bar{w} \cdot (\bar{\mu}_{y=1} - \bar{\mu}_{y=0}))^2}{\bar{w}^T (\Sigma_{y=0} + \Sigma_{y=1}) \bar{w}}$$

40

45 La LDA puede ser generalizada a un análisis discriminante de clase múltiple, donde  $y$  tiene  $N$  estados posibles, en lugar de sólo dos. Las medias y varianzas de clase se estiman de los valores contenidos en la base de datos para los marcadores elegidos. En una realización preferida, la matriz de la covarianza es ponderada por probabilidades anteriores iguales de cada tipo de tumor sometido a lo siguiente. Los pacientes masculinos se predicen por un modelo donde los anteriores son cero para cada grupo de tumor de órgano reproductor femenino. Igualmente, los pacientes femeninos se predicen por un modelo donde el anterior es cero para los órganos reproductores masculinos. En la presente invención, los anteriores son cero para pruebas en mujeres para próstata y cero para varones probados para mama y ovarios. Además, las muestras con un antecedente idéntico a una etiqueta de clase se prueban por un modelo donde la probabilidad anterior es cero para esa etiqueta de clase particular.

50 El problema anterior se puede ver como una maximización del cociente de Rayleigh manejado como un problema de valor propio generalizado. El subespacio reducido se usa en la clasificación calculando cada distancia de la muestra al centroide en el subespacio elegido. El modelo puede ser ajustado por probabilidad máxima, y las probabilidades posteriores se calculan usando el teorema de Bayes.

55 Un método alternativo puede incluir encontrar un mapa de un espacio característico  $n$ -dimensional, donde  $n$  es el número de variables usadas, a un conjunto de etiquetas de clasificación que implicará particionar el espacio característico en regiones, asignando después una clasificación para cada región. Las puntuaciones de estos algoritmos de tipo afines están relacionadas con la distancia entre los límites de decisión y no se traducen necesariamente en probabilidades de clase.

60 Si hay demasiadas variables de las que seleccionar, y muchas de ellas son ruido aleatorio, entonces la selección de variable y el riesgo del modelo ajustan en exceso el problema. Por lo tanto, se usan a menudo listas

5 clasificadas en varios puntos de corte como entradas a fin de limitar el número de variables. También se pueden usar algoritmos de búsqueda como un algoritmo genético para seleccionar un sub-conjunto de variables ya que ponen a prueba una función de costes. Se puede intentar hibridación simulada para limitar el riesgo de tomar la función de coste en un mínimo local. Sin embargo, estos procedimientos deben ser validados con muestras independientes al proceso de selección y de modelado.

10 También se pueden usar enfoques variables latentes. Se puede usar cualquier algoritmo de aprendizaje sin supervisar para estimar variedades de baja dimensión de espacio de alta dimensión para descubrir asociaciones entre las variables de entrada y lo bien que pueden encajar un conjunto más pequeño de variables latentes. Aunque las estimaciones de la efectividad de estas reducciones son subjetivas, se puede aplicar un algoritmo supervisado en el conjunto de variables reducido a fin de estimar la precisión de la clasificación. Por lo tanto un clasificador, que puede ser construido de las variables latentes, también se puede construir de un conjunto de variables correlacionadas significativamente con las variables latentes. Un ejemplo de esto incluiría usar variables correlacionadas con los componentes principales, de un análisis de componentes principales, como entradas a cualquier modelo de clasificación supervisado.

20 Estos algoritmos se pueden implementar en cualquier código de software que tenga métodos para introducir las variables, aprender las muestras con una función, probar una muestra basada en el modelo, y extraer los resultados a una consola. R, Octave, C, C++, Fortran, Java, Perl y Python tienen todas bibliotecas disponibles bajo licencia de código abierto para realizar muchas de las funciones enumeradas anteriormente. Los paquetes comerciales como S+ y Matlab también incluyen muchos de estos métodos.

25 El código realiza los siguientes pasos en el orden siguiente usando la versión R 2.2.1 (<http://www.r-project.org>) con la biblioteca MASS (Venables et al. (2002)) instalada. El término LDA se refiere a la función lda en el espacio de nombres MASS.

1) se almacenan valores de CT para 10 genes marcadores y 2 controles en un disco duro para todas las muestras del conjunto de aprendizaje disponibles.

30 2) Para cada muestra, restar la media específica de la muestra de los controles de cada marcador normaliza los 10 valores de genes marcadores.

3) El conjunto de datos de aprendizaje está compuesto de metástasis con sitios conocidos de origen donde cada muestra tiene al menos uno de sus marcadores objetivo específicos para el tejido etiquetado de origen con un valor de CT normalizado menor de 5.

35 4) El LDA construye 4 conjuntos de 2 modelos de LDA de los datos de aprendizaje en (3). En cada conjunto, un modelo es específico para varones, y tiene probabilidades anteriores para mama y ovario establecidas en cero así como las probabilidades anteriores para la próstata establecidas a los anteriores equivalentes de otras etiquetas de clase. El otro modelo en cada par es específico para mujeres con las probabilidades anteriores de próstata establecidas a cero, y con los anteriores para mama y ovario establecidos a los anteriores equivalentes encontrados en otras etiquetas de clase.

40 a. El primer conjunto se usa para probar muestras de CUP encontradas en el colon, las probabilidades anteriores para el colon están establecidas a cero y todas las otras etiquetas de clase no reproductoras están establecidas a los anteriores equivalentes.

45 b. Un segundo conjunto del modelo es específico para una CUP encontrado en el ovario, con las probabilidades anteriores para el ovario establecidas a cero y todas las otras etiquetas de clase no reproductoras establecidas a anteriores equivalentes.

50 c. Un tercer conjunto para un CUP encontrado en el pulmón, con las probabilidades anteriores para el pulmón establecidas a cero. Todas las otras etiquetas de clase no reproductoras tienen anteriores equivalentes.

d. EL modelo general usado para todos los otros tejidos de fondo. Todos los anteriores están establecidos equivalentemente con la excepción de las etiquetas de clase específicas reproductoras que están establecidas como se define en 4.

55 Para probar una muestra, ejecutamos un programa R que realiza lo siguiente.

1) Lee en un conjunto de datos de prueba.

2) Genera una media específica de la muestra de ambos controles.

3) Para cada muestra, usa la media específica de la muestra para restarla de cada marcador.

60 4) Reemplaza cualquier CT normalizado generado de un CT bruto de 40 con 12.

5) Para cada muestra en el conjunto de prueba se prueban los siguientes.

a. Si la media de ambos controles son mayores de 34 que la muestra es etiquetada con 'CTR\_FALLO' con ceros para probabilidades posteriores.

65 b. Se comprueban los antecedentes para colon, ovario o pulmón. Si se encuentra una coincidencia se comprueba también el género. El modelo específico de los antecedentes y género se usa entonces para

evaluar la muestra.

c. Si se encuentra mama, páncreas, pulmón, SCC o próstata como la etiqueta de antecedentes, entonces se da una etiqueta de "FALLO-muestra\_inelegible" a la muestra, y las probabilidades posteriores son todas establecidas en cero.

5 d. El modelo general ya sea para varón o mujer se usa para todas las otras muestras.

Los resultados son formateados y escritos en un archivo.

10 La presente divulgación incluye carpetas de expresión génica obtenidos por este proceso.

15 Los perfiles de expresión génica se pueden mostrar de varias maneras. La más común es organizar las intensidades de fluorescencia brutas o matriz de relación en un dendograma gráfico donde las columnas indican las muestras de prueba y las filas indican los genes. Los datos son organizados de tal manera que los genes que tienen perfiles génicos similares están próximos entre sí. La relación de la expresión para cada gen se visualiza como un color. Por ejemplo, una relación de menos de uno (regulación hacia abajo) aparece en la porción azul del espectro mientras que una relación mayor de uno (regulación hacia arriba) aparece en la porción roja del espectro. Hay disponibles programas de software de ordenador comercialmente disponibles para mostrar dichos datos incluyendo "GeneSpring" (Silicon Genetics, Inc.) y "Discovery" e "Infer" (Partek, Inc.).

20 Las mediciones de la abundancia de especies de ARN únicas se recogen de tumores primarios o tumores metastásicos de primarios de origen desconocido. Estas lecturas junto con los registros clínicos incluyendo, pero no limitado a, edad del paciente, género, sitio de origen del tumor primario, y sitio de metástasis (si es aplicable) se usan para generar una base de datos relacional. La base de datos se usa para seleccionar transcritos de ARN y factores clínicos que pueden ser usados como variables de marcadores para predecir el origen primario de un tumor metastásico.

30 En el caso de medir niveles de proteínas para determinar la expresión génica, es adecuado cualquier método conocido en la técnica siempre que resulte en sensibilidad y especificidad adecuadas. Por ejemplo, los niveles de proteínas pueden ser medidos enlazándolas con un anticuerpo o fragmento de anticuerpo específico para la proteína y midiendo la cantidad de proteína enlazada al anticuerpo. Los anticuerpos pueden ser etiquetados por reactivos radioactivos, fluorescentes u otros reactivos detectables para facilitar la detección. Los métodos de detección incluyen, sin limitación, ensayo inmunoabsorbente ligado a enzimas (ELISA) y técnicas de inmunotransferencia.

35 Los genes modulados usados en los métodos de la invención se describen en los Ejemplos. Los genes que son expresados diferencialmente son o regulados hacia arriba o regulados hacia abajo en pacientes con carcinomas de un origen particular en relación con los carcinomas de orígenes diferentes. Regulación hacia arriba y regulación hacia abajo son términos relativos que significan que se encuentra una diferencia detectable (más allá de la contribución del ruido en el sistema usado para medirla) en la cantidad de expresión de los genes en relación con alguna línea de base. En este caso, la línea de base se determina en base al algoritmo. Los genes de interés en las células enfermas son entonces o regulados hacia arriba o regulados hacia abajo en relación al nivel de la línea de base usando el mismo método de medición. Enfermo, en este contexto, se refiere a una alteración del estado de un cuerpo que interrumpe o perturba, o tiene el potencial de perturbar, el rendimiento adecuado de las funciones corporales como ocurre con la proliferación no controlada de células. Alguien es diagnosticado con una enfermedad cuando algún aspecto del genotipo o fenotipo de esa persona es consistente con la presencia de la enfermedad. Sin embargo, el acto de realizar un diagnóstico o pronóstico puede incluir la determinación de cuestiones de enfermedad/ estado como determinar la probabilidad de recaída, tipo de terapia y monitorización de la terapia. En la monitorización de la terapia, los juicios clínicos se hacen con respecto al efecto de un curso de terapia dado comparando la expresión de genes durante el tiempo para determinar si los perfiles de expresión génica han cambiado o están cambiando a patrones más consistentes con el tejido normal.

50 Los genes pueden ser agrupados de tal forma que la información obtenida sobre el conjunto de genes en el grupo proporcione una base sólida para hacer un juicio clínicamente relevante como un diagnóstico, pronóstico o elección de tratamiento. Estos conjuntos de genes componen las carpetas de la divulgación. Como con la mayoría de los Marcadores de diagnóstico, es a menudo deseable usar el menor número de Marcadores suficiente para hacer un juicio médico correcto. Esto evita un retraso en el tratamiento pendiente de análisis adicionales así como el uso improductivo de tiempo y recursos.

60 Un método de establecer carpetas de expresión génica es a través del uso de algoritmos de optimización como el algoritmo de varianza media usado ampliamente para establecer carteras de acciones. Este método se describe con detalle en la 20030194734. Esencialmente, el método designa el establecimiento de un conjunto de entradas (acciones en aplicaciones financieras, expresión medida por intensidad aquí) que optimizarán el retorno (por ejemplo, la señal que se genera) una vez que se recibe para su uso a la vez que se minimiza la variabilidad del retorno. Muchos programas de software comerciales están disponibles para realizar tales operaciones. Se prefiere el "Wagner Associates Mean-Variance Optimization Application," referida como "Wagner Software" a lo largo de esta



5 especificación. Este software usa funciones de la "Wagner Associates Mean-Variance Optimization Library" para determinar una frontera eficiente y carpetas óptimas en el sentido de Markowitz se prefiere. Markowitz (1952). El uso de este tipo de software requiere que los datos de micromatrices sean transformados para que puedan ser tratados como una entrada en la manera en que las mediciones de riesgo y retorno de acciones son usadas cuando el software se usa para sus propósitos de análisis financieros pretendidos.

10 El proceso de seleccionar una carpeta también puede incluir la aplicación de reglas heurísticas. Preferiblemente, dichas reglas se formulan en base a la biología y una comprensión de la tecnología usada para producir resultados clínicos. Más preferiblemente, se aplican a una salida del método de optimización. Por ejemplo el método de la varianza media de la selección de carpetas se puede aplicar a datos de micromatrices para un número de genes expresados diferencialmente en sujetos con cáncer. La salida del método será un conjunto optimizado de genes que podrían incluir algunos genes que son expresados en la sangre periférica así como en tejido enfermo. Si las muestras usadas en el método de prueba se obtienen de sangre periférica y ciertos genes expresados diferencialmente en casos de cáncer podrían ser también expresados diferencialmente en sangre periférica, entonces se puede aplicar una regla heurística en la que una carpeta es seleccionada de la frontera eficiente excluyendo las que se expresan diferencialmente en sangre periférica. Por supuesto, la regla puede ser aplicada antes de la formación de la frontera eficiente, por ejemplo, aplicando la regla durante la pre-selección de datos.

20 Se pueden aplicar otras reglas heurísticas que no están necesariamente relacionadas con la biología en cuestión. Por ejemplo, alguien puede aplicar una regla de que sólo un porcentaje prescrito de la carpeta puede estar representado por un gen particular o grupo de genes. El software comercialmente disponible como el Wagner Software acomoda fácilmente estos tipos de heurísticas. Esto puede ser útil, por ejemplo, cuando factores distintos de la exactitud y precisión (por ejemplo, derechos de licencia anticipados) tienen un impacto en la deseabilidad de incluir uno o más genes.

25 Los perfiles de expresión génica de esta divulgación también pueden ser usados en conjunción con otros métodos de diagnóstico no genéticos útiles en el diagnóstico, pronóstico o monitorización del tratamiento del cáncer. Por ejemplo, en algunas circunstancias es beneficioso combinar el poder diagnóstico de la expresión génica basado en métodos descritos anteriormente con datos de Marcadores convencionales como Marcadores de proteínas de suero (por ejemplo Antígeno del cáncer 27.29 ("27.29")). Existe una variedad de tales Marcadores incluyendo analitos como el CA27.29. En uno de dichos métodos, la sangre se toma periódicamente de un paciente tratado y se somete entonces a un inmunoensayo enzimático para uno de los Marcadores de suero descritos anteriormente. Cuando la concentración del Marcador sugiere la vuelta de tumores o el fallo de la terapia, se toma una fuente de muestra susceptible al análisis de expresión génica. Cuando existe una masa sospechosa, se toma un aspirado de aguja fina (FNA) y los perfiles de expresión génica de células tomadas de la masa se analizan como se ha descrito anteriormente. Alternativamente, se pueden tomar muestras de tejido de áreas adyacentes al tejido del que se ha retirado anteriormente el tumor. Este enfoque puede ser particularmente útil cuando otras pruebas producen resultados ambiguos.

40 Los kits hechos de acuerdo con la divulgación incluyen ensayos con formato para determinar los perfiles de expresión génica. Estos pueden incluir todos o algunos de los materiales necesarios para realizar los ensayos como reactivos e instrucciones y un medio a través del cual se ensayan los Biomarcadores.

45 Los artículos de esta divulgación incluyen representaciones de perfiles de expresión génica útiles para tratar, diagnosticar, pronosticar y evaluar de otras maneras enfermedades. Estas representaciones de perfiles se reducen a un medio que puede ser leído automáticamente por una máquina como un medio legible por ordenador (magnético, óptico y similar). Los artículos pueden incluir también instrucciones para evaluar los perfiles de expresión génica en dicho medio. Por ejemplo, los artículos pueden comprender un CD ROM que tiene instrucciones informáticas para comparar perfiles de expresión génica de las carpetas de genes descritas anteriormente. Los artículos pueden tener también perfiles de expresión génica grabados digitalmente en ellos de tal forma que pueden ser comparados con datos de expresión génica de muestras de pacientes. Alternativamente, los perfiles pueden ser grabados en un formato de representación diferente. Una inscripción gráfica es uno de dichos formatos. Los algoritmos de agrupación como los incorporados en el software "DISCOVERY" e "INFER" de Partek, Inc. mencionados anteriormente pueden ayudar mejor a la visualización de dichos datos.

55 Diferentes tipos de artículos de fabricación de acuerdo con la divulgación son medios o ensayos formateados usados para revelar los perfiles de expresión génica. Estos pueden comprender, por ejemplo, micromatrices en las que los complementos o sondas de secuencia están fijados a una matriz a la que las secuencias indicativas de genes de interés se combinan creando un determinante legible de su presencia. Alternativamente, los artículos de acuerdo con la divulgación pueden ser modelados en kits de reactivos para realizar hibridación, amplificación y generación de señal indicativos del nivel de expresión de los genes de interés para detectar cáncer.

60 Los siguientes ejemplos se proporcionan para ilustrar pero no limitar la invención reivindicada.

65

**Ejemplo 1**Materiales y Métodos5 *Descubrimiento de genes Marcadores de cáncer de páncreas*

El ARN se aisló de tejidos de tumor de páncreas, páncreas normal, pulmón, colon, mama y ovario usando Trizol. El ARN se usó entonces para generar ARN etiquetado, amplificado (Lipshutz et al. (1999)) que fue después hibridado en matrices Affymetrix U133A. Los datos fueron después analizados de dos maneras.

10 En el primer método, este conjunto de datos se filtró para conservar sólo aquellos genes con al menos dos designaciones presentes a través del conjunto de datos completo. El filtrado dejó 14.547 genes. Se determinó que 2.736 genes estaban sobreexpresados en el cáncer de páncreas frente al páncreas normal con un valor de p menor de 0,05. Cuarenta y cinco genes de los 2.736 estaban también sobreexpresados al menos dos veces en comparación con la intensidad máxima encontrada en tejidos de pulmón y colon. Finalmente, se encontraron seis conjuntos de sonda que estaban sobreexpresados al menos dos veces en comparación con la intensidad máxima de tejidos de pulmón, colon, mama y ovarios.

20 En el segundo método, este conjunto de datos se filtró para conservar sólo aquellos genes con no más de dos designaciones presentes en tejidos de mama, colon, pulmón y ovario. Este filtrado dejó 4.654 genes. Se descubrió que 160 genes de los 4.654 genes tenían al menos dos designaciones presentes en tejidos de páncreas (normal y cáncer). Finalmente, se seleccionaron ocho conjuntos de sondas que mostraron la expresión diferencial mayor entre el cáncer de páncreas y los tejidos normales.

25 *Muestras de tejido*

Se adquirieron un total de 260 tejidos primario y de metástasis de FFPE de una variedad de vendedores comerciales. Las muestras probadas incluían: 30 metástasis de mama, 30 metástasis colorrectal, 56 metástasis de pulmón, 49 metástasis de ovario, 43 metástasis de páncreas, 18 primarios de próstata y 2 metástasis de próstata y 32 de otros orígenes (6 de estómago, 6 de riñón, 3 de laringe, 2 de hígado, 1 de esófago, 1 de faringe, 1 de conducto biliar, 1 de pleura, 3 de vejiga, 5 de melanoma, 3 de linfoma).

30 *Extracción de ARN*

35 El aislamiento de ARN de secciones de tejido de parafina se basó en los métodos y reactivos descritos en el manual del High Pure RNA Paraffin Kit (Roche) con las siguientes modificaciones. Se seccionaron muestras de tejido embebidas en parafina de acuerdo al tamaño de la metástasis embebida (2-5mm = 9 X 10µm, 6-8mm = 6 X 10µm, 8 -> 10mm = 3 X 10 µm) y se colocaron en tubos Eppendorf de 1.5 ml de RNasa/DNasa. Las secciones se desparafinaron por incubación en 1ml de xileno durante 2-5 min a temperatura ambiente seguido por 10 a 20 segundos de agitación en vórtice. Los tubos fueron entonces centrifugados y se retiró el sobrenadante y se repitió el paso de desparafinización. Después de que el sobrenadante se hubo retirado, se añadió 1 ml de etanol y la muestra se agitó en vórtice durante 1 minuto, se centrifugó y se retiró el sobrenadante. Este proceso se repitió una vez adicional. Se retiró el etanol residual y el pellet se secó en un horno a 55° C durante 5-10 minutos y se resuspendió en 100 µl de tampón de lisis de tejido, 16 µl de 10% SDS y 80 µl de Proteinasa K. Las muestras se agitaron en vórtice y se incubaron en un conjunto termomezclador a 400 rpm durante 2 horas a 55° C. Se añadieron 325 µl de tampón de unión y 325 µl de etanol a cada muestra que fue entonces mezclada, centrifugada y se añadió el sobrenadante en la columna de filtro. Se centrifugaron la columna de filtro junto con el tubo de recogida durante 1 minuto a 8000 rpm y se descartó el flujo a través. Se realizaron una serie de lavados secuenciales (500 µl de Tampón de Lavado I → 500 µl de Tampón de Lavado II → 300 µl de Tampón de Lavado II) en los que se añadió cada solución a la columna, se centrifugó y se descartó el flujo a través. La columna fue entonces centrifugada a máxima velocidad durante 2 minutos, colocada en un tubo de 1,5 ml nuevo y se añadieron 90 µl de tampón de elución. El ARN se obtuvo después de 1 minutos de incubación a temperatura ambiente seguido por una centrifugación de 1 minuto a 8000 rpm. La muestra fue tratada con DNasa con la adición de 10 µl de tampón de incubación de DNasa, 2 µl de DNasa I y se incubó 30 minutos a 37° C. La DNasa fue inactiva después de la adición de 20 µl de tampón de lisis de tejido, 18 µl de 10% SDS y 40 µl de Proteinasa K. De nuevo, se añadieron 325 µl de tampón de unión y 325 µl de etanol a cada muestra que fue entonces mezclada, centrifugada y se añadió el sobrenadante en la columna de filtro. Los lavados secuenciales y la elución del ARN procedieron como se ha señalado anteriormente con la excepción de que se usaron 50µl de tampón de elución para eluir el ARN. Para eliminar la contaminación de fibra de vidrio arrastrada de la columna el ARN se centrifugó durante 2 minutos a toda velocidad y el sobrenadante se retiró a un tubo Eppendorf de 1,5 ml nuevo. Las muestras fueron cuantificadas por lecturas OD 260/280 obtenidas por un espectrómetro y las muestras se diluyeron a 50 ng/µl. El ARN aislado se almacenó en agua libre de RNasa a -80° C hasta su uso.

60 *Cebador TaqMan y Diseño de la Sonda*

65

Se usaron los números de acceso de la secuencia de referencia de ARNm apropiados en conjunción con Oligo 6.0 para desarrollar ensayos TaqMan® CUP (Marcadores pulmonares: proteína B asociada al pulmón surfactante humano (HUMPSPBA), factor de transcripción tiroideo 1 (TTF1), desmogleína 3 (DSG3), Marcador colorrectal: cadherina 17 (CDH17), Marcadores de mama: mamaglobina (MG), factor de transcripción ets derivado de próstata (PDEF), Marcador de ovario: tumor de Wilms 1 (WT1), Marcadores de páncreas: antígeno de células madre de próstata (LCSP), factor de coagulación V (F5), Marcador de próstata caliceína 3 (KLK3)) y ensayos domésticos de  $\beta$  actina, hidroximetilbilano sintasa (PBGD). Los cebadores y sondas de hidrólisis para cada ensayo se enumeran en la Tabla 2. La amplificación del ADN genómico se excluyó diseñando ensayos alrededor de sitios de corte y empalme exon-intron. Las sondas de hidrólisis se etiquetaron en el nucleótido 5' con FAM como el colorante indicador y en el nucleótido 3' con BHQ1-TT como el colorante de extinción interno.

#### *Reacción en Cadena de Polimerasa en Tiempo Real Cuantitativa*

La cuantificación de ARN específico del gen se llevo a cabo en una placa de 384 pocillos en el sistema de detección de secuencias ABI Prism 7900HT (Applied Biosystems). Para cada ejecución de termo-ciclador se amplificaron los calibradores y las curvas estándar. Los calibradores para cada Marcador consistieron de transcritos in vitro del gen objetivo que fueron diluidos en un ARN portador de riñón de rata a  $1 \times 10^5$  copias. Las curvas estándar para Marcadores domésticos consistieron de transcritos in vitro del gen objetivo que fueron diluidos en serie en ARN portador de riñón de rata a  $1 \times 10^7$ ,  $1 \times 10^5$  y  $1 \times 10^3$  copias. No se incluyeron controles del objetivo en cada ejecución del ensayo para asegurar la carencia de contaminación ambiental. Todas las muestras y controles se ejecutaron por duplicado. Se realizó qRT-PCR con reactivos de uso de laboratorio en general en una reacción de 10  $\mu$ l que contenía: Tampón de RT-PCR (50 nM Bicina/KOH pH 8,2, 115nM KAc, 8% de glicerol, 2,5 mM de  $MgCl_2$ , 3,5 mM de  $MnSO_4$ , 0,5 mM cada uno de dCTP, dATP, dGTP y dTTP), aditivos (2 mM de Tris-Cl pH 8, 0,2 mM de albúmina bovina, 150 mM de trehalosa, 0,002% de Tween 20), mezcla de enzimas (2U Tth (Roche), 0,4 mg/ $\mu$ l Ab TP6-25), Cebador y mezcla de sondas (0,2  $\mu$ M de Sondas, 0,5  $\mu$ M de cebadores). Siguieron los siguientes parámetros de ciclación; 1 ciclo a 95° C durante 1 minuto; 1 ciclo a 55° C durante 2 minutos; Desnivel 5%; 1 ciclo a 70° C durante 2 minutos; y 40 ciclos de 95° C durante 15 segundos; 58° C durante 30 segundos. Después de que se completó la reacción de PCR, se establecieron los valores de la línea de base y umbral en el software ABI 7900HT Prism y los valores de Ct calculados se exportaron a Microsoft Excel.

#### *Reacción de Un Paso frente a Dos Pasos*

La síntesis de la primera cadena se llevo a cabo usando o 100 ng de hexámeros aleatorios o cebadores específicos de genes por reacción. En el primer paso, se calentaron 11,5  $\mu$ l de Mezcla-1 (cebadores y lug de ARN total) a 65° C durante 5 minutos y después se refrigeraron en hielo. Se añadieron 8,5  $\mu$ l de Mezcla-2 (1x Tampón, 0,01 mM de DTT, 0,5 mM de cada dNTP, 0.25U/ $\mu$ l de RNasin®, 10 U/ $\mu$ l Superscript III) a Mezcla-1 y se incubó a 50° C durante 60 minutos seguido por 95° C durante 5 minutos. El ADNc fue almacenado a -20° C hasta que estaba listo para el uso. El qRT-PCR para el segundo paso de la reacción de 2 pasos se realizó como se ha señalado anteriormente con los parámetros de ciclación: 1 ciclo a 95° C durante 1 minuto; 40 ciclos de 95° C durante 15 segundos, 58° C durante 30 segundos. El qRT-PCR para la reacción de un paso se realizó exactamente como se ha señalado en el párrafo anterior. Las reacciones de tanto un paso como la de dos pasos se realizaron en 100 ng de plantilla (ARN/ADNc). Después de que se completó la reacción de PCR se establecieron los valores de línea de base y umbral en el software ABI 7900HT Prism y los valores de Ct calculados se exportaron a Microsoft Excel.

#### *Generación de un mapa de calor*

Para cada muestra, se calculo un  $\Delta Ct$  tomando el Ct medio para cada Marcador de CUP y restando el Ct medio de una media de los Marcadores domésticos ( $\Delta Ct = Ct(\text{Marcador de CUP}) - Ct(\text{Marcador Doméstico medio})$ ). El  $\Delta Ct$  mínimo para cada tejido del conjunto de Marcadores de origen (pulmón, mama, próstata, colon, ovario y páncreas) se determinó para cada muestra. El tejido de origen con el  $\Delta Ct$  mínimo general se obtuvo una vez y todos los otros tejidos de origen obtuvieron cero. Los datos fueron ordenados de acuerdo con el diagnóstico patológico. Se rellenó un Partek Pro con los datos de viabilidad modificados y se generó un gráfico de intensidad.

#### **Resultados.**

##### *Descubrimiento de nuevo tumor de origen de páncreas y Marcadores del estado del cáncer*

Primero se analizaron cinco candidatos para Marcador de páncreas: antígeno de células madre de próstata (PSCA), inhibidor de proteinasa de serina, miembro 1 de clade A (SERPINA1), citoqueratina 7 (KRT7), metaloproteasas de matriz 11 (MMP11), y mucin4 (MUC4) (Varadhachary et al (2004); Fukushima et al. (2004); Argani et al. (2001); Jones et al. (2004); Prasad et al. (2005); y Moniaux et al. (2004)) usando micromatrices de ADN y un panel de 13 adenocarcinomas ductales de páncreas, cinco tejidos de páncreas normales y 98 muestras de tumores de mama, colorrectales, pulmón y ovario. Solo el PSCA demostró sensibilidad moderada (se detectaron seis de trece o 46% de tumores) a un especificidad alta (se identificaron correctamente 91 de 98 o 93% como no ser de origen pancreático) (Figura 4A). Por el contrario KRT7, SERPINA1, MMP11 y MUC4 demostraron sensibilidades del

38%, 31%, 85% y 31% respectivamente, a especificidades del 66%, 91%, 82% y 81 % respectivamente. Estos datos se encontraban en buena concordancia con el qRTPCR realizado en 27 metástasis de origen pancreático y 39 metástasis de origen no pancreático para todos los Marcadores excepto el MMP11 que mostró una sensibilidad y especificidad más pobres con el qRTPCR y las metástasis. En conclusión, los datos de la micromatriz en tejido primario congelado rápidamente sirve como un buen indicador de la capacidad del Marcador para identificar a una metástasis de FFPE como pancreática en origen usando qRTPCR pero esos Marcadores adicionales pueden ser útiles para un rendimiento óptimo.

Como el adenocarcinoma ductal de páncreas se desarrolla de células epiteliales ductales que comprenden sólo un pequeño porcentaje de todas las células pancreáticas (con células acinares y células de los islotes que comprenden la mayoría) y como los tejidos de adenocarcinoma de páncreas contienen una cantidad significativa de tejido normal adyacente (Prasad et al. (2005) y Ishikawa et al. (2005)), ha sido difícil identificar Marcadores de cáncer de páncreas (es decir, regulados hacia arriba en cáncer) que también diferenciarían este órgano de los órganos. Para su uso en un panel CUP dicha diferenciación es necesaria. El primer método de consulta (ver Materiales y Métodos) devolvió seis conjuntos de sonda: factor de coagulación V (F5), una proteína hipotética FLJ22041 similar a las proteínas de unión a FK506 (FKBP10), integrina  $\beta$  6 (ITGB6), transglutaminasa 2 (TGM2), ribonucleoproteína heterogénea nuclear A0 (HNRPO), y BAX delta (BAX). El segundo método de consulta (ver Materiales y Métodos) devolvió ocho conjuntos de sonda: F5, TGM2, factor de transcripción de homeodominio tipo emparejado 1 (PITX1), isoforma trío ARNm (TRIO), ARNm para p73H (P73), una proteína desconocida para MGC: 10264 (SCD), y dos conjuntos de sonda para claudina18. F5 y TGM2 estaban presentes en ambos resultados de consulta y, de los dos, el F5 parecía el más prometedor (Figura 4B).

#### *Optimización de la preparación de las muestras y qRTPCR usando tejidos FFPE*

Después del aislamiento de ARN y qRTPCR los métodos se optimizaron usando tejidos fijados antes de examinar el rendimiento del panel de Marcadores. Primero se analizó el efecto de reducir el tiempo de incubación de la proteinasa K de dieciséis horas a 3 horas. No hubo efecto en el rendimiento. Sin embargo, algunas muestras mostraron fragmentos más largos de ARN cuando se usó el paso de proteinasa K más corto (Figura 5). Por ejemplo, cuando el ARN se aisló de un bloque de un año (c2"), no se observó diferencia en los electroferogramas. Sin embargo, cuando el ARN se aisló de un bloque de cinco años (C23), se observó una fracción más grande de ARNs de peso molecular más alto, según lo evaluado por la protuberancia en el desnivel, cuando se usó la digestión de proteinasa K más corta. Esta tendencia se mantuvo generalmente cuando se procesaron otras muestras, independientemente del órgano de origen para la metástasis de FFPE. En conclusión, acortar el tiempo de digestión de la proteinasa K no sacrifica rendimientos de ARN y puede ayudar a aislar ARN menos degradado, más largo.

Después, se compararon tres métodos diferentes de transcripción inversa: transcripción inversa con hexámeros aleatorios seguido por qPCR (dos pasos), transcripción inversa con un cebador específico del gen seguido por qPCR (dos pasos), y un qRTPCR de un paso usando cebadores específicos de los genes. El ARN se aisló de once metástasis y se comparó con valores de Ct a través de los tres métodos para  $\beta$ -actina, proteína surfactante humana B (HUMSPB) y factor de transcripción tiroideo (TTF) (Figura 6). Hubo diferencias estadísticamente significativas ( $p < 0,001$ ) para todas las comparaciones. Para los tres genes, la transcripción inversa sin hexámeros aleatorios seguido por qPCR (reacción de dos pasos) dieron los valores de Ct más altos mientras que la transcripción inversa con un cebador específico del gen seguido por qPCR (reacción de dos pasos) dio valores de Ct ligeramente (pero estadísticamente significativos) más bajos que los correspondientes a la reacción de 1 paso. Sin embargo, el REPCR de 2 pasos con cebadores específicos de los genes tenía un paso de transcripción inversa más largo. Cuando los valores de Ct de HUMSPB y TTF se normalizaron al valor de  $\beta$ -actina correspondiente para cada muestra, no hubo diferencias en los valores de Ct normalizados a través de los tres métodos. En conclusión, la optimización de las condiciones de reacción de RTPCR pueden generar valores de Ct más bajos, lo que puede ayudar a analizar bloques de parafina más viejos (Cronin et al (2004)), y una reacción de RTPCR de un paso con cebadores específicos de los genes puede generar valores de Ct comparables a los generados en la reacción de dos pasos correspondiente.

#### *Rendimiento de diagnóstico de un ensayo qRTPCR de CUP.*

Después se realizaron 12 reacciones de qRTPCR (10 Marcadores y dos genes domésticos) en 239 metástasis de FFPE. Los Marcadores usados para el ensayo se muestran en la Tabla 2. Los Marcadores del pulmón eran proteína B asociada al pulmón surfactante humana (HUMSPB), factor de transcripción tiroideo 1 (TTF1), y desmogleína 3 (DSG3). El Marcador colorrectal era cadherina 17 (CDH 17). Los Marcadores de mama eran mamaglobina (MG) y factor de transcripción derivado de la próstata Ets (PDEF). El Marcador de ovario era el tumor de Wilms 1 (WT1). Los Marcadores de páncreas eran antígeno de células madre de la próstata (PSCA) y factor de coagulación V (f5), y el Marcador de próstata era calicreína 3 (KLK3). Para la descripción de los genes, ver la Tabla 31.

Tabla 2. Secuencias de cebadores y sondas, números de acceso y longitudes de amplicones

Objetivo	SEQ ID NO	Secuencia (5'-3')	Descripción	SEQ ID NO
SP-B	59	cacagcccgcgaccttgatga	Cebador Directo	11
		ggfccagagcccctctca	Cebador Inverso	12
		agctgccagctgcaaggaaaagcc	Sonda*	13
		<b>cacagccccgaccttgatgagaactcagctgtccagctgcaaggaaaaagc</b> <b>caactgaacacccctctcttqqacc</b>	Amplicón	14
TTF1	60	ccaaccagaccgcgc	Cebador Directo	15
		cgcccatgcccctcatgttca	Cebador Inverso	16
		ccggccatctcccgcctcatg	Sonda*	17
		<b>ccaaccagaccgcgctccccccatctcccgcctcatggtggccccggcgagc</b> <b>ggcatgaacatgaggcgcatgggcg</b>	Amplicón	18
DSG3	61	gcagagaaggagaagataactcaa	Cebador Directo	19
		actccagagattcggtaggiga	Cebador Inverso	20
		attgccaaattacttcagattacca	Sonda*	21
		<b>gcagagaaggagaagataactcaaaaagaaccaccactgccaagattacttc</b> <b>agattaccaagcaaccagaaaatcacctaccgaaatctctggagt</b>	Amplicón	22
CDH17	62	tcctcggcagtggaagotta	Cebador Directo	23
		tcctcaaacctctgtgcccggta	Cebador Inverso	24
		ccaaaatcaatggtactcatgcccgcactg	Sonda*	25
		<b>tcctcggcagtggaagttacaacaacgacigggaaagttccaataatcaatggt</b> <b>actcatgcccgcactgtctaccaggcacacagagttgagga</b>	Amplicón	26
MG	63	agttcgtgatggtccctcatgc	Cebador Directo	27
		cactgtggattgatgtcttga	Cebador Inverso	28
		ccctctcccagcactgtctacgca	Sonda*	28
		<b>agttcgtgatggtccctcatgctggcgcccctctcccagcactgtctacgaggctct</b> <b>ggctgccccttattggagaatgtgattccaagacaatcaatccaacagtg</b>	Amplicón	30
PDEF	64	cgcccacctggacatctgga	Cebador Directo	31

(continuada).

Objetivo	SEQ ID NO	Secuencia (5'-3')	Descripción	SEQ ID NO
		cactgtcgaggccacagtagta	Cebador Inverso	32
		gtcagcgccctggatgaaagagcgg	Sonda*	33
		cgcccaccctggacatctggaagtgcgcccctggatgaaagagcggacttca ccctggggcgattcactactgctcctgaccagctg	Amplicón	34
WT1	65	gcgagcccccaatacagaatacac	Cebador Directo	35
		cggggtactccaggcaca	Cebador Inverso	36
		tcagaggcattcaggatgfcgacg	Sonda*	37
		gcgagcccccaatacagaatacacacgcgacgggtctcttcagaggcattcagga tgtcgcacgtgtcctggagtagccccg	Amplicón	38
PSCA	66	ctgtgatggcaggcttggc	Cebador Directo	39
		ttgctcacctggctttgca	Cebador Inverso	40
		gcagccaggcactgcccctgct	Sonda*	41
		ctgtgatggcaggcttggcccctgcagccaggcactgcccctgtgtgtactct gcaaagcccagggtgagcaa	Amplicón	42
F5	67	tgaagaaatatcctggattattca	Cebador Directo	43
		tatgtgtatcttctggaatacatca	Cebador Inverso	44
		acaaagggaaaacagatataagactc	Sonda*	45
		tgaagaaatatcctggattattcagaattgtacaaagggaaaacagatattgaa gactctgatgattccagaagataccacata	Amplicón	46
KLK3	68	ccccagtggtctcaca	Cebador Directo	47
		aggatgaaaacaagctgtgccga	Cebador Inverso	48
		caggaacaaaagcgtgatctgtcgtg	Sonda*	49
		ccccagtggtctcctcacagctgccacatcaggaacaaaagcgtgatct tgctgtggtcggcacagctgtttctctct	Amplicón	50
β actin	69	gccccagggcactctcca	Cebador Directo	51
		cggtatgtccacgtcacacttca	Cebador Inverso	52

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

(continuada)

Objetivo	SEQ ID NO	Secuencia (5'-3')	Descripción	SEQ ID NO
		cttcctcctgggcatggagtctg	Sonda*	53
		<b>gcccTgaggcactctccagcctcctcctcctgggcatggagTccTgTggcatccac gaaactaccTtcaactccatcatgaagTgacgTggacatccg</b>	Amplicón	54
PBGD	70	ccacacacagcctactttccaa	Cebador Directo	55
		taccacgcggaatcactctca	Cebador Inverso	56
		aacggcaatgcggctgcaacggcggaa	Sonda*	57
		<b>ccacacacagcctactttccaagcggagccatgctcTgTtaacggcaatgcggc TgcaacggcgggaagaaacagcccaaatgagagTgattcgcgTgggTta</b>	Amplicón	58

\* Las sondas son 5'FAM-3'BHQ1-TT

El análisis de los valores de Ct normalizados en un mapa de calor reveló la alta especificidad de los Marcadores de mama y próstata, especificidad moderada de colon, pulmón y ovario, y especificidad algo menor de los Marcadores de páncreas. Combinar los datos qRTPCR normalizados con refinamiento computacional mejora el rendimiento del panel de Marcadores. Los resultados se obtuvieron de los datos de qRTPCR normalizados combinados con el algoritmo y se precisó la precisión del ensayo qRTPCR.

## Discusión

En este ejemplo, la realización de perfiles de expresión basados en micromatrices se uso en tumores primarios para identificar Marcadores candidatos para su uso con metástasis. El hecho de que los tumores primarios puedan ser usados para descubrir tumores de Marcadores de origen para metástasis es consistente con varios descubrimientos recientes. Por ejemplo, Weigelt y sus colaboradores han demostrado que los perfiles de expresión génica de tumores de mama primarios se mantienen en metástasis distantes. Weigelt et al. (2003). Italiano y sus colaboradores descubrieron que el estado de EGFR, evaluado por IHC, era similar en 80 tumores colorrectales primarios y las 80 metástasis relacionadas. Italiano et al. (2005). Solo cinco de los 80 mostraron discordancia en el estado de EGFR. Italiano et al. (2005). Backus y sus colaboradores identificaron Marcadores putativos para detectar metástasis de cáncer de mama usando un análisis de expresión génica de genoma amplio de la mama y otros tejidos y demostraron que la mamaglobina y CK19 detectaron metástasis clínicamente procesable en ganglios linfáticos centinelas de mama con un 90% de sensibilidad y un 94% de especificidad. Backus et al. (2005).

Los estudios basados en micromatrices con tejido primario confirmaron la especificidad y sensibilidad de Marcadores conocidos. Como resultado, con la excepción de F5, todos los Marcadores usados tenían alta especificidad para los tejidos estudiados en la presente. Argani et al (2001; Backus et al. (2005); Cunha et al. (2005); Borgono et al. (2004); McCarthy et al. (2003); Hwang et al. (2004); Fleming et al. (2000); Nakamura et al. (2002) y Khor et al. (1997). Un estudio reciente determinó que usando IHC, el PSCA se sobreexpresa en la metástasis del cáncer de próstata. Lam et al. (2005). Dennis et al. (2002) también demostró que el PSCA podría ser usado como un tumor de Marcador de origen para páncreas y próstata. Como se muestra en la presente, la expresión fuerte del PSCA se encuentra en algunos tejidos de próstata a nivel de ARN pero, como se incluye el PSA en el ensayo, ahora se pueden segregar los cánceres de próstata y cáncer. Un descubrimiento nuevo de este estudio fue el uso del F5 como un Marcador complementario (a PSCA) para el tejido pancreático de origen. En tanto el conjunto de datos de micromatrices con tejido primario como el conjunto de datos de qRTPCR con metástasis de FFPE, el F5 complementó al PSCA (Figura 4 y Tabla 3).

Tabla 3 datos de viabilidad

	Mama	Colon	Pulmón	Otro	Ovario	Páncreas	Próstata	Total
Total probado	30	30	56	32	49	43	20	260
#Correcto	22	27	45	16	43	31	20	204
#Otro / No probado	1	1	3	n/a	1	4	0	10
#Incorrecto	7	2	8	16	5	8	0	46
% Probado	96.67	96.67	94.64	100	97.96	90.70	100	96.15
% Correcto de probados	75.86	93.10	84.91	0	89.58	79.49	100	81.60
Correcto del total (%)	73.33	90.00	80.36	50.00	87.76	72.09	100	78.46

Investigadores anteriores han generado ensayos de CUP usando IHC o micromatrices. Su et al. (2001); Ramaswamy et al. (2001) y Bloom et al. (2004). Más reciente, se ha acoplado SAGE a un panel de Marcadores qRTPCR pequeño. Dennis (2002) y Buckhaults et al. (2003). Este estudio es el primero en combinar perfiles de expresión basados en micromatrices con un panel pequeño de ensayos qRTPCR. Los estudios de micromatrices con tejido primario identificaron algunos, pero no todos, de los mismos Marcadores del tejido de origen como los identificados anteriormente por estudios SAGE. Algunos estudios han demostrado que existe un acuerdo modesto entre los datos de perfiles basados en micromatrices de ADN y SAGE y que la correlación mejora para genes con niveles de expresión más altos, van Ruissen et al. (2005) y Kim (2003). Por ejemplo, Dennis y sus colaboradores identificaron PSA, MG, PSCA y HUMSPB mientras Buckhaults y sus colaboradores (Dennis y otros, (2002)) identificaron PDEF. Se prefiere ejecutar el ensayo CUP usando qRTPCR ya que es una tecnología robusta y puede tener ventajas de rendimiento sobre el IHC. Al-Mulla et al. (2005); y Haas et al. (2005). Como se muestra en la presente, el protocolo de qRTPCR se mejoró a través del uso de cebadores específicos de genes en una reacción de un paso. Esta es la primera demostración del usos de cebadores específicos de genes en una reacción de qRTPCR de un paso con tejido de FFPE. Otros investigadores o han hecho un qRTPCR de dos pasos (síntesis de ADNc en una reacción seguido por qPCR) o han usado hexámeros aleatorios o truncado cebadores específicos de genes. Abrahamsen et al. (2003); Specht et al. (2001); Godfrey et al. (2000); Cronin et al. (2004) y Mikhitarian et al.



(2004).

**Ejemplo 2**

5 Protocolo de Aislamiento de ARN Total CUP FFPE  
(kit Highpure Cat#3270289)

**Propósito**

10 Aislamiento del ARN total del tejido de FFPE

**Procedimiento:**

**Preparación de soluciones de trabajo**

15 1. Proteinasa k (PK) en el kit  
Disolver liofilizado en 4.5 ml de Tampón de Elución. Alícuota y almacenar a -20° , estable durante 12 meses.  
20 PK-4x250mg (cat #3115852)  
Disolver liofilizado en 12,5 ml de Tampón de Elución (1x Tampón TE (pH 7,4-7). Alícuota y almacenar a -20° C.

2. Tampón de Lavado I

25 Añadir 60 ml de etanol absoluto a Tampón de Lavado I, almacenar a temperatura ambiente.

3. Tampón de Lavado II

30 Añadir 200 ml de etanol absoluto a Tampón de Lavado II, almacenar a temperatura ambiente.

4. DNasa I

35 Disolver liofilizado en 400 µl de Tampón de Elución. Alícuota y almacenar a -20° C, estable durante 12 meses.

**Seccionar Bloques de Parafina de -30-45 minutos para 12 bloques (12 bloques x 2 tubos = 24 tubos)**

Las secciones cortadas del bloque deberían ser procesadas inmediatamente para la extracción de ARN  
40 1. Usar una hoja de afeitar afilada limpia en Micrótopo para cortar secciones de 6 x 10 micras de grosor de bloques de tejido recortados (tamaño 3-4 x 5-10 mm).  
Nota: Bloque Nuevo-Descartar las secciones de cera hasta que se ha obtenido la sección del tejido. Bloques usados-Descartar las 3 primeras secciones de tejido  
2. Inmediatamente colocar el tejido cortado en tubos de microcentrifuga de 1,5 mL y tapar firmemente para minimizar la humedad.  
45 3. Se recomienda tomar el número de secciones en base al tamaño del tumor mostrado en la Tabla 4.

Tabla 4

Tamaño de MET	Secciones/Tubo
8-10 mm	6
6-8 mm	12
2-4 mm	18

**Desparafinización ~ 30-45 minutos**

60 1. Añadir 1.0 ml de xileno a cada muestra y agitar en vórtice vigorosamente durante 10-20 segundos e incubar RT durante 25 minutos. Centrifugar a toda velocidad 2 minutos. Retirar el sobrenadante con cuidado.  
2. Repetir el paso 1.  
3. Centrifugar a toda velocidad durante 2 minutos. Retirar el sobrenadante.  
65 4. Añadir 1 ml de etanol abs. y agitar en vórtice vigorosamente 1 min. Centrifugar a toda velocidad 2 minutos.

Retirar el sobrenadante.

5. Repetir el paso 4.

6. Secar el tubo brevemente en una toalla de papel para deshacerse de los residuos de etanol.

7. Secar el pellet seco durante 5-10 minutos a 55° C en un horno.

5 Nota: es crítico que el etanol sea retirado completamente y los pellets sean secados concienzudamente, el etanol residual puede inhibir la digestión de PK.

**Nota: si la PK está a -20C, calentarla en RT 20-30 minutos**

Extracción de ARN ~2,5-3 horas

10 1. Añadir 100 µl de Tampón de Lisis de Tejido, 16 µl de 10% SDS y 80 µl de solución de trabajo de Proteinasa K a un pellet de tejido, agitar en vórtice brevemente en varios intervalos e incubar **2 horas** a 55° C agitando a 400 rpm,

15 2. Añadir 325 µl de Tampón de Unión y 325 µl de etanol abs. Mezclar suavemente pipeteando arriba y abajo.

3. Centrifugar el lisado a toda velocidad durante 2 minutos.

4. Combinar el tubo de filtro y el tubo de colección (12 tubos), y pipetear el sobrenadante del lisado en el filtro.

5. Centrifugar durante 30 segundos a 800 rpm y descartar el exceso de flujo.

Nota: El Paso 4-5 se puede repetir, si el ARN necesita ser agrupado con 2 preparaciones de pellet de tejido más.

20 6. Repetir la centrifugación a 8000 rpm durante 30 segundos para secar el filtro.

7. Añadir 500 µl de solución de trabajo de Tampón de Lavado I a la columna y centrifugar durante 15-30 segundos a 8000 rpm, descartar el exceso de flujo.

8. Añadir 500 µl de solución de trabajo de Tampón de Lavado II. Centrifugar durante 15-30 segundos a 8000 rpm, descartar el exceso de flujo.

25 9. Añadir 300 µl de solución de trabajo de Tampón de Lavado II, centrifugar durante 15-30 segundos a 8000 rpm, descartar el exceso de flujo.

10. Centrifugar el filtro High Pure durante 2 minutos a máxima velocidad.

11. Colocar el tubo del filtro High Pure en un tubo de 1,5 ml nuevo y añadir 90 µl de Tampón de Elución. Incubar durante 1-2 minutos a temperatura ambiente. Centrifugar 1 minuto a 800 rpm. **Tratamiento de DNasa I ~ 1,5 horas.**

30 12. Añadir 10 µl de Tampón de Incubación de 10x DNasa y 1,0 µl de solución de trabajo de DNasa I al eluido y mezclar. Incubar durante 45 minutos a 37° C (o **2,0 µl de DNasa I durante 30 minutos**).

13. Añadir 20 µl de Tampón de Lisis de Tejido, 18 µl de 10% SDS y 40 µl de solución de trabajo de Proteinasa K. Agitar en vórtice brevemente. Incubar durante **30 minutos** (30-60 minutos) a 55° C

35 14. Añadir 325 µl de Tampón de Unión y 325 µl de etanol abs. Mezclar y pipetear en un tubo de filtro High Pure nuevo con tubo de recolección (12 tubos).

15. Centrifugar durante 30 segundos a 8000 rpm y descartar el exceso de flujo.

16. Repetir la centrifugación a 8000 rpm durante 30 segundos para secar el filtro.

40 17. Añadir 500 µl de solución de trabajo de Tampón de Lavado I a la columna. Centrifugar durante 15 segundos a 8000 rpm, descartar el exceso de flujo.

18. Añadir 500 µl de solución de trabajo de Tampón de Lavado II. Centrifugar durante 15 segundos a 8000 rpm, descartar el exceso de flujo.

19. Añadir 300 µl de solución de trabajo de Tampón de Lavado II. Centrifugar durante 15 segundos a 8000 rpm, descartar el exceso de flujo.

45 20. Centrifugar el filtro High Pure durante 2 minutos a máxima velocidad.

21. Colocar el tubo del filtro High Pure en un tubo de 1,5 ml nuevo. Añadir 50 µl de Tampón de Elución; incubar durante 1-2 minutos a temperatura ambiente. Centrifugar durante 1 minuto a 8000 rpm para recoger el ARN eluido.

50 22. Centrifugar el eluido durante 2 minutos a toda velocidad y transferir el sobrenadante a un nuevo tubo sin perturbar las fibras de vidrio en el fondo.

23. Tomar la lectura de 260/280 OD y diluir a 50 ng/µl. Almacenar a -80° C.

**Protocolo de Ensayo CUP ASR (ABI 7900)**

55 **Propósito:** Usar qRTPCR para determinar el tejido de origen de una muestra de CUP

**Configuración del Control:**

60 **1. Controles Positivos (Consultar la Tabla 5 y Placa C en la Configuración de Placas, Figura 7)**

65

Tabla 5 diluciones en Serie de IVT -5µl 1X10<sup>8</sup> en 470 µl de H<sub>2</sub>O + 25µl de 10000 ARNr

Control IVT	CE/µl	Muestra	Agua	ARNr de fondo
BACTIN	100E+05	50	425	25
CDH 17	100E+05	50	425	25
DSG3	100E+05	50	425	25
F5	100E+05	50	425	25
Hump	100E+05	50	425	25
MG	100E+05	50	425	25
PBGD	100E+05	50	425	25
PDEF	100E+05	50	425	25
PSCA	100E+05	50	425	25
TTF1	100E+05	50	425	25
WT1	100E+05	50	425	25
1E6. Tabla 5. Diluir 50,000 CE/µl de ARNr a 500 CE/µl - 5µl 50,000 CE/µl + 495 µl H <sub>2</sub> O				

Alícuotas 10 µl por tubo de tira (2 placas); Colocar la Mezcla a -80° C hasta que esté lista para el uso.

**2. Curvas estándar (Consultar la Tabla 6 y la Placa C en la Configuración de Placas, Figura 7)**

Paso 1: La curva estándar se configuro exactamente como se muestra en la Tabla 6.

Tabla 6. Solución Madre - 1X10<sup>8</sup> IVT. Diluir 50,000 CE/µl de ARNr a 500 CE/µl - 5ml 50,000 CE/µl + 495 µl H<sub>2</sub>O

Control IVT	CE/µl	Muestra	Agua	ARNr de fondo
BACTIN-1	100E+07	50	425	25
BACTIN-2	100E+06	50	425	25
BACTIN-3	100E+05	50	425	25
BACTIN-4	100E+04	50	425	25
BACTIN-5	100E+03	50	425	25
PBGD-1	100E+07	50	425	25
PBGD-2	100E+06	50	425	25
PBGD-3	100E+05	50	425	25
PBGD-4	100E+04	50	425	25
PBGD-5	100E+03	50	425	25

Alícuotas 10 µl por tubo de tira (2 placas); Colocar la Mezcla a -80° C hasta que esté lista para el uso.

**Mezcla de Enzimas:**

**1. Mezcla Maestra:** Enzima (Tth) / Anticuerpo (TP6-25), ver Tabla 7.

Tabla 7

Reactivo	2x
Enzima Tth (5U/ $\mu$ l)	600.00
Anticuerpo: TP6-25 (1mg/ml)	600.00
Agua	300.00
Total	1500.00

Alícuota 500  $\mu$ l/tubo y congelar a -20° C.

**Mezcla Maestra CUP:**

**1. 2.5X Mezcla Maestra CUP (Tablas 8-11):**

Tabla 8

ml	5x Aditivos	2.5x Conc.
0.50	1M Tris-Cl pH 8	5mM
1.25	40mg/ml de Albumina, bovina	500 $\mu$ g/ml
37.50	1M Trehalosa madre	375mM
2.5	20%v Tween 20	0.50%
7.00	ddH <sub>2</sub> O	
48.75		
<i>Permitir que el reactivo se mezcle completamente &gt; 15 minutos</i>		

Tabla 9

ml	5x Aditivos	2.5 x Conc
12.50	1M Bicina/Hidróxido de Potasio pH 8.2	125mM
5.75	5M Acetato de Potasio	287.5mM
20.00	Glicerol (V x D = M -> 19.6 x 1.26 = 24.6 g)	20%
1.25	500mM Cloruro de Magnesio	6.25mM
1.75	500mM Cloruro de Magnesio	8.75mM
5.00	ddH <sub>2</sub> O	
46.25		
<i>Permitir que el reactivo se mezcle completamente &gt; 15 minutos; Combinar las mezclas anteriores en un contenedor estéril - añadir lo siguiente</i>		

# ES 2 525 545 T3

Tabla 10

ml	5x Aditivos	2.5x Conc.
1.25	100mM dATP	1.25mM
1.25	100mM dCTP	1.25mM
1.25	100mM dTTP	1.25mM
1.25	100mM dGTP	1.25mM
100.00		
<i>Permitir que el reactivo se mezcle completamente &gt; 15 minutos; Alícuota 1.8ml/tubo y congelar a -20°C</i>		

Tabla 11

Cebador/Sonda	Madre (μM)	FC (μM)	ΣI
Cebador Directo	100	10	100.0
Cebador Inverso	100	10	100.0
Sonda (5'FAM/3'BHQ1-TT)	100	4	40.0
Agua DI			760.0
Total			1000.0

**Mezcla de Cebador y Sonda:**

*Alícuota 250 μl/tubo y congelar a -20° C*

**Mezcla de la Reacción:**

1. *Mezcla Maestra de CUP (CMM): (Consultar las Tablas 12-14 y la Placa A en la Configuración de Placas, Figura 7)*

Tabla 12

Reactivo	FC	X1 (10μl)	450
2.5 x Mezcla Maestra CUP	1X	4.00	1800
ROX	1x	0.20	90
2x MezclaTthAb	2U	1.00	450
Agua		2.3	1035
Total		7.50	3375

Preferiblemente, cada ejecución/placa no tendrá más de 356 reacciones: 12 muestras con 12 Marcadores (288 reacciones con 2 replicados para cada una) +10 controles de curva estándar por duplicado (20) + 2 controles positivos y 2 negativos para cada Marcador (4X12=48).

*Ajustar el agua para el volumen de muestra - 4,3 μl Muestra MAX; Mezclar Bien*

Tabla 13

Reactivo	FC	X1 (10μl)	34
Cebadores 10μM/Sonda 4μM	0.5 μM/0.2 μM	0.50	17
CMM	1x	7.50	255
Total		8.00	272

2. **Marcadores ToO: Mezclar Bien**

Tabla 14

Reactivo	FC	X1 (10µl)	44
Cebadores 10µM/Sonda 4µM	0.5µM/0.2µM	0.50	22
CMM	1x	7.50	330
Total		8.00	352

3. **Marcadores de β-Actina y PBGD: Mezclar bien**

**Configuración de la Muestra:**

Tabla 15

Muestra	ID de la Muestra	Conc	Agua Añadida = 50ng/µl
A1			
A2			
A3			
A4			
A5			
A6			
A7			
A8			
A9			
A10			
A11			
A12			

1. **Muestras de CUP: 12 muestras en Placa de 96 pocillos: A1-A12 (Consultar la Tabla 16 y la Placa B en la Configuración de Placas, Figura 7); Alícuota 50 µl de 50 ng/µl (2µl/rxn)**

**Placa de Carga:**

1. **Configuración de Placa de 384 Pocillos: (Consultar la Placa D en la Configuración de Placas, Figura 7)**

Se cargan 2 µl de muestra y 8 µl de CMM en la placa. (muestra=50ng/µl).

Se cargan 4 µl de muestra y 6 µl de CMM en la placa. (muestra=25ng/µl).

La placa se sella y se etiqueta. Centrifugar a 2000 rpm durante 1 minuto. **Configuración de ABI 7900 HT:** Colocar en el ABI 7900. Seleccionar el programa "CUP 384" y presionar comenzar.

Tabla 16

Condiciones de termociclado	
5	95C x 60s
	55C x 2m
	RAMP 5%
10	70C x 2m
	40 ciclos de
	95C x 15s
15	58C x 30s
	ROX Encendido

Se analizan los datos, los de Ct extraídos e insertados en el Algoritmo.

### Ejemplo 3

#### Algoritmo de CUP

Los valores de  $\Delta Ct$  normalizados de actina para HPT, MGB, PDEF, PSA, SP-B, TFF, DSG, WT1, PSCA, y F5 se colocan en 6 conjuntos basados en el tejido de origen para el que fueron originalmente seleccionados. Las constantes 9.00, 11.00, 7.50, 5.00, 10.00, 9.50, 6.50, 8.00, 9.00, y 8.00 se restan de cada  $\Delta Ct$  respectivamente. Después, para cada muestra se selecciona el valor de CT mínimo de cada uno de los 6 conjuntos (HPT, min (MGB, o PDEF), PSA, min (SP-B, TFF, o DSG), WT1, and min (PSCA, o F5)) como la variable representativa para el grupo.

Estas variables, y el sitio de metástasis se usan para clasificar la muestra usando discriminantes lineales. se deben construir dos modelos diferentes, uno para hombres y uno para mujeres, de los datos de aprendizaje usando la función 'lda' de la librería MASS (Venables et al. (2002) en R (version 2.0.1). Se calcula entonces una probabilidad posterior para cada ToO usando la función 'predecir2 para cada modelo masculino o femenino.

Las variables usadas en los modelos masculinos son HPT, PSA, el mínimo de ('SP-B', 'TFF', 'DSG3'), el mínimo de ('LCSP', 'F5'), y el sitio de metástasis. La categoría del sitio de metástasis tiene 4 niveles correspondientes a colon, pulmón, ovario y todos los otros tejido. Para los modelos femeninos, las variables son HPT, el mínimo de ('MGB', 'PDEF'), el mínimo de ('SP-B', 'TFF', 'DSG3'), WT1, el mínimo de ('LCSP', 'F5'), y el sitio de metástasis.

#### Ejemplo de Código R

```
#Training the male model
dat.m<-CUP2.MIN.NORM[,c
('HPT','PSA','SP.B.TTF.DSG3','PSCA.F5','Class','background')]
CUP.Ida.m<-
lda(Class~.,dat.m,prior=c(0,0.09,0.23,0.43,0.16,0.02)/sum(c
(0,0.09,0.23,0.43,0.16,0.02)))
#Training the female model
dat.f<-CUP2.MIN.NORM[,c
('HPT','MGB.PDEF','SP.B.TTF.DSG3','WT1','PSCA.F5','Class','back
ground')]
CUP.Ida.f<-
lda(Class~.,dat.f,prior=c(0.03,0.09,0.23,0.43,0.04,0.16,0)/sum(c
(0.03,0.09,0.23,0.43,0.04,0.16,0)))
#if unknown sample (i) is male
predict(CUP.Ida.m, CUP2.MIN.NORM.TEST[i,])
#if unknown sample (i) is female
predict(CUP.Ida.f, CUP2.MIN.NORM.TEST[i,])
```

Para ejecutar este código, un marco de datos llamado CUP2.MIN.NORM necesita contener los datos de

aprendizaje con el valor mínimo calculado para cada conjunto de tejido de origen como se ha descrito anteriormente.

La clase se corresponde con el tejido de origen, y el fondo se corresponde con los sitios de metástasis descritos anteriormente.

Los datos de la prueba pueden ser contenidos en el CUP2.MIN.NORM.TEST, y una muestra específica en la fila *i* puede ser probada usando la función predecir. De nuevo, los datos de prueba deben estar en el mismo formato que el conjunto de aprendizaje y tener también los ajustes del valor mínimo aplicados a él.

#### 10 Ejemplo 4

Muestras resueltas de CUP

Se compararon 48 muestras resueltas y no resueltas de CUP para determinar la correlación con las muestras de CUP verdaderas. Los métodos usados fueron los descritos en los Ejemplos 1-3. Los resultados obtenidos se presentan en la Tabla 17. Se probaron 11 muestras de CUP no resueltas, el diagnóstico se hizo en 8 muestras, 3 fueron de otra categoría.

Tabla 17

Categoría de la Muestra	Muestra #	Correcta	Incorrecta	sin prueba	% de precisión
ToO Conocido	15	11	3	1	79
CUP resuelto	22	17	4	1	81
CUP no resuelto	11	8	N/a	3	73

#### 20 Ejemplo 5

30 Límites del Ensayo de CUP

La figura 8 representa los resultados obtenidos, usando los métodos descritos en los Ejemplos 1-3, para determinar los límites de los ensayos de CUP. El rendimiento del ensayo se probó sobre un intervalo de concentraciones de ARN y se descubrió que los ensayos de CUP son eficientes en el intervalo de 100 -12.5 ng de ARN.

#### 35 Ejemplo 6

40 Ensayo de qRTPCR

**Materiales y Métodos.** *Muestras de Tejido congeladas para Análisis de Micromatrices.* Se usaron un total de 700 tejidos humanos primarios congelados para ella realización de perfiles de micromatrices de expresión génica. Las muestras se obtuvieron de una variedad de instituciones académicas, incluyendo la Universidad de Washington (St. Louis, MO), el Erasmus Medical Center (Rotterdam, Países Bajos), y compañías de banco de tejidos comerciales, incluyendo Genomics Collaborative, Inc (Cambridge, MA), Asterand (Detroit, MI), Oncomatrix (La Jolla, CA) y Clinomics Biosciences (Pittsfield, MA). Para cada espécimen, se recogió también información demográfica del paciente, clínica y patológica. Las características histopatológicas de cada muestra se revisaron para confirmar el diagnóstico y para estimar la conservación de la muestra y el contenido tumoral.

*Extracción de ARN e Hibridación del Chip Génico Affymetrix.* Se diseccionaron y homogeneizaron muestras de cáncer congeladas con más de un 70% de células tumorales, muestras benignas y normales con homogeneizador mecánico (UltraTurrex T8, Alemania) en reactivo Trizol (Invitrogen, Carlsbad, CA). El tejido fue homogeneizado en reactivo Trizol siguiendo el protocolo de Trizol estándar para el aislamientos de ARN de tejidos congelados (Invitrogen, Carlsbad, CA). Después de la centrifugación la fase de líquido superior se recogió y se precipitó el ARN total con alcohol isopropílico a -20° C. Los pellets de ARN se lavaron con 75% de etanol, se resolvieron en agua y se almacenaron a -80° C hasta su uso.

La calidad del ARN se examinó con un Agilent 2100 Bioanalyzer RNA 6000 Nano Assay (Agilent Technologies, Palo Alto, CA). El ARNc etiquetado se preparó e hibridó con el Chip Génico Hu133A de matriz de oligonucleótidos de alta densidad (Affymetrix, Santa Clara, CA) que contiene un total de 22.000 conjuntos de sonda de acuerdo con el protocolo del fabricante estándar. Las matrices fueron escaneadas usando protocolos y escáneres de Affymetrix. Para análisis posteriores, cada conjunto de sondas se consideró un gen separado. Los valores de expresión para cada gen se calcularon usando el software de análisis de Chip Génico de Affymetrix MAS 5.0. Todos los chips cumplían tres estándares de control de calidad: el porcentaje de designación "presente" para la matriz era mayor del 35%, el factor de escala era menor de 12 cuando se modificó a escala a una intensidad objetivo global de



600 y el nivel de fondo medio era menor de 150.

*Selección de Candidatos a Marcadores.* Para la selección de los candidatos a Marcadores del tejido de origen (ToO) para tejidos de pulmón, colon, mama, ovario y próstata, los niveles de expresión de los conjuntos de sondas se midieron en muestras de ARN que cubrían un total de 682 tejidos normales, benignos y cancerosos de mama, colon, pulmón, ovario, próstata. Los candidatos a Marcador específico del tejido fueron seleccionados en base al número de consultas estadísticas.

Para generar candidatos pancreáticos, se usaron perfiles de expresión génica de 13 adenocarcinomas ductales de páncreas primarios, 5 de páncreas normales y 98 especímenes de cáncer de pulmón, colon, mama y ovario para seleccionar los Marcadores de adenocarcinoma de páncreas. Se realizaron dos consultas. En la primera consulta, se creó un conjunto de datos que contenía 14547 genes con al menos 2 designaciones "presentes" en muestras de páncreas. Se identificó un total de 2736 genes que sobreexpresaban en cáncer de páncreas en comparación con el normal por prueba T ( $p < 0,05$ ). Se seleccionaron genes con expresión mínima en el percentil 11º de cáncer de páncreas que era al menos 2 veces más alto que el máximo en cáncer de colon y pulmón, haciendo 45 conjuntos de sondas. Como un paso final, se seleccionaron 6 genes con expresión máxima al menos 2 veces más alta que la expresión máxima en cánceres de colon, pulmón, mama y ovario. En una segunda consulta, se creó un conjunto de datos de 4654 conjuntos de sonda con un máximo de 2 designaciones "presentes" en especímenes de mama, colon, pulmón y ovario. Se seleccionaron un total de 160 genes que tenían al menos 2 designaciones "presentes" en muestras de cáncer y normal de páncreas. Fuera de los 160 genes, se seleccionaron 10 genes después de comparar su nivel de expresión entre tejidos de páncreas y normales. Los resultados de ambas consultas de páncreas se combinaron.

Además del análisis de los perfiles de expresión génica, se seleccionaron unos pocos Marcadores de la bibliografía. Los resultados de todas las consultas fueron combinados para hacer una lista corta de candidatos a Marcador de ToO para cada tipo de tejido. Se estimaron la sensibilidad y especificidad de cada Marcador. Los Marcadores que demostraron la mejor capacidad para diferenciar tejidos por su origen fueron nominados para pruebas de RT-PCR en base a la redundancia y complementariedad de los Marcadores.

*Carcinoma metastásico de FFPE de origen conocido y tejidos de CUP.* Se adquirieron un total de 386 carcinomas metastásicos de FFPE (Etapa III-IV) de origen conocido y 24 adenocarcinomas primarios de próstata de FFPE de una variedad de vendedores comerciales, incluyendo Proteogenex (Los Angeles, CA), Genomics Collaborative, Inc. (Cambridge, MA), Asterand (Detroit, MI), Ardais (Lexington, MA) y Oncomatrix (La Jolla, CA). se obtuvo un conjunto independiente de 48 carcinomas metastásicos de primario conocido y tejidos de CUP del Albany Medical College (Albany, NY). Para cada espécimen, se recogió información demográfica del paciente, clínica y patología. Se revisaron las características histopatológicas de cada muestra para confirmar el diagnóstico, y para estimar la conservación de la muestra y el contenido tumoral. Para las muestras metastásicas, se establecieron inequívocamente diagnósticos del carcinoma metastásico y ToO en base a la historia clínica del paciente y la evaluación histológica del carcinoma metastásico en comparación con los primarios correspondientes.

*Asilamiento de ADN de muestras de FFPE.* El aislamiento de ARN de secciones de tejido de parafina fue como se describe en el manual del High Pure RNA Paraffin Kit (Roche) con las siguientes modificaciones. Se seccionaron muestras de tejido embebidas en parafina de acuerdo al tamaño de la metástasis embebida (2-5mm = 9 X 10µm, 6-8mm = 6 X 10µm, 8 - ≥ 10mm = 3 X 10 µm). Las secciones fueron desparafinadas como se describe en el manual del Kit, el pellet de tejido se seco en un horno a 55º C durante 5-10 minutos y se resuspendió en 100 µl de tampón de lisis de tejido, 16µl de 10% SDS y 80 µl de Proteinasa K. Las mezclas se agitaron en vórtice y se incubaron en un conjunto termomezclador a 400 rpm durante 2 horas a 55º C. El procesamiento de muestras posterior se realizó de acuerdo al manual del High Pure RNA Paraffin Kit. Se cuantificaron las muestras por lecturas de OD 260/280 obtenidas por un espectrómetro y las muestras se diluyeron a 50 ng/µl . El ARN aislado fue almacenado en agua libre de RNasa a 80º C hasta su uso.

qRTPCR para *Pre-cribado de candidatos a Marcadores.* Un µg de ARN total de cada muestra fue sometido a transcripción inversa con hexámeros aleatorios usando la transcriptasa inversa Superscript II de acuerdo con las instrucciones del fabricante (Invitrogen, Carlsbad, CA). Se diseñaron cebadores y sondas MGB para los candidatos a Marcadores génicos probados y el gen de control ACTB usando software Primer Express (Applied Biosystems, Foster City, CA) o ABI Assay-on-Demand (Applied Biosystems, Foster City, CA). Todos los cebadores y sondas diseñados internamente se probaron para eficiencia de amplificación óptima por encima del 90%. La amplificación por RT-PCR se llevó a cabo en una mezcla de reacción de 20 ml que contenía 200 ng de ADNc de plantilla, 2 x mezcla maestra de PCR universal TaqMan® (10 ml) (Applied Biosystems, Foster City, CA), 500 nM de cebadores directos e inversos, y 250 nM de sonda. Las reacciones se ejecutaron en un Sistema de Detección de Secuencias ABI PRISM 7900HT (Applied Biosystems, Foster City, CA). Las condiciones de ciclación fueron: 2 minutos de activación de AmpErase UNG a 50º C, 10 minutos de activación de polimerasa a 95º C y 50 ciclos a 95º C durante 15 segundos y temperatura de hibridación (60º C) durante 60 segundos. En cada ensayo, el control "sin plantilla" junto con el ADNc de plantilla fue incluido por duplicado para tanto el gen de interés como el gen de control. La expresión relativa de cada gen objetivo fue representada como  $\Delta Ct$ , que es igual al Ct del gen objetivo restado por el

Ct del gen de control (ACTB).

*qRTPCR de un paso optimizado.* Se usaron los números de acceso de secuencias de ARNm apropiadas en conjunción con Oligo 6.0 para desarrollar ensayos de CUP TaqMan® (Marcadores de pulmón, surfactante humano, proteína B asociada al pulmón (HUMPSPBA), factor de transcripción tiroideo 1 (TTF1), desmogleína 3 (DSG3), marcador colorrectal: cadherina 17 (CDH17), marcadores de mama: mamaglobina (MG), factor de transcripción ets derivado de la próstata (PDEF), Marcador de ovario: tumor de Wilms 1 (WT1), marcadores de páncreas: antígeno de células madre de próstata (LCSP), factor de coagulación V (F5), calicreína del marcador de próstata 3 (KLK3)) y actina beta ( $\beta$ -actina) de ensayos domésticos, hidroximetilbilano sintasa (PBGD). Los cebadores específicos de los genes y las sondas de hidrólisis para el ensayo qRT-PCR de un paso optimizado se enumeran en la Tabla 2 (SEQ ID NOs: 11-58). La amplificación del ADN genómico fue excluida diseñando los ensayos alrededor de sitios de escisión exón-intrón. Las sondas de hidrólisis fueron etiquetadas en el nucleótido 5' con FAM como el colorante indicador y en el nucleótido 3' con BHQ1-TT como el colorante de extinción interno.

La cuantificación de ARN específico del gen se llevó a cabo en una placa de 384 pocillos en el sistema de detección de secuencias ABI Prism 7900HT (Applied Biosystems). Para cada ejecución de termo-ciclador se amplificaron los calibradores y las curvas estándar. Los calibradores para cada Marcador consistieron de transcritos in vitro del gen objetivo que fueron diluidos en ARN portador de riñón de rata a  $1 \times 10^5$  copias. Las curvas estándar para Marcadores domésticos consistieron de transcritos in vitro del gen objetivo que fueron diluidos en serie en ARN portador de riñón de rata a  $1 \times 10^7$ ,  $1 \times 10^5$  y  $1 \times 10^3$  copias. No se incluyeron tampoco controles del objetivo en cada ejecución del ensayo para asegurar la carencia de contaminación ambiental. Todas las muestras y controles se ejecutaron por duplicado. Se realizó qRTPCR con reactivos de uso de laboratorio en general en una reacción de 10  $\mu$ l que contenía: Tampón de RT-PCR (50 nM Bicina/KOH pH 8,2, 115nM KAc, 8% de glicerol, 2,5 mM de  $MgCl_2$ , 3,5 mM de  $MnSO_4$ , 0,5 mM cada uno de dCTP, dATP, dGTP y dTTP), aditivos (2 mM de Tris-Cl pH 8, 0,2 mM de albúmina bovina, 150 mM de trehalosa, 0,002% de Tween 20), Mezcla de Enzimas (2U Tth (Roche), 0,4 mg/ $\mu$ l Ab TP6-25), Cebador y mezcla de sondas (0,2  $\mu$ M de Sondas, 0,5 $\mu$ M de cebadores). Siguió los siguientes parámetros de ciclación; 1 ciclo a 95° C durante 1 minuto; 1 ciclo a 55° C durante 2 minutos; Desnivel 5%; 1 ciclo a 70° C durante 2 minutos; y 40 ciclos de 95° C durante 15 segundos; 58° C durante 30 segundos. Después de que se completó la reacción de PCR, se establecieron los valores de la línea de base y umbral en el software ABI 7900HT Prism y los valores de Ct calculados se exportaron a Microsoft Excel.

*Reacción de Un paso frente a Dos Pasos.* Para la comparación de las reacciones de RT-PCR de dos pasos con la de un paso, se llevó a cabo la síntesis de la primera cadena de la reacción de dos pasos usando 100 ng de hexámeros aleatorios o cebadores específicos del gen por reacción. En el primer paso, se calentaron 11,5  $\mu$ l de Mezcla-1 (cebadores y 1  $\mu$ g de ARN total) a 65° C durante 5 minutos y después se enfrió en hielo. Se añadieron 8,5  $\mu$ l de Mezcla-2 (1x Tampón, 0,01 mM de DTT, 0,5 mM de cada uno de los dNTP, 0,25 U/ $\mu$ l de RNasin®, 10 U/ $\mu$ l de Superscript III) a la Mezcla-1 y se incubó a 50° C durante 60 minutos seguido por 95° C durante 5 minutos. El ADNc fue almacenado a -20° C hasta que estuvo listo para el uso, el qRTPCR para el segundo paso de la reacción de dos pasos se realizó como se ha indicado anteriormente con los siguientes parámetros de ciclación: 1 ciclo a 95° C durante 1 minuto; 40 ciclos de 95° C durante 15 segundos, 58° C durante 30 segundos. El qRTPCR para la reacción de un paso se realizó exactamente como se ha señalado en el párrafo anterior. Las reacciones de tanto un paso como la de dos pasos se realizaron en 100 ng de plantilla (ARN/ADNc). Después de que se completó la reacción de PCR se establecieron los valores de línea de base y umbral en el software ABI 7900HT Prism y los valores de Ct calculados se exportaron a Microsoft Excel.

*Desarrollo del Algoritmo.* se construyeron discriminadores lineales usando la función 'idea' de la biblioteca MASS (Venables y Ripley) en el lenguaje R (versión 2.1.1). El modelo usado es dependiente del tejido del que se extrajo la metástasis, así como del género del paciente. Cuando se encuentra un sitio de metástasis de pulmón, colon u ovario, el antecedente de clase es establecido a cero para la clase que es equivalente al sitio de metástasis. Además, las probabilidades anteriores son establecidas a cero para la clase de mama y ovario en pacientes masculinos, mientras que en pacientes femeninos, el antecedente de la clase de próstata es establecido a cero. Todas las otras probabilidades anteriores usadas en los modelos son equivalentes. Además, la clasificación para cada muestra se basa en la probabilidad posterior más alta determinada por el modelo para cada clase. Para estimar el rendimiento del modelo, se realizó una validación cruzada dejando uno fuera. Además de esto, los conjuntos de datos fueron divididos aleatoriamente en mitades., conservando la relación proporcional entre cada clase, en conjuntos de aprendizaje y prueba. Esta división aleatorio se repitió tres veces.

**Resultados.** El objetivo de este estudio era desarrollar un ensayo qRTPCR para predecir el tejido de origen de carcinoma metastásico. El trabajo experimental consistió de dos partes principales. La primera parte incluía la nominación de candidatos a Marcador específicos del tejido, su validación en tejidos de carcinoma metastásico de FFPE, y la selección de diez Marcadores para el ensayo (Figura 9A). La segunda parte incluía la optimización del ensayo qRTPCR seguido por la implementación del ensayo en otro conjunto de carcinomas metastásicos de FFPE, la construcción de un algoritmo de predicción, su validación cruzada y validación en un conjunto de muestras independiente. (Figura 9B).

*Características de la Muestra.* Se usó el ARN de un total de 700 muestras de tejido primario congeladas para la realización del perfil de expresión génica e identificación del gen específico del tipo de tejido. Las muestras incluían 545 carcinomas primarios (296 de pulmón, 13 de páncreas, 315 de mama, 128 colorrectales, 38 de próstata, 22 de ovario), 37 lesiones benignas (1 de pulmón, 4 colorrectales, 6 de mama, 26 de próstata) y 118 (36 de pulmón, 5 de páncreas, 36 colorrectales, 14 de mama, 3 de próstata, 24 de ovario) tejidos normales.

Se usaron en este estudio un total de 375 carcinomas metastásicos de origen conocido (Etapa III-IV) y 26 muestras de adenocarcinoma primario de próstata. Los carcinomas metastásicos se originaron a partir de pulmón, páncreas, colorrectal, de ovario, de próstata, así como otros tipos de cáncer. La categoría de muestra "otro" consistió de metástasis derivada de tejidos distintos de pulmón, páncreas, colon, mama, ovario y próstata. Las características de los pacientes se resumen en la Tabla 18.

Tabla 18

		CUP Metastásico	Conjunto de la Muestra
	Número Total	401	48
	Edad Media	57.8 ± 11*	62.13 ± 11.7
	Género		
	Femenino	241	20
	Masculino	160	28
	Tejido de Origen		
	Pulmón	65	9
	Páncreas	63	2
	Colorrectal	61	4
	Mama	63	5
	Ovario	82	2
	Próstata	27	2
	Riñón	8	8
	Estomago	7	0
	Otro**	25	5
	Carcinoma de Primario Desconocido		11
	Diagnóstico Histopatológico		
	Adenocarcinoma, moderadamente/bien diferenciado	306	27
	Adenocarcinoma, pobremente diferenciado	49	4
	carcinoma de células escamosas	16	5
	Carcinoma pobremente diferenciado	16	10
	Carcinoma de células pequeñas	3	
	Melanoma	5	
	Linfoma	3	
	Carcinoma Hepatocelular	2	
	Mesotelioma	1	
	Otro***	14	2
	Sitio de Metástasis		
	Nódulos Linfáticos	73	1
	Cerebro	17	14
	Pulmón	20	7
	Hígado	75	11
	Región pélvica (ovario, vejiga, trompas de Falopio)	53	2
	Abdomen (epiplón (epiplón, mesenterio, colon, peritoneo)	91	5
	Otro (piel, tiroides, pared torácica, ombligo)	44	8
	Desconocido	2	
	Primario (próstata)	26	

\* La edad es desconocida para 26 pacientes

\*\* esófago, vejiga, pleura, vesícula biliar del hígado, conductos biliares, laringe, faringe, linfoma no Hodgkin

\*\*\* de células pequeñas, mesotelioma, hepatocelular, melanoma, linfoma

Las muestras fueron separadas en dos conjuntos: el conjunto de validación (205 especímenes) que fue usado para validar la expresión diferencial específica del tejido de los candidatos a Marcador y el conjunto de aprendizaje (260 especímenes) que fue usado para probar el procedimiento qRT-PCR de un paso optimizado y para el aprendizaje de un algoritmo de predicción. El primer conjunto de 205 muestras incluía 25 de pulmón, 41 de páncreas, 31 colorrectales, 33 de mama, 33 de ovario, 1 de próstata, 23 otras metástasis de cáncer y 18 cánceres primarios de próstata. El segundo conjunto consistía de 260 muestras incluidos 56 de pulmón, 43 de páncreas, 30 colorrectales, 30 de mama, 49 de ovario, 32 otras metástasis de cáncer y 20 cánceres primarios de próstata. Sesenta y cuatro especímenes, incluyendo 16 de pulmón, 21 de páncreas, 15 otros metastásicos y 12 carcinomas primarios de próstata fueron del mismo paciente en ambos conjuntos.

El conjunto de muestra independiente obtenido del Albany Medical College estaba comprendido de 33 especímenes de CUP con un primario sugerido para 22 de ellos, y 15 carcinomas metastásicos de origen conocido. Para los CUPs que tenían un primario sugerido, se hizo un diagnóstico en base a características morfológicas y/o resultados de pruebas con un panel de Marcadores IHC. Las características demográficas del paciente, clínicas y de patología se presentan en la Tabla 18.

*Selección de candidatos a Marcador.* El análisis de los perfiles de expresión génica de 5 tipos de tejido primario (pulmón, colon, mama, ovario, próstata) resultó en la nominación de 13 candidatos a Marcador específicos del tejido para las pruebas qRT-PCR. Los candidatos principales han sido identificados en estudios previos de cánceres in situ. Argani et al. (2001); Backus et al. (2005); Cunha et al. (2005); Borgono et al. (2004); McCarthy et al. (2003); Hwang et al. (2004); Fleming et al. (2000); Nakamura et al. (2002); y Khoor et al. (1997). Además del análisis de los datos de micromatrices, se seleccionaron dos Marcadores de la bibliografía, incluyendo un DSG3 Marcador de carcinoma de células escamosas de pulmón complementario y el Marcador de mama PDEF. Backus et al. (2005). Los datos de las micromatrices confirmaron la alta sensibilidad y especificidad de estos Marcadores.

Se usó un enfoque especial para identificar los Marcadores específicos del páncreas. Primero, se analizaron cinco candidatos a Marcador de páncreas: antígeno de células madre de próstata (PSCA), inhibidor de proteinasa serina, miembro de clade A 1 (SERPINA1), citoqueratina 7 (KRT7), metaloproteasas de matriz 11 (MMP11), y mucina 4 (MUC4) (Varadhachary et al. (2004); Argani et al. (2001); Jones et al. (2004); Prasad et al. (2005) y Moniaux et al. (2004)) usando micromatrices de ADN y un panel de 13 adenocarcinomas ductales de páncreas, cinco tejidos de páncreas normal y 98 muestras de tumores de mama, colorrectales, pulmón, y ovario. Sólo el PSCA demostró sensibilidad moderada (se detectaron seis de 13 o el 46% de los tumores de páncreas) a una alta especificidad (94 de 98 o el 93% fueron correctamente identificados como no siendo de origen pancreático). Por el contrario, KRT7, SERPINA1, MMP11, y MUC4 demostraron sensibilidades del 38%, 31%, 85% y 31%, respectivamente, a especificidades del 66%, 91%, 82% y 81%, respectivamente. Estos datos se encontraban en concordancia con el qRT-PCR realizado en 27 metástasis de origen pancreático y 39 metástasis de origen no pancreático para todos los Marcadores excepto el MMP11 que mostró sensibilidad y especificidad más pobres con el qRT-PCR que las metástasis. En conclusión los datos de las micromatriz en tejido primario congelado rápidamente sirve como un buen indicador de la capacidad del Marcador para identificar una metástasis de FFPE como pancreática en origen usando qRT-PCR pero esos Marcadores adicionales pueden ser útiles para un rendimiento óptimo.

El adenocarcinoma ductal pancreático se desarrolla de células epiteliales ductales que comprende sólo un pequeño porcentaje de todas las células pancreáticas (con células acinares y células de los islotes que comprenden la mayoría) en el páncreas normal. Además, los tejidos de adenocarcinoma de páncreas contienen una cantidad significativa de tejido normal adyacente (Prasad et al. (2005) y Ishikawa et al. (2005)). Debido a esto los Marcadores de páncreas candidatos han sido enriquecidos para genes elevados en adenocarcinoma de páncreas en relación a las células de páncreas normales. El primer método de consulta devolvió seis conjuntos de sonda: factor de coagulación V (F5), una proteína hipotética FLJ22041 similar a las proteínas de unión a FK506 (FKBP10), integrina beta 6 (ITGB6), transglutaminasa 2 (TGM2), ribonucleoproteína heterogénea nuclear A0 (HNRPO), y BAX delta (BAX). El segundo método de consulta (ver Materiales y Métodos para los detalles) devolvió ocho conjuntos de sonda: F5, TGM2, factor de transcripción de homeodominio tipo emparejado 1 (PITX1), isoforma trío ARNm (TRIO), ARNm para p73H (P73), una proteína desconocida para MGC: 10264 (SCD), y dos conjuntos de sonda para claudina18.

Se seleccionaron un total de 23 candidatos a Marcador específicos del tejido para validación por RT-PCR adicional en tejidos de FFPE de carcinoma metastásicos por qRT-PCR. Los candidatos a Marcador se probaron en 205 carcinomas metastásicos de FFPE, de pulmón, páncreas, colon, mama, ovario, próstata y carcinomas primarios de próstata. La Tabla 19 proporciona los símbolos génicos de los Marcadores específicos de tejido seleccionados de la validación por RT-PCR y también resume los resultados de las pruebas realizadas con estos Marcadores.

Tabla 19

5	Tipo de Tejido	SEQ ID NOs	Método de ID		Filtros de selección de Marcador			Marcador adecuado?
			Micro matriz	Lit	Baja expresión en Tejido met correspondiente	Redundancia del Marcador	Reactividad cruzada del Tejido	
10	Pulmón	1/59	X	X				X
		60	X	X				X
		61		X		X		X
15	Páncreas	66		X				X
		67	X					X
		71	X			X		
		72	X		X			
20		73		X				
		74		X				
		75		X				
25		76		X				
	Colon	4/85	X	X				X
		77	X	X				
30		78	X	X		X		
		79	X	X		X		
35	Próstata	9/86	X	X				X
		80	X	X		X		
	Mama	63	X	X				X
		81	X	X			X	
40		64		X				X
	Ovario	82	X	X			X	
		83	X	X			X	
45		65	X	X				X

De los 23 marcadores probados, trece fueron rechazados en base a su reactividad cruzada, bajo nivel de expresión en los tejidos metastásicos correspondientes, o redundancia. Diez Marcadores se seleccionaron para la versión final del ensayo. Los Marcadores de pulmón eran proteína B asociada al pulmón surfactante humana (HUMPSPBA), factor de transcripción tiroideo 1 (TTF1) y desmogleína 3 (DSG3). Los Marcadores de páncreas eran antígeno de células madre de próstata (PSCA) y factor de coagulación V (F5), y el Marcador de próstata era calicleína 3 (KLK3). El Marcador colorrectal era cadherina 17 (CDH17) Lo Marcadores de mama eran mamaglobina (MG) y factor de transcripción ets derivado de próstata (PDEF). El Marcador de ovario era tumor de Wilms 1 (WT1). Los valores de expresión relativos normalizados medios de los Marcadores seleccionados en diferentes tejidos metastásicos se presentan en la Figura 10.

*Optimización de la preparación de la muestra y qRT-PCR usando tejidos de FFPE.* Después los métodos de aislamiento de ARN y qRT-PCR se optimizaron usando tejidos fijados antes de examinar el rendimiento del panel de Marcadores. Primero se analizó el efecto de reducir el tiempo de incubación de la proteinasa K de dieciséis horas a 3 horas. No hubo efecto en el rendimiento. Sin embargo, algunas muestras mostraron fragmentos más largos de ARN cuando su uso el paso de proteinasa K más corto (Figura 11A, B). Por ejemplo, cuando el ARN fue aislado de un bloque de un año (C22), no se observó diferencia en los electroferogramas. Sin embargo, cuando el ARN se aisló de un bloque de cinco años (C23), se observó una fracción más grande de ARNs de peso molecular más alto, según lo evaluado por la protuberancia en el desnivel, cuando se usó la digestión de proteinasa K más corta. Esta

tendencia se mantuvo generalmente cuando se procesaron otras muestras, independientemente del órgano de origen para la metástasis de FFPE. En conclusión, acortar el tiempo de digestión de la proteinasa K no sacrifica rendimientos de ARN y puede ayudar a aislar ARN menos degradado, más largo.

5 Después se compararon tres métodos diferentes de transcripción inversa: transcripción inversa con hexámeros aleatorios seguida por qPCR (dos pasos), transcripción inversa con un cebador específico del gen seguida por qPCR (dos pasos) y una qRTPCR de un paso usando cebadores específicos de los genes. El ARN se aisló de once metástasis y se comparó con los valores de Ct a través de tres métodos con  $\beta$ -actina, HUMSPB (Figuras 11C, D) y TTF. Los resultados mostraron diferencias estadísticamente significativas ( $p < 0,001$ ) para todas las comparaciones. Para ambos genes, la transcripción inversa con hexámeros aleatorios seguida por qPCR (reacción de dos pasos) dio los valores de Ct más altos mientras que la transcripción inversa con un cebador específico del gen seguida por qPCR (reacción de dos pasos) dio valores de Ct ligeramente más bajos (pero estadísticamente significativos) que la reacción de un paso correspondiente. Sin embargo, la RTPCR de dos pasos con cebadores específicos de los genes tuvo un paso de transcripción inversa más largo. Cuando los valores de Ct de HUMSPB se normalizaron al valor de  $\beta$ -actina correspondiente para cada muestra, no hubo diferencias en los valores de Ct normalizados a través de los tres métodos. En conclusión, la optimización de las condiciones de reacción de RTPCR pueden generar valores de Ct más bajos, lo que ayuda a analizar bloques de parafina más viejos (Cronin et al (2004)), y una reacción de RTPCR de un paso con cebadores específicos de los genes puede generar valores de Ct comparables a los generados en la reacción de dos pasos correspondiente.

20 *Rendimiento de diagnóstico de un ensayo qRTPCR de CUP.* Después se realizaron 12 reacciones de qRTPCR (10 Marcadores y dos genes domésticos) en un nuevo conjunto de 260 metástasis de FFPE. Veintiún muestras dieron valores de Ct altos para los genes domésticos por lo que sólo se usaron 238 en el análisis de mapa de calor. El análisis de valores de Ct normalizados en un mapa de calor reveló la alta especificidad de los Marcadores de mama y próstata, especificidad moderada de colon, pulmón y ovario, y especificidad algo más baja de los Marcadores de páncreas (Figura 12). Combinar los datos de qRTPCR normalizados con el refinamiento computacional mejora el rendimiento del panel de Marcadores.

30 Usando valores de expresión, normalizados a medios de la expresión de dos genes domésticos, se desarrolló un algoritmo para predecir el tejido de origen de la metástasis combinando los datos de qRTPCR normalizados con el algoritmo y se determinó la precisión del ensayo de qRTPCR realizando un test de validación cruzada dejando uno fuera (LOOCV). Para los seis tipos de tejido incluidos en el ensayo, se estimó separadamente tanto el número de designaciones falso-positivas, cuando se predijo una muestra erróneamente como otro tipo de tumor incluido en el ensayo (páncreas como colon, por ejemplo, como el número de veces que una muestra no fue predicha como de los incluidos en los tipos de tejido del ensayo (otros). Los resultados del LOOCV se presentan en la Tabla 20.

Tabla 20  
Tejido de Origen

Predicción	Mama	Colon	Pulmón	Ovario	Páncreas	Próstata	Otro	Total
Mama	22	0	2	1	1	0	0	
Colon	1	27	3	2	4	0	4	
Pulmón	1	2	45	2	3	0	5	
Otro	1	1	3	1	4	0	16	
Ovario	5	0	0	43	0	0	1	
Páncreas	0	0	3	0	31	0	6	
Próstata	0	0	0	0	0	20	0	
Total	30	30	56	49	43	20	32	260
# Correctos	22	27	45	43	31	20	16	204
Precisión	72.3	90.0	87.8	87.8	72.1	100.0	50.0	78.5

60 El tejido de origen se predijo correctamente para 204 de las 260 muestras probadas con una precisión total del 78%. Una proporción significativa de las designaciones de falso positivo se debieron a la reactividad cruzada de los Marcadores en tejidos histológicamente similares. Por ejemplo, tres carcinomas metastásicos de células escamosas originados de faringe, laringe y esófago fueron previstos erróneamente como de pulmón debido a la

expresión de DSG3 en estos tejidos. La expresión positiva de CDH17 en carcinomas GI distintos de colon, incluyendo estómago y páncreas, causo la clasificación falsa de 4 de las 6 de estomago probadas y de 3 de las 43 metástasis de cáncer de páncreas probadas como de colon.

5 Además de un test LOOCV, los datos fueron divididos aleatoriamente en 3 pares separados de conjuntos de aprendizaje y prueba. Cada división contenía aproximadamente un 50% de las muestras para cada clase. A 50/50 divisiones en tres pares separados de conjuntos de aprendizaje y prueba, las precisiones de clasificación totales del ensayo fueron del 77%, 71% y 75%, confirmando la estabilidad del rendimiento del ensayo.

10 Por último, se probó otro conjunto independiente de 48 carcinomas metastásicos de FFPE que incluía carcinomas metastásicos de primario conocido, especímenes de CUP con un diagnóstico del tejido de origen suministrado por evaluación patológica incluyendo IHC, y especímenes de CUP que permanecieron CUP después de las pruebas IHC. La precisión del predicción del tejido de origen fue estimada separadamente para cada categoría de muestras. La Tabla 21 resume los resultados del ensayo.

Tabla 21

	Probados	correctos	Precisión
Mets conocidas	15	11	73.3
CUP Resueltos	22	17	77.3
CUP sin resolver	11		

25 La predicción del tejido de origen fue, con solo unas pocas excepciones, consistente con el primario conocido o diagnóstico del tejido de origen evaluado por evaluación clínica/patológica incluyendo IHC. De manera similar al conjunto de aprendizaje, el ensayo no fue capaz de diferenciar carcinomas de células escamosas originados de fuentes diferentes y los predijo falsamente como de pulmón.

30 El ensayo también hizo diagnósticos de tejido de origen putativo para ocho de las once muestras que permanecieron como CUP después de la prueba de diagnóstico estándar. Uno de los casos de CUP fue especialmente interesante. Un paciente masculino con una historia de cáncer de próstata fue diagnosticado con carcinoma metastásico en el pulmón y la pleura. Las pruebas de PSA en suero e IHC con anticuerpos de PAS en tejido metastásico fueron negativas, por lo que el diagnóstico del patólogo fue CUP con una inclinación hacia tumores gastrointestinales. El ensayo predijo fuertemente (probabilidad posterior de 0,99) el tejido de origen como colon. **Discusión:** En este estudio, la realización de perfiles de expresión basados en micromatrices en tumores primarios se usó para identificar candidatos a Marcador para su uso con metástasis. El hecho de que los tumores primarios puedan ser usado para descubrir Marcadores de tumores de origen para metástasis es consistente con varios descubrimientos recientes. Por ejemplo, Weigelt y sus colaboradores han demostrado que los perfiles de expresión génica de tumores de mama primarios se mantienen en metástasis distantes. Weigelt et al. (2003). Backus y sus colaboradores identificaron Marcadores putativos para detectar metástasis de cáncer de mama usando un análisis de expresión génica de genoma amplio de la mama y otros tejidos y demostraron que la mamaglobina y CK19 detectaron metástasis clínicamente procesable en ganglios linfáticos centinelas de mama con un 90% de sensibilidad y un 94% de especificidad. Backus et al. (2005).

45 Durante el desarrollo del ensayo, la selección se enfocó en seis tipos de cáncer, incluyendo pulmón, páncreas y colon que están entre los más prevalentes en CUP (Ghosh et al. (2005); y Pavlidis et al. (2005)) y mama, ovario y próstata para los que el tratamiento podría ser más beneficioso para los pacientes. Ghosh et al. (2005). Sin embargo, los tipos de tejidos y Marcadores adicionales pueden ser añadidos al panel siempre que la precisión total del ensayo no se comprometa y, si es aplicable, las lógicas de las reacciones de RTPCR no se entorpezcan.

50 Los estudios basados en micromatrices con tejido primario confirmaron la especificidad y sensibilidad de los Marcadores conocidos. Como resultado, la mayoría de los Marcadores específicos de tejido tienen alta especificidad para los tejidos aquí estudiados. Un estudio reciente descubrió que, usando IHC, el PSCA se sobreexpresa en metástasis de cáncer de próstata. Lam et al. (2005). Dennis et al. (2002) también demostraron que el PSCA podría ser usado como un Marcador de tumor de origen para el páncreas y próstata. La fuerte expresión del PSCA en algunos tejidos de próstata a nivel de ARN estaba presente pero, debido a la inclusión de PSA en el ensayo, los cánceres de próstata y páncreas pueden ser ahora segregados. Un nuevo descubrimiento de este estudio fue el uso de F5 como un Marcador complementario (al PSCA) para tejido de origen pancreático. En tanto el conjunto de datos de micromatrices con tejido primario como el conjunto de datos de qRTPCR con metástasis de FFPE, se descubrió que el F5 complementa al PSCA.

65 Investigadores anteriores han generado ensayos de CUP usando IHC (Brown et al. (1997); DeYoung et al. (2000); y Dennis et al. (2005a)) o micromatrices. Su et al. (2001); Ramaswamy et al. (2001); y Bloom et al. (2004). Más recientemente, el SAGE ha sido acoplado a un panel de Marcadores de qRTPCR pequeño. Dennis et al.

(2002); y Buckhaults et al. (2003). Este estudio es el primero en combinar realización de perfiles de expresión basados en micromatrices con un panel pequeño de ensayos de qRTPCR. Los estudios de micromatrices con tejido primario identificaron algunos, pero no todos, de los mismos Marcadores de tejido de origen que los identificados previamente por estudios SAGE. Este descubrimiento no es sorprendente dados los estudios que han demostrado que existe una concordancia modesta entre los datos de realización de perfiles basados en micromatrices de ADN y SAGE y que la correlación mejora para los genes con niveles de expresión más altos. van Ruissen et al. (2005); y Kim et al. (2003). Por ejemplo, Dennis y sus colaboradores identificaron el PSA, MG, PSCA y HUMSPC mientras que Buckhaults y sus colaboradores (Buckhaults et al. (2003)) identificaron el PDEF.

La ejecución de un ensayo CUP es preferible por qRTPCR porque es una tecnología robusta y puede tener ventajas de rendimiento sobre el IHC. Al-Mulla et al. (2005); y Haas et al. (2005). Además, como se demuestra en la presente, el protocolo qRTPCR ha sido mejorado a través del uso de cebadores específicos de genes en una reacción de un paso. Esta es la primera demostración del uso de cebadores específicos de genes en una reacción de qRTPCR de un paso con tejido de FPPE. Otros investigadores han hecho o una qRTPCR de dos pasos (síntesis de ADNc en una reacción seguida por qPCR) o han usado hexámeros aleatorios o cebadores específicos de genes truncados. Abrahamsen et al. (2003); Specht et al. (2001); Godfrey et al. (2000); Cronin et al.(2004); y Mikhitarian et al. (2004).

En resumen, la precisión total del 78% del ensayo para seis tipos de tejido se compara favorablemente con otros estudios. Brown et al. (1997); DeYoung et al. (2000); Dennis et al. (2005a); Su et al. (2001); Ramaswamy et al. (2001); y Bloom et al. (2004).

**Ejemplo 7**

En este estudio se construyeron clasificadores usando carpetas de marcadores génicos eligiéndolos de MVO y usando este clasificador para predecir el tejido de origen y el estado del cáncer para cinco tipos de cáncer principales incluyendo mama, colon, pulmón, ovario y próstata. Se analizaron trescientos setenta y ocho cánceres primarios, 23 lesiones epiteliales proliferativas benignas y 103 especímenes de tejido humano congelado rápidamente normales para sustraer la expresión génica enmascarada potencialmente por la co-expresión en células de fondo de leucocitos. Se desarrolló un método bioinformático basado en OMV nuevo para seleccionar las carpetas de marcadores génicos para tejido de origen y estado del cáncer. Los datos demostraron que un panel de 26 genes podría ser usado como un clasificador para predecir con precisión el tejido de origen y estado del cáncer entre los 5 tipos de cáncer. Por lo tanto se puede obtener un método de clasificación multi-cáncer determinando los perfiles de expresión génica de un número razonablemente pequeño de marcadores génicos. La Tabla 22 muestra los Marcadores identificados para los orígenes de tejido indicados. Para las descripciones génicas ver la Tabla 31.

Tabla 22

Tejido	SEQ ID NO:	Nombre
Pulmón	59	SP-B
	60	TTF1
	61	DSG3
Páncreas	66	PSCA
	67	F5
	71	ITGB6



(continuada)

5

10

15

20

25

Tejido	SEQ ID NO:	Nombre
	72	TGM2
	84	HNRPA0
Colon	85	HPT1
	77	FABP1
	78	CDX1
	79	GUCY2C
Próstata	86	PSA
	80	hKLN2
Mama	63	MGB1
	81	PIP
	64	PDEF
Ovario	82	HE4
	83	PAX8
	65	WT1

30

El conjunto de la muestra incluía un total de 299 muestras de carcinomas de colon, mama, páncreas, ovario, próstata, pulmón y otros metastásicos y cáncer primario de próstata. Se implementó el QC basado en evaluación histológica, el rendimiento de ARN y la expresión de beta-actina del gen de control. Otra categoría de muestras incluía metástasis originadas de carcinomas de estomago (5), riñón (6), colangio/vesícula biliar (4), hígado (2), cabeza y cuello (4), íleon (1) y un mesotelioma. La Tabla 23 resume los resultados.

35

Tabla 23

Tipo de tejido	Recogido	QC Histología	QC Aislamiento ARN	QC Punto de corte ACTB
Pulmón	41	37	36	25
Páncreas	6	57	49	41
Colon	4	42	42	31
Mama	40	35	35	34
Ovario	37	36	35	33
Próstata	2	27	25	19
Otros	4	34	29	23
Total	299	268	251	205

40

45

50

Probar las muestras anteriores resultó en la reducción del conjunto de Marcadores a los de la Tabla 24 con los resultados observados en la Tabla 25

55

Tabla 24

Tabla de Marcadores Finales		
Pulmón	Proteína asociada a surfactante	SP-B
	factor de transcripción tiroideo 1	TTF1
	desmogleína 3	DSG3
Páncreas	antígeno de células madre de próstata	PSCA

60

65

(continuada)

Tabla de Marcadores Finales		
	factor de coagulación 5	F5
Colon	Transportador asociado a péptido intestinal	HPT1
Próstata	prostate-specific antigen	PSA
Mama	mamaglobina	MGB
	Factor de transcripción Ets	PDEF
Ovario	tumor de Wilms 1	WT1

Tabla 25

Cáncer	Muestras#	Marcador	Correctas	% de Sens.	Incorrectas	% de Espec
Pulmón	25/180	SP-B	13/25	52	0/180	100
		TTF	12/25	48	1/180	99
		DSG3	5/25	20	0/180	100
Páncreas	41/164	PSCA	24/41	59	6/164	96
		F5	6/41	15	4/164	98
Colon	31/174	HPT1	22/31	71	2/174	99
Mama	33/172	MGB	23/33	70	3/172	98
		PDEF	16/3	48	1/172	99
Próstata	19/186	PSA	19/19	100	0/186	100
		PDEF	19/19	100	2/186	99
Ovario	33/172	WT1	24/33	71	1/172	99
Total	205					

Los resultados mostraron que de 205 tumores metastásicos embebidos en parafina; 166 muestras (81%) tenían resultados de ensayo conclusivos, Tabla 26.

Tabla 26

	Candidato	Correctos	Incorrectos	Nº	Precisión (%)
Pulmón	SP-B + TTF+DSG3	19	0	6	76
Páncreas	PSCA+F5	27	1	13	66
Colon	HPT1	24	2	5	78
Próstata	PSA	19	0	0	100
Mama	MGB + PDEF	23	3	7	70
Ovario	WT1	23	2	8	70
Otros		20	3		87
Total		155	11	39	76

De los resultados de falso positivo, muchos falsos se derivaron de tejidos histológicamente y embriológicamente similares, Tabla 27.

Tabla 27

ID de la muestra	Diagnóstico	Previsto
OV_26	Ovario	Mama
Br_24	Mama	Colon
Br_37	Mama	Colon
CRC_25	Colon	Ovario
Pn_59	Páncreas	Colon
Cont_27	Estomago	páncreas
Cont_34	Estomago	Colon
Cont_35	Estomago	Colon
Cont_43	Conducto biliar	Páncreas
Cont_44	Conducto biliar	Páncreas
Cong_25	Hígado	páncreas

Se consideraron los siguientes parámetros para el desarrollo del modelo:

Separar los Marcadores en conjuntos femeninos y masculinos y calcular la probabilidad de CUP de forma separada para pacientes masculinos y femeninos. El conjunto masculino incluía: SP\_B, TTF1, DSG3, PSCA, F5, PSA, HPT1; el conjunto femenino incluía: SP\_B, TTF1, DSG3, PSCA, F5, HPT1, MGB, PDEF, WT1. La expresión de fondo se excluyó de los resultados del ensayo: Pulmón: SP\_B, TTF1, DSG3; Ovario: WT1; y Colon: HPT1.

El modelo CUP se ajusto a la prevalencia de CUP (5). pulmón 23, páncreas 16, colorrectal 9, mama 3, ovario 4, próstata 2, otros 43. La prevalencia de mama y ovario ajustada a 0% para pacientes masculinos, y la de próstata ajustada a 0% para pacientes femeninos.

Se tomaron los siguientes pasos: colocar marcadores en escala similar; reducir el número de variables de 12 a 8 seleccionando el valor mínimo de cada conjunto específico de tejido; dejar fuera 1 muestra; construir el modelo de las muestras restantes; probar la muestra dejada fuera; repetir hasta que el 100% de las muestras estén probadas dejas fuera aleatoriamente ~50% de las muestras (~50% por tejido); construir el modelo de las muestras restantes; probar el ~50% de las muestras y repetir para 3 divisiones aleatorias diferentes.

La precisión de la clasificación se ajusto a la prevalencia de tipos de cáncer para producir los resultados resumidos en la Tabla 28 con los datos brutos mostrados en la Tabla 29.

ES 2 525 545 T3

Tabla 28

	Mama	Colon	Pulmón	Otros	Ovario	Panc	próstata	Total	Ajustado
5	Correcto	23	29	22	19	24	35	19	171
	No Probado	3	2	2		2	3	0	12
	Incorrecto	7	0	1	4	7	3	0	22
10	Prevalencia	0.03	0.09	0.23	0.43	0.04	0.16	0.02	
	Probado/total%	91	94	92	100	94	93	100	94
15	Correcto/total%	70	94	88	83	73	85	100	89
	No Probado%	9	6	8	n/a	6	7	0	6
20	Correcto	23	25	19	20	20	24	19	150
	No Probado%	7	6	5		10	15	0	43
	Incorrecto	3	0	1	3	3	2	0	12
25	Prevalencia	0.03	0/09	0.23	0.43	0.04	0.16	0.02	
	Probado/total%	79	81	80	100	70	63	100	79
	Correcto/total%	70	81	76	87	61	59	100	73
30	Correcto/probado %	88	100	95	87	87	92	100	93
	No Probado%	21	19	20	n/a	30	37	0	21

35

40

45

50

55

60

65

Tabla 29  
Muestra

ID	Género	Origen	BK	Predicción	BACTIN	PBGD	Media	CDH17	DSG3	F5	HUMP	KLK3	MG	PDEF	PSCA	TTF1	WT1
128	f	mama	pulmón	otro	23.37	30.04	26.71	40.00	37.78	35.74	22.19	40.00	40.00	30.36	29.96	29.39	34.85
134	f	mama	uk	mama	19.60	27.00	23.30	40.00	31.27	30.83	40.00	40.00	29.51	25.07	24.67	40.00	34.13
166	f	mama	uk	mama	23.47	27.95	25.71	40.00	40.00	26.66	40.00	28.20	24.78	25.19	30.69	40.00	35.32
331	f	mama	ovario	mama	25.12	31.40	28.26	40.00	40.00	40.00	40.00	40.00	22.26	26.01	40.00	40.00	40.00
356	f	mama	uk	mama	28.59	33.89	31.24	40.00	34.01	40.00	40.00	40.00	35.73	33.19	30.72	40.00	40.00
163	f	colon	uk	colon	24.69	30.34	27.52	29.39	40.00	26.52	40.00	40.00	40.00	37.72	40.00	40.00	36.17
184	m	colon	uk	colon	22.47	28.63	25.55	26.22	33.26	28.76	40.00	40.00	40.00	34.07	33.44	40.00	31.64
339	f	colon	uk	colon	28.35	34.29	31.32	33.76	40.00	40.00	40.00	40.00	40.00	35.99	40.00	40.00	40.00
346	m	colon	pulmón	colon	23.15	28.77	25.96	26.36	40.00	32.64	20.89	40.00	40.00	32.47	40.00	26.75	30.58
363	m	colon	uk	colon	24.46	30.62	27.54	26.20	31.84	29.98	34.44	40.00	40.00	30.45	35.00	40.00	30.35
101	m	pulmón	uk	pulmón	24.68	28.79	26.74	40.00	40.00	39.34	21.57	40.00	40.00	28.21	27.47	40.00	35.76
106	m	pulmón	uk	pulmón	22.05	27.50	24.78	40.00	40.00	32.24	23.68	40.00	40.00	25.79	25.02	26.42	37.27
110	m	pulmón	uk	pulmón	29.19	32.32	30.76	40.00	40.00	40.00	21.21	40.00	40.00	32.77	32.43	30.70	36.13
112	m	pulmón	uk	otro	22.48	27.79	25.14	40.00	37.05	37.38	36.08	40.00	40.00	37.12	36.04	40.00	37.45
199	f	pulmón	uk	pulmón	21.21	27.07	24.14	35.65	25.56	31.23	40.00	40.00	28.94	32.19	27.95	32.14	31.60
200	m	pulmón	uk	pulmón	22.16	26.94	24.55	40.00	24.53	33.69	40.00	40.00	40.00	36.67	38.34	38.61	33.55
313	m	pulmón	uk	otro	24.76	30.05	27.41	38.40	40.00	40.00	40.00	40.00	40.00	40.00	40.00	40.00	35.11
323	m	pulmón	uk	páncreas	23.82	30.24	27.03	32.43	31.82	33.81	40.00	40.00	40.00	33.60	28.12	40.00	31.87
325	m	pulmón	uk	pulmón	22.09	27.97	25.03	40.00	26.84	34.88	38.61	40.00	38.04	34.29	27.31	39.21	31.23
335	m	pulmón	uk	otro	24.89	29.73	27.31	40.00	29.62	38.00	40.00	40.00	40.00	39.23	40.00	31.12	32.12
347	m	pulmón	uk	pulmón	23.40	29.08	26.24	40.00	26.72	37.21	40.00	40.00	40.00	36.10	30.76	40.00	39.44
374	m	pulmón	uk	pulmón	22.50	28.23	25.37	40.00	40.00	38.76	21.38	40.00	37.26	26.56	38.26	24.86	36.60
385	f	pulmón	uk	pulmón	21.65	26.44	24.05	37.05	40.00	34.51	19.89	40.00	40.00	27.36	40.00	23.72	37.09
114	f	otro	pulmón	otro	24.80	30.56	27.68	40.00	40.00	28.16	21.51	40.00	40.00	35.76	37.85	28.19	37.21
129	m	otro	pulmón	otro	21.49	28.25	24.87	39.47	40.00	28.86	20.65	40.00	40.00	32.98	40.00	28.14	31.11
179	f	otro	uk	otro	23.97	30.45	27.21	40.00	40.00	29.79	40.00	40.00	40.00	40.00	40.00	40.00	32.64
194	m	otro	uk	otro	25.28	32.47	28.88	40.00	40.00	28.90	40.00	40.00	40.00	40.00	40.00	34.75	35.41

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

302	f	otro	colon	mama	25.67	31.47	28.57	34.17	40.00	40.00	40.00	40.00	30.55	32.47	40.00	38.20
305	m	otro	uk	otro	23.80	29.74	26.77	29.64	40.00	34.06	40.00	40.00	31.82	40.00	40.00	40.00
317	m	otro	uk	pancreas	25.90	30.62	28.26	40.00	40.00	27.75	40.00	40.00	31.89	33.06	40.00	35.12
333	f	otro	uk	otro	22.45	28.82	25.64	30.54	40.00	37.01	40.00	40.00	37.85	40.00	40.00	40.00
334	m	otro	uk	otro	22.14	29.20	25.67	31.79	40.00	36.27	40.00	40.00	34.69	40.00	40.00	40.00
342	f	otro	uk	pancreas	27.32	31.37	29.35	32.36	40.00	29.24	40.00	40.00	32.89	40.00	40.00	38.18
382	m	otro	uk	otro	25.04	30.22	27.63	40.00	40.00	36.13	40.00	40.00	38.30	40.00	40.00	34.91
404	m	otro	uk	otro	23.27	30.16	26.72	40.00	39.36	34.75	40.00	40.00	39.02	40.00	40.00	34.24
354	f	ovario	uk	ovario	24.62	31.54	28.08	40.00	40.00	34.90	40.00	40.00	36.62	40.00	40.00	29.71
148	f	ovario	uk	pancreas	23.55	29.88	26.72	40.00	40.00	30.60	40.00	40.00	32.12	31.76	40.00	38.59
417	f	pancreas	uk	pancreas	23.42	29.46	26.44	28.28	38.96	29.05	37.01	40.00	30.15	30.23	40.00	30.69
136	m	prostata	pulmon	prostata	22.37	26.95	24.66	40.00	40.00	29.47	23.69	21.38	40.00	24.70	24.28	30.89
407	m	prostata	pulmon	prostata	28.20	31.87	30.04	40.00	40.00	40.00	27.70	25.98	40.00	27.65	40.00	39.13
116	f	CUP	uk	pulmonSCC	21.66	27.31	24.49	28.95	27.86	31.06	40.00	40.00	33.49	29.31	40.00	38.11
123	m	CUP	pulmon	colon	27.09	30.59	28.84	27.92	36.01	40.00	40.00	40.00	40.00	40.00	40.00	36.65
157	m	CUP	uk	pancreas	26.81	31.94	29.38	40.00	40.00	26.82	40.00	40.00	36.68	40.00	40.00	40.00
177	m	CUP	uk	pancreas	25.44	31.52	28.48	40.00	40.00	27.15	40.00	40.00	39.67	40.00	40.00	34.71
306	m	CUP	uk	pulmon	23.15	28.38	25.77	37.30	40.00	34.94	19.71	40.00	30.81	40.00	25.45	39.28
360	m	CUP	uk	otro	21.14	27.43	24.29	33.97	36.98	32.72	40.00	40.00	27.75	40.00	40.00	40.00
372	f	CUP	uk	ovario	23.16	29.12	26.14	40.00	40.00	34.07	40.00	40.00	32.93	40.00	40.00	25.28
187	f	CUP	uk	colon	24.44	29.80	27.12	26.83	35.91	26.32	30.55	40.00	40.00	40.00	29.75	40.00

**Ejemplo 8**

Estudio de la firma del gen prospectivo de cáncer metastásico de CUP con sitio primario desconocido para predecir el tejido de origen

5 El objetivo específico de este estudio era determinar la capacidad de la firma de 10 genes para predecir el tejido de origen de carcinoma metastásico en pacientes con carcinoma de primario desconocido (CUP).

10 **Objetivo primario:** Confirmar la factibilidad de realizar un análisis génico de muestras de biopsia del núcleo en pacientes consecutivos con CUP.

15 **Objetivo secundario:** Correlacionar los resultados del ensayo de RT-PCR de la firma de 10 genes con el trabajo de diagnóstico hecho en el M.D. Anderson Cancer Center (MDACC). Tercer objetivo: Correlacionar la prevalencia de 6 tipos de cáncer previstos por ensayo con la prevalencia derivada de la bibliografía y la experiencia del MDACC.

20 El método descrito en la presente se usó para realizar un análisis de expresión génica de micromatrices de 700 carcinomas primarios congelados, y especímenes benignos y normales y candidatos a marcador génicos identificados, específicos para carcinomas de pulmón, páncreas, colon, mama, próstata y ovario. Los candidatos a marcador génico fueron probados por RT-PCR en 205 especímenes fijados en formalina, embebidos en parafina (FFPE) de carcinoma metastásico (Etapa III-IV) originados de pulmón, páncreas, colon, mama, ovario y próstata así como metástasis originadas de otros tipos de cáncer para el control de especificidad. Otros tipos de cáncer metastásico incluían carcinomas gástrico, de células renales, hepatocelular, colangio vesícula biliar y de cabeza y cuello. Los resultados permitieron seleccionar la firma de 10 genes que predijo el tejido de origen de carcinoma metastásico y dio una precisión total del 76%. El CV medio para mediciones repetidas en experimentos RT-PCR es del 1,5%, calculado en base a 4 puntos de datos replicados. La beta-actina (ACTB) se usó como gen doméstico y su expresión mediana fue similar en muestras metastásicas de origen diferente (CV=5,6%).

30 El objetivo específico para este estudio era validar la capacidad de la firma de 10 genes de predecir el tejido de origen de carcinoma metastásico en los pacientes CUP en comparación con el trabajo de diagnóstico exhaustivo.

Elegibilidad del Paciente

35 El paciente debe ser de al menos 18 años de edad con un estado de rendimiento ECOG de 0-2. Se aceptaron los pacientes con adenocarcinoma de diagnóstico o diagnóstico de carcinoma pobremente diferenciado. El grupo de pacientes con adenocarcinoma incluye tumores, bien, moderadamente y pobremente diferenciados.

40 Los pacientes han cumplido los criterios para el CUP: sin primario detectado después de una evaluación completa que es definida como la historia completa y examen físico, examen de laboratorio detallado, estudios de imágenes y estudios invasivos de síntomas o dirigidos al sitio. Solo se permitieron pacientes no tratados en el estudio.

45 Si un paciente ha sido tratado con quimioterapia o radiación, la participación en el estudio se permite si hay tejido anterior (al tratamiento) disponible como bloques archivados dentro de un periodo temporal de 10 años.

Los pacientes proporcionaron consentimiento/autorización escrita para participar en este estudio.

Diseño del Estudio

50 A los pacientes con diagnóstico de CUP que han experimentado una biopsia de escisión o con aguja gruesa de la lesión metastásica más accesible se les permitió en el estudio. Los pacientes con una biopsia FNA sólo no fueron elegibles. Los primeros 60 pacientes presentados consecutivos que cumplieron con los criterios de inclusión y consintieron con el estudio fueron inscritos. Si se requiere biopsia repetida en el MDACC para propósitos de diagnóstico para su tratamiento, se obtuvo tejido adicional para el estudio si el paciente dio consentimiento. 55 Todos los participantes fueron registrados en el protocolo en el institucional Protocol Data Management System (PDMS).

60 El trabajo de diagnóstico completo, incluyendo las evaluaciones clínicas y patológicas, se realizó en todos los pacientes inscritos de acuerdo con los estándares del MDACC. La parte de patología del trabajo de diagnóstico puede tener incluidos ensayos inmunohistoquímicos (IHC) con marcadores incluyendo CK-7, CK-20, TTF-1 y otros según se considere indicado por el patólogo. Esto es parte del trabajo rutinario de todos los pacientes que se presentan con CUP.

Recogida de la muestra de tejido

65

El estudio incluyó especímenes de carcinoma metastásico fijados en formalina, embebidos en parafina recogidos de pacientes con CUP.

5 Se usaron seis secciones de 10  $\mu\text{m}$  para el aislamiento del ARN, especímenes de tejido más pequeños requerirán nueve secciones de 10  $\mu\text{m}$ . El diagnóstico histopatológico y el contenido tumoral se confirmaron para cada muestra usada para el aislamiento del ARN en una sección adicional teñida con hematoxilina y eosina (HE). La muestra del tumor debería haber tenido un contenido tumoral mayor del 30% en la sección HE.

10 Los datos clínicos fueron suministrados de forma anónima a Veridex e incluyen la edad del paciente, el sexo, la histología del tumor mediante microscopía óptica, el grado tumoral (diferenciación), sitio de la metástasis, la fecha de obtención del espécimen, la descripción del trabajo de diagnóstico realizado para cada paciente individual.

#### Procesamiento del tejido y experimentos RT-PCR

15 Se extrajo el ARN total de cada muestra de tejido usando el protocolo descrito anteriormente. Sólo las muestras que produjeron más de 1  $\mu\text{m}$  de ARN total de la cantidad estándar de tejido se usaron para pruebas de RT-PCR posteriores. Las muestras con menos rendimiento de ADN se consideraron degradadas y excluidas de experimentos posteriores. Se implementó el control de integridad del ARN en base a la expresión doméstica para excluir las muestras con ARN degradado, de acuerdo con el procedimiento Veridex estándar.

20 El ensayo RT-PCR que incluye el panel de 10 genes y 1-2 genes de control se usó para el análisis de las muestras de ARN. La transcripción inversa y el ensayo PCR se completaron usando los protocolos descritos anteriormente.

25 El valor de la expresión relativa para cada gen probado presentado como  $\Delta\text{Ct}$ , que es igual al CT del gen objetivo sustraído por Ct de los genes de control, se calculó y usó para la predicción del tejido de origen.

#### Tamaño de la muestra e interpretación de datos

30 Se estudió un tamaño de la muestra limitado de 60 pacientes debido a la naturaleza exploratoria del estudio piloto. Hasta la fecha, se han probado 22 pacientes. Una de las muestras de pacientes no consiguió producir suficiente ARN para la prueba RT-PCR y 3 no consiguieron pasar el control de calidad evaluado por el RE-PCR con los genes de control. Se usaron un total de 18 pacientes para determinar la probabilidad de lesión metastásica del paciente.

35 El modelo estadístico se usó para determinar la probabilidad del tejido de origen del carcinoma metastásico de la siguientes siete categorías: pulmón, páncreas, colon, mama, próstata, ovario y sin prueba (otros). Para cada muestra, las probabilidades para cada categoría se calculan de un modelo de clasificación lineal. Los resultados del ensayo se resumen en la Tabla 30.

40 La probabilidad de una lesión metastásica del paciente (con primarios conocidos) proviniendo de un estos 6 sitios (colon, páncreas, pulmón, próstata, ovario, mama) es de alrededor del 76%. Este número está derivado de la bibliografía dada la incidencia de varios cánceres y potencial de esparcimiento y datos no publicados generados en el M.D. Anderson de registros tumorales. Para las muestras probadas, la prevalencia de los 6 sitios fue del 67% (12 de 18 muestras probadas), que es muy consistente con observaciones anteriores.

50

55

60

65



Tabla 30

Datos del paciente			Probabilidad posterior de ToO (%)							
ID	M/F	predicción	Mama	Colon	Pulmón	PulmónSCC	Otro	Ovario	Páncreas	próstata
1	M	Otro	0.00	0.00	0.81	0.00	98.68	0.00	0.51	0.00
4	F	Colon	0.00	99.70	0.00	0.00	0.09	0.20	0.01	0.00
5	M	Pulmón	0.00	33.29	52.27	0.01	13.30	0.00	1.13	0.00
6	F	Colon	0.00	99.91	0.00	0.00	0.09	0.00	0.00	0.00
2	M	Colon	0.00	93.19	0.01	0.00	2.90	0.00	3.90	0.00
10	F	Otro	0.02	2.04	0.03	0.03	61.43	1.12	35.34	0.00
16	F	Colon	0.00	48.59	0.01	1.57	47.62	0.17	2.05	0.00
22	M	PulmónSCC	0.00	8.85	0.01	71.69	11.84	0.00	7.62	0.00
23	M	Colon	0.00	99.27	0.01	0.00	0.72	0.00	0.00	0.00
24	F	Colon	0.00	90.59	0.00	0.00	2.36	0.00	7.04	0.00
26	F	Pulmón	0.00	0.00	99.93	0.00	0.06	0.00	0.01	0.00
17	M	Otro	0.00	0.07	0.02	0.09	94.06	0.00	5.77	0.00
19	F	Otro	0.02	0.11	0.04	0.22	76.36	23.24	0.01	0.00
21	F	Páncreas	0.00	6.97	0.00	0.00	2.37	8.43	82.23	0.00
27	F	Otro	0.00	0.04	0.04	0.59	99.06	0.14	0.13	0.00
11	M	Otro	0.00	0.23	0.07	0.09	99.52	0.00	0.09	0.00
32	F	Ovario	0.00	0.01	0.00	0.00	7.23	92.63	0.13	0.00
34	M	PulmónSCC	0.00	0.03	0.00	65.64	7.96	0.00	26.38	0.00
3	F	Fallo ctr	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	M	Fallo ctr	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20	F	Fallo ctr	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Aunque la invención anterior se ha descrito con cierto detalle a modo de ilustración y ejemplo para propósitos de claridad de comprensión, las descripciones y ejemplos no deben interpretarse como limitantes del ámbito de la invención.

Tabla 31

Nombre	SEQ ID NOs	Acceso	Descripción
CDH17	62	NM_004063	Cadherina 17
CDX1	78	NM_001804	Factor de transcripción de caja homeopática 1
DSG3	61/3	NM_001944	Desmogleína 3
F5	67/6	NM_000130	Factor de coagulación V
FABP1	71	NM_001443	Proteína de Unión de Ácidos grasos 1, hígado
GUCY2C	79	NM_004963	Guanilato ciclasa 2C
HE4	82	NM_006103	Marcador de carcinoma de ovario Putativo
KLK2	80	BC005196	Calicreína 2, de próstata
HNRPA0	84	NM_006805	Ribonucleoproteína nuclear heterogénea A0

ES 2 525 545 T3

(continuada)

5	HPT1	85/4	U07969	Transportador asociado a péptidos intestinal
	ITGB6	71	NM_000888	Integrina, beta 6
	KLK3	68	NM_001648	Calicreína 3
	MGB1	63/7	NM_002411	Mamaglobina 1
10	PAX8	83	BC001060	Gen de caja emparejado 8
	PBGD	70	NM_000190	Hidroximetilbilano sintasa
15	PDEF	64/8	NM_012391	Factor de transcripción Ets que contiene el dominio
	PIP	81	NM_002652	Proteína inducida por prolactina
	PSA	86/9	U17040	Precursor del antígeno específico de la próstata
20	PSCA	66/5	NM_005672	Antígeno de células madre de próstata
	SP-B	59/1	NM_198843	Proteína B asociada a surfactante pulmonar
	TGM2	72	NM_004613	Transglutaminasa 2
25	TTF1	60/2	NM_003317	Similar al factor de transcripción tiroideo 1
	WT1	65/10	NM_024426	Tumor de Wilms 1
	β-actina	69	NM_001101	β-actina
30	KRT6F	87	L42612	Isoforma de queratina 6
	p73H	88	AB010153	Proteína p53-relacionada
	SFTPC	89	NM_003018	Proteína C asociada al pulmón, surfactante
35	KLK10	90	NM_002776	Calicreína 10
	CLDN18	91	NM_016369	Claudina 18
	TR10	92	BD280579	Receptor del factor de necrosis tumoral
40	B305D	93		
	B726	94		
	GABA-pi	95	BC109105	Receptor del ácido gamma-aminobutírico A, pi
45	StAR	96	NM_01007243	regulador agudo esteroideogénico
	EMX2	97	NM_004098	homólogo de espiráculos vacío 2 (Drosophila)
	NGEP	98	AY617079	Variante larga de NGEP
50	NPY	99	NM_000905	Neuropéptido Y
	SERPINA1	100	NM_000295	inhibidor de peptidasa serpina, miembro 1 de clado A
55	KRT7	101	NM_005556	queratina 7
	MMP11	102	NM_005940	metalopeptidasa de la matriz 11 (estromalisina 3)
	MUC4	103	NM_018406	Mucina 4 asociada a la superficie celular
60	FLJ22041	104	AK025694	
	BAX	105	NM_138763	Δ de la variante de transcripción de proteínas X asociada a BCL2
	PITX1	106	NM_002653	Factor trans del homeodominio tipo emparejado 1
65	MGC:10264	107	BC005807	estearoil-CoA desaturasa (Δ-9-desaturasa)

REFERENCIAS

Publicaciones de solicitudes de patentes y patentes US

5	5242974	5700637	20030194733
	5350840	5786148	20030198970
	5384261	6004755	20030215803
	5405783	6136182	20030215835
	5412087	6218114	20030219760
10	5424186	6218122	20030219767
	5429807	6225051	20030232350
	5436327	6232073	20030235820
	5445934	6261766	20040005563
	5472672	6271002	20040009154
15	5527681	6339148	20040009489
	5529756	20010029020	20040018969
	5532128	20020055627	20040029114
	5545531	20020068288	20040076955
	5554501	20020168647	20040126808
20	5556752	20030044859	20040146862
	5561071	20030087818	20040219572
	5571639	20030104448	20040219575
	5593839	20030124128	20050037010
	5599695	20030124579	20050059008
25	5624711	20030138793	20060094035
	5658734	20030190656	

Publicaciones de patentes y patentes extranjeras

30	WO1998040403	WO2001073032	WO2004030615
	WO1998056953	WO2002046467	WO2004031412
	WO2000006589	WO2002073204	WO2004063355
	WO2000055320	WO2002101357	WO2004077060
	WO2001031342	WO2004018999	WO2005005601

35

Artículos de revistas

Abrahamsen et al. (2003) Towards quantitative mRNA analysis in paraffin-embedded tissues using real-time reverse transcriptase-polymerase chain reaction *J Mol Diag* 5:34-41

40 Al-Mulla et al. (2005) BRCA1 gene expression in breast cancer: a correlative study between real-time RT-PCR and immunohistochemistry *J Histochem Cytochem* 53:621-629

Argani et al. (2001) Discovery of new Markers of cancer through serial analysis of gene expression: prostate stem cell antigen is overexpressed in pancreatic adenocarcinoma *Cancer Res* 61:4320-4324

45 Autiero et al. (2002) Intragenic amplification and formation of extrachromosomal small circular DNA molecules from the PIP gene on chromosome 7 in primary breast carcinomas *Int J Cancer* 99:370-377

Backus et al. (2005) Identification and characterization of optimal gene expression Markers for detection of breast cancer metastasis *J Mol Diagn* 7:327-336

Bentov et al. (2003) The WT1 Wilms' tumor suppressor gene: a novel target for insulin-like growth factor-I action *Endocrinol* 144:4276-4279

50 Bera et al. (2004) NGEF, a gene encoding a membrane protein detected only in prostate cancer and normal prostate *Proc Natl Acad Sci USA* 101:3059-3064

Bibikova et al (2004) Quantitative gene expression profiling in formalin-fixed, paraffin-embedded tissues using universal bead arrays *Am j Pathol* 165:1799-1807

Bloom et al. (2004) Multi-platform, multi-site, microarray-based human tumor classification *Am J Pathol* 164:9-16

55 Borchers et al. (1997) Heart-type fatty acid binding protein - involvement in growth inhibition and differentiation *Prostaglandins Leukot Essent Fatty Acids* 57:77-84

Borgono et al. (2004) Human tissue kallikreins: physiologic roles and applications in cancer *Mol Cancer Res* 2:257-280

Brookes (1999) The essence of SNPs *Gene* 23:177-186

Brown et al. (1997) Immunohistochemical identification of tumor Markers in metastatic adenocarcinoma. A diagnostic adjunct in the determination of primary site *Am J Clin Pathol* 107:12-19

60 Buckhaults et al. (2003) Identifying tumor origin using a gene expression-based classification map *Cancer Res* 63:4144-4149

Chan et al. (1985) Human liver fatty acid binding protein cDNA and amino acid sequence. Functional and evolutionary implications *J Biol Chem* 260:2629-2632

Chen et al. (1986) Human liver fatty acid binding protein gene is located on chromosome 2 *Somat Cell Mol Genet* 12:303-306

Cheung et al. (2003) Detection of the PAX8-PPAR gamma fusion oncogene in both follicular thyroid carcinomas and adenomas *J Clin Endocrinol Metab* 88:354-357

65 Clark et al. (1999)

- The potential role for prolactin-inducible protein (PIP) as a Marker of human breast cancer micrometastasis *Br J Cancer* 81:1002-1008
- Cronin et al. (2004) Measurement of gene expression in archival paraffin-embedded tissue *Am J Pathol* 164:35-42
- 5 Cunha et al. (2006) Tissue-specificity of prostate specific antigens: Comparative analysis of transcript levels in prostate and non-prostatic tissues *Cancer Lett* 236:229-238
- Dennis et al. (2002) Identification from public data of molecular Markers of adenocarcinoma characteristic of the site of origin *Can Res* 62:5999-6005
- Dennis et al. (2005a) Hunting the primary: novel strategies for defining the origin of tumors *J Pathol* 205:236-247
- 10 Dennis et al. (2005b) Markers of adenocarcinoma characteristic of the site of origin: development of a diagnostic algorithm *Clin Can Res* 11:3766-3772
- DeYoung et al. (2000) Immunohistologic evaluation of metastatic carcinomas of unknown origin: an algorithmic approach *Semin Diagn Pathol* 17:184-193
- Di Palma et al. (2003) The paired domain-containing factor Pax8 and the homeodomain-containing factor TTF-1 directly interact and synergistically activate transcription *Biol Chem* 278:3395-3402
- 15 Dwight et al. (2003) Involvement of the PAX8 peroxisome proliferator-activated receptor gamma rearrangement in follicular thyroid tumors *J Clin Endocrinol Metab* 88:4440-4445
- Feldman et al. (2003) PDEF expression in human breast cancer is correlated with invasive potential and altered gene expression *Cancer Res* 63:4626-4631
- Fleming et al. (2000) Mammaglobin, a breast-specific gene, and its utility as a Marker for breast cancer *Ann N Y Acad Sci* 923:78-89
- 20 Fukushima et al. (2004) Characterization of gene expression in mucinous cystic neoplasms of the pancreas using oligonucleotide microarrays *Oncogene* 23:9042-9051
- Ghosh et al (2005) Management of patients with metastatic cancer of unknown primary *Curr Probl Surg* 42:12-66
- Giordano et al. (2001) Organ-specific molecular classification of primary lung, colon, and ovarian adenocarcinomas using gene expression profiles *Am J Pathol*.159:1231-1238
- 25 Glasser et al (1988) cDNA, deduced polypeptide structure and chromosomal assignment of human pulmonary surfactant proteolipid, SPL(pVal) *J Biol Chem* 263:9-12
- Godfrey et al. (2000) Quantitative mRNA expression analysis from formalin-fixed, paraffin-embedded tissues using 5' nuclease quantitative reverse transcription-polymerase chain reaction *J Mol Diag* 2:84-91
- Goldstein et al. (2002) WT1 immunoreactivity in uterine papillary serous carcinomas is different from ovarian serous carcinomas *Am J Clin Pathol* 117:541-545
- 30 Gradi et al. (1995) The human steroidogenic acute regulatory (StAR) gene is expressed in the urogenital system and encodes a mitochondrial polypeptide *Biochim Biophys Acta* 1258:228-233
- Greco et al. (2004) Carcinoma of unknown primary site: sequential treatment with paclitaxel/carboplatin/etoposide and gemcitabine/irinotecan: A Minnie Pearl cancer research network phase II trial *The Oncologist* 9:644-652
- 35 Haas et al. (2005) Combined application of RT-PCR and immunohistochemistry on paraffin embedded sentinel lymph nodes of prostate cancer patients *Pathol Res Pract* 200:763-770
- Hwang et al. (2004) Wilms tumor gene product: sensitive and contextually specific Marker of serous carcinomas of ovarian surface epithelial origin *Appl Immunohistochem Mol Morphol* 12:122-126
- 40 Ishikawa et al. (2005) Experimental trial for diagnosis of pancreatic ductal carcinoma based on gene expression profiles of pancreatic ductal cells *Cancer Sci* 96:387-393
- Italiano et al. (2005) Epidermal growth factor receptor (EGFR) status in primary colorectal tumors correlates with EGFR expression in related metastatic sites: biological and clinical implications *Ann Oncol* 16:1503-1507
- Jones et al. (2004) Comprehensive analysis of matrix metalloproteinase and tissue inhibitor expression in pancreatic cancer: increased expression of matrix metalloproteinase-7 predicts poor survival *Clin Cancer Res* 10:2832-2845
- 45 Jones et al. (2005) Thyroid transcription factor 1 expression in small cell carcinoma of the urinary bladder: an immunohistochemical profile of 44 cases *Hum Pathol* 36:718-723
- Khoor et al. (1997) Expression of surfactant protein B precursor and surfactant protein B mRNA in adenocarcinoma of the lung *Mod Pathol* 10:62-67
- 50 Kim (2003) Comparison of oligonucleotide-microarray and serial analysis of gene expression (SAGE) in transcript profiling analysis of megakaryocytes derived from CD34+ cells *Exp Mol Med* 35:460-466
- Kim et al. (2003) Steroidogenic acute regulatory protein expression in the normal human brain and intracranial tumors *Brain Res* 978:245-249
- Lam et al. (2005) Prostate stem cell antigen is overexpressed in prostate cancer metastases *Clin Can Res* 11:2591-2596
- 55 Lembersky et al. (1996) Metastases of unknown primary site *Med Clin North Am.* 80:153-171
- Lewis et al. (2001) Unlocking the archive-gene expression in paraffin-embedded tissue *J Pathol* 195:66-71
- Lipshutz et al. (1999) High density synthetic oligonucleotide arrays *Nature Genetics* 21:S20-24
- Lowe et al. (1985) Human liver fatty acid binding protein. Isolation of a full length cDNA and comparative sequence analyses of orthologous and paralogous proteins *J Biol Chem* 260:3413-3417
- 60 Ma et al. (2006) Molecular classification of human cancers using a 92-gene real-time quantitative polymerase chain reaction assay *Arch Pathol Lab med* 130:465-473
- Magklara et al. (2002) Characterization of androgen receptor and nuclear receptor co-regulator expression in human breast cancer cell lines exhibiting differential regulation of kallikreins 2 and 3 *Int J Cancer* 100:507-514
- 65 Markowitz (1952) Portfolio Selection *J Finance* 7:77-91

- Marques et al. (2002) Expression of PAX8-PPAR gamma 1 rearrangements in both follicular thyroid carcinomas and adenomas *J Clin Endocrinol Metab* 87:3947-3952
- Masuda et al. (1999) Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples *Nucl Acids Res* 27:4436-4443
- 5 McCarthy et al. (2003) Novel Markers of pancreatic adenocarcinoma in fine-needle aspiration: mesothelin and prostate stem cell antigen labeling increases accuracy in cytologically borderline cases *Appl Immunohistochem Mol Morphol* 11:238-243
- Mikhitarian et al. (2004) Enhanced detection of RNA from paraffin-embedded tissue using a panel of truncated gene-specific primers for reverse transcription *BioTechniques* 36:1-4
- 10 Mintzer et al. (2004) Cancer of unknown primary: changing approaches, a multidisciplinary case presentation from the Joan Karnell Cancer Center of Pennsylvania Hospital *The Oncologist* 9:330-338
- Moniaux et al. (2004) Multiple roles of mucins in pancreatic cancer, a lethal and challenging malignancy *Br J Cancer* 91:1633-1638
- Murphy et al. (1987) Isolation and sequencing of a cDNA clone for a prolactin-inducible protein (PIP). Regulation of PIP gene expression in the human breast cancer cell line, T-47D *J Biol Chem* 262:15236-15241
- 15 Myal et al. (1991) The prolactin-inducible protein (PIPGCDFP-15) gene: cloning, structure and regulation *J Mol Cell Endocrinol* 80:165-175
- Nakamura et al. (2002) Expression of thyroid transcription factor-1 in normal and neoplastic lung tissues *Mod Pathol* 15:1058-1067
- Noonan et al. (2001) Characterization of the homeodomain gene EMX2: sequence conservation, expression analysis, and a search for mutations in endometrial cancers *Genomics* 76:37-44
- 20 Oettgen et al. (2000) PDEF, a novel prostate epithelium-specific Ets transcription factor, interacts with the androgen receptor and activates prostate-specific antigen gene expression *J Biol Chem* 275:1216-1225
- Oji et al. (2003) Overexpression of the Wilms' tumor gene WT1 in head and neck squamous cell carcinoma *Cancer Sci* 94:523-529
- 25 Pavlidis et al. (2003) Diagnostic and therapeutic management of cancer of an unknown primary *Eur J Can* 39: 990-2005
- Pilot-Mathias et al. (1989) Structure and organization of the gene encoding human pulmonary surfactant proteolipid SP-B *DNA* 8:75-86
- Pilozzi et al. (2004) CDX1 expression is reduced in colorectal carcinoma and is associated with promoter hypermethylation *J Pathol* 204:289-295
- 30 Poleev et al. (1992) PAX8, a human paired box gene: isolation and expression in developing thyroid, kidney and Wilms' tumors *Development* 116:611-623
- Prasad et al. (2005) Gene expression profiles in pancreatic intraepithelial neoplasia reflect the effects of Hedgehog signaling on pancreatic ductal epithelial cells *Cancer Res* 65:1619-1626
- 35 Ramaswamy (2004) Translating cancer genomics into clinical oncology *N Engl J Med* 350:1814-1816
- Ramaswamy et al. (2001) Multiclass cancer diagnosis using tumor gene expression signatures *Proc Natl Acad Sci USA* 98:15149-15154
- Rauscher (1993) The WT1 Wilms tumor gene product: a developmentally regulated transcription factor in the kidney that functions as a tumor suppressor *FASEB J* 7:896-903
- 40 Reinholz et al. (2005) Evaluation of a panel of tumor Markers for molecular detection of circulating cancer cells in women with suspected breast cancer *Clin Cancer Res* 11:3722
- Schlag et al. (1994) Cancer of unknown primary site *Ann Chir Gynaecol* 83:8-12
- Senoo et al. (1998) A second p53-related protein, p73L, with high homology to p73 *Biochem Biophys Res Comm* 248:603-607
- 45 Specht et al. (2001) Quantitative gene expression analysis in microdissected archival formalin-fixed and paraffinembedded tumor tissue *Amer J Pathol* 158:419-429
- Su et al. (2001) Molecular classification of human carcinomas by use of gene expression signatures *Cancer Res* 61:7388-7393
- Takahashi et al. (1995) Cloning and characterization of multiple human genes and cDNAs encoding highly related type II keratin 6 isoforms *J Biol Chem* 270:18581-18592
- 50 Takamura et al. (2004) Reduced expression of liver-intestine cadherin is associated with progression and lymph node metastasis of human colorectal carcinoma *Cancer Lett* 212:253-259
- Tohill et al. (2005) An expression-based site of origin diagnostic method designed for clinical application to cancer of unknown origin *Can Res* 65:4031-4040
- van Ruissen et al. (2005) Evaluation of the similarity of gene expression data estimated with SAGE and Affymetrix GeneChips *BMC Genomics* 6:91
- 55 Varadhachary et al. (2004) Diagnostic strategies for unknown primary cancer *Cancer* 100:1776-1785
- Venables et al. (2002) *Modern Applied Statistics with S*. Fourth edition. Springer
- Wallace et al. (2005) Accurate Molecular detection of non-small cell lung cancer metastases in mediastinal lymph nodes sampled by endoscopic ultrasound-guided needle aspiration *Cest* 127:430-437
- 60 Wan et al. (2003) Desmosomal proteins, including desmoglein 3, serve as novel negative Markers for epidermal stem cell-containing population of keratinocytes *J Cell Sci* 116:4239-4248
- Watson et al. (1996) Mammaglobin, a mammary-specific member of the uteroglobin gene family, is overexpressed in human breast cancer *Cancer Res* 56:860-865
- 65 Watson et al. (1998) Structure and transcriptional regulation of the human mammaglobin gene, a breast cancer associated member of the uteroglobin gene family localized to chromosome 11q13 *Oncogene* 16:817-824

Weigelt et al. (2003) Gene expression profiles of primary breast tumors maintained in distant metastases Proc Natl Acad Sci USA 100:15901-15905

Zapata-Benavides et al. (2002) Downregulation of Wilms' tumor 1 protein inhibits breast cancer proliferation Biochem Biophys Res Commun 295:784-790

5

LISTADO DE SECUENCIAS

<110> Veridex, LLC

10

<120> Métodos y Materiales para Identificar el origen de un Carcinoma de origen Primario Desconocido

<130> P054566EP

15

<150> US60/718,501

<151> 2005-09-19

<150> US60/725,680

<151> 2005-10-12

20

<160> 107

<170> PatentIn version 3.3

25

<210> 1

<211> 476

<212> ADN

<213> Homo sapiens

30

<400> 1

```

gaaaaaccag ccactgcttt acaggacagg gggttgaagc tgagccccgc ctcacacca      60
cccccatgca ctcaaagatt ggattttaca gctacttgca attcaaaatt cagaagaata      120
aaaaatggga acatacagaa ctctaaaaga tagacatcag aaattgttaa gttaagcttt      180
ttcaaaaaat cagcaattcc ccagcgtagt caagggtgga cactgcacgc tctggcatga      240
tgggatggcg accgggcaag ctttcttctt cgagatgctc tgctgcttga gagctattgc      300
tttgttaaga tataaaaagg ggtttctttt tgtctttctg taaggtggac ttccagattt      360
tgattgaaag tcctaggggtg attctatttc tgctgtgatt tatctgctga aagctcagct      420
ggggttgtgc aagctagggga cccattcctg tgtaatacaa tgtctgcacc aatgct      476
    
```

45

<210> 2

<211> 493

50

<212> ADN

<213> Homo sapiens

<400> 2

55

60

65

ES 2 525 545 T3

	gtgattcaaa tgggttttcc acgctagggc ggggcacaga ttggagaggg ctctgtgctg	60
	acatggctct ggactctaaa gaccaaactt cactctgggc acactctgcc agcaaagagg	120
5	actcgttgt aaataccagg atttttttt ttttttgaag ggaggacggg agctggggag	180
	aggaaagagt cttcaacata acccacttgt cactgacaca aaggaagtgc cccctccccg	240
	gcaccctctg gccgcctagg ctcagcggcg accgccctcc gcgaaaatag tttgtttaat	300
10	gtgaacttgt agctgtaaaa cgctgtcaaa agttggacta aatgcctagt ttttagtaat	360
	ctgtacattt tgttgtaaaa agaaaaacca ctcccagtcc ccagcccttc acatttttta	420
	tgggcattga caaatctgtg tatattattt ggcagtttgg tatttgcggc gtcagtcctt	480
15	ttctgttgta act	493
	<210> 3	
20	<211> 545	
	<212> ADN	
	<213> Homo sapiens	
	<400> 3	
25	ccatcccata gaagtccagc agacaggatt tgtaagtgc cagactttgt caggaagtca	60
	aggagcttct gctttgtccg cctctgggtc tgtccagcca gctgtttcca tcctgacct	120
	tctgcagcat ggtaactatt tagtaacgga gacttactcg gcttctgggt ccctcgtgca	180
30	accttccact gcaggctttg atccacttct cacacaaaat gtgatagtga cagaaaggg	240
	gatctgtccc atttccagtg ttcttgcaa cctagctggc ccaacgcagc tacgagggtc	300
	acatactatg ctctgtacag aggatccttg ctcccgtcta atatgaccag aatgagctgg	360
35	aataccacac tgaccaaac tggatctttg gactaaagta ttcaaatag catagcaaag	420
	ctcactgtat tgggctaata atttggcact tattagcttc tctcataaac tgatcacgat	480
40	tataaattaa atgtttgggt tcatacccca aaagcaatat gttgtcactc ctaattctca	540
	agtac	545
	<210> 4	
45	<211> 284	
	<212> ADN	
	<213> Homo sapiens	
	<400> 4	
50	ctgcaccac ctacttagat atttcatgtg ctatagacat tagagagatt tttcattttt	60
	ccatgacatt tttcctctct gcaaatggct tagctacttg tgtttttccc ttttggggca	120
55	agacagactc attaaatatt ctgtacattt tttctttatc aaggagatat atcagtgttg	180
	tctcatagaa ctgcctggat tccatttatg ttttttctga ttccatcctg tgtccccttc	240
	atccttgact ctttgggtat ttcactgaat ttcaaacatt tgtc	284
60		
	<210> 5	
	<211> 394	
65	<212> ADN	
	<213> Homo sapiens	

5 <220>  
<221> misc\_feature  
<222> (58)..(58)  
<223> n es a, c, g, o t

10 <220>  
<221> misc\_feature  
<222> (95)..(95)  
<223> n es a, c, g, o t

15 <220>  
<221> misc\_feature  
<222> (99)..(99)  
<223> n es a, c, g, o t

20 <220>  
<221> misc\_feature  
<222> (119)..(119)  
<223> n es a, c, g, o t

25 <220>  
<221> misc\_feature  
<222> (123)..(123)  
<223> n es a, c, g, o t

30 <220>  
<221> misc\_feature  
<222> (130)..(130)  
<223> n es a, c, g, o t

35 <220>  
<221> misc\_feature  
<222> (151)..(151)  
<223> n es a, c, g, o t

40 <220>  
<221> misc\_feature  
<222> (155)..(155)  
<223> n es a, c, g, o t

45 <220>  
<221> misc\_feature  
<222> (161)..(161)  
<223> n es a, c, g, o t

50 <220>  
<221> misc\_feature  
<222> (212)..(212)  
<223> n es a, c, g, o t

<400> 5

55

60

65



ES 2 525 545 T3

5      ttcctgaggc acatcctaac gcaagtttga ccatgtatgt ttgcaccctt tttccccnaa      60  
       ccctgacctt cccatgggcc ttttccagga ttccnaccng gcagatcagt tttagtana      120  
       canatccgn tgcagatggc ccctccaacc nttntgttg ntgtttccat ggcccagcat      180  
       tttccaccct taaccctgtg ttcaggcact tnttcccca ggaagccttc cctgcccacc      240  
 10     ccatttatga attgagccag gtttgggccg tgggtgtccc cgcaccagc aggggacagg      300  
       caatcaggag ggcccagtaa aggctgagat gaagtggact gagtagaact ggaggacaag      360  
       agttgacgtg agttcctggg agtttccaga gatg                              394

15     <210> 6  
       <211> 470  
       <212> ADN  
       <213> Homo sapiens

20     <220>  
       <221> misc\_feature  
       <222> (61)..(61)  
       <223> n es a, c, g, o t

25     <220>  
       <221> misc\_feature  
       <222> (82)..(82)  
       <223> n es a, c, g, o t

30     <400> 6

      atcctctaca gccagatgtc acagggatac gtctactttc acttggtgct ggagaattca      60  
       naagtcaaga acatgctaag cntaaggac ccaaggtaga aagagatcaa gcagcaaagc      120  
 35     acaggttctc ctggatgaaa ttactagcac ataaagttgg gagacaccta agccaagaca      180  
       ctggttctcc ttccggaatg aggccctggg aggaccttcc tagccaagac actggttctc      240  
       cttccagaat gaggccctgg aaggaccctc ctagtgatct gttactctta aaacaaagta      300  
       actcatctaa gattttggtt gggagatggc atttggcttc tgagaaagggt agctatgaaa      360  
       taatccaaga tactgatgaa gacacagctg ttaacaattg gctgatcagc ccccagaatg      420  
 45     cctcacgtgc ttggggagaa agcaccctc ttgccaacaa gcctggaaag              470

50     <210> 7  
       <211> 396  
       <212> ADN  
       <213> Homo sapiens

55     <400> 7

60

65

ES 2 525 545 T3

	gcagcagcct caccatgaag ttgctgatgg tcctcatgct ggcggccctc tcccagcact	60
	gctacgcagg ctctggctgc cccttattgg agaatgtgat ttccaagaca atcaatccac	120
5	aagtgtctaa gactgaatac aaagaacttc ttcaagagtt catagacgac aatgccacta	180
	caaatgccat agatgaattg aaggaatggt ttcttaacca aacggatgaa actctgagca	240
	atgttgaggt gtttatgcaa ttaatatatg acagcagtct ttgtgattta ttttaacttt	300
10	ctgcaagacc tttggctcac agaactgcag ggtatggtga gaaaccaact acggattgct	360
	gcaaaccaca ctttctcttt cttatgtctt tttact	396
15	<210> 8 <211> 491 <212> ADN <213> Homo sapiens	
20	<400> 8	
	gagtggggcc cttaaactgg attcaaaaaa tgctctaaac ataggaatgg ttgaagaggt	60
	cttgcagtct tcagatgaaa ctaaactctt agaagaggca caagaatggc taaagcaatt	120
25	catccaaggg ccaccggaag taattagagc tttgaaaaaa tctgtttggt caggcagaga	180
	gctatatattg gaggaagcat tacagaacga aagagatctt ttaggaacag tttggggtag	240
	gcctgcaaat ttagaggcta ttgctaagaa aggaaaattt aataaataat tggtttttcg	300
30	tgtggatgta ctccaagtaa agctccagtg actaatatgt ataaatgta aatgatatta	360
	aatatgaaca tcagttaaaa aaaaaattct ttaaggctac tattaatag cagacttact	420
	tttaatcatt tgaaatctga actcatttac ctcatctt gccaattact cccttgggta	480
35	tttactgcgt a	491
40	<210> 9 <211> 265 <212> ADN <213> Homo sapiens	
45	<400> 9	
	tgggtgtaatt ttgtcctctc tgtgtcctgg ggaatactgg ccatgcctgg agacatatca	60
	ctcaatttct ctgaggacac agataggatg ggggtgtctgt gttatttggt gggtagagag	120
	atgaaagagg ggtgggatcc aactgagag agtggagagt gacatgtgct ggacactgct	180
50	catgaagcac tgagcagaag ctggaggcac aacgcaccag aactcacag caaggatgga	240
	gctgaaaaca taaccactc tgtcc	265
55	<210> 10 <211> 441 <212> ADN <213> Homo sapiens	
60	<400> 10	
65		

ES 2 525 545 T3

	atagatgtac atacctcctt gcacaaatgg aggggaattc attttcatca ctgggagtgt	60
	ccttagtgta taaaaccat gctggtatat ggcttcaagt tgtaaaaatg aaagtgactt	120
5	taaaagaaaa taggggatgg tccaggatct cactgataa gactgttttt aagtaactta	180
	aggacctttg ggtctacaag tatatgtgaa aaaaatgaga cttactgggt gaggaaatcc	240
	attgtttaaa gatggtcgtg tgtgtgtgtg tgtgtgtgtg tgtgttgtgt tgtgttttgt	300
10	tttttaaggg agggaattta ttatttaccg ttgcttgaaa ttactgtgta aatatatgtc	360
	tgataatgat ttgctctttg acaactaaaa ttaggactgt ataagtacta gatgcatcac	420
	tgggtgttga tcttacaaga t	441
15	<210> 11 <211> 21 <212> ADN <213> Homo sapiens	
20	<400> 11 cacagccccg accttgatg a 21	
25	<210> 12 <211> 19 <212> ADN <213> Homo sapiens	
30	<400> 12 ggtcccagag cccgtctca 19	
35	<210> 13 <211> 26 <212> ADN <213> Homo sapiens	
40	<400> 13 agctgtccag ctgcaaagga aaagcc 26	
45	<210> 14 <211> 75 <212> ADN <213> Homo sapiens	
50	<400> 14 cacagccccg acctttgatg agaactcagc tgtccagctg caaaggaaaa gccaaagtgag	60
	acgggctctg ggacc	75
55	<210> 15 <211> 17 <212> ADN <213> Homo sapiens	
60	<400> 15 ccaaccaga cccgcgc 17	
65	<210> 16 <211> 21 <212> ADN <213> Homo sapiens	
	<400> 16	

ES 2 525 545 T3

cgcccatgcc gctcatgttc a 21

5 <210> 17  
<211> 21  
<212> ADN  
<213> Homo sapiens

10 <400> 17  
cccgccatct cccgcttcat g 21

15 <210> 18  
<211> 78  
<212> ADN  
<213> Homo sapiens

<400> 18

20 ccaaccaga cccgcgcttc cccgccatct cccgcttcat gggcccggcg agcggcatga 60  
acatgagcgg catgggcg 78

25 <210> 19  
<211> 23  
<212> ADN  
<213> Homo sapiens

30 <400> 19  
gagagaagga gaagataact caa 23

35 <210> 20  
<211> 22  
<212> ADN  
<213> Homo sapiens

40 <400> 20  
actccagaga ttcgtaggt ga 22

45 <210> 21  
<211> 26  
<212> ADN  
<213> Homo sapiens

50 <400> 21  
attgccaaga ttactcaga ttacca 26

55 <210> 22  
<211> 97  
<212> ADN  
<213> Homo sapiens

60 <400> 22

65

ES 2 525 545 T3

	gcagagaagg agaagataac tcaaaaagaa acccaattgc caagattact tcagattacc	60
5	aagcaacca gaaaatcacc taccgaatct ctggagt	97
	<210> 23 <211> 21 <212> ADN <213> Homo sapiens	
10	<400> 23 tccctcggca gtggaagctt a 21	
15	<210> 24 <211> 24 <212> ADN <213> Homo sapiens	
20	<400> 24 tcctcaaact ctgtgtgctt ggta 24	
25	<210> 25 <211> 29 <212> ADN <213> Homo sapiens	
30	<400> 25 ccaaaatcaa tggactcat gcccgactg 29	
35	<210> 26 <211> 95 <212> ADN <213> Homo sapiens	
	<400> 26	
	tccctcggca gtggaagctt acaaaacgac tgggaagttt ccaaaaatcaa tggactcat	60
40	gcccgactgt ctaccaggca cacagagttt gagga	95
	<210> 27 <211> 21 <212> ADN <213> Homo sapiens	
45	<400> 27 agttgctgat gtcctcatg c 21	
50	<210> 28 <211> 24 <212> ADN <213> Homo sapiens	
55	<400> 28 cactgtgga ttgattgtct tggga 24	
60	<210> 29	
65		

ES 2 525 545 T3

<211> 23  
 <212> ADN  
 <213> Homo sapiens  
 5 <400> 29  
 ccctctcca gcactgctac gca 23  
 <210> 30  
 <211> 107  
 10 <212> ADN  
 <213> Homo sapiens  
 <400> 30  
 15 agttgctgat ggtcctcatg ctggcggccc tctcccagca ctgctacgca ggctctggct 60  
 gcccttatt ggagaatgtg atttccaaga caatcaatcc acaagtg 107  
 <210> 31  
 <211> 20  
 <212> ADN  
 <213> Homo sapiens  
 20 <400> 31  
 cgcccactg gacatctgga 20  
 <210> 32  
 <211> 23  
 <212> ADN  
 30 <213> Homo sapiens  
 <400> 32  
 cactggtcga ggcacagtag tga 23  
 35 <210> 33  
 <211> 25  
 <212> ADN  
 <213> Homo sapiens  
 40 <400> 33  
 gtcagcggcc tggatgaaag agcgg 25  
 <210> 34  
 <211> 86  
 45 <212> ADN  
 <213> Homo sapiens  
 <400> 34  
 50 cgcccacctg gacatctgga agtcagcggc ctggatgaaa gagcggactt cacctggggc 60  
 gattcactac tgtgcctcga ccagtg 86  
 <210> 35  
 <211> 23  
 <212> ADN  
 <213> Homo sapiens  
 55  
 60  
 65

ES 2 525 545 T3

	<400> 35 gcgagccca atacagaata cac 23	
5	<210> 36 <211> 19 <212> ADN <213> Homo sapiens	
10	<400> 36 cggggtact ccaggcaca 19	
15	<210> 37 <211> 25 <212> ADN <213> Homo sapiens	
20	<400> 37 tcagggcat tcaggatgtg cgacg 25	
25	<210> 38 <211> 80 <212> ADN <213> Homo sapiens	
30	<400> 38 <b>gcgagccca atacagaata cacacgcacg gtgtcttcag aggcattcag gatgtgcgac</b> <b>gtgtgcctgg agtagccccg</b>	<b>60</b> <b>80</b>
35	<210> 39 <211> 20 <212> ADN <213> Homo sapiens	
40	<400> 39 ctgtgatgg caggcttggc 20	
45	<210> 40 <211> 20 <212> ADN <213> Homo sapiens	
50	<400> 40 ttgctcacct gggcttgca 20	
55	<210> 41 <211> 21 <212> ADN <213> Homo sapiens	
60	<400> 41 gcagccaggc actgcctgc t 21	
65	<210> 42 <211> 74 <212> ADN <213> Homo sapiens	
	<400> 42	

ES 2 525 545 T3

	<b>ctgtttgatgg caggcttggc cctgcagcca ggcactgccc tgctgtgcta ctcctgcaaa</b>	<b>60</b>
	<b>gcccaggtga gcaa</b>	<b>74</b>
5	<210> 43 <211> 25 <212> ADN <213> Homo sapiens	
10	<400> 43 tgaagaaata tcctgggatt attca 25	
15	<210> 44 <211> 27 <212> ADN <213> Homo sapiens	
20	<400> 44 tatgtggtat ctctggaat atcatca 27	
25	<210> 45 <211> 27 <212> ADN <213> Homo sapiens	
30	<400> 45 acaaagggaa acagatattg aagactc 27	
35	<210> 46 <211> 87 <212> ADN <213> Homo sapiens	
	<400> 46	
	<b>tgaagaaata tcctgggatt attcagaatt tgtacaaagg gaaacagata ttgaagactc</b>	<b>60</b>
	<b>tgatgatatt ccagaagata ccacata</b>	<b>87</b>
40	<210> 47 <211> 19 <212> ADN <213> Homo sapiens	
45	<400> 47 ccccagtg gtcctaca 19	
50	<210> 48 <211> 22 <212> ADN <213> Homo sapiens	
55	<400> 48 aggatgaaac aagctgtgcc ga 22	
	<210> 49	
60		
65		



ES 2 525 545 T3

<211> 26  
 <212> ADN  
 <213> Homo sapiens  
 5 <400> 49  
 caggaacaaa agcgtgatct tgctgg 26  
 <210> 50  
 <211> 82  
 10 <212> ADN  
 <213> Homo sapiens  
 <400> 50  
 15 cccccagtgg gtcctcacag ctgccactg catcaggaac aaaagcgtga tcttgctggg 60  
 tcggcacagc ttgtttcatc ct 82  
 20 <210> 51  
 <211> 19  
 <212> ADN  
 <213> Homo sapiens  
 25 <400> 51  
 gccctgaggc actttcca 19  
 <210> 52  
 <211> 22  
 30 <212> ADN  
 <213> Homo sapiens  
 <400> 52  
 cggatgtcca cgtcacactt ca 22  
 35 <210> 53  
 <211> 25  
 <212> ADN  
 <213> Homo sapiens  
 40 <400> 53  
 cttccttct gggcatggag tctg 25  
 <210> 54  
 <211> 100  
 <212> ADN  
 <213> Homo sapiens  
 45 <400> 54  
 50 gccctgaggc actttccag cttccttcc tgggcatgga gtcctgtggc atccacgaaa 60  
 ctaccttcaa ctccatcatg aagtgtgacg tggacatccg 100  
 55 <210> 55  
 <211> 22  
 <212> ADN  
 <213> Homo sapiens  
 60  
 65

ES 2 525 545 T3

<400> 55  
ccacacacag cctactttcc aa 22

5 <210> 56  
<211> 21  
<212> ADN  
<213> Homo sapiens

10 <400> 56  
taccacgcg aatcactctc a 21

15 <210> 57  
<211> 27  
<212> ADN  
<213> Homo sapiens

20 <400> 57  
aacggcaatg cggctgcaac ggcggaa 27

25 <210> 58  
<211> 103  
<212> ADN  
<213> Homo sapiens

30 <400> 58  
ccacacacag cctactttcc aagcggagcc atgtctggta acggcaatgc ggctgcaacg 60  
gcggaagaaa acagcccaaa gatgagagtg attcgcgtgg gta 103

35 <210> 59  
<211> 2724  
<212> ADN  
<213> Homo sapiens

40 <400> 59

45

50

55

60

65

ES 2 525 545 T3

ggtgccatgg ctgagtcaca cctgctgcag tggctgctgc tgctgctgcc cacgctctgt 60  
 ggcccaggca ctgctgcctg gaccacctca tccttggcct gtgcccaggg ccctgagttc 120  
 5 tggtgccaaa gcctggagca agcattgcag tgcagagccc tagggcattg cctacaggaa 180  
 gtctggggac atgtgggagc cgatgaccta tgccaagagt gtgaggacat cgtccacatc 240  
 10 cttaacaaga tggccaagga ggccattttc caggacacga tgaggaagtt cctggagcag 300  
 gagtgcaacg tcctcccctt gaagctgctc atgccccagt gcaaccaagt gcttgacgac 360  
 tacttcccc tggtcatcga ctacttccag aaccagactg actcaaacgg catctgtatg 420  
 15 cacctgggcc tgtgcaaadc ccggcagcca gagccagagc aggagccagg gatgtcagac 480  
 cccctgcca aacctctgcg ggaccctctg ccagaccctc tgctggacaa gctcgtcctc 540  
 cctgtgctgc ccggggccct ccaggcgagg cctgggctc acacacagga tctctccgag 600  
 20 cagcaattcc ccattcctct ccctatttgc tggctctgca gggctctgat caagcggatc 660  
 caagccatga ttcccaaggg tgcgctagct gtggcagtgg cccaggtgtg ccgctgggta 720  
 25 cctctggtgg cgggcggcat ctgccagtgc ctggctgagc gctactccgt catcctgctc 780  
 gacacgctgc tgggccgcat gctgccccag ctggtctgcc gcctcgtcct ccggtgctcc 840  
 atggatgaca gcgctggccc aaggctgccg acaggagaat ggctgccgag agactctgag 900  
 30 tgccacctct gcatgtccgt gaccaccag gccgggaaca gcagcgagca ggccatacca 960  
 caggcaatgc tccaggcctg tgttggctcc tggctggaca gggaaaagtg caagcaattt 1020  
 gtggagcagc acacgcccc gctgctgacc ctggtgcca ggggctggga tgccccaccc 1080  
 35 acctgccagg ccctcggggt gtgtgggacc atgtccagcc ctctccagtg tatccacagc 1140  
 cccgaccttt gatgagaact cagctgtcca gaaaaagaca ccgtccttta aagtgtgca 1200  
 40 gtatggccag acgtggtggc tcacacctgc aatcccagca ccttaggagg ccgaggcagg 1260  
 aggatccttg aggtcaggag ttcgagacca gcctcgcca catggtgaaa cccattttct 1320  
 actaaaaata caaaaaatta gccaaagtgt gtggcatatg cctgtaatcc caactactca 1380  
 45 gaaggccgag gcaggagaat tacttgaacg caggagaatc actgcagccc aggaggcaga 1440  
 ggttgcagtg agccgagatt gcaccactgc actccagcct gggtgacaga gcaagactcc 1500  
 50 atctcagtaa ataaataaat aaataaaaag cgctgcagta gctgtggcct caccctgaag 1560  
 tcagcgggcc caggcctacc tcaactcttc ccttggcaga gaagcagacg tccatagctc 1620  
 ctctccctca caagcgtcc cagcctgcc tccagctgct gctctcccct cccagtctct 1680  
 55 actcactggg atgaggttag gtcatgagga caccaaaaac ctaaaaaata acaaaaagcc 1740  
 aaacaagcct tagcttttct taaagactga aatgcctgga agtgtccctt tatttataaa 1800  
 60 ataacttttg tcatatttct tatacatggt tcttgaaga aattcagaaa ctacagacaa 1860  
 agagagtgga aattaccac tgtcaggcct ctgagcccaa gctaagccat catatcccct 1920

65

ES 2 525 545 T3

gtgccctgca cgtatacacc cagatggcct gaagcaactg aagatccaca aaagaagtga 1980  
 aaatagccag ttcctgcctt aactgatgac attccaccat tgtgatttgt tcctgccccca 2040  
 5 ccctaactga tcaattgacc ttgtgacaat acaccttccc cacccttgag aagggtgcttt 2100  
 gtaatattct cccacccac cccacgcccg cacccccga cccttaagaa ggtattttgt 2160  
 aatattctct cgcattga gaatgtgctt tgtaagatcc acccctgcc cacaaaaaat 2220  
 10 tgctcctaac tccaccgctt atcccaaacc tacaagaact aatgataatc ccaccaccct 2280  
 ttgctgactc tttttggact cagcccact gcaccaggt gattaanaag ctttattggt 2340  
 cacacaaagc ctgtttggta gtctcttcac agggaagcat gtgacacca caatcccacc 2400  
 tagcccagga gagagctacg gcaggggtgtg tgttttgaca ctgagcttgg ggctttttcc 2460  
 atcttctccc cacagcctct ggctccacac ctccaccgtt caagcgccag aaagagctgt 2520  
 20 ctatgcagcc tgctcttggg cctggggatg agacacacaa ttcattggct cctggatttt 2580  
 aagtagacat ttgtaaact atagctaact actgtcctta aagccattgt ttccattaca 2640  
 aaatccaact ctctgagaga aaaggggtgtt ttaaatttaa aaaaataaaa acaaaaaagt 2700  
 25 ttgattgaga aaaaaaaaaa aaaa 2724

<210> 60  
 <211> 2352  
 <212> ADN  
 <213> Homo sapiens  
  
 <400> 60

35

40

45

50

55

60

65

ES 2 525 545 T3

	gaaacttaaa ggtgtttacc ttgtcatcag catgtaagct aattatctcg ggcaagatgt	60
	aggcttctat tgtcttggtg ctttagcgct tacgccccgc ctctggtggc tgcctaaaac	120
5	ctggcgccgg gctaaaacaa acgcgaggca gccccgagc ctccactcaa gccaattaag	180
	gaggactcgg tccactccgt tacgtgtaca tccaacaaga tcggcgtaa ggtaacacca	240
	gaatatttgg caaagggaga aaaaaaagc agcgaggctt cgccttcccc ctctcccttt	300
10	tttttctcc tcttcttcc tcctccagcc gccgccgaat catgtcgatg agtccaaagc	360
	acacgactcc gttctcagtg tctgacatct tgagtcccct ggaggaaagc tacaagaaag	420
15	tgggcatgga gggcggcggc ctcggggctc cgctggcggc gtacaggcag ggccaggcgg	480
	caccgccaac agcggccatg cagcagcacg ccgtggggca ccacggcgcc gtcaccgccg	540
	cctaccacat gacggcggcg ggggtgcccc agctctcgca ctccgccgtg gggggctact	600
20	gcaacggcaa cctgggcaac atgagcgagc tgccgccgta ccaggacacc atgaggaaca	660
	gcgctctgg ccccgatgg tacggcgcca acccagacc gcgcttcccc gccatctccc	720
	gcttcatggg cccggcgagc ggcataaaca tgagcggcat gggcggcctg ggctcgctgg	780
25	gggacgtgag caagaacatg gccccgtgc caagcgcgcc gcgcaggaag cgccgggtgc	840
	tcttctcgca ggcgcaggtg tacgagctgg agcgacgctt caagcaacag aagtacctgt	900
30	cggcgccgga gcgcgagcac ctggccagca tgatccacct gacgcccacg caggtcaaga	960
	tctggttcca gaaccaccgc taaaaatga agcggccaggc caaggacaag gcggcgcagc	1020
35		
40		
45		
50		
55		
60		
65		

ES 2 525 545 T3

5 agcaactgca gcaggacagc ggcggcggcg ggggcggcgg gggcaccggg tgccccgagc 1080  
 agcaacagggc tcagcagcag tcgccgcgac gcgtggcggg gccggtcctg gtgaaagacg 1140  
 gcaaaccgtg ccaggcgggt gccccgcgc cgggcgccgc cagcctaca ggcacgcgc 1200  
 10 agcagcagggc gcagcaccag gcgcaggccg cgcaggcggc ggcagcggcc atctccgtgg 1260  
 gcagcgggtg cgccggcctt ggcgcacacc cgggccacca gccaggcagc gcagggcagt 1320  
 ctccggacct ggcgcaccac gccgccagcc ccgcggcgct gcagggccag gtatccagcc 1380  
 tgtccacct gaactcctcg ggctcggact acggcaccat gtctgtctc accttgctat 1440  
 15 acggtcggac ctggtgagag gacgccgggc cggccctagc ccagcgtct gcctcaccgc 1500  
 ttccctcctg cccgccacac agaccaccat ccaccgctgc tccacgcgct tcgacttttc 1560  
 20 ttaacaacct ggccgcgttt agaccaagga acaaaaaaac cacaaaggcc aaactgctgg 1620  
 acgtctttct tttttcccc ccctaaaatt tgtgggtttt ttttttaaa aaaagaaaat 1680  
 gaaaaacaac caagcgcac caatctcaag gaatcttta gcagagaagg gcataaaaca 1740  
 25 gctttggggg gtcttttttt ggtgattcaa atgggttttc cacgctaggg cggggcacag 1800  
 attggagagg gctctgtgct gacatggctc tggactctaa agaccaaact tactctggg 1860  
 cacactctgc cagcaaagag gactcgttg taaataccag gattttttt ttttttgaa 1920  
 30 gggaggacgg gagctgggga gaggaaagag tcttcaacat aaccacttg tactgacac 1980  
 aaaggaagtg ccccctcccc ggcaccctct ggccgcctag gctcagcggc gaccgcctc 2040  
 35 cgcgaaaata gtttgtttaa tgtgaacttg tagctgtaaa acgctgtcaa aagttggact 2100  
 aaatgcctag ttttagtaa tctgtacatt ttgttgtaaa aagaaaaacc actcccagtc 2160  
 cccagccctt cacatTTTTT atgggcattg acaaactctgt gtatattatt tggcagttg 2220  
 40 gtatttgcg cgtcagtctt tttctgttg aacttatgta gatatttggc ttaaataatag 2280  
 ttccctaagaa gcttctaata aattatacaa attaaaaaga ttctttttct gattaaaaaa 2340  
 aaaaaaaaa aa 2352

45  
 <210> 61  
 <211> 3336  
 <212> ADN  
 <213> Homo sapiens  
 50  
 <400> 61

55

60

65

ES 2 525 545 T3

	ttttcttaga cattaactgc agacggctgg caggatagaa gcagcggctc acttggactt	60
	tttcaccagg gaaatcagag acaatgatgg ggctcttccc cagaactaca ggggctctgg	120
5	ccatcttcgt ggtggtcata ttggttcatg gagaattgcg aatagagact aaaggccaat	180
	atgatgaaga agagatgact atgcaacaag ctaaaagaag gcaaaaacgt gaatgggtga	240
	aatttgccaa accctgcaga gaaggagaag ataactcaaa aagaaacca attgccaaga	300
10	ttacttcaga ttaccaagca acccagaaaa tcacctaccg aatctctgga gtgggaatcg	360
	atcagccgcc ttttgaatc tttgttggtg acaaaaacac tggagatatt aacataacag	420
15	ctatagtcga ccgggaggaa actccaagct tcctgatcac atgtcgggct ctaaagccc	480
20		
25		
30		
35		
40		
45		
50		
55		
60		
65		

ES 2 525 545 T3

aaggactaga thtagagaaa ccacttatac taacgggtaa aatdddggat attaatgata 540  
 atcctccagt atdddcaaa caatdddca tgggtgaaat tgaagaaaat agtgcctcaa 600  
 5 actcactggg gatgatacta aatgccacag atgcagatga accaaaccac ttgaattcta 660  
 aaattgcctt caaaattgtc tctcaggaac cagcaggcac acccatgttc ctcctaagca 720  
 gaaacactgg ggaagtccgt acttdgacca attctcttga ccgagagcaa gctagcagct 780  
 10 atcgtctggg tgtgagtggg gcagacaaaag atggagaagg actatcaact caatgtgaat 840  
 gtaatattaa agtgaaagat gtcaacgata acttdccaat gtttagagac tctcagtatt 900  
 15 cagcacgtat tgaagaaaat attdtaagtt ctgaattact tcgatttcaa gtaacagatt 960  
 tggatgaaga gtacacagat aatggcctg cagtatattt ctttacctct gggaatgaag 1020  
 gaaattgggt tgaatacaa actgatccta gaactaatga aggcacctg aaagtgggga 1080  
 20 aggcctctaga ttatgaacaa ctacaaaagc tgaacttag tattgtgtc aaaaacaaag 1140  
 ctgaatttca ccaatcagtt atctctcgat accgagttca gtcaacccca gtcacaattc 1200  
 aggtaataaa tghtaagaaa ggaattgcat tccgtcctgc ttccaagaca tttactgtgc 1260  
 25 aaaaaggcat aagtagcaa aaattgggtg attatatact gggaaacatat caagccatcg 1320  
 atgaggacac taacaaagct gcctcaaatg tcaaatatgt catgggacgt aacgatggg 1380  
 gatacctaat gattgattca aaaactgctg aatcaaat tgtcaaaaat atgaaccgag 1440  
 30 attctacttt catagttaac aaaacaatca cagctgaggt tctggccata gatgaataca 1500  
 cgggtaaaac ttctacaggc acggtatatg ttagagtacc cgatttcaat gacaattgtc 1560  
 caacagctgt cctcgaaaaa gatgcagttt gcagttcttc acctccctg gttgtctccg 1620  
 35 ctagaacact gaataataga taaactggcc cctatacatt tgcactggaa gatcaacctg 1680  
 taaagttgcc tgccgatgg agtatcaaa ccctcaatgc tacctcggcc ctcctcagag 1740  
 40 cccaggaaca gatacctcct ggagtatacc acatctcct ggtacttaca gacagtcaga 1800  
 acaatcggg tggatgcca cgcagcttga cactggaagt ctgtcagtgt gacaacaggg 1860  
 gcatctgtgg aacttcttac ccaaccacaa gccctgggac caggatggc aggccgcact 1920  
 45 caggaggct gggcctgccc gccatcggcc tgctgtcct tggctcctg ctgctgtgt 1980  
 tggccccct tctgtgttg acctgtgact gtggggcagg ttctactggg ggagtgacag 2040  
 gtggttttat cccagttcct gatggctcag aaggaacaat tcatcagtgg ggaattgaag 2100  
 50 gagcccatcc tgaagacaag gaaatcaaa atattdgtgt gcctcctgta acagccaatg 2160  
 gagccgattt catggaaagt tctgaagttt gtacaaatac gtatgccaga ggcacagcgg 2220  
 55 tggaaaggcac ttcaggaatg gaaatgacca ctaagcttgg agcagccact gaatctggag 2280  
 gtgctgcagg ctttgcaaca gggacagtgt caggagctgc ttcaggattc ggagcagcca 2340  
 ctggagttgg catctgttcc tcagggcagt ctggaacct gagaacaagg cattccactg 2400  
 60 gaggaaccaa taaggactac gctgatgggg cgataagcat gaattdctg gactcctact 2460  
 ttdctcagaa agcatttgc tgtgcggagg aagacgatgg ccaggaagca aatgactgct 2520

65



ES 2 525 545 T3

	tgttgatcta tgataatgaa ggcgcagatg ccaactggttc tcctgtgggc tccgtggggtt	2580
	gttgcagttt tattgctgat gacctggatg acagcttctt ggactcactt ggacccaaat	2640
5	ttaaaaaact tgacagagata agccttggtg ttgatggtga aggcaaagaa gttcagccac	2700
	cctctaaaga cagcggttat gggattgaat cctgtggcca tcccatagaa gtccagcaga	2760
	caggatttgt taagtgccag actttgtcag gaagtcaagg agcttctgct ttgtccgcct	2820
10	ctgggtctgt ccagccagct gtttccatcc ctgaccctct gcagcatggt aactatttag	2880
	taacggagac ttactcggct tctggttccc tcgtgcaacc ttccactgca ggctttgatc	2940
	cacttctcac acaaaatgtg atagtgcag aaagggatg ctgtcccatt tccagtgttc	3000
15	ctggcaacct agctggcca acgcagctac gagggtcaca tactatgctc tgtacagagg	3060
	atccttgctc ccgtctaata tgaccagaat gagctggaat accacactga ccaaatctgg	3120
	atctttggac taaagtattc aaaatagcat agcaaagctc actgtattgg gctaataatt	3180
20	tggcacttat tagcttctct cataaactga tcacgattat aaattaaatg tttgggttca	3240
	taccccaaaa gcaatatggt gtcactccta attctcaagt actattcaaa ttgtagtaaa	3300
25	tcttaaagtt tttcaaaacc ctaaaatcat attcgc	3336

	<210> 62
	<211> 3697
30	<212> ADN
	<213> Homo sapiens
	<400> 62

35

40

45

50

55

60

65

ES 2 525 545 T3

	agggagtgtt	cccgggggag	atactccagt	cgtagcaaga	gtctcgacca	ctgaatggaa	60
	gaaaaggact	ttaaccacc	atthttgtgac	ttacagaaag	gaatttgaat	aaagaaaact	120
5	atgatacttc	aggcccatct	tcactccctg	tgtcttctta	tgctttatth	ggcaactgga	180
	tatggccaag	aggggaagtt	tagtggaacc	ctgaaacca	tgacatthtc	tatthtatgaa	240
	ggccaagaac	cgagtcaaat	tatattccag	ttaaggcca	atcctcctgc	tgtgacttht	300
10	gaactaactg	gggagacaga	caacatath	gtgatagaac	gggagggact	tctgtattac	360
	aacagagcct	tggacagga	aacaagatct	actcacaatc	tccaggttgc	agccctggac	420
	gctaattgga	ttatagtga	gggtccagtc	cctatcacca	tagaagtgaa	ggacatcaac	480
15	gacaatcgac	ccacgtthct	ccagtcaaag	tacgaaggct	cagtaaggca	gaactctcgc	540
	ccaggaaagc	ccttcttgta	tgtcaatgcc	acagacctgg	atgatccggc	cactcccaat	600
	ggccagctth	attaccagat	tgtcatccag	cttcccatga	tcaacaatgt	catgtactth	660
20	cagatcaaca	acaaaacggg	agccatctct	cttaccggag	agggatctca	ggaattgaat	720
	cctgctaaga	atccttccta	taatctgggtg	atctcagtga	aggacatggg	aggccagagt	780
25	gagaattcct	tcagtgatac	cacatctgtg	gatatcatag	tgacagagaa	tatthtgaaa	840
	gcaccaaaaac	ctgtggagat	gggtggaaaac	tcaactgatc	ctcaccccat	caaaatcact	900
	caggtgcgggt	ggaatgatcc	cggtgcacaa	tatthcttag	ttgacaaaaga	gaagctgcca	960
30	agattcccat	thtcaattga	ccaggaagga	gatathtacg	tgactcagcc	cttggaccga	1020

35

40

45

50

55

60

65

ES 2 525 545 T3

gaagaaaagg atgcatatgt tttttatgca gttgcaaagg atgagtacgg aaaaccactt 1080  
 tcatatccgc tggaaattca tgtaaaagtt aaagatatta atgataatcc acctacatgt 1140  
 5 ccgtcaccag taaccgtatt tgagggtccag gagaatgaac gactgggtaa cagtatcggg 1200  
 acccttactg cacatgacag ggatgaagaa aatactgcc aacagttttct aaactacagg 1260  
 attgtggagc aaactcccaa acttcccatg gatggactct tcctaatacca aacctatgct 1320  
 10 ggaatgttac agttagctaa acagtccttg aagaagcaag atactcctca gtacaactta 1380  
 acgatagagg tgtctgacaa agatttcaag accctttgtt ttgtgcaaat caacgttatt 1440  
 gatataaatg atcagatccc catctttgaa aaatcagatt atggaaacct gactcttgct 1500  
 15 gaagacacaa acattgggtc caccatctta accatccagg ccaactgatgc tgatgagcca 1560  
 tttactggga gttctaaaat tctgtatcat atcataaagg gagacagtga gggacgcctg 1620  
 20 ggggttgaca cagatcccca taccaacacc ggatatgtca taattaaaaa gcctcttgat 1680  
 tttgaaacag cagctgtttc caacattgtg ttcaaagcag aaaatcctga gcctctagtg 1740  
 tttgggtgta agtacaatgc aagttctttt gccaaagtca cgcttattgt gacagatgtg 1800  
 25 aatgaagcac ctcaattttc ccaacacgta ttccaagcga aagtcagtga ggatgtagct 1860  
 ataggcacta aagtgggcaa tgtgactgcc aaggatccag aaggctctgga cataagctat 1920  
 tcaactgagg gagacacaag aggttggtt aaaattgacc acgtgactgg tgagatcttt 1980  
 30 agtgtggctc cattggacag agaagccgga agtccatctc gggtaacaagt ggtggccaca 2040  
 gaagtagggg ggtcttcctt gagctctgtg tcagagtcc acctgatcct tatggatgtg 2100  
 aatgacaacc ctcccaggct agccaaggac tacacgggct tgttcttctg ccatcccctc 2160  
 35 agtgcacctg gaagtctcat tttcgaggct actgatgatg atcagcactt atttcggggg 2220  
 ccccatthta cattttccct cggcagtgga agcttacaaa acgactggga agtttccaaa 2280  
 40 atcaatggta ctcatgcccg actgtctacc aggacacag agtttgagga gagggagtat 2340  
 gtcgtcttga tccgatcaa tgatgggggt cggccaccct tggaaaggcat tgtttcttta 2400  
 ccagttacat tctgcagttg tgtggaagga agttgtttcc ggccagcagg tcaccagact 2460  
 45 gggataccca ctgtgggcat ggcagttggt atactgctga ccacccttct ggtgattggt 2520  
 ataattttag cagttgtgtt tatccgcata aagaaggata aaggcaaaga taatgttgaa 2580  
 agtgtctcaag catctgaagt caaacctctg agaagctgaa tttgaaaagg aatgtttgaa 2640  
 50 tttatatagc aagtgtctatt tcagcaacaa ccatctcatc ctattacttt tcatctaacg 2700  
 tgcattataa ttttttaaac agatattccc tcttgtcctt taatatttgc taaatatttc 2760  
 ttttttgagg tggagtcttg ctctgtcgcc caggctggag tacagtgggtg tgatcccagc 2820  
 55 tcaactgcaac ctccgcctcc tgggttcaca tgattctcct gcctcagctt cctaagtagc 2880  
 tgggtttaca ggcaccacc accatgccc a gtaattttt gtatttttaa tagagacggg 2940  
 gtttcgcat ttggccaggc tggctctgaa ctctgacgt caagtgatct gcctgccttg 3000  
 60 gtctcccaat acaggcatga accactgcac ccacctactt agatatttca tgtgctatag 3060  
 acattagaga gatttttcat ttttccatga catttttctt ctctgcaaat ggcttagcta 3120  
 65

ES 2 525 545 T3

5 cttgtgtttt tcccttttgg ggcaagacag actcattaaa tattctgtac attttttctt 3180  
 tatcaaggag atatatcagt gttgtctcat agaactgcct ggattccatt tatgtttttt 3240  
 ctgattccat cctgtgtccc cttcatcctt gactcctttg gtatttctact gaatttcaaa 3300  
 catttgtcag agaagaaaaa cgtgaggact caggaaaaat aaataaataa aagaacagcc 3360  
 10 ttttccctta gtattaacag aatgtttct gtgtcattaa ccatctttaa tcaatgtgac 3420  
 atgttgctct ttggctgaaa ttcttcaact tggaaatgac acagaccac agagggtgtt 3480  
 caaacacaac ctactctgca aaccttggtg aaggaaccag tcagctggcc agatttcctc 3540  
 15 actacctgcc atgcatacat gctgcgcatg ttttcttcat tcgtatgtta gtaaagtttt 3600  
 ggttattata tatttaacat gtggaagaaa acaagacatg aaaagagtgg tgacaaatca 3660  
 agaataaaca ctggtttag tagttttgt ttgttaa 3697

20  
 <210> 63  
 <211> 503  
 <212> ADN  
 <213> Homo sapiens  
 25  
 <400> 63

30 gacagcggct tccttgatcc ttgccaccg cgactgaaca ccgacagcag cagcctcacc 60  
 atgaagttgc tgatggctct catgctggcg gccctctccc agcactgcta cgcaggctct 120  
 ggctgccctt tattggagaa tgtgatttcc aagacaatca atccacaagt gtctaagact 180  
 35 gaatacaaag aacttcttca agagttcata gacgacaatg cactacaaa tgccatagat 240  
 gaattgaagg aatgttttct taaccaaacg gatgaaactc tgagcaatgt tgagggtgtt 300  
 atgcaattaa tatatgacag cagtctttgt gatttatttt aactttctgc aagacctttg 360  
 40 gctcacagaa ctgcagggta tggtagagaaa ccaactacgg attgctgcaa accacacctt 420  
 ctctttctta tgtcttttta ctacaaacta caagacaatt gttgaaacct gctatacatg 480  
 tttattttta taaattgatg gca 503

45  
 <210> 64  
 <211> 1894  
 <212> ADN  
 <213> Homo sapiens  
 50  
 <400> 64

55

60

65

ES 2 525 545 T3

gtctgacttc ctcccagcac attcctgcac tctgccgtgt ccacactgcc ccacagaccc 60  
 agtcctccaa gcctgctgcc agtcacctgc aagccccctca ggttgggcct tgccacggtg 120  
 5 ccagcaggca gccctgggct gggggtaggg gactccctac aggcacgcag ccctgagacc 180  
 tcagagggcc accccttgag ggtggccagg cccccagtgg ccaacctgag tgctgcctct 240  
 gccaccagcc ctgctggccc ctggttccgc tggcccccca gatgcctggc tgagacacgc 300  
 10 cagtggcctc agctgcccac acctcttccc ggccccctgaa gttggcactg cagcagacag 360  
 ctccctgggc accaggcagc taacagacac agccgccagc ccaaacagca gcggcatggg 420  
 cagcgcacgc ccgggtctga gcagcgtatc ccccagccac ctctgctgc cccccgacac 480  
 15 ggtgtcgcgg acaggcttgg agaaggcggc agcgggggca gtgggtctcg agagacggga 540  
 ctggagtccc agtccacccg ccacgcccga gcagggcctg tccgccttct acctctccta 600  
 20 ctttgacatg ctgtaccctg aggacagcag ctgggcagcc aaggccccctg gggccagcag 660  
 tcgggaggag ccacctgagg agcctgagca gtgcccggtc attgacagcc aagccccagc 720  
 gggcagcctg gacttgggtc ccggcgggct gaccttggag gagcactcgc tggagcaggt 780  
 25 gcagtccatg gtgggtggcg aagtgtcaa ggacatcgag acggcctgca agctgtctaa 840  
 catcaccgca gatcccatgg actggagccc cagcaatgtg cagaagtggc tcctgtggac 900  
 agagcaccaa taccggctgc cccccatggg caaggccttc caggagctgg cgggcaagga 960  
 30 gctgtgcgcc atgtcggagg agcagttccg ccagcgcctc cccctgggtg gggatgtgct 1020  
 gcacgcccac ctggacatct ggaagtgcgc ggccctggatg aaagagcgga cttcacctgg 1080  
 35 ggcgattcac tactgtgcct cgaccagtga ggagagctgg accgacagcg aggtggactc 1140  
 atcatgctcc gggcagccca tccacctgtg gcagttctc aaggagttgc tactcaagcc 1200  
 ccacagctat ggccgcttca ttaggtggct caacaaggag aagggcatct tcaaaattga 1260  
 40 ggactcagcc caggtggccc ggctgtgggg catccgcaag aaccgtcccg ccatgaacta 1320  
 cgacaagctg agccgctcca tccgccagta ttacaagaag ggcacatcc ggaagccaga 1380  
 catctcccag cgcctcgtct accagttcgt gcaccccatc tgagtgcctg gccagggcc 1440  
 45 tgaaacccgc cctcaggggc ctctctcctg cctgccctgc ctcagccagg ccctgagatg 1500  
 ggggaaaacg ggcagtctgc tctgtgctc tgaccttcca gagcccaagg tcagggaggg 1560  
 gcaaccaact gccccagggg gatatgggtc ctctggggcc ttcgggacca tggggcaggg 1620  
 50 gtgcttctc ctcaggccca gctgctccc tggaggacag agggagacag ggctgctccc 1680  
 caacacctgc ctctgacccc agcatttcca gagcagagcc tacagaagg cagtgactcg 1740  
 55 acaaaggcca caggcagtc aggcctctct ctgctccatc cccctgcctc ccattctgca 1800  
 ccacacctgg catggtgcag ggagacatct gcaccctga gttgggcagc caggagtgcc 1860  
 cccgggaatg gataataaag atactagaga actg 1894  
 60 <210> 65  
 <211> 3029  
 <212> ADN  
 65

ES 2 525 545 T3

<213> Homo sapiens

<400> 65

5	ccaggcagct ggggtaagga gttcaaggca gcgcccacac ccgggggctc tccgcaacct	60
	gaccgcctgt ccgctcccc acttcccgcc ctccctccca cctactcatt cacccacca	120
	cccaccaga gccgggacgg cagcccaggc gcccgggccc cgccgtctcc tcgcccgat	180
10	cctggacttc ctcttgctgc aggaccggc ttccacgtgt gtcccggagc cggcgtctca	240
	gcacacgctc cgctccgggc ctgggtgcct acagcagcca gagcagcagg gaggccgga	300
15	cccgggaggc atctgggcca agttaggcgc cgccgaggcc agcgctgaac gtctccagg	360
	ccggaggagc cgcggggcgt ccgggtctga gccgcagcaa atgggctccg acgtgcggga	420

20

25

30

35

40

45

50

55

60

65

ES 2 525 545 T3

cctgaacgcg ctgctgcccg ccgtcccctc cctgggtggc ggcggcggct gtgccctgcc 480  
 tgtgagcggc gcggcgcagt gggcgccggt gctggacttt gcgcccccg gcgcttcggc 540  
 5 ttacgggtcg ttgggcggcc ccgcgccgcc accggctccg ccgccacccc cgccgccgcc 600  
 gcctcactcc ttcataaac aggagccgag ctggggcggc gcggagccgc acgaggagca 660  
 gtgcctgagc gccttcactg tccacttttc cggccagttc actggcacag ccggagcctg 720  
 10 tcgctacggg cccttcggtc ctctccgcc cagccaggcg tcatccggcc aggccaggat 780  
 gtttcctaac gcgccctacc tgcccagctg cctcgagagc cagccccgta ttcgcaatca 840  
 gggttacagc acggtcacct tcgacgggac gccagctac ggtcacacgc cctcgcacca 900  
 15 tgcggcgcag ttccccaacc actcattcaa gcatgaggat cccatgggcc agcagggctc 960  
 gctgggtgag cagcagtact cggtgccgcc cccgggtctat ggctgccaca cccccaccga 1020  
 cagctgcacc ggcagccagg ctttgctgct gaggacgcc tacagcagtg acaatttata 1080  
 20 ccaaagaca tcccagcttg aatgcatgac ctggaatcag atgaacttag gagccacctt 1140  
 aaagggagt gctgctggga gctccagctc agtgaatgg acagaagggc agagcaacca 1200  
 25 cagcacaggg tacgagagcg ataaccacac aacgccatc ctctgcggag cccaatacag 1260  
 aatacacagc cacggtgtct tcagaggcat tcaggatgtg cgacgtgtgc ctggagtagc 1320  
 cccgactctt gtacggtcgg catctgagac cagtgagaaa cccccctca tgtgtgctta 1380  
 30 cccaggctgc aataagagat attttaagct gtcccactta cagatgcaca gcaggaagca 1440  
 cactgggtgag aaaccatacc agtgtgactt caaggactgt gaacgaaggt tttctcgttc 1500  
 agaccagctc aaaagacacc aaaggagaca tacagggtgtg aaaccattcc agtgtaaaac 1560  
 35 ttgtcagcga aagttctccc ggtccgacca cctgaagacc cacaccagga ctatacagg 1620  
 taaaacaagt gaaaagccct tcagctgtcg gtggccaagt tgtcagaaaa agtttgcccg 1680  
 40 gtcagatgaa ttagtccgcc atcacaacat gcatcagaga aacatgacca aactccagct 1740  
 ggcgctttga ggggtctccc tcggggaccg ttcagtgtcc caggcagcac agtgtgtgaa 1800  
 ctgctttcaa gtctgactct cactcctcc tactaaaaa ggaaactca gttgatcttc 1860  
 45 ttcaccaac ttccaagaca agataccggt gcttctggaa actaccaggt gtgcctggaa 1920  
 gagttggtct ctgccctgcc tacttttagt tgactcacag gccctggaga agcagctaac 1980  
 aatgtctggt tagttaaag cccattgcca tttggtgtgg attttctact gtaagaagag 2040  
 50 ccatagctga tcatgtcccc ctgacccttc ccttctttt ttatgctcgt tttcgtggy 2100  
 gatggaatta ttgtaccatt ttctatcatg gaatatttat aggccagggc atgtgtatgt 2160  
 gtctgctaata gtaaactttg tcatggtttc catttactaa cagcaacagc aagaaataaa 2220  
 55 tcagagagca aggcacgagg ggtgaatctt gtctaacatt cccgaggtca gccaggctgc 2280  
 taacctggaa agcaggatgt agttctgcca ggcaactttt aaagctcatg catttcaagc 2340  
 agctgaagaa aaaatcagaa ctaaccagta cctctgtata gaaatctaaa agaattttac 2400  
 60 cattcagtta attcaatgtg aacctggca cactgctctt aagaaactat gaagatctga 2460  
 gatttttttg tgtatgtttt tgactctttt gagtggtaat catatgtgtc tttatagatg 2520  
 65

ES 2 525 545 T3

5      tacatacctc cttgcacaaa tggaggggaa ttcattttca tcaactgggag tgtccttagt    2580  
       gtataaaaac catgctggta tatggcttca agttgtaaaa atgaaagtga ctttaaaaga    2640  
       aaatagggga tgggccagga tctccactga taagactggt ttttaagtaac ttaaggacct    2700  
       ttgggtctac aagtatatgt gaaaaaatg agacttactg ggtgaggaaa tccattgttt    2760  
 10     aaagatggtc gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtttgtgt    2820  
       ttttaaggga ggggaatttat tatttaccgt tgcttgaaat tactgtgtaa atatatgtct    2880  
       gataatgatt tgctctttga caactaaaat taggactgta taagtactag atgcatcact    2940  
 15     ggggtgtgat cttacaagat attgatgata acacttaaaa ttgtaacctg catttttcac    3000  
       tttgctctca attaaagtct attcaaaag    3029

20     <210> 66  
       <211> 1064  
       <212> ADN  
       <213> Homo sapiens  
  
       <400> 66

25     tttgaggcca tataaagtca cctgaggccc tctccaccac agcccaccag tgaccatgaa        60  
       ggctgtgctg cttgccctgt tgatggcagg cttggccctg cagccaggca ctgccctgct    120  
 30     gtgctactcc tgcaaagccc aggtgagcaa cgaggactgc ctgcaggtgg agaactgcac    180  
       ccagctgggg gagcagtgtt ggaccgcgcg catccgcgca gttggcctcc tgaccgtcat    240  
       cagcaaaggc tgcagcttga actgcgtgga tgactcacag gactactacg tgggcaagaa    300  
 35     gaacatcacg tgctgtgaca ccgacttgtg caacgccagc ggggcccattg ccctgcagcc    360  
       ggctgtgctc atccttgcgc tgctccctgc actcggcctg ctgctctggg gaccgggcca    420  
       gctctaggct ctggggggcc ccgctgcagc ccacactggg tgtggtgccc caggcctctg    480  
 40     tgccactcct cacacacccg gccagtgagg agcctgtcct ggttcctgag gcacatccta    540  
       acgcaagtct gaccatgtat gtctgcgccc ctgtccccc aacctgacct cccatggccc    600  
       tctccaggac tcccacccgg cagatcggct ctattgacac agatccgcct gcagatggcc    660  
 45     cctccaacc tctctgctgc tgtttccatg gcccagcatt ctccaccctt aacctgtgc    720  
       tcaggcacct cttccccag gaagccttc ctgcccacc catctatgac ttgagccagg    780  
 50     tctggtccgt ggtgtcccc gcaccagca ggggacaggc actcaggagg gcccgtaaa    840  
       ggctgagatg aagtggactg agtagaactg gaggacagga gtcgacgtga gttcctggga    900  
       gtctccagag atggggcctg gaggcctgga ggaaggggcc aggcctcaca ttcgtggggc    960  
 55     tccctgaatg gcagcctcag cacagcgtag gcccttaata aacacctgtt ggataagcca    1020  
       aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaa                            1064

60     <210> 67  
       <211> 6962  
       <212> ADN  
       <213> Homo sapiens

65



<400> 67

5

10

15

20

25

30

35

40

45

50

55

60

65

ES 2 525 545 T3

gcaagaactg caggggagga ggacgctgc acccacagcc tctagagctc attgcagctg 60  
 ggacagcccg gagtgtgggt agcagctcgg caagcgctgc ccaggtcctg ggggtggggc 120  
 5 agccagcggg agcaggaaag gaagcatggt cccaggctgc ccacgcctct gggtcctggt 180  
 ggtcttgggc accagctggg taggctgggg gagccaaggg acagaagcgg cacagctaag 240  
 gcagttctac gtggctgctc agggcatcag ttggagctac cgacctgagc ccacaaactc 300  
 10 aagtttgaat ctttctgtaa cttcctttaa gaaaattgtc tacagagagt atgaaccata 360  
 ttttaagaaa gaaaaaccac aatctaccat ttcaggactt cttgggccta ctttatatgc 420  
 tgaagtcgga gacatcataa aagttcactt taaaaataag gcagataagc ctttgagcat 480  
 15 ccatcctcaa ggaattaggt acagtaaatt atcagaaggt gcttcttacc ttgaccacac 540  
 attccctgcg gagaagatgg acgacgctgt ggctccaggg cgagaataca cctatgaatg 600  
 20 gagtatcagt gaggacagtg gaccaccca tgatgaccct ccatgcctca cacacatcta 660  
 ttactcccat gaaaatctga tcgaggattt caactcgggg ctgattgggc ccctgcttat 720  
 ctgtaaaaaa gggaccctaa ctgagggtgg gacacagaag acgtttgaca agcaaactcgt 780  
 25 gctactatth gctgtgtttg atgaaagcaa gagctggagc cagtcatcat ccctaatgta 840  
 cacagtcaat ggatatgtga atgggacaat gccagatata acagtttgtg cccatgacca 900  
 catcagctgg catctgctgg gaatgagctc ggggccagaa ttattctcca ttcatttcaa 960  
 30 cggccaggtc ctggagcaga accatcataa ggtctcagcc atcacccttg tcagtgtctac 1020  
 atccactacc gcaaatatga ctgtgggccc agaggggaaag tggatcatat cttctctcac 1080  
 cccaaaacat ttgcaagctg ggatgcaggg ttacattgac attaaaaact gcccaaagaa 1140  
 35 aaccaggaat cttagaaaaa taactcgtga gcagagggcg cacatgaaga ggtgggaata 1200  
 cttcattgct gcagaggaag tcatttggga ctatgcacct gtaataccag cgaatatgga 1260  
 caaaaaatac aggtctcagc atttggataa tttctcaaac caaattggaa aacattataa 1320  
 40 gaaagtatg tacacacagt acgaagatga gtccttcacc aaacatacag tgaatcccaa 1380  
 tatgaaagaa gatgggattt tgggtcctat tatcagagcc caggtcagag acacactcaa 1440  
 45 aatcgtgttc aaaaatatgg ccagccgccc ctatagcatt taccctcatg gagtgacctt 1500  
 ctgccttat gaagatgaag tcaactctc tttcacctca ggcaggaaca acaccatgat 1560  
 cagagcagtt caaccagggg aaacctatac ttataagtgg aacatcttag agtttgatga 1620  
 50 acccacagaa aatgatgccc agtgctaac aagaccatac tacagtgacg tggacatcat 1680  
 gagagacatc gcctctgggc taataggact acttctaate tgtaagagca gatccctgga 1740  
 caggcgagga atacagaggg cagcagacat cgaacagcag gctgtgtttg ctgtgtttga 1800  
 55 tgagaacaaa agctggtacc ttgaggacaa catcaacaag ttttgtaaa atcctgatga 1860  
 ggtgaaacgt gatgacccca agttttatga atcaaacatc atgagcacta tcaatggcta 1920  
 tgtgcctgag agcataacta ctcttgatt ctgctttgat gacactgtcc agtggcactt 1980  
 60 ctgtagtgtg gggaccaga atgaaatth gaccatccac ttcactgggc actcattcat 2040  
 ctatggaaag aggcattgagg acacctgac cctcttcccc atgctgtggag aatctgtgac 2100

65

ES 2 525 545 T3

	ggtcacaatg gataatggtg gaacttggat gttaacttcc atgaattcta gtccaagaag	2160
5	caaaaagctg aggctgaaat tcagggatgt taaatgtatc ccagatgatg atgaagactc	2220
	atatgagatt tttgaacctc cagaatctac agtcatggct acacggaaaa tgcattgatcg	2280
	tttagaacct gaagatgaag agagtgatgc tgactatgat taccagaaca gactggctgc	2340
10	agcattagga atcaggatcat tccgaaactc atcattgaat caggaagaag aagagttcaa	2400
	tcttactgcc ctagctctgg agaatggcac tgaattcggt tcttcaaaca cagatataat	2460
	tgttggttca aattattctt cccaagtaa tattagtaag ttcactgtca ataaccttgc	2520
15	agaacctcag aaagcccctt ctaccaaca agccaccaca gctggttccc cactgagaca	2580
	cctcattggc aagaactcag ttctcaattc ttccacagca gagcattcca gcccatattc	2640
20	tgaagaccct atagaggatc ctctacagcc agatgtcaca gggatacgtc tactttcact	2700
	tggtgctgga gaattcaaaa gtcaagaaca tgctaagcat aagggaccca aggtagaaag	2760
	agatcaagca gcaagcaca ggttctcctg gatgaaatta ctagcacata aagttgggag	2820
25	acacctaagc caagacactg gttctcctc cggaatgagg ccctgggagg accttctag	2880
	ccaagacact ggttctcctt ccagaatgag gccctggaag gaccctccta gtgatctggt	2940
	actcttaaaa caaagtaact catctaagat tttggttggg agatggcatt tggttctga	3000
30	gaaaggtagc tatgaaataa tccaagatac tgatgaagac acagctgtta acaattggct	3060
	gatcagcccc cagaatgcct cacgtgcttg gggagaaagc acccctcttg ccaacaagcc	3120
	tggaagcagc agtggccacc caaagtttcc tagagttaga cataaatctc tacaagtaag	3180
35	acaggatgga ggaaagagta gactgaagaa aagccagttt ctcattaaga cacgaaaaaa	3240
	gaaaaaagag aagcacacac accatgctcc tttatctccg aggaccttc accctctaag	3300
40	aagtgaagcc tacaacacat ttctcagaag aagacttaag cattcgttgg tgcttcataa	3360
	atccaatgaa acatctcttc ccacagacct caatcagaca ttgccctcta tggattttgg	3420
	ctggatagcc tcaattctg accataatca gaattcctca aatgacactg gtcaggcaag	3480
45	ctgtctcca ggtctttatc agacagtgcc cccagaggaa cactatcaaa cattccccat	3540
	tcaagaccct gatcaaatgc actctacttc agacccagct cacagatcct cttctccaga	3600
	gctcagtga atgcttgagt atgaccgaag tcacaagtcc ttccccacag atataagtca	3660
50	aatgtcccct tctcagaac atgaagtctg gcagacagtc atctctccag acctcagcca	3720
	ggtgaccctc tctccagaac tcagccagac aaacctctct ccagacctca gccacacgac	3780
	tctctctcca gaactcattc agagaaacct ttccccagcc ctcggtcaga tgcccatttc	3840
55	tccagacctc agccatacaa ccctttctcc agacctcagc catacaacct tttctttaga	3900
	cctcagccag acaaacctct ctccagaact cagtcagaca aacctttctc cagccctcgg	3960
60	tcagatgccc ctttctccag acctcagcca tacaacctt tctctagact tcagccagac	4020
	aaacctctct ccagaactca gccatatgac tctctctcca gaactcagtc agacaaacct	4080
	ttccccagcc ctcggtcaga tgcccatttc tccagacctc agccatacaa ccctttctct	4140
65		

ES 2 525 545 T3

agacttcagc cagacaaacc tctctccaga actcagtcaa acaaaccttt ccccagccct 4200  
 cggtcagatg cccctttctc cagaccccag ccatacaacc ctttctctag acctcagcca 4260  
 gacaaacctc tctccagaac tcagtcagac aaacctttcc ccagacctca gtgagatgcc 4320  
 5 cctctttgca gatctcagtc aaattcccct taccacagac ctcgaccaga tgacactttc 4380  
 tccagacctt ggtgagacag atctttcccc aaactttggt cagatgtccc tttccccaga 4440  
 cctcagccag gtgactctct ctccagacat cagtgcacc acccttctcc cggatctcag 4500  
 10 ccagatatca cctcctccag accttgatca gatattctac ctttctgaat ctagtctcagc 4560  
 attgcttctt caagaattta atgagtcttt tccttatcca gaccttggtc agatgccatc 4620  
 tccttcatct cctactctca atgatacttt tctatcaaag gaatttaatc cactgggtat 4680  
 15 agtgggcctc agtaaagatg gtacagatta cattgagatc attccaaagg aagaggtcca 4740  
 gagcagtga gatgactatg ctgaaattga ttatgtgcc tatgatgacc cctacaaaac 4800  
 20 tgatgttagg acaaacatca actcctccag agatcctgac aacattgcag catgggtacct 4860  
 ccgagcaac aatggaaaca gaagaaatta ttacattgct gctgaagaaa tatcctggga 4920  
 ttattcagaa tttgtacaaa gggaaacaga tattgaagac tctgatgata ttccagaaga 4980  
 25 taccacatat aagaaagtag tttttcgaaa gtacctgac agcactttta ccaaactgga 5040  
 tcctcgaggg gagtatgaag agcatctcgg aattcttggt cctattatca gagctgaagt 5100  
 ggatgatggt atccaagttc gttttaaaaa tttagcatcc agaccgtatt ctctacatgc 5160  
 30 ccatggactt tcctatgaaa aatcatcaga gggaaagact tatgaagatg actctcctga 5220  
 atggtttaag gaagataatg ctgttcagcc aaatagcagt tatacctacg tatggcatgc 5280  
 cactgagcga tcagggccag aaagtcctgg ctctgcctgt cgggcttggg cctactactc 5340  
 35 agctgtgaac ccagaaaaag atattcactc aggcttgata ggtcccctcc taatctgcca 5400  
 aaaaggaata ctacataagg acagcaacat gcctatggac atgagagaat ttgtcttact 5460  
 atttatgacc tttgatgaaa agaagagctg gtactatgaa aagaagtccc gaagtcttg 5520  
 40 gagactcaca tcctcagaaa tgaaaaaatc ccatgagttt cagccatta atgggatgat 5580  
 ctacagcttg cctggcctga aatgtatga gcaagagtgg gtgaggttac acctgctgaa 5640  
 45 cataggcggc tccaagaca ttcacgtggt tcactttcac ggccagacct tgctggaaaa 5700  
 tggcaataaa cagcaccagt taggggtctg gcccttctg cctggttcat ttaaaactct 5760  
 tgaaatgaag gcatcaaac ctggctggtg gtcctaaac acagagggtg gagaaaacca 5820  
 50 gagagcaggg atgcaaacgc ctttcttat catggacaga gactgtagga tgccaatggg 5880  
 actaagcact ggtatcatat ctgattcaca gatcaaggct tcagagtttc tgggttactg 5940  
 ggagcccaga ttagcaagat taacaatgg tggatcttat aatgcttggg gtgtagaaaa 6000  
 55 acttgagca gaatttgctt ctaaacttg gatccaggtg gacatgcaa aggaagtcat 6060  
 aatcacaggg atccagacct aaggtgcaa aactacctg aagtcctgct ataccacaga 6120  
 gttctatgta gcttacagtt ccaaccagat caactggcag atcttcaaag ggaacagcac 6180  
 60 aaggaatggt atgtatttta atggcaattc agatgcctct acaataaaag agaatcagtt 6240

65

ES 2 525 545 T3

	tgaccacct attgtggcta gatatattag gatctctcca actcgagcct ataacagacc	6300
	tacccttcga ttggaactgc aaggttgtga ggtaaattgga tgtccacac ccctgggtat	6360
5	ggaaaatgga aagatagaaa acaagcaaat cacagcttct tcgtttaaga aatcttggtg	6420
	gggagattac tgggaaccct tccgtgcccg tctgaatgcc cagggacgtg tgaatgcctg	6480
	gcaagccaag gcaacaaca ataagcagtg gctagaaatt gatctactca agatcaagaa	6540
10	gataacggca attataacac agggctgcaa gtctctgtcc tctgaaatgt atgtaaagag	6600
	ctataccatc cactacagtg agcagggagt ggaatggaaa ccatacaggc tgaaatcctc	6660
	catggtggac aagatTTTTG aaggaaatac taataccaaa ggacatgtga agaactTTTT	6720
15	caacccccca atcatttcca ggtttatccg tgcattcct aaaacatgga atcaaagtat	6780
	tgcacttcgc ctggaactct ttggctgtga tatttactag aattgaacat tcaaaaacc	6840
20	ctggaagaga ctctttaaga cctcaaacca tttagaatgg gcaatgtatt ttacgctgtg	6900
	ttaaattgta acagttttcc actatttctc tttcttttct attagtgaat aaaattttat	6960
	ac	6962
25	<210> 68 <211> 1464 <212> ADN <213> Homo sapiens	
30	<400> 68	
35		
40		
45		
50		
55		
60		
65		

ES 2 525 545 T3

agccccaagc ttaccacctg cacccggaga gctgtgtcac catgtgggtc cgggttgtct 60  
 tcctcacct gtccgtgacg tggattggtg ctgcaccct catcctgtct cggattgtgg 120  
 5 gaggctggga gtgcgagaag cattcccaac cctggcaggt gcttgtggcc tctcgtggca 180  
 gggcagtctg cggcgggtgt ctgggtcacc cccagtgggt cctcacagct gccactgca 240  
 tcaggaacaa aagcgtgatc ttgctgggtc ggcacagcct gtttcatcct gaagacacag 300  
 10 gccaggattt tcaggctcagc cacagcttcc cacacccgct ctacgatatg agcctcctga 360  
 agaatcgatt cctcaggcca ggtgatgact ccagccacga cctcatgctg ctccgcctgt 420  
 cagagcctgc cgagctcacg gatgctgtga aggtcatgga cctgcccacc caggagccag 480  
 15 cactggggac cacctgctac gcctcaggct ggggcagcat tgaaccagag gagttcttga 540  
 ccccaaagaa acttcagtgt gtggacctcc atgttatttc caatgacgtg tgtgcgcaag 600  
 20 ttcaccctca gaaggtgacc aagttcatgc tgtgtgctgg acgctggaca gggggcaaaa 660  
 gcacctgctc ggggtattct gggggccac ttgtctgtaa tgggtgtgctt caaggtatca 720  
 cgtcatgggg cagtgaacca tgtgccctgc ccgaaaggcc ttccctgtac accaaggtgg 780  
 25 tgcattaccg gaagtggatc aaggacacca tcgtggccaa cccctgagca cccctatcaa 840  
 cccctattg tagtaaactt ggaaccttgg aatgaccag gccaaagactc aagcctcccc 900  
 agttctactg acctttgtcc ttaggtgtga ggtccagggt tgctaggaaa agaaatcagc 960  
 30 agacacaggt gtagaccaga gtgtttctta aatgggtgtaa ttttgtcctc tctgtgtcct 1020  
 ggggaatact ggccatgcct ggagacatat cactcaatth ctctgaggac acagatagga 1080  
 35 tgggggtgtct gtgttatttg tgggttacag agatgaaaga ggggtgggat ccacactgag 1140  
 agagtggaga gtgacatgtg ctggacactg tccatgaagc actgagcaga agctggaggc 1200  
 acaacgcacc agacactcac agcaaggatg gagctgaaaa cataaccac tctgtcctgg 1260  
 40 aggcactggg aagcctagag aaggctgtga gccaaaggagg gagggctctc ctttggcatg 1320  
 ggatggggat gaagtaagga gagggactgg accccctgga agctgattca ctatgggggg 1380  
 aggtgtattg aagtcctcca gacaaccctc agatttgatg atttcctagt agaactcaca 1440  
 45 gaaataaaga gctgttatac tgtg 1464

<210> 69  
 <211> 1793  
 <212> ADN  
 50 <213> Homo sapiens  
 <400> 69

55

60

65

ES 2 525 545 T3

cgcgctccgcc ccgcgagcac agagcctcgc ctttgccgat ccgcccggccg tccacacccg 60  
 ccgccagctc accatggatg atgatatcgc cgcgctcgtc gtcgacaacg gctccggcat 120  
 5 gtgcaaggcc ggcttcgcgg gcgacgatgc cccccgggcc gtcttcccct ccatcgtggg 180  
 gcgccccagg caccagggcg tgatgggtggg catgggtcag aaggattcct atgtgggcga 240  
 cgaggcccag agcaagagag gcatacctcac cctgaagtac cccatcgagc acggcatcgt 300  
 10 caccaactgg gacgacatgg agaaaatctg gcaccacacc ttctacaatg agctgcgtgt 360  
 ggctcccag gagcaccgcc tgctgctgac cgaggcccc ctgaaccca aggccaaccg 420  
 cgagaagatg acccagatca tgtttgagac cttcaacacc ccagccatgt acgttgctat 480  
 15 ccaggctgtg ctatccctgt acgcctctgg ccgtaccact ggcatcgtga tggactccgg 540  
 tgacggggtc acccactctg tgcccatacga cgaggggtat gccctcccc atgccatcct 600  
 20 gcgtctggac ctggctggcc gggacctgac tgactacctc atgaagatcc tcaccgagcg 660  
 cggctacagc ttcaccacca cggccgagcg ggaaatcgtg cgtgacatta aggagaagct 720  
 gtgctacgtc gccctggact tcgagcaaga gatggccacg gctgcttcca gctcctcct 780  
 25 ggagaagagc tacgagctgc ctgacggcca ggtcatcacc attggcaatg agcggttccg 840  
 ctgccctgag gactcttcc agccttcctt cctgggcatg gactcctgtg gcatccacga 900  
 aactacctc aactccatca tgaagtgtga cgtggacatc cgcaaagacc tgtacgcca 960  
 30 cacagtgtg tctggcggca ccaccatgta ccctggcatt gccgacagga tgcagaagga 1020  
 gatcactgcc ctggcaccga gcacaatgaa gatcaagatc attgctctc ctgagcgcaa 1080  
 35 gtactccgtg tggatcggcg gctccatcct ggcctcgtg tccaccttcc agcagatgtg 1140  
 gatcagcaag caggagatg acgagtcgg cccctccatc gtccaccgca aatgcttcta 1200  
 ggcggactat gacttagttg cgttacacc tttcttgaca aaacctaact tgcgcagaaa 1260  
 40 acaagatgag attggcatgg ctttatttgt tttttttgtt ttgttttggg ttttttttt 1320  
 tttttggctt gactcaggat ttaaaaactg gaacggtgaa ggtgacagca gtcggttgga 1380  
 gcgagcatcc ccaaagttc acaatgtggc cgaggacttt gattgcacat tgttgttttt 1440  
 45 ttaatagtca ttccaaatat gagatgcatt gttacaggaa gtcccttgcc atcctaaaag 1500  
 ccacccact tctctctaag gagaatggcc cagtcctctc ccaagtccac acaggggagg 1560  
 tgatagcatt gctttcgtgt aaattatgta atgcaaaatt tttttaatct tcgccttaat 1620  
 50 acttttttat tttgttttat tttgaatgat gagccttcgt gccccccctt ccccctttt 1680  
 gtcccccaac ttgagatgta tgaaggcttt tggctctcct gggagtggtt ggaggcagcc 1740  
 55 agggcttacc tgtactactga cttgagacca gttgaataaa agtgcacacc tta 1793

<210> 70  
 <211> 1526  
 <212> ADN  
 60 <213> Homo sapiens

<400> 70

65

ES 2 525 545 T3

5 ccggaagtga cgcgaggctc tgcggagacc aggagtcaga ctgtaggacg acctcgggtc 60  
 ccacgtgtcc ccggtactcg ccggccggag cccccggctt cccggggccg ggggacctta 120  
 gcggcacca cacacagcct actttccaag cggagccatg tctggtaacg gcaatgcggc 180  
 10 tgcaacggcg gaagaaaaca gcccaaagat gagagtgatt cgcgtgggta cccgcaagag 240  
 ccagcttgct cgcatacaga cggacagtgt ggtggcaaca ttgaaagcct cgtaccctgg 300  
 cctgcagttt gaaatcattg ctatgtccac cacaggggac aagattcttg atactgcaact 360  
 15 ctctaagatt ggagagaaaa gcctgtttac caaggagcct gaacatgcc tggagaagaa 420  
 tgaagtggac ctggttgttc actccttgaa ggacctgcc actgtgcttc ctctggctt 480  
 caccatcgga gccatctgca agcgggaaaa ccctcatgat gctgttgtct ttcacccaaa 540  
 20 atttgttggg aagaccctag aaaccctgcc agagaagagt gtggtgggaa ccagctccct 600  
 gcgaagagca gccagctgc agagaaagtt cccgcatctg gagttcagga gtattcgggg 660  
 25 aaacctcaac acccggcttc ggaagctgga cgagcagcag gagttcagtg ccatcatcct 720  
 ggcaacagct ggccctgcagc gcatgggctg gcacaaccgg gtggggcaga tcctgcaccc 780  
 tgaggaatgc atgtatgctg tgggccaggg ggccctgggc gtggaagtgc gagccaagga 840  
 30 ccaggacatc ttggatctgg tgggtgtgct gcacgatccc gagactctgc ttcgctgcat 900  
 cgctgaaagg gccttcctga ggcacctgga aggaggctgc agtgtgccag tagccgtgca 960  
 35 tacagctatg aaggatgggc aactgtacct gactggagga gtctggagtc tagacggctc 1020  
 agatagcata caagagacca tgcaggctac catccatgtc cctgcccagc atgaagatgg 1080  
 ccctgaggat gaccacagt tggtaggcat cactgctcgt aacattccac gagggcccca 1140  
 40 gttggctgcc cagaacttg gcatcagcct ggccaacttg ttgctgagca aaggagccaa 1200  
 aaacatcctg gatgttgac ggcagcttaa cgatgcccat taactggttt gtggggcaca 1260  
 45 gatgcctggg ttgctgctgt ccagtgcta catcccgggc ctcagtgccc cattctcact 1320  
 gctatctggg gagtgattac cccgggagac tgaactgcag ggttcaagcc ttccagggat 1380  
 ttgcctcacc ttggggcctt gatgactgcc ttgcctctc agtatgtggg ggcttcatct 1440  
 50 ctttagagaa gtccaagcaa cagccttga atgtaaccaa tcctactaat aaaccagttc 1500  
 tgaagtgta aaaaaaaaa aaaaaa 1526

55 <210> 71  
 <211> 2397  
 <212> ADN  
 <213> Homo sapiens

60 <400> 71

65



ES 2 525 545 T3

5 gcaagaactg aaacgaatgg ggattgaact gctttgcctg ttctttctat ttctaggaag 60  
 gaatgatcac gtacaagggtg gctgtgcccct gggagggtgca gaaacctgtg aagactgcct 120  
 gcttattgga cctcagtgtg cctgggtgtgc tcaggagaat ttactcatc catctggagt 180  
 10 tggcgaaagg tgtgataccc cagcaaacct tttagctaaa ggatgtcaat taaacttcat 240  
 cgaaaacct gtctccaag tagaaatact taaaaataag cctctcagtg taggcagaca 300  
 gaaaaatagt tctgacattg ttcagattgc gcctcaaagc ttgatcctta agttgagacc 360  
 15 aggtggtgcg cagactctgc aggtgcatgt ccgccagact gaggactacc cgggtgattt 420  
 gtattacctc atggacctct ccgcctccat ggatgacgac ctcaacacaa taaaggagct 480  
 20 gggctcccgg ctttccaaag agatgtctaa attaaccagc aactttagac tgggcttcgg 540  
 atcttttgtg gaaaaacctg tatccccctt cgtgaaaaca acaccagaag aaattgccaa 600  
 cccttgcaat agtattccat acttctgttt acctacattt ggattcaagc acattttgcc 660  
 25 attgacaaat gatgctgaaa gattcaatga aattgtgaag aatcagaaaa tttctgctaa 720  
 tattgacaca cccgaagggtg gatttgatgc aattatgcaa gctgctgtgt gtaaggaaaa 780  
 aattggctgg cggaatgact ccctccacct cctggctctt gtgagtgatg ctgatttca 840  
 30 ttttggaatg gacagcaaac tagcaggcat cgtcattcct aatgacgggc tctgtcactt 900  
 ggacagcaag aatgaatact ccatgtcaac tgtcttgaa tatccaacaa ttggacaact 960  
 35 cattgataaa ctggtacaaa acaacgtgtt attgatctt gctgtaacc agaacaagt 1020  
 tcatttatat gagaattacg caaaacttat tcctggagct acagtaggtc tacttcagaa 1080  
 ggactccgga aacattctcc agctgatcat ctcagcttat gaagaactgc ggtctgaggt 1140  
 40 ggaactggaa gtattaggag aactgaagg actcaacttg tcatttacag ccatctgtaa 1200  
 caacggtacc ctctccaac accaaaagaa atgctctcac atgaaagtgg gagacacagc 1260  
 45 ttccttcagc gtgactgtga atatcccaca ctgagagaga agaagcaggc acattatcat 1320  
 aaagcctgtg gggctggggg atgccctgga attacttgtc agcccagaat gcaactgcga 1380  
 ctgtcagaaa gaagtggaag tgaacagctc caaatgtcac cacgggaacg gctctttcca 1440  
 50 gtgtgggggtg tgtgcctgcc accctggcca catggggcct cgctgtgagt gtggcgagga 1500  
 catgctgagc acagattcct gcaaggaggc cccagatcat ccctcctgca gcggaagggg 1560  
 tgactgctac tgtgggcagt gtatctgcca cttgtctccc tatggaaaca tttatgggcc 1620  
 55 ttattgccag tgtgacaatt tctcctgcgt gagacacaaa gggctgctct gcggaggtaa 1680  
 cggcgactgt gactgtggtg aatgtgtgtg caggagcggc tggactggcg agtactgcaa 1740  
 60 ctgcaccacc agcacggact cctgcgtctc tgaagatgga gtgctctgca gcgggcgagg 1800

65

ES 2 525 545 T3

5 ggactgtggt tgtggcaagt gtgtttgcac aaaccctgga gcctcaggac caacctgtga 1860  
 acgatgtcct acctgtgggtg acccctgtaa ctctaaacgg agctgcattg agtgccacct 1920  
 gtcagcagct ggccaagccc gagaagaatg tgtggacaag tgcaaactag ctggtgcgac 1980  
 catcagtga gaagaagatt tctcaaagga tggttctgtt tcctgctctc tgcaaggaga 2040  
 10 aatgaatgt cttattacat tcctaataac tacagataat gaggggaaaa ccatcattca 2100  
 cagcatcaat gaaaaagatt gtccgaagcc tccaaacatt cccatgatca tgtaggggt 2160  
 ttccctggct attcttctca tcggggttgt cctactgtgc atctggaagc tactggtgtc 2220  
 15 atttcatgat cgtaaagaag ttgccaaatt tgaagcagaa cgatcaaaag ccaagtggca 2280  
 aacgggaacc aatccactct acagaggatc cacaagtact tttaaaaatg taacttataa 2340  
 20 acacagggaa aaacaaaagg tagaccttc cacagattgc tagaactact ttatgca 2397

25 <210> 72  
 <211> 2118  
 <212> ADN  
 <213> Homo sapiens  
 <400> 72

30

35

40

45

50

55

60

65

ES 2 525 545 T3

	tggggagccc aagcagaaac gcaagctggt ggctgaggtg tccctgcaga acccgctccc	60
	tgtggccctg gaaggetgca ctttactgt ggagggggcc ggcctgactg aggagcagaa	120
5	gacggtggag atcccagacc ccgtggaggc aggggaggaa gttaaggtga gaatggacct	180
	gctgccgctc cacatgggccc tccacaagct ggtggtgaac ttcgagagcg acaagctgaa	240
10	ggctgtgaag ggcttccgga atgtcatcat tggccccgcc taagggacct ctgctcccag	300
	cctgctgaga gccccacct tgatcccaat ctttatccca agctagttag caaaatatgc	360
	cccttcttgg gccccagacc ccagggcagg gtgggcagcc tatgggggct ctcggaaatg	420
15	gaatgtgccc ctggcccatc tcagcctcct gagcctgtgg gtccccactc accccctttg	480
	ctgtgaggaa tgctctgtgc cagaaacagt gggagccctg accttggctg actggggctg	540
	gggtgagaga ggaaagacct acattccctc tcctgcccag atgccctttg gaaagccatt	600
20	gaccaccac catattgttt gatctacttc atagctcctt ggagcaggca aaaaaggac	660
	agcatgcccc ttggctggat cagggaatcc agctccctag actgcatccc gtacctctc	720
25	ccatgactgc acccagctcc aggggccctt gggacagcca gagctgggtg gggacagtga	780
	taggccaag gtcccccca catcccagca gcccaagctt aatagccctc cccctcaacc	840
	tcaccattgt gaagcaccta ctatgtgctg ggtgcctccc acacttgctg gggctcacgg	900
30	ggcctccaac ccatttaatc accatgggaa actgttgtgg gcgctgcttc caggataagg	960
	agactgaggc ttagagagag gaggcagccc cctccacacc agtggcctcg tggttattag	1020
35	caaggctggg taatgtgaag gcccaagagc agagtctggg cctctgactc tgagtccact	1080
	gctccattta taaccccagc ctgacctgag actgtcggag aggctgtctg gggcctttat	1140
	caaaaaaaga ctgagccaag acaaggaggt agagagggga ctgggggact gggagtcaga	1200
40	gccctggctg ggttcaggtc ccacgtctgg ccaggcactg cttctcctc tctgggcctt	1260
45		
50		
55		
60		
65		

ES 2 525 545 T3

	tgtttccttg ttggtcagag gagtgattga accagctcat ctccaaggat cctctccact	1320
	ccatgtttgc aatgctttta tatggcccag ccttgtaa at aaccacaagg tccactccct	1380
5	gctccacgaa gccttaagcc ataggcccag gatatttctg agagtgaaac catgactgtg	1440
	accaccttct gtccccagcc ctgtcctggg tccttcctat gcccaggtag cacccttcag	1500
10	accccagttc taggggagaa gagccctgga caccctgct ctacccatga gcctgcccgc	1560
	tgcaatgcct agacttccca acagccttag ctgccagtgc tggtcactaa ccaacaaggt	1620
	tggcacccca gctaccctt ctttgcaggg ctaaggcccc caaacatagc ccctgccccg	1680
15	gaggaagctt ggggaacca tgagttgtca gctttgactt tatctcctgc tctttctaca	1740
	tgactgggcc tcccttgggc tggaagaatt ggggattctc tattggaggt gagatcacag	1800
	cctccagggc cccccaaatc ccaggggaagg acttggagag aatcatgctg ttgcatttag	1860
20	aactttctgc tttgcacagg aaagagtcac acaattaatc aacatgtata ttttctctat	1920
	acatagagct ctatttctct acggttttat aaaagccttg ggttccaacc aggcagtaga	1980
25	tgtgcttctg aaccgcaagg agcaaacact gaaataaaat agtttatttt tcacactcaa	2040
	aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	2100
	aaaaaaaaa aaaaaaaaaa	2118
30	<210> 73 <211> 2832 <212> ADN <213> Homo sapiens	
35	<400> 73	
40		
45		
50		
55		
60		
65		

ES 2 525 545 T3

	aaagctcaaa	ccgacaccct	cacgcagatg	atgacatcaa	ctcttttttc	ttccccaagt	60
	gtacacaatg	tgatggagac	tgttacgcag	gagacagctc	ctccagatga	aatgaccaca	120
5	tcattttccct	ccagtgtcac	caacacactc	atgatgacat	caaagactat	aacaatgaca	180
	acctccacag	actccactct	tggaaacaca	gaagagacat	caacagcagg	aactgaaagt	240
	tctaccccag	tgacctcagc	agtctcaata	acagctggac	aggaaggaca	atcacgaaca	300
10	acttcctgga	ggacctctat	ccaagacaca	tcagcttctt	ctcagaacca	ctggactcgg	360
	agcacgcaga	ccaccaggga	atctcaaacc	agcaccctaa	cacacagaac	cacttcaact	420
15	cctttctttct	ctccaagtgt	acacaatgtg	acagggactg	tttctcagaa	gacatctcct	480
	tcaggtgaaa	cagctacctc	atccctctgt	agtgtcacia	acacatccat	gatgacatca	540
	gagaagataa	cagtgacaac	ctccacaggc	tccactcttg	gaaaccagg	ggagacatca	600
20	tcagtacctg	ttactggaag	tcttatgcca	gtcacctcag	cagccttagt	aacagttgat	660
	ccagaaggac	aatcaccagc	aactttctca	aggacttcta	ctcaggacac	aacagctttt	720
25	tctaagaacc	accagactca	gagcgtggag	accaccagag	tatctcaaat	caacaccctc	780
	aacaccctca	caccggttac	aacatcaact	gttttatcct	caccaagtgg	attcaaccca	840
	agtggaacag	tttctcagga	gacattccct	tctggtgaaa	caaccatctc	atccccttcc	900
30	agtgtcagca	atacattcct	ggtaacatca	aagggtgttca	gaatgccaat	ctccagagac	960
35							
40							
45							
50							
55							
60							
65							

ES 2 525 545 T3

	tctactcttg gaaacacaga ggagacatca ctatctgtaa gtggaacat ttctgcaatc	1020
	acttccaaag tttcaacat atggtggtca gacactctgt caacagcact ctccccagt	1080
5	tctctacctc caaaaatata cacagctttc cacaccagc agagtgaagg tgcaagacc	1140
	acaggacggc ctcatgagag gagctcattc tctccagggtg tgtctcaaga aatatttact	1200
	ctacatgaaa caacaacatg gccttcctca ttctccagca aaggccacac aacttggca	1260
10	caaacagAAC tgccctcaac atcaacaggt gctgccacta ggcttgtcac aggaaatcca	1320
	tctacagggA cagctggcac tattccaagg gtccccctta aggtctcagc aataggggaa	1380
	ccaggagagc ccaccacata ctctcccac agcacaactc tccccaaaac aacaggggca	1440
15	ggcggccaga cacaatggac acaagaaacg gggaccactg gagaggctct tctcagcagc	1500
	ccaagctaca gtgtgactca gatgataaaa acggccacat ccccatcttc ttcacctatg	1560
20	ctggatagac acacatcaca acaaattaca acggcaccat caacaaatca ttcaacaata	1620
	cattccacaa gcacctctcc tcaggaatca ccagctgttt cccaaagggg tcacactcaa	1680
	gccccgcaga ccacacaaga atcacaaacc acgagggtccg tctccccat gactgacacc	1740
25	aagacagtca ccaccccagg ttcttccttc acagccagtg ggcactcgc ctcagaaatt	1800
	gttcctcagg acgcaccac cataagtgca gcaacaacct ttgccccagc tcccaccggg	1860
	gatggtcaca caaccaggc cccgaccaca gcaactgcagg cagcaccag cagccatgat	1920
30	gccaccctgg ggccctcagg aggcacgtca ctttccaaa caggtgccct tactctggcc	1980
	aactctgtag tgtcaacacc agggggcca gaaggacaat ggacatcagc ctctgccagc	2040
	acctcacctg acacagcagc agccatgacc catacccacc aggtgagag cacagaggcc	2100
35	tctggacaaa cacagaccag cgaaccggcc tcctcagggt cacgaaccac ctcagcgggc	2160
	acagctacc cttctctatc cggggcgagt ggcacaacac cttcaggaag cgaaggaata	2220
40	tccacctcag gagagacgac aagggtttca tcaaaccct ccagggacag tcacacaacc	2280
	cagtcaacaa ccgaattgct gtccgcctca gccagtcattg gtgccatccc agtaagcaca	2340
	ggaatggcgt cttcgatcgt ccccgccacc tttcatcca ccctctctga ggcctccact	2400
45	gcagggagac cgacaggaca gtcaagcca acttctcca gtgcctctcc tcaggagaca	2460
	gccgccattt cccgatggc ccagactcag aggacaagaa ccagcagagg gtctgacact	2520
	atcagcctgg cgtcccaggc aaccgacacc ttctcaacag tcccaccac acctccatcg	2580
50	atcacatcca ctgggcttac atctccaca acccagacc acactctgtc accttcaggg	2640
	tctggtaaaa cttcaccac ggcctctatc agcaacgcca cccctcttc tgtcacctac	2700
	gcttctcgg catccacagg tcacaccacc cctcttcatg tcaccgatgc ttcctcagta	2760
55	tccacaggtc acgccacccc tcttctgtc accagccctt cctcagtatc cacaggtcac	2820
	accaccctc tt	2832
60	<210> 74 <211> 1607 <212> ADN <213> Homo sapiens	
65		

ES 2 525 545 T3

<400> 74

5     aatgactcct ttcggtaagt gcagtggaag ctgtacactg cccaggcaaa gcgtccgggc     60  
        agcgtaggcg ggcgactcag atcccagcca gtggacttag cccctgtttg ctccctccgat   120  
        aactgggggtg accttgggta atattcacca gcagcctccc ccgttgcccc tctggatcca   180  
 10     ctgcttaaat acggacgagg acagggccct gtctcctcag cttcaggcac caccactgac   240  
        ctgggacagt gaatcgacaa tgccgtcttc tgtctcgtgg ggcacccctc tgctggcagg   300  
        cctgtgctgc ctggctcctg tctccctggc tgaggatccc cagggagatg ctgccagaa   360  
 15     gacagataca tcccaccatg atcaggatca cccaaccttc aacaagatca cccccaacct   420  
        ggctgagttc gccttcagcc tataaccgcca gctggcacac cagtccaaca gcaccaatat   480  
        cttcttctcc ccagtgagca tcgctacagc ctttgcaatg ctctccctgg ggaccaaggc   540  
 20     tgacactcac gatgaaatcc tggagggcct gaatttcaac ctcacggaga ttccggaggc   600  
        tcagatccat gaaggcttcc aggaactcct ccgtagcctc aaccagccag acagccagct   660  
        ccagctgacc accggcaatg gcctgttctt cagcgagggc ctgaagctag tggataagtt   720  
 25     tttgagggat gttaaaaagt tgtaccactc agaagccttc actgtcaact tcggggacac   780  
        cgaagaggcc aagaaacaga tcaacgatta cgtggagaag ggtactcaag ggaaaattgt   840  
        ggatttggtc aaggagcttg acagagacac agtttttgct ctgggtgaatt acatcttctt   900  
 30     taaaggcaaa tgggagagac cttttgaagt caaggacacc gaggaagagg acttccacgt   960  
        ggaccaggtg accaccgtga aggtgcctat gatgaagcgt ttaggcatgt ttaacatcca 1020  
        gcactgtaag aagctgtcca gctgggtgct gctgatgaaa tacctgggca atgccaccgc   1080  
 35     catcttcttc ctgcctgatg aggggaaact acagcacctg gaaaatgaac tcaccacga   1140  
        tatcatcacc aagttcctgg aaaatgaaga cagaaggtct gccagcttac atttacccaa   1200  
 40     actgtccatt actggaacct atgatctgaa gagcgtcctg ggtcaactgg gcatcactaa   1260  
        ggctcttcagc aatggggctg acctctccgg ggtcacagag gaggcacccc tgaagctctc   1320  
        caaggccgtg cataaggctg tgctgaccat cgacgagaaa gggactgaag ctgctggggc   1380  
 45     catgttttta gaggccatac ccatgtctat ccccccgag gtcaagttca acaaaccctt   1440  
        tgtcttctta atgattgaac aaaataccaa gtctcccctc ttcattggaa aagtggtgaa   1500  
 50     tcccaccaa aaataactgc ctctcgtcc tcaaccctc ccctccatcc ctggccccct   1560  
        ccctggatga cattaagaa gggttgagct ggtccctgcc tgcaaaa             1607

<210> 75

<211> 1753

<212> ADN

<213> Homo sapiens

<400> 75

60

65

ES 2 525 545 T3

cagccccgcc cctacctgtg gaagcccagc cgcccgctcc cgcgataaa aggcgcggag 60  
 tgtccccgag gtcagcgagt gcgcgctcct cctcgcccgc cgctaggtcc atccccggcc 120  
 5 agccaccatg tccatccact tcagctcccc ggtattcacc tcgcgctcag ccgccttctc 180  
  
 gggccgcggc gccaggtgc gcctgagctc cgctcgcccc ggcggccttg gcagcagcag 240  
 10 cctctacggc ctcgggcgct cacggccgcg cgtggccgtg cgctctgcct atggggggcc 300  
 ggtgggcgcc ggcattccgc aggtcaccat taaccagagc ctgctggccc cgctgcggct 360  
 ggacgccgac ccctccctcc agcgggtgcg ccaggaggag agcgagcaga tcaagaccct 420  
 15 caacaacaag tttgcctcct tcatcgacaa ggtgcggttt ctggagcagc agaacaagct 480  
 gctggagacc aagtggacgc tgctgcagga gcagaagtcg gccaagagca gccgcctccc 540  
 agacatcttt gaggcccaga ttgctggcct tcggggctcag cttgaggcac tgcaggtgga 600  
 20 tggggggccgc ctggaggcgg agctgcggag catgcaggat gtggtggagg acttcaagaa 660  
 taagtacgaa gatgaaatta accaccgcac agctgctgag aatgagtttg tggtgctgaa 720  
 gaaggatgtg gatgctgcct acatgagcaa ggtggagctg gaggccaagg tggatgcctt 780  
 25 gaatgatgag atcaacttcc tcaggaccct caatgagacg gagttgacag agctgcagtc 840  
 ccagatctcc gacacatctg tgggtgctgc catggacaac agtcgctccc tggacctgga 900  
 30 cggcatcatc gctgaggta aggcgcagta tgaggagatg gccaaatgca gccgggctga 960  
 ggctgaagcc tggtagcaga ccaagtttga gaccctccag gccaggctg ggaagcatgg 1020  
 ggacgacctc cggaataccc ggaatgagat ttcagagatg aaccgggcca tccagaggct 1080  
 35 gcaggctgag atcgacaaca tcaagaacca gcgtgccaaag ttggaggccg ccattgccga 1140  
 ggctgaggag cgtggggagc tggcgtcaa ggatgctcgt gccaaagcagg aggagctgga 1200  
 agccgccctg cagcggggca agcaggatat ggcacggcag ctgcgtgagt accaggaact 1260  
 40 catgagcgtg aagctggccc tggacatcga gatcgccacc taccgcaagc tgctggaggg 1320  
 cgaggagagc cggttggctg gagatggagt gggagccgtg aatatctctg tgatgaattc 1380  
 45 cactggtggc agtagcagtg gcggtggcat tgggctgacc ctcgggggaa ccatgggcag 1440  
 caatgccctg agcttctcca gcagtgcggg tcctgggctc ctgaaggctt attccatccg 1500  
 gaccgcatcc gccagtcgca ggagtgcccg cgactgagcc gcctcccacc actccactcc 1560  
 50 tccagccacc accacaatc acaagaagat tcccaccctt gcctcccatg cctggtccca 1620  
 agacagtgag acagtctgga aagtgatgtc agaatagctt ccaataaagc agcctcattc 1680  
 tgaggcctga gtgatccacg tgaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1740  
 55 aaaaaaaaaa aaa 1753

<210> 76  
 <211> 2255  
 60 <212> ADN  
 <213> Homo sapiens  
  
 <400> 76

65



ES 2 525 545 T3

gatggctccg gccgcctggc tccgcagcgc ggccgcgcgc gccctcctgc ccccgatgct 60  
gctgctgctg ctccagccgc cgccgctgct ggcccgggct ctgccgccgg acgcccacca 120  
5 cctccatgcc gagaggaggg ggccacagcc ctggcatgca gccctgcca gtagcccggc 180  
acctgccctt gccacgcagg aagccccccg gcctgccagc agcctcaggc ctccccgctg 240

10

15

20

25

30

35

40

45

50

55

60

65

ES 2 525 545 T3

	tggcgtgccc gacccatctg atgggctgag tgcccgcaac cgacagaaga ggttcgtgct	300
	ttctggcggg cgctgggaga agacggacct cacctacagg atccttcggt tcccatggca	360
5	gttggtgcag gagcaggtgc ggcagacgat ggagaggcc ctaaaggat ggagcgatgt	420
	gacgccactc acctttactg aggtgcacga gggccgtgct gacatcatga tcgacttcgc	480
	caggtactgg catggggacg acctgccgtt tgatgggcct gggggcatcc tggcccatgc	540
10	cttctcccc aagactcacc gagaagggga tgtccacttc gactatgatg agacctggac	600
	tatcggggat gaccagggca cagacctgct gcaggtggca gcccatgaat ttggccacgt	660
	gctggggctg cagcacacaa cagcagccaa ggccctgatg tccgccttct acaccttcg	720
15	ctacccactg agtctcagcc cagatgactg caggggcggt caacacctat atggccagcc	780
	ctggcccact gtcacctcca ggaccccagc cctgggcccc caggctggga tagacaccaa	840
20	tgagattgca ccgctggagc cagacgcccc gccagatgcc tgtgaggcct cctttgacgc	900
	ggtctccacc atccgaggcg agctcttttt cttaaagcg ggcttttgtg ggcgcctccg	960
	tgggggccag ctgcagcccc gctaccacagc attggcctct cgccactggc agggactgcc	1020
25	cagccctgtg gacgctgcct tcgaggatgc ccagggccac atttggttct tccaagggtc	1080
	tcagtactgg gtgtacgacg gtgaaaagcc agtcctgggc cccgcacccc tcaccgagct	1140
	gggcctggtg aggttcccgg tccatgctgc cttggtctgg ggtcccgaga agaacaagat	1200
30	ctacttcttc cgaggcaggg actactggcg tttccacccc agcaccgggc gtgtagacag	1260
	tcccgtgccc cgcagggcca ctgactggag aggggtgccc tctgagatcg acgctgcctt	1320
	ccaggatgct gatggctatg cctacttctc gcgcggccgc ctctactgga agtttgacc	1380
35	tgtgaaggty aaggctctgg aaggcttccc ccgtctcgtg ggtcctgact tctttgctg	1440
	tgccgagcct gccaacactt tcctctgacc atggcttggg tgccctcagg ggtgctgacc	1500
40	cctgccaggc cacgaatatac aggctagaga cccatggcca tctttgtggc tgtgggcacc	1560
	aggcatggga ctgagcccat gtctcctcag ggggatgggg tggggtacaa ccaccatgac	1620
	aactgccggg agggccacgc aggtcgtggt cacctgccag cgactgtctc agactgggca	1680
45	gggaggcttt ggcattgactt aagaggaagg gcagtcctgg gcccgctatg caggtcctgg	1740
	caaacctggc tgccctgtct ccatccctgt ccctcagggc agcaccatgg caggactggg	1800
	ggaactggag tgtccttgct gtatccctgt tgtgaggttc cttccagggg ctggcactga	1860
50	agcaagggtg ctggggcccc atggccttca gccctggctg agcaactggg ctgtagggca	1920
	gggccacttc ctgaggtcag gtcttggtag gtgcctgcat ctgtctgcct tctggctgac	1980
	aatcctggaa atctgttctc cagaatccag gccaaaaagt tcacagtcaa atggggaggg	2040
55	gtattcttca tgcaggagac cccaggccct ggaggctgca acatacctca atcctgtccc	2100
	aggccgatc ctctgaagc ccttttcgca gcactgctat cctccaaagc cattgtaaat	2160
60	gtgtgtacag tgtgtataaa ctttcttctt cttttttttt ttttaactg aggattgtca	2220
	ttaaacacag ttgttttcta aaaaaaaaaa aaaaa	2255

65

ES 2 525 545 T3

<210> 77  
 <211> 462  
 <212> ADN  
 <213> Homo sapiens  
 5  
 <400> 77  
 agctctattg ccaccatgag tttctccggc aagtaccaac tgcagagcca ggaaaacttt 60  
 10 gaagccttca tgaaggcaat cggctcgccg gaagagctca tccagaaggg gaaggatatac 120  
 aaggggggtgt cggaaatcgt gcagaatggg aagcacttca agttcaccat caccgctggg 180  
 tccaaagtga tccaaaacga attcacgggtg ggggaggaat gtgagctgga gacaatgaca 240  
 15 ggggagaaag tcaagacagt gggtcagttg gaaggtgaca ataaactggt gacaactttc 300  
 aaaaacatca agtctgtgac cgaactcaac ggcgacataa tcaccaatac catgacattg 360  
 20 ggtgacattg tcttcaagag aatcagcaag agaatttaaa caagtctgca tttcatatta 420  
 ttttagtgty taaaattaat gtaataaagt gaactttggt tt 462  
 <210> 78  
 <211> 2108  
 <212> ADN  
 <213> Homo sapiens  
 25  
 <400> 78  
 30  
 35  
 40  
 45  
 50  
 55  
 60  
 65

ES 2 525 545 T3

	gggaccgcct	cggaggcaga	agagccgcga	ggagccagcg	gagcaccgcg	ggctggggcg	60
5	cagccacccg	ccgctcctcg	agtcccctcg	cccccttccc	ttcgtgcccc	ccggcagcct	120
	ccagcgtcgg	tccccaggca	gcatgggtgag	gtctgtctccc	ggaccctcgc	caccatgtac	180
	gtgagctacc	tcctggacaa	ggacgtgagc	atgtacccta	gctccgtgcg	ccactctggc	240
10	ggcctcaacc	tggcgccgca	gaacttcgtc	agccccccgc	agtaccggga	ctacggcggg	300
	taccacgtgg	cggccgcagc	tgcagcggca	gcgaacttgg	acagcgcgca	gtccccgggg	360
	ccatcctggc	cggcagcgta	tggcgcccca	ctccgggagg	actggaatgg	ctacgcgccc	420
15	ggaggcgccg	cggccgcccg	caacgccgtg	gctcacggcc	tcaacggtgg	ctccccggcc	480
	gcagccatgg	gctacagcag	ccccgcagac	taccatccgc	accaccacc	gcatcaccac	540
	ccgcaccacc	cggccgcccg	gccttctctg	gcttctgggc	tgctgcaaac	gctcaacccc	600
20	ggcctcctcg	ggcccgccgc	caccgctgcc	gccgagcagc	tgtctcccgg	cggccagcgg	660
	cggaacctgt	gcgagtggat	gcggaagccg	gcgcagcagt	ccctcggcag	ccaagtgaaa	720
	accaggacga	aagacaaata	tcgagtggtg	tacacggacc	accagcggct	ggagctggag	780
25	aaggagtttc	actacagtcg	ctacatcacc	atccggagga	aagccgagct	agccgccacg	840
	ctggggctct	ctgagaggca	ggttaaaatc	tggtttcaga	accgcagagc	aaaggagagg	900
30	aaaatcaaca	agaagaagtt	gcagcagcaa	cagcagcagc	agccaccaca	gccgcctccg	960
	ccgccaccac	agcctcccca	gcctcagcca	ggtcctctga	gaagtgtccc	agagcccttg	1020
	agtccggtgt	cttcctgca	agcctcagtg	tctggctctg	tcctgggggt	tctggggcca	1080
35	actggggggg	tgctaaacc	caccgtcacc	cagtgacca	ccgggggtctg	cagcggcaga	1140
	gcaattccag	gctgagccat	gaggagcgtg	gactctgcta	gactcctcag	gagagacccc	1200

40

45

50

55

60

65

ES 2 525 545 T3

	tccccccca cccacagcca tagacctaca gacctggctc tcagaggaaa aatgggagcc	1260
	aggagtaaga caagtgggat ttggggcctc aagaaatata ctctcccaga tttttacttt	1320
5	ttcccatctg gctttttctg ccaactgagga gacagaaagc ctccgctggg cttcattccg	1380
	gactggcaga agcattgcct ggactgacca caccaaccag gccttcatcc tcctccccag	1440
	ctctttctctt cctagatctg caggctacac ctctggctag agccgagggg agagagggac	1500
10	tcaagggaaa ggcaagcttg aggccaagat ggctgctgcc tgctcatggc cctcggaggt	1560
	ccagctgggc ctctgcctc cgggcaggca aggtttacac tgcggaagcc aaaggcagct	1620
	aagatagaaa gctggactga ccaagactg cagaaccccc aggtggcctg cgtctttttt	1680
	ctcttccctt cccagaccag gaaaggcttg gctgggtgat gcacaggggtg tggatgagg	1740
	gggtggttat tggactccag gcctgaccag gggggccgaa cagggacttg tttagagagc	1800
20	ctgtcaccag agcttctctg ggctgaatgt atgtcagtgc tataaatgcc agagccaacc	1860
	tggacttctt gtcattttca caatcttggg gctgatgaag aagggggtgg ggggagtttg	1920
	tgttggtggt gctgctgttt gggttggttg tctgtgtaac atccaagcca gagtttttaa	1980
25	agccttctgg atccatgggg ggagaagtga tatggtgaag ggaagtgggg agtatttgaa	2040
	cacagttgaa ttttttctaa aaagaaaaag agataaatga gctttccaga aaaaaaaaaa	2100
30	aaaaaaaa	2108
	<210> 79	
	<211> 3745	
	<212> ADN	
35	<213> Homo sapiens	
	<400> 79	
40		
45		
50		
55		
60		
65		

ES 2 525 545 T3

	cgcaaagcaa	gtgggcacaa	ggagtatggt	tctaactgta	ttggggtcat	gaagacgttg	60
	ctgttgact	tggctttgtg	gtcactgctc	ttccagcccg	ggtggctgtc	ctttagttcc	120
5	caggtgagtc	agaactgcca	caatggcagc	tatgaaatca	gcgtcctgat	gatgggcaac	180
	tcagcctttg	cagagcccct	gaaaaacttg	gaagatgcgg	tgaatgaggg	gctggaaata	240
	gtgagaggac	gtctgcaaaa	tgctggccta	aatgtgactg	tgaacgctac	tttcatgtat	300
10	tcggatggtc	tgattcataa	ctcagggcag	tgccggagta	gcacctgtga	aggcctcgac	360
	ctactcagga	aaatttcaaa	tgcaaacgg	atgggctgtg	tcctcatagg	gccctcatgt	420
	acatactcca	ccttccagat	gtaccttgac	acagaattga	gctaccccat	gatctcagct	480
15	ggaagttttg	gattgtcatg	tgactataaa	gaaaccttaa	ccaggctgat	gtctccagct	540
	agaaagttga	tgtacttctt	ggttaacttt	tggaaacca	acgatctgcc	cttcaaaact	600
20	tattcctgga	gcacttcgta	tgtttacaag	aatggtacag	aaactgagga	ctgtttctgg	660
	taccttaatg	ctctggaggc	tagcgtttcc	tatttctccc	acgaactcgg	ctttaagggtg	720
	gtgtaagac	aagataagga	gtttcaggat	atcttaatgg	accacaacag	gaaaagcaat	780
25	gtgattatta	tgtgtggtgg	tccagagttc	ctctacaagc	tgaaggggta	ccgagcagtg	840
	gctgaagaca	ttgtcattat	tctagtggat	cttttcaatg	accagtactt	ggaggacaat	900
30							
35							
40							
45							
50							
55							
60							
65							

ES 2 525 545 T3

gtcacagccc ctgactatat gaaaaatgtc cttgttctga cgctgtctcc tgggaattcc 960  
 cttctaaata gctcttttctc caggaatcta tcaccaacaa aacgagactt tgctcttgcc 1020  
 5 tatttgaatg gaatcctgct ctttggacat atgctgaaga tatttcttga aatggagaa 1080  
 aatattacca cccccaat tgctcatgct ttcaggaatc tcacttttga aggttatgac 1140  
 ggtccagtga ccttggatga ctgggggat gttgacagta ccatgggtgct tctgtatacc 1200  
 10 tctgtggaca ccaagaaata caaggttctt ttgacctatg ataccacgt aaataagacc 1260  
 tatcctgtgg atatgagccc cacattcact tggagaact ctaaacttcc taatgatatt 1320  
 acaggccggg gccctcagat cctgatgatt gcagtcttca ccctcactgg agctgtggtg 1380  
 15 ctgctcctgc tcgtcgtctt cctgatgctc agaaaatata gaaaagatta tgaacttcgt 1440  
 cagaaaaaat ggtcccacat tcctcctgaa aatatctttc ctctggagac caatgagacc 1500  
 20 aatcatgtta gcctcaagat cgatgatgac aaaagacgag atacaatcca gagactacga 1560  
 cagtgcaaat acgacaaaaa gcgagtgatt ctcaaagatc tcaagcacia tgatggtaat 1620  
 ttactgaaa aacagaagat agaattgaac aagttgcttc agattgacta ttacaacctg 1680  
 25 accaagttct acggcacagt gaaactgat accatgatct tcggggtgat agaatactgt 1740  
 gagagaggat ccctccggga agttttaaata gacacaattt cctaccctga tggcacattc 1800  
 atggattggg agtttaagat ctctgtcttg tatgacattg ctaagggaat gtcatatctg 1860  
 30 cactccagta agacagaagt ccatggctcg ctgaaatcta ccaactgcgt agtggacagt 1920  
 agaatgggtg tgaagatcac tgattttggc tgcaattcca ttttacctcc aaaaaaggac 1980  
 ctgtggacag ctccagagca cctccgcaa gccaacatct ctcagaaagg agatgtgtac 2040  
 35 agctatggga tcatcgcaca ggagatcatt ctgcggaaag aaaccttcta cactttgagc 2100  
 tgtcgggacc ggaatgagaa gattttcaga gtggaaaatt ccaatggaat gaaaccttc 2160  
 cgccagatt tattcttggg aacagcagag gaaaaagagc tagaagtga cctacttgta 2220  
 40 aaaaactggt gggaggaaga tccagaaaag agaccagatt tcaaaaaaat tgagactaca 2280  
 cttgccaaaga tatttggact ttttcatgac caaaaaatg aaagctatat ggataccttg 2340  
 45 atccgacgtc tacagctata ttctcgaac ctggaacatc tggtagagga aaggacacag 2400  
 ctgtacaagg cagagagga cagggctgac agacttaact ttatgttgct tccaaggcta 2460  
 gtggtaaagt ctctgaagga gaaaggctt gtggagccg aactatatga ggaagttaca 2520  
 50 atctacttca gtgacattgt aggtttcact actatctgca aatacagcac ccccatggaa 2580  
 gtggtggaca tgcttaatga catctataag agttttgacc acattgttga tcatcatgat 2640  
 gtctacaagg tggaaacat cggtgatgcy tacatggtgg ctagtgttt gcctaagaga 2700  
 55 aatggcaatc ggcattgcaat agacattgcc aagatggcct tggaaatcct cagcttcag 2760  
 gggaccttg agctggagca tcttctggc ctccaatat ggattcgcat tggagttcac 2820  
 60 tctggtccct gtgctgctgg agttgtggga atcaagatgc ctcgttattg tctatttggg 2880  
 gatacggtca acacagcctc taggatggaa tccactggcc tcctttgag aattcacgtg 2940

65

ES 2 525 545 T3

agtggctcca ccatagccat cctgaagaga actgagtgcc agttccttta tgaagtgaga 3000  
 ggagaaacat acttaaaggg aagaggaaat gagactacct actggctgac tgggatgaag 3060  
 5 gaccagaaat tcaacctgcc aaccctcct actgtggaga atcaacagcg tttgcaagca 3120  
 gaattttcag acatgattgc caactcttta cagaaaagac aggcagcagg gataagaagc 3180  
 caaaaacca gacgggtagc cagctataaa aaaggcactc tggataactt gcagctgaat 3240  
 10 accacagaca aggagagcac ctatttttaa acctaaatga ggtataagga ctcacacaaa 3300  
 ttaaaataca gctgcactga ggcagcgacc tcaagtgtcc tgaaagctta cattttcctg 3360  
 agacctcaat gaagcagaaa tgtacttagg cttggctgcc ctgtctggaa catggacttt 3420  
 15 cttgcatgaa tcagatgtgt gttctcagtg aaataactac cttccactct ggaaccttat 3480  
 tccagcagtt gttccagga gcttctacct ggaaaagaaa agaatgaat agactatcta 3540  
 gaacttgaga agattttatt cttatttcat ttattttttg tttgtttatt tttatcgttt 3600  
 20 ttgtttactg gctttccttc tgtattcata agatttttta aattgtcata attatatttt 3660  
 aaatacccat cttcattaa gtatatttaa ctcataattt ttgcagaaaa tatgctatat 3720  
 25 attaggcaag aataaaagct aaagg 3745

30 <210> 80  
 <211> 901  
 <212> ADN  
 <213> Homo sapiens  
 <400> 80

35

40

45

50

55

60

65



ES 2 525 545 T3

agccccaac tcaccacctg gccgtggaca cctgtgtcag catgtgggac ctggttctct 60  
 ccatcgccctt gtctgtgggg tgcactggtg ccgtgcccct catccagtct cggattgtgg 120  
 5 gaggctggga gtgtgagaag cattccaac cctggcaggt ggctgtgtac agtcatggat 180  
 gggcacactg tgggggtgtc ctggtgcacc cccagtgggt gtcacagct gccattgccc 240  
 taaagaagaa tagccaggtc tggctgggtc ggcacaacct gtttgagcct gaagacacag 300  
 10 gccagagggt ccctgtcagc cacagcttcc cacacccgct ctacaatatg agccttctga 360  
 agcatcaaag ccttagacca gatgaagact ccagccatga cctcatgctg cttcgccctgt 420  
 cagagcctgc caagatcaca gatgttgtga aggtcctggg cctgcccacc caggagccag 480  
 15 cactggggac cacctgctac gcctcaggct ggggcagcat cgaaccagag gagttcttgc 540  
 gccccaggag tcttcagtgt gtgagcctcc atctcctgtc caatgacatg tgtgctagag 600  
 cttactctga gaaggtgaca gagttcatgt tgtgtgctgg gctctggaca ggtggtaaag 660  
 20 acacttgtgg ggggtattct gggggccac ttgtctgtaa tgggtgtgctt caaggatatca 720  
 catcatgggg ccctgagcca tgtgccctgc ctgaaaagcc tgctgtgtac accaaggtgg 780  
 25 tgcattaccg gaagtggatc aaggacacca tcgcagccaa cccctgagtg cccctgtccc 840  
 acccctacct ctagtaaatt taagtccacc tcaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 900  
 a 901

30 <210> 81  
 <211> 618  
 <212> ADN  
 <213> Homo sapiens

35 <400> 81  
 ggggaccact tctctgggac acattgcctt ctgttttctc cagcatggcg ttgctccagc 60  
 tcctgttcag ggccagccct gccaccctgc tcctggttct ctgcctgcag ttggggggcca 120  
 40 acaaagctca ggacaacact cggaagatca taataaagaa ttttgacatt cccaagtcag 180  
 tacgtccaaa tgacgaagtc actgcagtc ttgcagttca aacagaattg aaagaatgca 240  
 45 tggtyggtta aacttacctc attagcagca tccctctaca aggtgcattt aactataagt 300  
 atactgcctg cctatgtgac gacaatccaa aaaccttcta ctgggacttt tacaccaaca 360  
 gaactgtgca aattgcagcc gtcgttgatg ttattcggga attaggcatc tgccctgatg 420  
 50 atgctgctgt aatccccatc aaaaacaacc ggttttatac tattgaaatc ctaaaggtag 480  
 aataatggaa gccctgtctg tttgccacac ccaggtgatt tcctctaaag aaacttggct 540  
 ggaatttctg ctgtggtcta taaaataaac ttcttaacat gcttctacaa aaaaaaaaaa 600  
 55 aaaaaaaaaa aaaaaaaaaa 618

60 <210> 82  
 <211> 594  
 <212> ADN  
 <213> Homo sapiens

65

ES 2 525 545 T3

<400> 82

5 gtcggttag gactttctgc ctccactatt gctatcggta ctggaatagc aggcatttca 60  
 acatctgtca cgaccttcca tagcctatat aatgacttat ctgctagcat cacagacata 120  
 tcacaaactt tadcagtcct ccaggcccaa gttgaatctt tagctgcagt tgccttcaa 180  
 aaccgccgag gccttgactt acttactgct taaagaggag gactctgcat attcttaaat 240  
 10 gaggagtgtt gtttttacat aatcaatct ggcttgggtg atgacaacat aaaaaaattc 300  
 aaggatagag cccaaaaact taccaaccaa gcaagtaatt tcaactgaacc cccttgggca 360  
 15 ctccctaatt ggggtgcctg ggtcctccca attcttagtc ctttaatacc catttttctc 420  
 ctctttttat tcagaccttg tatcttctgt ttagcttctc aattcatcca aaaccatc 480  
 caggccatca ccaatcattc tatacgacaa atgtttctta taacatcccc acaatatcac 540  
 20 cccttaccac aagacctccc ttcaacttaa tctctcccga tataggttcc caca 594

<210> 83

<211> 1372

<212> ADN

<213> Homo sapiens

25

<400> 83

30 gaattcggcg atgcctcaca actccatcag atctggccat ggagggctga accagctggg 60  
 aggggccttt gtgaatggca gacctctgcc ggaagtggtc cgccagcgca tcgtagacct 120  
 35 ggcccaccag ggtgtaaggc cctgcgacat ctctcgccag ctccgctca gccatgggtg 180  
 cgtcagcaag atccttgga ggtactacga gactggcagc atccggcctg gagtgatagg 240

40

45

50

55

60

65

ES 2 525 545 T3

5 gggctccaag cccaaggtgg ccacccccaa ggtggtggag aagattgggg actacaaacg 300  
 ccagaaccct accatgtttg cctgggagat ccgagaccgg ctccctggctg agggcgtctg 360  
 tgacaatgac actgtgcca gtgtcagctc cattaataga atcatccgga ccaaagtgca 420  
 10 gcaaccattc aacctcccta tggacagctg cgtggccacc aagtccctga gtcccggaca 480  
 cacgctgatc cccagctcag ctgtaactcc cccggagtca cccagtcgg attccctggg 540  
 ctccacctac tccatcaatg ggctcctggg catcgtcag cctggcagcg acaagaggaa 600  
 15 aatggatgac agtgatcagg atagctgccg actaagcatt gactcacaga gcagcagcag 660  
 cggaccccg aagcacctc gcacggatgc cttcagccag caccacctc agccgctcga 720  
 gtgcccattt gagcggcagc actaccaga ggccatgccc tccccagcc acaccaaagg 780  
 20 cgagcagggc ctctaccgc tgcccttgct caacagcacc ctggacgacg ggaaggccac 840  
 cctgaccct tccaacacgc cactggggcg caacctctc actcaccaga cctaccccgt 900  
 25 ggtggcagat cctcactcac ccttcgcat aaagcaggaa accccgagg tgtccagttc 960  
 tagctccacc ccttcctctt tatctagctc cgcctttttg gatctgcagc aagtcggctc 1020  
 cgggggtcccg cccttcaatg cctttcccca tgctgcctcc gtgtacgggc agttcacggg 1080  
 30 ccaggccctc ctctcagggc gagagatggg ggggcccacg ctgcccggat acccaccaca 1140  
 catccccacc agcggacagg gcagctatgc ctctctgcc atcgcaggca tgggtggcagg 1200  
 35 aagtgaatac tctggcaatg cctatggcca cccccctac tcctcctaca gcgaggcctg 1260  
 gcgcttcccc aactccagct tgctgagttc cccatattat tacagttcca catcaaggcc 1320  
 gagtgcaccg cccaccactg ccacggcctt tgaccatctg tagttgaagc tt 1372

40  
 <210> 84  
 <211> 2983  
 <212> ADN  
 <213> Homo sapiens  
 45  
 <400> 84

50

55

60

65

ES 2 525 545 T3

	gcccagatag	gggagcggag	gtggcggcgg	cggcggtagc	ggtggccttg	gttgtcttcc	60
5	agtctcctcg	gctcgccctt	tagccggcac	cgctcccctt	ccctcccctt	tcctctcttc	120
	cttccttccc	tccccttccc	tttttcctt	ccccgtcgg	gagcggcggg	ggtggctcca	180
	gcaacggctg	ggcccaagct	gtgtagaggc	cttaaccaac	gataacggcg	gcgacggcga	240
10	aacctcggag	ctcgcagggc	gggggcaagg	cccgggcctt	ggagatggag	aattctcagt	300
	tgtgtaagct	gttcatcggc	ggcctcaatg	tgcagacgag	tgagtcgggc	ctgcgcggcc	360
	actttgaggc	ctttgggact	ctgacggact	gcgtggtggt	ggtgaatccc	cagaccaagc	420
15	gctcccgttg	ctttggcttc	gtgacctact	ccaatgtgga	ggaggcggac	gccgccatgg	480
	ccgcctcgcc	ccatgccgtg	gacggcaaca	ctgtggagct	gaagcgggcg	gtgtcccggg	540
20	aggattcggc	gcggcccgg	gcccacgcca	aggttaagaa	gctctttgtc	ggaggcctta	600
	aaggagacgt	ggctgagggc	gacctgatcg	agcacttctc	gcagtttggc	accgtggaaa	660
25	aggccgagat	tattgccgac	aagcagtccg	gcaagaagcg	tggattcggc	ttcgtgtatt	720
30							
35							
40							
45							
50							
55							
60							
65							

ES 2 525 545 T3

	tccagaatca	cgacgcggca	gacaaggccg	cggtggtcaa	gttccatccg	attcagggcc	780
	atcgcgtgga	ggtgaagaaa	gcagtcccca	aggaggatat	ctactccggt	gggggtggag	840
5	gcggtccccg	atcctccccg	ggcggccgag	gcggccgggg	gcgcggcggg	ggtcggagacc	900
	agaacggcct	ttccaagggc	ggcggcggcg	gttacaacag	ctacggtggt	tacggcggcg	960
	gcggagggcg	cggctacaat	gcctacggag	gcggcggcgg	cggttcgtcc	tacggtggga	1020
10	gcgactacgg	taacggcttc	ggcggcttcg	gcagctacag	ccagcatcag	tcctcctatg	1080
	ggcccatgaa	gagcggcggc	ggcggcggcg	gtggaggcag	tagctggggc	ggtcgcagta	1140
	atagtggacc	ttacagaggc	ggctatggcg	gtgggggtgg	ctatggaggc	agctccttct	1200
15	aaaagaaaat	ttaaaatgcc	tgggagtggc	tataggggta	gctctttcca	acagcccaag	1260
	tggggtaaac	tcctaagccc	cacccctca	cacacaccgc	cttccctggt	ttgcccttgg	1320
	gggagccact	tctaaggctg	cttacccttg	gggggtttcc	tctatttgcc	tgccacctct	1380
20	cttgtctctc	cctctgaaga	tggactcggc	cccacataca	catttttggtg	ttacagtcac	1440
	tgatggactc	tattttttta	ttattacttg	gaccttggtc	gtttttatac	tagcaaaatg	1500
	tcttgtttta	atttgtgttt	tttgggggga	gggagggagt	gaacttgctg	attctgtagc	1560
25	aaaacctggg	tgggggttgg	ggtggggggt	agtttacttt	gttgtaagga	cttgataacc	1620
	tggctacagc	gttttctatg	aaatctactt	ggatcccatg	cctgaaattt	ggaagcatat	1680
30	gtacaaaaat	cattttttacg	ttttatTTTT	aataaatcat	tgtgtttgac	cgtagatgtc	1740
	taacatTTTT	tttctaggat	ccattccgta	ccgTTTTTTA	agggatattt	gtttaagact	1800
	ttacgtgtta	attctttatt	cttgatgtgt	acttagagaa	acttaagagg	tcctgtggtt	1860
35	TTTTTCCCT	ctcctgttgc	cctgctagtt	gcgtgttgaa	ttatatccct	tacaggcaaa	1920
	acttttgaag	tggtggatgt	ggctTTTTAA	actcttaagt	ttctgtgcat	ccatctcttg	1980
	tactaagcga	attgtttatc	atcttgacat	ggttggtcat	ttctatgaca	atttacttca	2040
40	aactgtgtac	tgtgtagttc	tatatagttt	gtgTTAAGCA	tgTcattcat	ataaactggt	2100
	taaaatTTTT	cagatggcct	agtttcatcc	ctcttactgg	tttgtctgta	atgaatggtt	2160
	aaaaataagg	gttatatTTT	accctcaaat	gcgTTTTTGT	actttcagag	caggtTTAAA	2220
45	cgTTTTTTTT	TTTTTTTTCC	tatatccgaa	ctgTTGGCCT	catggaaatc	cTTTTCCCGA	2280
	tctttgtagc	accatctact	ggcagaatgg	cagagtagct	gcgaaacaat	ttgtttaaaa	2340
	acttgcttaa	gacaattgca	tcagatttgg	aagTTTTGCC	atcaaaattc	tttgcagaat	2400
50	tggaagttaa	cacatttgct	tgtaactgag	atgggcttca	caggaatgta	gttgccagtt	2460
	catatcacia	tagccctttc	tatatgaggt	ttgaaaatgt	aaactgctat	gcatagcttg	2520
	ggcaatagcc	ctaaattgct	atgacaacta	atgaaccagc	tacgtatact	ggtatTTTAG	2580
55	gtgcaagtTg	taaagcaaaa	tatctgtgta	ttctgcttgg	ttaacaaatg	tatatTTTgta	2640
	gccctttcct	gcaatagcat	tcaagttggt	gtttataaga	gaagaacaaa	agtgataata	2700
60	ggtgaaaatt	gcctttctgg	atagaaatag	agaatagcaa	cgtttatgga	tatcaciaat	2760
	aaagaattca	attctttaca	tgattgagtg	agagtatgta	taacctggtg	ggtgggttca	2820

65

ES 2 525 545 T3

gagtaccttt taatctagta tgcttaactt gatgtaata ttttaacttaa atatttgact 2880  
tacatgttga cgttgaaggc tcaaagctat actaagaagc tttctgaaag attgggcttt 2940  
5 aaaataaaat aatatttttaa tattgaaaaa aaaaaaaaaa aaa 2983

10 <210> 85  
<211> 3345  
<212> ADN  
<213> Homo sapiens

15 <400> 85

20

25

30

35

40

45

50

55

60

65

ES 2 525 545 T3

gaattccgtc tcgaccactg aatggaagaa aaggactttt aaccaccatt ttgtgactta 60  
 cagaaaggaa tttgaataaa gaaaactatg atacttcagg cccatcttca ctccctgtgt 120  
 5 cttcttatgc tttatattggc aactggatat ggccaagagg ggaagtttag tggaccctg 180  
 aaacccatga ctttttctat ttatgaaggc caagaaccga gtcaaattat attccagttt 240  
 aaggccaatc ctctgctgt gacttttgaa ctaactgggg agacagacaa catatattgtg 300  
 10 atagaacggg agggacttct gtattacaac agagccttgg acagggaaac aagatctact 360  
 cacaatctcc aggttgacgc cctggacgct aatggaatta tagtggaggg tccagtccct 420  
 atcaccatag aagtgaagga catcaacgac aatcgacca cgtttctcca gtcaaagtac 480  
 15 gaaggctcag taaggcagaa ctctcgccca ggaaagccct tcttgatgt caatgccaca 540  
 gacctggatg atccggccac tccaatggc cagctttatt accagattgt catccagctt 600  
 20 cccatgatca acaatgtcat gtactttcag atcaacaaca aaacgggagc catctctctt 660  
 acccgagagg gatctcagga attgaatcct gctaagaatc cttcctataa tctgggtgatc 720  
 tcagtgaagg acatgggagg ccagagtgag aattccttca gtgataccac atctgtggat 780  
 25 atcatagtga cagagaatat ttgaaagca ccaaacctg tggagatggt ggaaaactca 840  
 actgatcctc accccatcaa aatcactcag gtgcggtgga atgatcccg tgcacaatat 900  
 30 tccttagttg acaaagagaa gctgccaaga ttcccatttt caattgacca ggaaggagat 960  
 atttacgtga ctcagccctt ggaccgagaa gaaaaggatg catatgtttt ttatgcagtt 1020  
 gcaaaggatg agtacggaaa accactttca tatccgctgg aaattcatgt aaaagttaa 1080  
 35 gatattaatg ataatccacc tacatgtccg tcaccagtaa ccgtatttga ggtccaggag 1140  
 aatgaacgac tgggtaacag tatcgggacc cttactgcac atgacagggg tgaagaaaat 1200  
 actgccaaca gttttctaaa ctacaggatt gtggagcaaa ctcccaaact tcccatggat 1260  
 40 ggactcttcc taatccaaac ctatgctgga atgttacagt tagctaaaca gtccttgaag 1320  
 aagcaagata ctctcagta caacttaacg atagaggtgt ctgacaaaga tttcaagacc 1380  
 45 ctttgttttg tgcaaatcaa cgttattgat atcaatgatc agatcccat ctttgaaaaa 1440  
 tcagattatg gaaacctgac tcttgctgaa gacacaaaca ttgggtccac catcttaacc 1500  
 atccaggcca ctgatgctga tgagccattt actgggagtt ctaaaattct gtatcatatc 1560  
 50 ataaagggag acagtgaggg acgcctgggg gttgacacag atcccatac caacaccgga 1620  
 tatgtcataa ttaaaaagcc tcttgatttt gaaacagcag ctgtttccaa cattgtgttc 1680

55

60

65

ES 2 525 545 T3

aaagcagaaa atcctgagcc tctagtgttt ggtgtgaagt acaatgcaag ttcttttgcc 1740  
 aagttcacgc ttattgtgac agatgtgaat gaagcacctc aattttccca acacgtattc 1800  
 5 caagcгааag tcagtgagga tgtagctata ggcactaaag tgggcaatgt gactgccaag 1860  
 gatccagaag gtctggacat aagctattca ctgagggggag acacaagagg ttggcttaa 1920  
 attgaccacg tgactgggtga gatcttttagt gtggctccat tggacagaga agccggaagt 1980  
 10 ccatatcggg tacaagtggg ggccacagaa gtaggggggt cttccttaag ctctgtgtca 2040  
 gagttccacc tgatccttat ggatgtgaat gacaaccctc ccaggctagc caaggactac 2100  
 acgggcttgt tcttctgcca tcccctcagt gcacctggaa gtctcatttt cgaggctact 2160  
 15 gatgatgatc agcacttatt tcgggggtccc cattttacat tttccctcgg cagtggaagc 2220  
 ttacaaaacg actgggaagt ttccaaaatc aatggtactc atgcccgact gtctaccagg 2280  
 20 cacacagact ttgaggagag ggcgtatgtc gtcttgatcc gcatcaatga tgggggtcgg 2340  
 ccacccttgg aaggcattgt ttctttacca gttacattct gcagtttgtgt ggaaggaagt 2400  
 tgtttccggc cagcaggtca ccagactggg ataccactg tgggcatggc agttggtata 2460  
 25 ctgctgacca cccttctggt gattggtata attttagcag ttgtgtttat ccgcataaag 2520  
 aaggataaag gcaaagataa tgttgaaagt gctcaagcat ctgaagtcaa acctctgaga 2580  
 agctgaattt gaaaaggaat gtttgaattt atatagcaag tgctatttca gcaacaacca 2640  
 30 tctcatccta ttacttttca tctaacgtgc attataattt tttaaacaga tattccctct 2700  
 tgtcctttaa tatttgctaa atatttcttt tttgaggtgg agtcttgctc tgtcgcccag 2760  
 35 gctggagtac agtgggtgtga tcccagctca ctgcaaccctc cgctcctgg gttcacatga 2820  
 ttctcctgcc tcagcttcct aagtagctgg gtttacaggc acccaccacc atgcccagct 2880  
 aatttttgta tttttaatag agacgggggt tgcgccatttg gccaggctgg tcttgaactc 2940  
 40 ctgacgtcaa gtgatctgcc tgccttggtc tcccaataca ggcataaacc actgcacca 3000  
 cctacttaga tatttcatgt gctatagaca ttagagagat ttttcatttt tccatgacat 3060  
 ttttctctc tgcaaatggc ttagctactt gtgtttttcc cttttggggc aagacagact 3120  
 45 cattaatat tctgtacatt ttttctttat caaggagata tatcagtgtt gtctcataga 3180  
 actgcctgga ttccatttat gttttttctg attccatcct gtgtcccctt catccttgac 3240  
 50 tcctttggta tttcactgaa tttcaaacat ttgtcagaga agaaaaaagt gaggactcag 3300  
 gaaaaataaa taaataaaag aacagccttt tgcggccgcg aattc 3345

55 <210> 86  
 <211> 990  
 <212> ADN  
 <213> Homo sapiens

60 <400> 86

65



ES 2 525 545 T3

agccccaagc ttaccacctg cacccggaga gctgtgtcac catgtgggtc ccggttgtct 60  
 tcctcacct gtccgtgacg tggattggtg ctgcaccct catcctgtct cggattgtgg 120  
 5 gaggctggga gtgcgagaag cattcccaac cctggcaggt gcttgtggcc tctcgtggca 180  
 gggcagtctg cggcgggtgtt ctggtgcacc cccagtgggt cctcacagct gccactgca 240  
 10 tcaggaacaa aagcgtgac ttgctgggtc ggcacagcct gtttcatcct gaagacacag 300  
 gccaggtatt tcaggtcagc cacagcttcc cacacccgct ctacgatatg agcctcctga 360  
 agaatcgatt cctcaggcca ggtgatgact ccagccacga cctcatgctg ctccgcctgt 420  
 15 cagagcctgc cgagctcag gatgctgtga aggtcatgga cctgcccacc caggagccag 480  
 cactggggac cacctgctac gcctcaggct ggggcagcat tgaaccagag gagttcttga 540  
 ccccaaagaa acttcagtgt gtggacctcc atgttatttc caatgacgtg tgtgcgcaag 600  
 20 ttcaccctca gaaggtgacc aagttcatgc tgtgtgctgg acgctggaca gggggcaaaa 660  
 gcacctgctc ggggtgattct gggggcccac ttgtctgtaa tgggtgtgctt caaggtatca 720  
 25 cgtcatgggg cagtgaacca tgtgccctgc ccgaaaggcc ttccctgtac accaaggtgg 780  
 tgcattaccg gaagtggatc aaggacacca tcgtggccaa cccctgagca cccctatcaa 840  
 ccccctattg tagtaaactt ggaaccttgg aatgaccag gccaagactc aagcctcccc 900  
 30 agttctactg acctttgtcc ttaggtgtga ggtccagggt tgctaggaaa agaaatcagc 960  
 agacacaggt gtagaccaga gtgtttctta 990

35 <210> 87  
 <211> 1805  
 <212> ADN  
 <213> Homo sapiens

40 <400> 87

45

50

55

60

65

ES 2 525 545 T3

	gcgcacactc tcctaagccc tctcatctcc tggaaccatg gccagcacat ccaccacat	60
	caggagccac agcagcagcc gccggggttt cagtgccaac tcagccaggc tccctggggt	120
5	cagccgctct ggcttcagca gcatctccgt gtcccgtcc aggggcagtg gtggcctggg	180
	tggcgcagtg ggaggagctg gctttggcag ccgcagtctg tatggcctgg ggggctccaa	240
	gaggatctcc attggagggg gcagctgtgc catcagtggc ggctatggca gcagagccgg	300
10	aggcagctat ggctttggtg gcgccgggag tggatttggg ttcggtggtg gagccggcat	360
	tggctttggt ctgggtggtg gagccggcct tgctggtggc tttgggggcc ctggcttccc	420
	tgtgtgcccc cctggaggca tccaagaggc cactgtcaac cagagtctcc tgactcccct	480
15	caacctgcaa attgaccccg ccatccagcg ggtgcgggcc gaggagcgtg agcagatcaa	540
	gacctcaac aacaagtttg cctccttcat cgacaagggtg cggttcctag agcagcagaa	600
20	caaggttctg gacaccaagt ggaccctgct gcaggagcag ggcaccaaga ctgtgaggca	660
	gaacctggag ccggtgttcg agcagtacat caacaacctc aggaggcagc tggacaacat	720
	cgtgggggaa cggggtcgtc tggactcggg gctgagaaac atgcaggacc tgggtgagga	780
25	cctcaagaac aaatatgagg atgaaatcaa caagcgaca gcagcagaga atgaatttgt	840
	gactctgaag aaggatgtgg atgctgccta catgaacaag gttgaactgc aagccaaggc	900
30	agacactctt acagatgaga tcaacttctt gagagccttg tatgatgcag agctgtccca	960
	gatgcagacc cacatctcag acacatccgt ggtgctatcc atggacaaca accgcaacct	1020
	ggacctggac agcatcatcg ctgagggtcaa ggccaatat gaggagattg ctgagaggag	1080
35	cagggctgag gctgagtcct ggtaccagac aaagtacgag gagctgcaga tcacagcagg	1140
	cagacatggg gacgacctgc gcaacaccaa gcaggagatt gctgagatca accgcatgat	1200
	ccagaggctg agatctgaga tcgaccacgt caagaagcag tgtgccaacc tacaggccgc	1260
40	cattgctgat gctgagcagc gtggggagat ggccctcaag gatgctaaga acaagctgga	1320
	agggctggag gatgccctgc agaaggccaa gcaggacctg gcccggtgc tgaaggagta	1380
45	ccaggagctg atgaacgtca agctggccct ggacgtggag atcgccacct accgcaagct	1440
	gctggagggc gaggagtgcg ggctgaatgg cgaaggcgtt ggacaagtca acatctctgt	1500
	agtgcagtcc accgtctcca gtggctatgg cgggtccagc ggtgtcggca gtggcttagg	1560
50	cctgggtgga ggaagcagct actcctatgg cagtggctct ggcggttgag gcggctttag	1620
	ttccagcagc ggcagagcca ctgggggtgg cctcagctct gttggaggcg gcagttccac	1680
	catcaagtac accaccacct cctcctccag caggaagagc tacaagcact gaagtcgtgc	1740
55	cgccagctct cagtcccaca gctctcaggc ccctctctgg cagcagagcc ctctcctcag	1800
	gttgc	1805
60	<210> 88 <211> 2820 <212> ADN <213> Homo sapiens	
65		

ES 2 525 545 T3

<400> 88

5	tggcaaaatc ctggagccag aagaaaggac agcagcattg atcaatctta cagctaacat	60
	gttgtacctg gaaaacaatg cccagactca atttagtgag ccacagtaca cgaacctggg	120
	gctcctgaac agcatggacc agcagattcg gaacggctcc tcgtccacca gtcctataa	180
10	cacagaccac gcgcagaaca gcgtcacggc gccctcgccc tacgcacagc ccagccccac	240
	cttcgatgct ctctctccat caccgcctat cccctccaac accgactacc caggccccga	300
	cagttccgac gtgtccttcc agcagtcgag caccgccaag tcggccacct ggacgtattc	360
15	cactgaactg aagaaactct actgccaaat tgcaaagaca tgccccatcc agatcaaggt	420
	gatgaccca cctcctcagg gagctggtat ccgcgccatg cctgtctaca aaaaagctga	480
	gcacgtcacg gaggtggtga agcggtgccc caaccatgag ctgagccgtg agttcaacga	540
20	gggacagatt gccctccta gtcatttgat tcgagtagag gggaacagcc atgccagta	600
	tgtagaagat cccatcacag gaagacagag tgtgctggta cttatgagc caccccaggt	660
25	tggcactgaa ttcacgacag tcttgtacaa tttcatgtgt aacagcagtt gtggtggagg	720
	gatgaaccgc cgtccaattt taatcattgt tactctggaa accagagatg ggcaagtcct	780
	gggccgacgc tgctttgagg cccggatctg tgcttgccca ggaagagaca ggaaggcgga	840
30	tgaagatagc atcagaaagc agcaagtttc ggacagtaca aagaacggtg atggtacgaa	900
	gcgcccgttt cgtcagaaca cacatggtat ccagatgaca tccatcaaga aacgaagatc	960
35	cccagatgat gaactggtat acttaccagt gaggggcccgt gagacttatg aaatgctggt	1020
40		
45		
50		
55		
60		
65		

ES 2 525 545 T3

	gaagatcaaa	gagtccttg	aactcatgca	gtaccttct	cagcacacaa	ttgaaacgta	1080
	caggcaacag	caacagcagc	agcaccagca	cttacttcag	aaacagacct	caatacagtc	1140
5	tccatcttca	tatggtaaca	gctccccacc	tctgaacaaa	atgaacagca	tgaacaagct	1200
	gccttctgtg	agccagctta	tcaaccctca	gcagcgcaac	gccctcactc	ctacaacat	1260
10	tcctgatggc	atgggagcca	acattcccat	gatgggcacc	cacatgcca	tggctggaga	1320
	catgaatgga	ctcagcccca	cccaggcact	ccctcccca	ctctccatgc	catccacctc	1380
	ccactgcaca	ccccacctc	cgtatcccac	agattgcagc	attgtcagtt	tcttagcgag	1440
15	gttgggctgt	tcatcatgtc	tggactat	cacgaccag	gggctgacca	ccatctatca	1500
	gattgagcat	tactccatgg	atgatctggc	aagtctgaaa	atccctgagc	aatttcgaca	1560
20	tgcgatctgg	aagggcatcc	tggaccaccg	gcagctccac	gaattctcct	ccccttctca	1620
	tctcctgcgg	acccaagca	gtgcctctac	agtcagtggt	ggctccagtg	agaccgggg	1680
	tgagcgtggt	attgatgctg	tgcgattcac	cctccgccag	accatctctt	tcccaccccg	1740
25	agatgagtgg	aatgacttca	actttgacat	ggatgctcgc	cgcaataagc	aacagcgc	1800
	caaagaggag	ggggagtgag	cctcaccatg	tgagctcttc	ctatccctct	cctaactgcc	1860
	agccccctaa	aagcactcct	gcttaatctt	caaagccttc	tccttagctc	ctccccttc	1920
30	tcttgtctga	tttcttaggg	gaaggagaag	taagaggcta	cctcttacct	aacatctgac	1980
	ctggcatcta	attctgattc	tggctttaag	ccttcaaaac	tatagcttgc	agaactgtag	2040
35	ctgccatggc	taggtagaag	tgagcaaaaa	agagttgggt	gtctccttaa	gctgcagaga	2100
	tttctcattg	acttttataa	agcatgttca	cccttatagt	ctaagactat	atatataaat	2160
	gtataaatat	acagtataga	tttttgggtg	gggggcattg	agtattgttt	aaaatgtaat	2220
40	ttaaataaaa	gaaaattgag	ttgcacttat	tgaccat	tttaatttact	tgttttggat	2280
	ggcttgtcta	tactccttcc	cttaaggggt	atcatgtatg	gtgataggta	tctagagctt	2340
45	aatgctacat	gtgagtgcga	tgatgtacag	attctttcag	ttctttggat	tctaaataca	2400
	tgccacatca	aacctttgag	tagatccatt	tccattgctt	attatgtagg	taagactgta	2460
	gatatgtatt	cttttctcag	tgttgggtata	ttttatatta	ctgacatttc	ttctagtgat	2520
50	gatgggtcac	gttgggggtga	tttaatccag	ttataagaag	aagttcatgt	caaacggtc	2580
	ctcttttagtt	tttgggtggg	aatgaggaaa	attcttaaaa	ggcccatagc	agccagttca	2640
	aaaacacccg	acgtcatgta	tttgagcata	tcagtaacc	ccttaaattt	aataccaga	2700
55	taccttatct	tacaatggtg	attgggaaaa	catttgctgc	ccattacaga	ggtattaaaa	2760
	ctaaatttca	ctactagatt	gactaactca	aatacacatt	tgctactggt	gtaagaattc	2820
60		<210> 89	<211> 991	<212> ADN	<213> Homo sapiens		
65		<400> 89					

ES 2 525 545 T3

cttatctcgg cttcgtttct ggagggccag gaacaaacag gcttcaaagc caagggcttg 60  
 gctggcacac agggggcttg gtccttcacc tctgtcccct ctcctacgg acacataata 120  
 5 gaccctggtc acacctggga gaggaggaga ggagagcata gcacctgcag caagatggat 180  
 gtgggcagca aagaggtcct gatggagagc ccgccggact actccgcagc tccccggggc 240  
 10 cgatttgga ttcctgctg cccagtgcac ctgaaacgcc ttcttatcgt ggtggtggtg 300  
 gtggtcctca tcgtcgtggt gattgtggga gccctgctca tgggtctcca catgagccag 360  
 aaacacacgg agatggttct ggagatgagc attggggcgc cggaagccca gcaacgcctg 420  
 15 gccctgagtg agcacctggt taccactgcc accttctcca tcggctccac tggcctcgtg 480  
 gtgtatgact accagcagct gctgatcgcc tacaagccag cccctggcac ctgctgctac 540  
 atcatgaaga tagctccaga gagcatcccc agtcttgagg ctctcaatag aaaagtccac 600  
 20 aacttccaga tggaatgctc tctgcaggcc aagcccgcag tgcctacgtc taagctgggc 660  
 caggcagagg ggcgagatgc aggctcagca ccctccggag gggacccggc cttcctgggc 720  
 25 atggccgtga acaccctgtg tggcgaggtg ccgctctact acatctagga cgcctccggt 780  
 gagcagggtc agtgggaagcc ccaacgggaa aggaaacgcc ccgggcaaag ggtcttttgc 840  
 agcttttgca gacgggcaag aagctgcttc tgcccacacc gcagggacaa accctggaga 900  
 30 aatgggagct tggggagagg atgggagtgg gcagaggtgg caccagggg cccgggaact 960  
 cctgccacaa cagaataaag cagcctgatt g 991

35 <210> 90  
 <211> 1580  
 <212> ADN  
 <213> Homo sapiens

40 <400> 90

45

50

55

60

65

ES 2 525 545 T3

catcctgccca cccctagcct tgctggggac gtgaaccctc tccccgcgcc tgggaagcct 60  
 tcttggcacc gggaccgga gaatccccac ggaagccagt tccaaaaggg atgaaaaggg 120  
 5 ggcgtttcgg gcaactggag aagcctgtat tccagggccc ctcccagagc aggaatctgg 180  
 gaccaggag tgccagcctc acccacgcag atcctggcca tgagagctcc gcacctccac 240  
 ctctccgccg cctctggcgc ccgggctctg gcgaagctgc tgccgctgct gatggcgcaa 300  
 10 ctctggggccg cagaggcggc gctgctcccc caaaacgaca cgcgcttggga ccccgaagcc 360  
 tatggctccc cgtgcgcgcg cggctcgcag ccctggcagg tctcgctctt caacggcctc 420  
 tcgttccact gcgcggtgtt cctggtggac cagagttggg tgctgacggc cgcgcaactgc 480  
 15 ggaacaagc cactgtgggc tcgagtaggg gatgaccacc tgctgcttct tcaggagag 540  
 cagctccgcc ggaccactcg ctctgttgtc catcccaagt accaccaggg ctccaggcccc 600  
 20 atcctgcca ggcgaacgga tgagcacgat ctcatgttgc tgaagctggc caggccccgta 660  
 gtgctggggc cccgctccg ggcctgcag ctccctacc gctgtgctca gccgggagac 720  
 cagtccagg ttgctggctg gggcaccacg gccgcccgga gagtgaagta caacaagggc 780  
 25 ctgacctgct ccagcatcac tatcctgagc cctaaagagt gtgaggtctt ctaccctggc 840  
 gtggtcacca acaacatgat atgtgctgga ctggaccggg gccaggacct ttgccagagt 900  
 gactctggag gccccctggt ctgtgacgag accctccaag gcatcctctc gtggggtgtt 960  
 30 taccctgtg gctctgcca gcatccagct gtctacacc agatctgcaa atacatgtcc 1020  
 tggatcaata aagtcatacg ctccaactga tccagatgct acgctccagc tgatccagat 1080  
 35 gttatgctcc tgctgatcca gatgcccaga ggctccatcg tccatcctct tctctcccag 1140  
 tcggctgaac tctccccttg tctgcaactgt tcaaacctct gccgcccctc acaccttaa 1200  
 acatctcccc tctcacctca tccccacc tatccccatt ctctgcctgt actgaagctg 1260  
 40 aatgcagga agtgggtggca aaggtttatt ccagagaagc caggaagccg gtcacaccc 1320  
 agcctctgag agcagttact ggggtcacc aacctgactt cctctgccac tcctgtctgt 1380  
 gtgactttgg gcaagccaag tgccctctct gaacctcagt ttctcatct gcaaaatggg 1440  
 45 aacaatgacg tgccctacct ttagacatgt tgtgaggaga ctatgatata acatgtgtat 1500  
 gtaaactctc atgggtgatt tcatgtaagg cttaacacag tgggtggtga gttctgacta 1560  
 50 aaggttacct gttgtcgtga 1580

<210> 91  
 <211> 3359  
 <212> ADN  
 55 <213> Homo sapiens  
 <400> 91

60

65

ES 2 525 545 T3

	cacaccttcg	gcagcaggag	ggcggcagct	tctcgcaggc	ggcagggcgg	gcggccagga	60
	tcatgtccac	caccacatgc	caagtgggtg	cgttcctcct	gtccatcctg	gggctggccg	120
5	gctgcatcgc	ggccaccggg	atggacatgt	ggagcaccca	ggacctgtac	gacaacccccg	180
	tcacctccgt	gttccagtac	gaagggctct	ggaggagctg	cgtagaggcag	agttcaggct	240
	tcaccgaatg	caggccctat	ttcaccatcc	tgggacttcc	agccatgctg	caggcagtg	300
10	gagccctgat	gatcgtaggc	atcgtcctgg	gtgccattgg	cctcctggta	tccatctttg	360
	ccctgaaatg	catccgcatt	ggcagcatgg	aggactctgc	caaagccaac	atgacactga	420
15	cctccgggat	catgttcatt	gtctcaggtc	tttgtgcaat	tgctggagtg	tctgtgtttg	480
	ccaacatgct	ggtgactaac	ttctggatgt	ccacagctaa	catgtacacc	ggcatgggtg	540
	ggatgggtgca	gactgttcag	accaggtaca	catttggtgc	ggctctgttc	gtgggctggg	600
20	tcgctggagg	cctcacacta	attgggggtg	tgatgatgtg	catcgcctgc	cggggcctgg	660
	caccagaaga	aaccaactac	aaagccgttt	cttatcatgc	ctcaggccac	agtgttgctt	720
	acaagcctgg	aggcttcaag	gccagcactg	gctttgggtc	caacaccaaa	aacaagaaga	780
25	tatacgatgg	aggtgcccgc	acagaggacg	aggtacaatc	ttatccttcc	aagcacgact	840
	atgtgtaatg	ctctaagacc	tctcagcacg	ggcgggaagaa	actcccggag	agctcaccca	900
30	aaaaacaagg	agatcccatc	tagatttctt	cttgcttttg	actcacagct	ggaagttaga	960
	aaagcctcga	tttcatcttt	ggagaggcca	aatggcttta	gcctcagtct	ctgtctctaa	1020
	atattccacc	ataaaacagc	tgagttattt	atgaattaga	ggctatagct	cacattttca	1080
35	atcctctatt	tcttttttta	aatataactt	tctactctga	tgagagaatg	tggttttaat	1140
40							
45							
50							
55							
60							
65							

ES 2 525 545 T3

ctctctctca cattttgatg atttagacag actccccctc ttctctctag tcaataaacc 1200  
 5 cattgatgat ctatttccca gcttatcccc aagaaaactt ttgaaaggaa agagtagacc 1260  
 caaagatggt attttctgct gtttgaattt tgtctcccca cccccaactt ggctagtaat 1320  
 10 aaacacttac tgaagaagaa gcaataagag aaagatattt gtaatctctc cagcccatga 1380  
 tctcggtttt cttacactgt gatcttaaaa gttaccaaac caaagtcatt ttcagtttga 1440  
 ggcaacccaaa cttttctact gctggtgaca tcttcttatt acagcaacac cattctagga 1500  
 15 gtttcctgag ctctccactg gagtcctctt tctgtcgcgg gtcagaaatt gtccttagat 1560  
 gaatgagaaa attatttttt ttaatttaag tcctaaatat agttaaata aataatgttt 1620  
 tagtaaatg atacactatc tctgtgaaat agcctcacc ctacatgtgg atagaaggaa 1680  
 20 atgaaaaaat aattgctttg acattgtcta tatggtactt tgtaaagtca tgcttaagta 1740  
 caaattccat gaaaagctca ctgatcctaa ttctttccct ttgaggctc tatggtctg 1800  
 attgtacatg atagtaagt taagccatgt aaaaagtaa taatgtctgg gcacagtggc 1860  
 25 tcacgcctgt aatcctagca ctttgggagg ctgaggagga aggatcactt gagcccagaa 1920  
 gttcgagact agcctgggca acatggagaa gccctgtctc taaaaatac agagagaaaa 1980  
 aatcagccag tcatggtggc ctacacctgt agtcccagca ttccgggagg ctgagggtggg 2040  
 aggatcactt gagcccaggg aggttggggc tgcagtgagc catgatcaca cactgact 2100  
 ccagccaggt gacatagcga gatcctgtct aaaaaataa aaaataaata atggaacaca 2160  
 35 gcaagtccta ggaagtaggt taaaactaat tctttaaaaa aaaaaaaag ttgagcctga 2220  
 attaaatgta atgtttccaa gtgacaggta tccacatttg catggttaca agccactgcc 2280  
 agttagcagt agcactttcc tggcactgtg gtcggttttg ttttgtttg ctttgtttag 2340  
 40 agacggggtc tactttcca ggctggcctc aaactcctgc actcaagcaa ttcttctacc 2400  
 ctggcctccc aagtagctgg aattacaggt gtgcgccatc acaactagct ggtggtcagt 2460  
 45 tttgttactc tgagagctgt tcaattctct gaattcacct agagtggttg gaccatcaga 2520  
 tgtttgggca aaactgaaag ctctttgcaa ccacacacct tcctgagct tacatcactg 2580  
 cccttttgag cagaaagtct aaattccttc caagacagta gaattccatc ccagtaccaa 2640  
 50 agccagatag gccccctagg aaactgaggt aagagcagtc tctaaaaact acccacagca 2700  
 gcattggtgc aggggaactt ggccattagg ttattatttg agaggaaagt cctcacatca 2760  
 atagtacata tgaaagtgac ctccaagggg attggtgaat actcataagg atcttcaggc 2820  
 55 tgaacagact atgtctgggg aaagaacgga ttatgcccc ttaaataaca agttgtgttc 2880  
 aagagtcaga gcagtgagct cagaggccct tctcactgag acagcaacat ttaaaccaaa 2940  
 60 ccagaggaag tatttgtgga actcactgcc tcagtttggg taaaggatga gcagacaagt 3000  
 caactaaaga aaaaagaaaa gcaaggagga gggttgagca atctagagca tggagtttgt 3060  
 taagtgtctc ctggatttga gttgaagagc atccatttga gttgaaggcc acagggcaca 3120  
 65 atgagctctc cttctacca ccagaaagtc cctggtcagg tctcaggtag tgcggtgtgg 3180



ES 2 525 545 T3

	ctcagctggg tttttaatta gcgcattctc tatccaacat ttaattgttt gaaagcctcc	3240
	atatagttag attgtgcttt gtaattttgt tgttggtgct ctatcttatt gtatatgcat	3300
5	tgagtattaa cctgaatggt ttgttactta aatattaaaa acactgttat cctacagtt	3359
	<210> 92	
	<211> 733	
10	<212> ADN	
	<213> Homo sapiens	
	<400> 92	
15	gggatccgga gcccaaactt tctgacaaaa ctcacacatg cccaccgtgc ccagcacctg	60
	aattcgaggg tgcaccgtca gtcttctctt tcccccaaaa acccaaggac accctcatga	120
	tctcccggac tcttgaggtc acatgcgtgg tgggtggacgt aagccacgaa gaccctgagg	180
20	tcaagttcaa ctggtacgtg gacggcgtgg aggtgcataa tgccaagaca aagccgcggg	240
	aggagcagta caacagcacg taccgtgtgg tcagcgtcct caccgtcctg caccaggact	300
	ggctgaatgg caaggagtac aagtgcaagg tctccaacaa agccctccca acccccatcg	360
25	agaaaacat ctcaaagcc aaagggcagc cccgagaacc acaggtgtac accctgcccc	420
	catcccggga tgagctgacc aagaaccagg tcagcctgac ctgcctggtc aaaggcttct	480
30	atccaagcga catcgccgtg gagtgggaga gcaatgggca gccggagaac aactacaaga	540
	ccacgcctcc cgtgctggac tccgacggct ccttcttctt ctacagcaag ctcaccgtgg	600
	acaagagcag gtggcagcag gggaaactct tctcatgctc cgtgatgcat gaggctctgc	660
35	acaaccacta cacgcagaag agcctctccc tgtctccggg taaatgagtg cgacggccgc	720
	gactctagag gat	733
40	<210> 93	
	<211> 1076	
	<212> ADN	
	<213> Homo sapiens	
45	<400> 93	
50		
55		
60		
65		

ES 2 525 545 T3

	atggtgggtg aggttgattc catgccggct gcctcttctg tgaagaagcc atttggtctc	60
	aggagcaaga tgggcaagtg gtgctgccgt tgcttcccct gctgcagga gagcggcaag	120
5	agcaacgtgg gcacttctgg agaccacgac gactctgcta tgaagacact caggagcaag	180
	atgggcaagt ggtgccgcca ctgcttcccc tgctgcaggg ggagtggcaa gagcaacgtg	240
	ggcgcttctg gagaccacga cgactctgct atgaagacac tcaggaacaa gatgggcaag	300
10	tggtgctgcc actgcttccc ctgctgcagg gggagcggca agagcaaggt gggcgcttgg	360
	ggagactacg atgacagtgc cttcatggag cccagggtacc acgtccgtgg agaagatctg	420
15	gacaagctcc acagagctgc ctggtggggg aaagtcccca gaaaggatct catcgtcatg	480
	ctcagggaca ctgacgtgaa caagaaggac aagcaaaaga ggactgctct acatctggcc	540
	tctgccaatg ggaattcaga agtagtaaaa ctctgctgg acagacgatg tcaacttaat	600
20	gtccttgaca acaaaaagag gacagctctg ataaaggccg tacaatgcca ggaagatgaa	660
	tgtgcgtaa tgttgctgga acatggcact gatccaaata ttccagatga gtatggaaat	720
25	accactctgc actacgctat ctataatgaa gataaattaa tggccaaagc actgctctta	780
	tatggtgctg atatcgaatc aaaaaacaag catggcctca caccactggt acttggtgta	840
	catgagcaaa aacagcaagt cgtgaaatth ttaatcaaga aaaaagcgaa tttaaatgca	900
30	ctggatagat atggaaggac tgctctcata cttgctgtat gttgtggatc agcaagtata	960
	gtcagccttc tacttgagca aaatattgat gtatcttctc aagatctatc tggacagacg	1020
	gccagagagt atgctgtttc tagtcatcat catgtaatth gccagttact ttctga	1076

35

<210> 94  
 <211> 3675  
 <212> ADN  
 <213> Homo sapiens  
 <400> 94

40

45

50

55

60

65

ES 2 525 545 T3

	tccgagctga	ttacagacac	caaggaagat	gctgtaaaga	gtcagcagcc	acagccctgg	60
	ctagctggcc	ctgtgggcat	ttattagtaa	agttttaatg	acaaaagctt	tgagtcaaca	120
5	cacccgtggg	taattaacct	ggatcatcccc	accctggaga	gccatcctgc	ccatgggtga	180
	tcaaagaagg	aacatctgca	ggaacacctg	atgaggctgc	acccttggcg	gaaagaacac	240
	ctgacacagc	tgaaagcttg	gtggaaaaaa	cacctgatga	ggctgcaccc	ttggtggaaa	300
10	gaacacctga	cacggctgaa	agcttgggtg	aaaaaacacc	tgatgaggct	gcatccttgg	360
	tggagggAAC	atctgacaaa	attcaatggt	tggagaaagc	gacatctgga	aagttcgaac	420
15	agtcagcaga	agaaacacct	agggaaatta	cgagtcctgc	aaaagaaaca	tctgagaaat	480
	ttacgtggcc	agcaaaagga	agacctagga	agatcgcctg	ggagaaaaaa	gaagacacac	540
	ctagggaaat	tatgagtccc	gcaaaagaaa	catctgagaa	atttacgtgg	gcagcaaaag	600
20	gaagacctag	gaagatcgca	tgggagaaaa	aagaaacacc	tgtaaagact	ggatgcgtgg	660
	caagagtaac	atctaataaa	actaaagttt	tggaaaaagg	aagatctaag	atgattgcat	720
	gtcctacaaa	agaatcatct	acaaaagcaa	gtgccaatga	tcagaggttc	ccatcagaat	780
25	ccaaacaaga	ggaagatgaa	gaatattctt	gtgattctcg	gagtctcttt	gagagttctg	840
	caaagattca	agtgtgtata	cctgagtcta	tatatcaaaa	agtaatggag	ataaatagag	900
30	aagtagaaga	gcctcctaag	aagccatctg	ccttcaagcc	tgccattgaa	atgcaaaact	960
	ctgttccaaa	taaagccttt	gaattgaaga	atgaacaaac	attgagagca	gatccgatgt	1020
	tcccaccaga	atccaaacaa	aaggactatg	aagaaaattc	ttgggattct	gagagtctct	1080
35	gtgagactgt	ttcacagaag	gatgtgtggt	taccaaggc	tacacatcaa	aaagaaatag	1140
	ataaaataaa	tggaaaatta	gaagagtctc	ctaataaaga	tggtcttctg	aaggctacct	1200
	gcggaatgaa	agtttctatt	ccaactaaag	ccttagaatt	gaaggacatg	caaactttca	1260
40	aagcagagcc	tccggggaag	ccatctgcct	tcgagcctgc	cactgaaatg	caaaagtctg	1320
	tcccaataa	agccttggaa	ttgaaaaatg	aacaaacatt	gagagcagat	gagatactcc	1380
45	catcagaatc	caaacaaaag	gactatgaag	aaagttcttg	ggattctgag	agtctctgtg	1440
	agactgtttc	acagaaggat	gtgtgtttac	ccaaggctcc	atcaaaaaga	aatagataaa	1500

50

55

60

65

ES 2 525 545 T3

5 ataaatggaa aattagaagg gtctcctggt aaagatggtc ttctgaaggc taactgcgga 1560  
 atgaaagttt ctattccaac taaagcctta gaattgatgg acatgcaaac tttcaaagca 1620  
 gagcctcccg agaagccatc tgccttcgag cctgccattg aaatgcaaaa gtctgttcca 1680  
 10 aataaagcct tgggaattgaa gaatgaacaa acattgagag cagatgagat actcccatca 1740  
 gaatccaaac aaaaggacta tgaagaaagt tcttgggatt ctgagagtct ctgtgagact 1800  
 gtttcacaga aggatgtgtg tttaccaag gctccatcaa aaagaaatag ataaaataaa 1860  
 15 tggaaaatta gaagagtctc ctgataatga tggttttctg aaggctccct gcagaatgaa 1920  
 agtttctatt ccaactaaag ccttagaatt gatggacatg caaactttca aagcagagcc 1980  
 tcccgagaag ccatctgcct tcgagcctgc cattgaaatg caaaagtctg ttccaaataa 2040  
 20 agccttggaa ttgaagaatg aacaacatt gagagcagat cagatgttcc cttcagaatc 2100  
 aaaacaaaag aagttgaaga aaattcttgg gattctgaga gtctccgtga gactgtttca 2160  
 25 cagaaggatg tgtgtgtacc caaggctaca catcaaaaag aaatggataa aataagtgga 2220  
 aaattagaag attcaactag cctatcaaaa atcttggata cagttcattc ttgtgaaaga 2280  
 gcaagggAAC ttcaaaaaga tcaactgtgaa caacgtacag gaaaaatgga acaaatgaaa 2340  
 30 aagaagtttt gtgtactgaa aaagaaactg tcagaagcaa aagaaataaa atcacagtta 2400  
 gagaaccaa aagttaaatt ggaacaagag ctctgcagtg tgaggtttct cacactcatg 2460  
 35 aaaatgaaaa ttatctctta catgaaaatt gcatgttgaa aaaggaaatt gccatgctaa 2520  
 aactggaaat agccacactg aaacaccaat accaggaaaa ggaaaataaa tactttgagg 2580  
 acattaagat tttaaaagaa aagaatgctg aacttcagat gaccctaaaa ctgaaagagg 2640  
 40 aatcattaac taaaagggca tctcaatata gtgggcagct taaagtctg atagctgaga 2700  
 acacaatgct cacttctaaa ttgaaggaaa aacaagacaa agaaatacta gaggcagaaa 2760  
 ttgaatcaca ccatcctaga ctggcttctg ctgtacaaga ccatgatcaa attgtgacat 2820  
 45 caagaaaaag tcaagaacct gctttccaca ttgcaggaga tgcttgttg caagaaaaa 2880  
 tgaatgttga tgtgagtagt acgatataata acaatgaggt gctccatcaa ccactttctg 2940  
 50 aagctcaaag gaaatccaaa agcctaaaaa ttaatctcaa ttatgctggag atgctctaag 3000  
 agaaaataca ttggtttcag aacatgcaca aagagaccaa cgtgaaacac agtgtcaaat 3060  
 gaaggaagct gaacacatgt atcaaacga acaagataat gtgaacaaac aactgaaca 3120  
 55 gcaggagtct ctagatcaga aattatttca actacaaagc aaaaatatgt ggcttcaaca 3180  
 gcaattagtt catgcacata agaaagctga caacaaaagc aagataacaa ttgatattca 3240  
 ttttcttgag aggaaaatgc aacatcatct cctaaaagag aaaaatgagg agatatttaa 3300  
 60 ttacaataac catttaaaaa accgtatata tcaatatgaa aaagagaaag cagaaacaga 3360  
 aaactcatga gagacaagca gtaagaaact tcttttggag aaacaacaga ccagatcttt 3420  
 65 actcacaact catgctagga ggccagtcct agcatcacct tatgttgaaa atcttaccaa 3480  
 tagtctgtgt caacagaata cttattttag aagaaaaatt catgatttct tcctgaagcc 3540

ES 2 525 545 T3

tacagacata aaataacagt gtgaagaati acttgttcac gaattgcata aagctgcaca 3600  
ggattcccat ctaccctgat gatgcagcag acatcattca atccaaccag aatctcgctc 3660  
5 tgcactcag gctgg 3675

<210> 95  
<211> 2658  
10 <212> ADN  
<213> Homo sapiens

<400> 95

15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

ES 2 525 545 T3

	accagaaga	ccgtgccttg	cctggaagtc	ctgcctgtag	gcctgaagga	cttgccttaa	60
5	cagagcctca	acaactacct	ggtgattcct	acttcagccc	cttgggtgtga	gcagcttctc	120
	aacatgaact	acagcctcca	cttggccttc	gtgtgtctga	gtctcttcac	tgagaggatg	180
	tgcatccagg	ggagtcagtt	caacgtcgag	gtcggcagaa	gtgacaagct	ttccctgcct	240
10	ggctttgaga	acctcacagc	aggatataac	aaatttctca	ggccaattt	tggtggagaa	300
	cccgtacaga	tagcgctgac	tctggacatt	gcaagtatct	ctagcatttc	agagagtaac	360
15	atggactaca	cagccacat	atacctccga	cagcgctgga	tggaccagcg	gctgggtgtt	420
	gaaggcaaca	agagcttcac	tctggatgcc	cgctcgtgg	agttcctctg	ggtgccagat	480
	acttacattg	tggagtccaa	gaagtccttc	ctccatgaag	tcactgtggg	aaacaggctc	540
20	atccgcctct	tctccaatgg	cacggtcctg	tatgccctca	gaatcacgac	aactgttgca	600
	tgtaacatgg	atctgtctaa	atacccatg	gacacacaga	catgcaagtt	gcagctggaa	660
25	agctggggct	atgatggaaa	tgatgtggag	ttcacctggc	tgagagggaa	cgactctgtg	720
	cgtggactgg	aacacctgcg	gcttgctcag	tacaccatag	agcggtat	caccttagtc	780
	accagatcgc	agcaggagac	aggaaattac	actagattgg	tcttacagtt	tgagcttcgg	840
30	aggaatgttc	tgtatttcat	tttggaaacc	tacgttcctt	ccactttcct	ggtgggtgtg	900
	tcctggggtt	cattttggat	ctctctcgat	tcagtccctg	caagaacctg	cattggagtg	960
35	acgaccgtgt	tatcaatgac	cacactgatg	atcgggtccc	gcacttctct	tccaacacc	1020
	aactgcttca	tcaaggccat	cgatgtgtac	ctggggatct	gctttagctt	tgtgtttggg	1080
	gccttgctag	aatatgcagt	tgctcactac	agttccttac	agcagatggc	agccaaagat	1140
40	agggggacaa	caaaggaagt	agaagaagtc	agtattacta	atatcatcaa	cagctccatc	1200
	tccagcttta	aacggaagat	cagctttgcc	agcattgaaa	ttccagcga	caacgttgac	1260
	tacagtgact	tgacaatgaa	aaccagcgac	aagttaaagt	ttgtcttccg	agaaaagatg	1320
45	ggcaggattg	ttgattattt	cacaattcaa	aaccccagta	atggtgatca	ctattccaaa	1380
	ctactgtttc	ctttgatttt	tatgctagcc	aatgtatttt	actgggcata	ctacatgtat	1440
50	ttttgagtca	atgttaaatt	tcttgcagtc	cataggtcct	caacaggaca	agataatgat	1500
	gtaaattgga	ttttaggcca	agtgtgcacc	cacatccaat	ggtgctacaa	gtgactgaaa	1560
	taatatttga	gtctttctgc	tcaaagaatg	aagctccaac	cattgttcta	agctgtgtag	1620
55	aagtcctagc	attataggat	cttgtaatag	aaacatcagt	ccattcctct	ttcatcttaa	1680
60							
65							

ES 2 525 545 T3

	tcaaggacat tcccatggag cccaagattā caaatgtact cagggctggt tattcgggtgg	1740
5	ctccctgggt tgcatttacc tcatataaag aatgggaagg agaccattgg gtaaccctca	1800
	agtgtcagaa gttgttttcta aagtaactat acatgttttt tactaaatct ctgcagtgct	1860
	tataaaatac attgtttgcct atttagggag taacattttc tagtttttgt ttctgggtta	1920
10	aatgaaatat gggcttatgt caattcattg gaagtcaatg cactaactca ataccaagat	1980
	gagtttttaa ataatgaata ttatttaata ccacaacaga attatcccca atttccaata	2040
	agtcctatca ttgaaaattc aaatataagt gaagaaaaaa ttagtagatc aacaatctaa	2100
15	acaaatccct cggttctaag atacaatgga ttccccatac tggaaggact ctgaggcttt	2160
	attccccac tatgcatatc ttatcatttt attattatac acacatccat cctaaactat	2220
20	actaaagccc ttttcccatg catggatgga aatggaagat ttttttttaa cttgttctag	2280
	aagtcttaat atgggctggt gccatgaagg cttgcagaat tgagtccatt ttctagctgc	2340
	ctttattcac atagtgacgg ggtactaaaa gtactggggt gactcagaga gtcgctgtca	2400
25	ttctgtcatt gctgctactc taacactgag caaactctc ccagtggcag atcccctgta	2460
	tcattccaag aggagcattc atccccttgc tctaatgatc aggaatgatg cttattagaa	2520
30	aacaaactgc ttgaccagcagg aacaagtggc ttagcttaag taaacttggc tttgctcaga	2580
	tcctgatcc ttccagctgg tctgctctga gtggcttatc ccgcatgagc aggagcgtgc	2640
	tggccctgag tactgaac	2658
35		
	<210> 96	
	<211> 2531	
	<212> ADN	
40	<213> Homo sapiens	
	<400> 96	
45		
50		
55		
60		
65		

ES 2 525 545 T3

5	gcagtgtag gcaatcgctc tacccttgac cccttccttt gcacagtgag tgatggcggt	60
	tttatctcct gatgatgatg cacagccttc agcggggggac atttaagacg cagaacacca	120
	ggtccaggct gcagctgcgg gactcagagg cgaagcttga ggggctcagg aaggacgaag	180
10	aaccaccctt gagagaagag gcagcagcag cggcggcagc agcagcggca gcgacccac	240
	cactgccaca ttgcccagga aacaatgctg cttagcgacat tcaagctgtg cgctgggagc	300
15	tcctacagac acatgcgcaa catgaagggg ctgaggcaac aggctgtgat ggccatcagc	360
	caggagctga accggagggc cctggggggc cccacccta gcacgtggat taaccagggt	420
	cggcggcgga gctctctact cggttctcgg ctggaagaga ctctctacag tgaccaggag	480
20	ctggcctatc tccagcaggg ggaggaggcc atgcagaagg ccttgggcat ccttagcaac	540
	caagagggct ggaagaagga gagtcagcag gacaatgggg acaaagtgat gagtaaagtg	600
	gtcccagatg tgggcaaggt gttccggctg gaggtcgtgg tggaccagcc catggagagg	660
25	ctctatgaag agctcgtgga gcgcatggaa gcaatggggg agtggaacc caatgtcaag	720
	gagatcaagg tcctgcagaa gatcgggggg ccccgctgact ttgtgagcgt gcgctgtgcc	780
30	aagcggcgag gctccacctg tgtgctggct ggcatggcca cagacttcgg gaacatgcct	840

35

40

45

50

55

60

65



ES 2 525 545 T3

gagcagaagg gtgtcatcag ggcggagcac ggtcccactt gcatggtgct tcacccgttg 900  
 gctggaagtc cctctaagac caaacttacg tggctactca gcatcgacct caaggggtgg 960  
 5 ctgccaaga gcatcatcaa ccaggtcctg tcccagaccc aggtggattt tgccaaccac 1020  
 ctgcgcaagc gcctggagtc ccaccctgcc tctgaagcca ggtgttgaag accagcctgc 1080  
 tgttcccaac tgtgcccagc tgcactggta cacacgctca tcaggagaat ccctactgga 1140  
 10 agcctgcaag tctaagatct ccatctggtg acagtgggat gggggtgggtt cgtgtttaga 1200  
 gtatgacact aggattcaga ttggtgaaag tttttagtac caagaaaaca gggatgaggc 1260  
 15 tcttggatta aaaggtaact tcattcactg attagctatg acatgagggg tcaggcccct 1320  
 aaaataattg taaaactttt tttctgggcc cttatgtacc cacctaaaac catctttaa 1380  
 atgctagtgg ctgatatggg tgtgggggat gctaaccaca gggcctgaga agtcttgctt 1440  
 20 tatgggctca agaatgccat gcgctggcag tacatgtgca caaagcagaa tctcagaggg 1500  
 tctctgcag ccctctgctc ctcccggccg ctgcacagca acaccacaga acaagcagca 1560  
 cccacagtg ggtgccttcc agaaatatag tccaagcttt ctctgtggaa aaagacaaaa 1620  
 25 ctcatagta gacatgtttc cctattgctt tcataggcac cagtcagaat aaagaatcat 1680  
 aattcacaca aacatcagtc tttgttttaa tattgtactg ttaaaaaaat ctatgcagct 1740  
 ggggtgcagtg gctcacgcct gtaatcccag cattttggga ggctgaggta ggcggatcac 1800  
 30 aaggctcagga gatcgagacc atcctggcca acatggtgaa acccctctc tactaaaata 1860  
 caaaaaatta gctgggtgtg gtggcgcaaa cctgtagtcg tagctacttg ggaggctgag 1920  
 35 gcaggggaat cacttgaacc ccggaggcgg aggttgtagt gggccgagat tgtgccactg 1980  
 cgctccagcg tgggcgacag agtgagactc catctcaaaa aaaaaaaaaa aaaaatctat 2040  
 gctagtagat tacaacttca cactagagga gttctggaca aagcttttaa ttagtcaaac 2100  
 40 taaattaagg ctcatataaa ggaaaggaac tactgggaaa ttatgcaatt caataattta 2160  
 gactctgtta ccaggatctt tcataaaaat ttaatttcca taatcataac ctaaattgagt 2220  
 45 tcttaaagaa ttctataagc aatagctgat taatgggcc tggagatga agattataac 2280  
 tgtttattta cctaattaaa aggaaaggca gtgccaaata tgagaggata aacaatatta 2340  
 gttaacattt ctgttattta tgatgccaat tagtagtaag ataattccac agctgtcaac 2400  
 50 tttgtttggg gctggcaact tctctgtta aacaggctaa aagtttagta ttctgggaga 2460  
 agtggctgga agaaggggta atatggtgaa agcaaattcc ctttcccagg agtcaagaga 2520  
 atttatgtga g 2531  
 55  
 <210> 97  
 <211> 6164  
 <212> ADN  
 <213> Homo sapiens  
 60  
 <220>  
 <221> misc\_feature  
 <222> (2851)..(2851)  
 65

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

<220>

<221> misc\_feature

<222> (2856)..(2856)

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

<400> 97

5

10

15

20

25

30

35

40

45

50

55

60

65

ES 2 525 545 T3

cgggcgccgc aggagcgagt gagctgggag cgaggggcca aggcgcggag aagccccgcc 60  
 gcccggtggg cggcagaagg ctacgccgag gcggcgccgc cgactccgtt cactctcgg 120  
 5 cccggatcca ggcctccggg ttcccaggcg ctacactccc tctgacgcac tttaaagagt 180  
 ctccccctt ccactcagg gcgagtaata gcgaccaatc atcaagccat ttaccaggct 240  
 tcggaggaag ctgtttatgt gatccccgca ctaattaggc tcatgaacta acaaatcggt 300  
 10 tgcacaactt gtgaagaagc gaacacttcc atggattgtc cttggactta gggcgccctg 360  
 cccgcctttt gcagaggaga aaaaactttt tttttttttt gcctcccccg agaactttcc 420  
 ccccttctcc tccctgcctc taactccgat cccccacgc catctcgcca aaaaaaaaaa 480  
 15 aaaaaaaaaa aaagaaaaaa aaagaaaaaa aaagaaaaaa aattacccca atccacgcct 540  
 gcaaattctt ctggaaggat tttccccct ctcttcaggc tgggcgcgtt tgggtgcaaga 600  
 20 ttctcgggat cctcggcttt gcctctccct ctccctcccc cctcctttcc tttttccttt 660  
 cttttccttt ctttcttctt ttcttcccc ccacccccac cccaccccca aacaaacgag 720  
 tcccgaattc tcgtccgtcc tcgccgcggg cagcgggcgg cggaggcagc gtgcggcggt 780  
 25 cgccaggagc tgggagccca gggcgccgc tcctcggcgc agcatgttcc agccggcgcc 840  
 caagcgtgc ttcaccatcg agtcgctggt ggccaaggac agtcccctgc ccgcctcgcg 900  
 30 ctccgaggac cccatccgtc ccgcggcact cagctacgct aactccagcc ccataaatcc 960  
 gttcctcaac ggcttccact cggccgccgc cgccgccgcc ggtaggggcg tctactccaa 1020  
 cccggacttg gtgttcgccg aggcggctct gcacccgcc aaccccgccg tgccagtgca 1080  
 35 cccggtgccg ccgccgcag ccttggccgc ccaccccta cctcctcgc actcgccaca 1140  
 ccccctattc gcctcgcagc agcgggatcc gtccacctc taccctggc tcatccaccg 1200  
 ctaccgatat ctgggtcatc gcttccaagg gaacgacact agccccgaga gtttcctttt 1260  
 40 gcacaacgcg ctggcccgaag agcccaagcg gatccgaacc gccttctccc cgtcccagct 1320  
 tctaaggctg gaacacgcct ttgagaagaa tctactcgtg gtgggcgccg aaaggaagca 1380  
 45 gctggcacac agcctcagcc tcacggaaac tcaggtaaaa gtatggtttc agaaccgaag 1440  
 aacaaagttc aaaaggcaga agctggagga agaaggctca gattcgcaac aaaagaaaaa 1500  
 agggacgcac catattaacc ggtggagaat cgccaccaag caggcgagtc cggaggaaat 1560  
 50 agacgtgacc tcagatgatt aaaaacataa acctaacccc acagaaacgg acaacatgga 1620  
 gcaaaagaga caggagagg tggagaagga aaaaacccta caaaacaaaa acaaaccgca 1680  
 55 tacacgttca ccgagaaagg gagaggggat cggagggagc agcggaatgc ggcgaagact 1740  
 ctggacagcg agggcacagg gtcccaaacc gaggccgcgc caagatggca gaggatggag 1800  
 gctccttcat caacaagcga cctcgtcta aagaggcagc tgagtgagag acacagagag 1860

60

65

ES 2 525 545 T3

	aaggagaaag	agggagggag	agagagaaaag	agagagaaag	agagagagag	agagagagag	1920
	agaaagctga	acgtgcactc	tgacaagggg	agctgtcaat	caaacaccaa	accggggaga	1980
5	caagatgatt	ggcaggtatt	ccgtttatca	cagtccactt	aaaaaatgat	gatgatgata	2040
	aaaaccacga	cccaaccagg	cacaggactt	ttttgttttt	tgcacttcgc	tgtgtttccc	2100
	ccccatcttt	aaaaataatt	agtaataaaa	aacaaaaatt	ccatatctag	ccccatccca	2160
10	cacctgtttc	aaatccttga	aatgcatgta	gcagttgttg	ggcgaatggt	gtttaaagac	2220
	cgaaaatgaa	ttgtaatfff	cttttccttt	taaagacagg	ttctgtgtgc	tttttatfff	2280
	gatttttttt	cccaagaaat	gtgcagtctg	taaacacttt	ttgatacctt	ctgatgtcaa	2340
15	agtgattgtg	caagctaaat	gaagtaggct	cagcgatagt	ggtcctctta	cagagaaacg	2400
	gggagcagga	cgacgggggg	gctgggggtg	gcgggggagg	gtgccacaa	aaagaatcag	2460
20	gacttgact	gggaaaaaaa	cccctaaatt	aattatattt	cttggacatt	ccctttccta	2520
	acatcctgag	gcttaaaacc	ctgatgcaa	cttctccttt	cagtggttgg	agaaattggc	2580
	cgagttcaac	cattcactgc	aatgcctatt	ccaaacttta	aatctatcta	ttgcaaaacc	2640
25	tgaaggactg	tagttagcgg	ggatgatggt	aagtgtggcc	aagcgcacgg	cggcaagttt	2700
	tcaagcactg	agtttctatt	ccaagatcat	agacttacta	aagagagtga	caaatgcttc	2760
	cttaatgtct	tctataccag	aatgtaaata	tttttgtggt	ttgtgttaat	ttgttagaat	2820
30	tctaacacac	tataacttc	caagaagtad	nahmanaaaa	gatagatcct	gctccaggag	2880
	ccgggaagcc	tcgccctggc	cagctgtgct	gggcacctcc	cctgcctgct	tcctggccca	2940
	cttgacaggca	aggtgagggc	atgcgaatgg	ctgccactgc	ctgggcgggg	ctccaagggc	3000
35	caccctccc	caccctctgt	cccgcagtga	ggacgggact	ctactgccga	gaccaggctc	3060
	acgctgagag	gtgggccatg	acctccgaga	cctcttcggt	aagccactgt	gccaggagca	3120
40	ggatgctgcg	gcgacggggc	caggaagagg	acagcacctg	cctgatcgat	gtgagcccc	3180
	ctgaggcaga	gaagaggggc	tcttacggga	gcacagccca	cgctcggag	ccaggtggac	3240
	agcaagcggc	cgctcgcaga	gctggggagtc	ctgccaagcc	ccggatcgca	gacttcgtcc	3300
45	tcgtttggga	ggaggacctg	aagctagaca	ggcagcagga	cagtgccgcc	cgggacagaa	3360
	cagacatgca	caggacctgg	cgggagactt	ttctggataa	tcttcgtgcg	gctgggctgt	3420
	gtgtagacca	gcaggacgtc	caggacggga	acaccacagt	gcactacgcc	ctcctcagcg	3480
50	cctcctgggc	tgtgctctgc	tactacgccg	aagacctgcg	cctgaagctg	cccttcagag	3540
	agttacccaa	ccaggcctcc	aactggtcgg	ccggcctgct	ggcatggctg	ggcatcccca	3600
	acgtcctgct	ggaggttgtg	ccagacgtac	cccccgagta	ctactcctgc	cggttcagag	3660
55	tgaacaagct	gccacgcttc	ctcgggagtg	acaaccagga	caccttcttc	acaagcacca	3720
	agaggcacca	aattctgttt	gagatcctgg	ccaagacccc	gtatggccac	gagaagaaaa	3780
60	acctgcttgg	gatccaccag	ctgctggcag	aggggtgcct	cagtgccgcc	ttccccctgc	3840
	atgacggccc	cttcaagacg	ccccagagg	gcccgcaggc	tccacgcctc	aaccagcgcc	3900
	aagtcctttt	ccagcactgg	gcgcgctggg	gcaagtggaa	caagtaccag	cccctggacc	3960
65							

ES 2 525 545 T3

acgtgcgcag gtacttcggg gagaaggtgg ccctctactt cgcctggctc gggttttaca 4020  
 caggctggct cctgccagcg gcagtggtgg gcacactggt gttcctgggt ggctgcttcc 4080  
 5 tgggtgttctc agacataccc acgcaggaac tgtgtggcag caaggacagc ttcgagatgt 4140  
 gccactttg cctcgactgc cttttctggc tgctctccag cgcctgtgcc ctggcccagg 4200  
 ccggccggct gttcgaccac ggcggcaccg tgttcttcag cttgttcatg gactgtggg 4260  
 10 ccgtgctgct gctggagtac tggaagcggg agagcggcac gctggcctac cgctgggact 4320  
 gctctgacta cgaggacact gaggagaggc ctcggcccca gtttgccgcc tcagcccca 4380  
 tgacagcccc gaaccccatc acgggtgagg acgagcccta cttccctgag aggagccgg 4440  
 15 cgcgccgcat gctggccggc tctgtgtgga tcgtggtgat ggtggccgtg gtggtcatgt 4500  
 gcctcgtgtc tatcatcctg taccgtgcca tcatggccat cgtggtgtcc aggtcgggca 4560  
 acacccttct cgcagcctgg gcctctcgca tcgccagcct cacggggtct gtagtgaacc 4620  
 20 tcgtcttcat cctcatcctc tccaagatct atgtatcctt ggcccacgtc ctgacacgat 4680  
 gggaaatgca ccgcaccag accaagtctg aggacgcctt caccctcaag gtgttcatct 4740  
 tccagttcgt caacttctac tcctcaccgg tctacattgc cttcttcaag ggcaggtttg 4800  
 tgggataccc aggcaactac cacaccttgt ttggagtccg caatgaggag tgcgcggctg 4860  
 gaggctgcct gatcgagctg gcacaggagc tcctgggtcat catggtgggc aagcaggta 4920  
 30 tcaacaacat gcaggaggtc ctcatcccga agctaaaggg ctggtggcag aagtccggc 4980  
 ttcgctccaa gaagaggaag gcgggagctt ctgcaggggc tagccagggg ccctgggagg 5040  
 acgactatga gcttgtgccc tgtgagggtc tgtttgacga gtacctgaa atggtgctgc 5100  
 35 agttcggctt cgtcaccatc ttcgtggccg cctgtccgct cgcgccgctc ttcgccctgc 5160  
 tcaacaactg ggtggagatc cgcttgacg cgcgcaagtt cgtctgcgag taccggcgcc 5220  
 ctgtggccga gcgcgccag gacatcggca tctggttcca catcctggcg ggcctcacgc 5280  
 40 acctggcggg catcagcaac gccttctctc tggccttctc gtccgacttc ctgccgcgcg 5340  
 cctactaccg gtggaccgcg gccacgacc tgcgcggctt cctcaacttc acgctggcgc 5400  
 gagccccgtc ctcttctgcc gccgcgaca accgcacgtg caggatcgg gctttccggg 5460  
 45 atgacgatgg acattattcc cagacctact ggaatcttct tgccatccgc ctggccttcg 5520  
 tcattgtggt tgagcatgtg gttttctccg ttggccgcct cctggacctc ctggtgcctg 5580  
 50 acatcccaga gtctgtggag atcaaagtga agcgggagta ctacctggct aagcaggcac 5640  
 tggtgagaa tgaggttctt tttggaacga acggaacaaa ggatgagcag cccaagggct 5700  
 cagagctcag ctcccactgg acacccttca cggttcccaa ggccagccag ctgcagcagt 5760  
 55 gacgcctgga aggacatctg gtggtcctta ggggagtggc ccctcctgag ccctcgcgagc 5820  
 agcgtccttt tcctcttccc tcaggcagcg gctgtgtgaa ccgctggctg ctgttgtgcc 5880  
 60 tcatctctgg gcacattgcc tgcttcccc cagcgcgggc ttctctctc agagcgctg 5940  
 tcaactccatc cccggcaggg agggaccgtc agctcacaag gccctctttg tttcctgctc 6000

65

ES 2 525 545 T3

	ccagacataa gcccgaagggg cccctgcacc caagggaccc tgtccctcgg tggcctcccc	6060
	aggcccctgg acacgacagt tctcctcagg caggtgggct ttgtggctct cgccgcccct	6120
5	ggccacatcg cctctcctc ttacacctgg tgaccttca atgt	6164
10	<210> 98 <211> 551 <212> ADN <213> Homo sapiens  <400> 98	
15	accccatccg ctggctctca cccctcggag acgctcggcc gacagcatag tacttgccgc	60
	ccagccacgc ccgcgcgcca gccaccatgc taggtaacaa gcgactgggg ctgtccggac	120
20	tgaccctcgc cctgtccctg ctcgtgtgcc tgggtgcgct ggccgaggcg taccctcca	180
	agccggacaa cccgggagag gacgcaccag cggaggacat ggccagatac tactcggcgc	240
	tgcgacacta catcaacctc atcaccaggc agagatatgg aaaacgatcc agcccagaga	300
25	caactgatttc agacctcttg atgagagaaa gcacagaaaa tgttcccaga actcggcttg	360
	aagaccctgc aatgtgggtga tgggaaatga gacttgctct ctggcctttt cctattttca	420
30	gcccataatt catcgtgtaa aacgagaatc cacccatcct accaatgcat gcagccactg	480
	tgctgaattc tgcaatgttt tcctttgtca tcattgtata tatgtgtgtt taaataaagt	540
	atcatgcatt c	551
35	<210> 99 <211> 1607 <212> ADN <213> Homo sapiens  <400> 99	
40		
45		
50		
55		
60		
65		

ES 2 525 545 T3

aatgactcct ttcggaagt gcagtggaag ctgtacactg cccaggcaaa gcgtccgggc 60  
 agcgtaggcg ggcgactcag atcccagcca gtggacttag cccctgtttg ctctccgat 120  
 5 aactgggggtg accttggtta atattcacca gcagcctccc ccgttgcccc tctggatcca 180  
 ctgcttaaat acggacgagg acagggccct gtctcctcag cttcaggcac caccactgac 240  
 ctgggacagt gaatcgacaa tgccgtcttc tgtctcgtgg ggcacccctc tgctggcagg 300  
 10 cctgtgctgc ctggtccctg tctccctggc tgaggatccc caggagatg ctgcccagaa 360  
 gacagataca tcccaccatg atcaggatca cccaaccttc aacaagatca cccccaacct 420  
 15 ggctgagttc gccttcagcc tataccgcca gctggcacac cagtccaaca gcaccaatat 480  
 cttcttctcc ccagtgagca tcgctacagc ctttgcaatg ctctccctgg ggaccaaggc 540  
 tgacactcac gatgaaatcc tggagggcct gaatttcaac ctcacggaga ttccggaggc 600  
 20 tcagatccat gaaggcttcc aggaactcct ccgtacctc aaccagccag acagccagct 660  
 ccagctgacc accggcaatg gcctgttctt cagcgagggc ctgaagctag tggataagtt 720  
 tttggaggat gttaaaaagt tgtaccactc agaagccttc actgtcaact tcggggacac 780  
 25 cgaagaggcc aagaaacaga tcaacgatta cgtggagaag ggtactcaag ggaaaattgt 840  
 ggatttggtc aaggagcttg acagagacac agtttttgct ctggtgaatt acatcttctt 900  
 30 taaaggcaaa tgggagagac cttttgaagt caaggacacc gaggaagagg acttccacgt 960  
 ggaccaggty accaccgtga aggtgcctat gatgaagcgt ttaggcatgt ttaacatcca 1020  
 35 gcactgtaag aagctgtcca gctgggtgct gctgatgaaa tacctgggca atgccaccgc 1080  
 catcttcttc ctgcctgatg aggggaaact acagcacctg gaaaatgaac tcaccacga 1140  
 tatcatcacc aagttcctgg aaaatgaaga cagaaggctt gccagcttac atttacccaa 1200  
 40 actgtccatt actggaacct atgatctgaa gagcgtcctg ggtcaactgg gcatcactaa 1260  
 ggtcttcagc aatggggctg acctctccgg ggtcacagag gaggcacccc tgaagctctc 1320  
 caaggccgtg cataaggctg tgctgaccat cgacgagaaa gggactgaag ctgctggggc 1380  
 45 catgttttta gaggccatac ccatgtctat ccccccgag gtcaagttca acaaaccctt 1440  
 tgtcttctta atgattgaac aaaataccaa gtctccccct tcatgggaa aagtggtgaa 1500  
 50 tcccaccaa aaataactgc ctctcgtctc tcaaccctc ccctccatcc ctggccccct 1560  
 ccctggatga cattaagaa gggttgagct ggtccctgcc tgcaaaa 1607

55 <210> 100  
 <211> 1753  
 <212> ADN  
 <213> Homo sapiens

60 <400> 100

65

65

ES 2 525 545 T3

5 cagccccgcc cctacctgtg gaagcccagc cgcccgctcc cgcgataaa aggcgaggag 60  
 tgtccccgag gtcagcgagt gcgcgctcct cctcgcccgc cgctagggtcc atcccggccc 120  
 agccaccatg tccatccact tcagctcccc ggtattcacc tcgcgctcag ccgccttctc 180  
 10 gggccgaggc gccaggtgc gcctgagctc cgctcgcccc ggcggccttg gcagcagcag 240  
 cctctacggc ctcgggcctt cacggccgag cgtggccgtg cgctctgcct atgggggccc 300  
 ggtgggagcc ggcattccgc aggtcaccat taaccagagc ctgctggccc cgctgaggct 360  
 ggacgccgac ccctccctcc agcgggtgag ccaggaggag agcgagcaga tcaagaccct 420  
 15 caacaacaag ttgacctctt tcatcgacaa ggtgaggttt ctggagcagc agaacaagct 480  
 gctggagacc aagtggagcg tgctgcagga gcagaagtcg gccaaagaca gccgcctccc 540  
 20 agacatcttt gaggcccaga ttgctggcct tcgggggtcag cttgaggcac tgcagggtgga 600  
 tgggggcccgc ctggaggcgg agctgaggag catgcaggat gtggtggagg acttcaagaa 660  
 taagtacgaa gatgaaatta accaccgcac agctgctgag aatgagtttg tgggtgctgaa 720  
 25 gaaggatgtg gatgctgcct acatgagcaa ggtggagctg gaggccaagg tggatgcctt 780  
 gaatgatgag atcaacttcc tcaggaccct caatgagagc gagttgacag agctgcagtc 840  
 ccagatctcc gacacatctg ttgtgctgtc catggacaac agtcgctccc tggacctgga 900  
 30 cggcatcatc gctgagggtca aggcgcagta tgaggagatg gccaaatgca gccgggctga 960  
 ggctgaagcc tgggtaccaga ccaagtgtga gaccctccag gccagggtg ggaagcatgg 1020  
 35 ggacgacctc cggaataccc ggaatgagat ttcagagatg aaccgggcca tccagaggct 1080  
 gcaggctgag atcgacaaca tcaagaacca gcgtgccaag ttggaggccg ccattgccga 1140  
 ggctgaggag cgtgggggagc tggcgctcaa ggatgctcgt gccaaagcagg aggagctgga 1200  
 40 agccgccctg cagcggggca agcaggatat ggcacggcag ctgctgagat accaggaact 1260  
 catgagcgtg aagctggccc tggacatcga gatcgccacc taccgcaagc tgctggaggg 1320  
 cgaggagagc cgggttgctg gagatggagt gggagccgtg aatatctctg tgatgaattc 1380  
 45 cactggtggc agtagcagtg gcggtggcat tgggctgacc ctggggggaa ccatgggcag 1440  
 caatgccctg agcttctcca gcagtgcggg tcctgggctc ctgaaggctt attccatccg 1500  
 50 gaccgcatcc gccagtcgca ggagtgcccg cgactgagcc gcctcccacc actccactcc 1560  
 tccagccacc accacaatc acaagaagat tcccaccctt gcctcccatt cctgggtcca 1620  
 agacagtgag acagtctgga aagtgatgtc agaatagctt ccaataaagc agcctcattc 1680  
 55 tgaggcctga gtgatccacg tgaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1740  
 aaaaaaaaaa aaa 1753

60 <210> 101  
 <211> 2276  
 <212> ADN  
 <213> Homo sapiens  
 65 <400> 101



ES 2 525 545 T3

aagcccagca gccccggggc ggatggctcc ggccgcctgg ctccgcagcg cggcccgcgcg 60  
 cgccctcctg cccccgatgc tgctgctgct gctccagccg ccgcccgtgc tggccccgggc 120  
 5 tctgccgccg gacgcccacc acctccatgc cgagaggagg gggccacagc cctggcatgc 180  
 agccctgccc agtagcccgg cacctgcccc tgccacgcag gaagcccccc ggctgcccag 240  
 cagcctcagg cctccccgct gtggcgtgcc cgacccatct gatgggctga gtgcccgcaa 300  
 10 ccgacagaag aggttcgtgc tttctggcgg gcgctgggag aagacggacc tcacctacag 360  
 gatccttcgg ttcccatggc agttgggtgca ggagcagggtg cggcagacga tggcagaggc 420  
 15 cctaaaggta tggagcgtat tgacgccact cacctttact gaggtgcacg agggccgtgc 480  
 tgacatcatg atcgacttcg ccaggtaacty gcatggggac gacctgccgt ttgatggggc 540  
 tgggggcatc ctggcccatg ctttcttccc caagactcac cgagaagggg atgtccactt 600  
 20 cgactatgat gagacctgga ctatcgggga tgaccagggc acagacctgc tgcaggtggc 660  
 agcccatgaa tttggccacg tgctggggct gcagcacaca acagcagcca aggccctgat 720  
 gtccgccttc tacacctttc gctaccact gagtctcagc ccagatgact gcaggggcgt 780  
 25 tcaacaccta tatggccagc cctggcccac tgtcacctcc aggaccccag ccctggggccc 840  
 ccaggctggg atagacacca atgagattgc accgctggag ccagacgccc cgccagatgc 900  
 30 ctgtgaggcc tcctttgacg cggctctccac catccgaggc gagctctttt tcttcaaagc 960  
 gggctttgtg tggcgcctcc gtggggggcca gctgcagccc ggctacccag cattggcctc 1020  
 tcgccactgg cagggactgc ccagccctgt ggacgctgcc ttcgaggatg cccagggcca 1080  
 35 catttggttc ttccaagggtg ctcagtactg ggtgtacgac ggtgaaaagc cagtcctggg 1140  
 ccccgacacc ctcaccgagc tgggcctggt gaggttcccg gtccatgctg ccttggtctg 1200  
  
 40  
  
 45  
  
 50  
  
 55  
  
 60  
  
 65

ES 2 525 545 T3

5 ggggtcccgag aagaacaaga tctacttctt ccgaggcagg gactactggc gtttccaccc 1260  
 cagcaccggg cgtgtagaca gtcccgtgcc ccgcagggcc actgactgga gaggggtgcc 1320  
 ctctgagatc gacgctgcct tccaggatgc tgatggctat gcctacttcc tgcgcggccg 1380  
 cctctactgg aagtttgacc ctgtgaaggc gaaggctctg gaaggcttcc cccgtctcgt 1440  
 10 gggtcctgac ttctttggct gtgccgagcc tgccaacact ttcctctgac catggccttg 1500  
 atgccctcag ggggtgctgac ccctgccagg ccacgaatat caggctagag acccatggcc 1560  
 atctttgtgg ctgtgggcac caggcatggg actgagccca tgtctcctca gggggatggg 1620  
 15 gtgggggtaca accaccatga caactgccgg gagggccacg caggctcgtg tcacctgcca 1680  
 gcgactgtct cagactgggc agggaggctt tggcatgact taagaggaag ggcagtcttg 1740  
 ggcccgtat gcaggctctg gcaaacctgg ctgccctgtc tccatccctg tccctcaggg 1800  
 20 tagcaccatg gcaggactgg gggaaactgga gtgtccttgc tgtatccctg ttgtgaggtt 1860  
 ccttccaggg gctggcactg aagcaagggt gctggggccc catggccttc agccctggct 1920  
 gagcaactgg gctgtagggc agggccactt cctgagggtca ggtccttgga ggtgcctgca 1980  
 25 tctgtctgcc ttctggctga caatcctgga aatctgttct ccagaatcca ggccaaaaag 2040  
 ttcacagtca aatggggagg ggtattcttc atgcaggaga cccagggccc tggaggctgc 2100  
 30 aacatacctc aatcctgtcc caggccggat cctcctgaag cccttttcgc agcactgcta 2160  
 tcctccaaag ccattgtaaa tgtgtgtaca gtgtgtataa accttcttct tctttttttt 2220  
 tttttaact gaggattgtc attaaacaca gttgttttct aaaaaaaaaa aaaaaa 2276

35

<210> 102  
 <211> 7381  
 <212> ADN  
 <213> Homo sapiens  
 <400> 102

40

45

50

55

60

65

ES 2 525 545 T3

	tacagcccca	aggtcgctcc	ctctggggcc	ctttcttccc	cattcttccc	agcagcccaa	60
	agctctggtg	ggacaggggc	agcccctggg	gagggaggag	aggacccagg	aaccgggcta	120
5	ggaggggtggc	ccaccattt	ccagtgtgac	ctgttcccat	tccccatgt	ctcctcccat	180
	ccctcccgcc	actcagctca	ggctgatgag	aagcagagca	acgggtgtat	cggtgttttc	240
	tttcctggtg	gggtagtggg	gtggggctga	ggagagaaaa	gggtgattag	cgtggggccc	300
10	cgccctcttt	tgtcctcttc	ccaggttccc	tggccccttc	ggagaaacgc	acttggttcg	360
	ggccagccgc	ctgaggggac	gggctcacgt	ctgctcctca	cactgcagct	gctggggccgt	420
15	ggagcttccc	cagggagcca	gggggacttt	tgccgcagcc	atgaaggggg	cacgctggag	480
	gagggtcccc	tgggtgtccc	tgagctgcct	gtgtctctgc	ctccttccgc	atgtggtccc	540
	aggaaccaca	gaggacacat	taataactgg	aagtaaaact	cctgccccag	tcacctcaac	600
20	aggctcaaca	acagcgacac	tagagggaca	atcaactgca	gcttcttcaa	ggacctctaa	660
	tcaggacata	tcagcttcat	ctcagaacca	ccagactaag	agcacggaga	ccaccagcaa	720
25	agctcaaacc	gacaccctca	cgcagatgat	gacatcaact	cttttttctt	ccccaagtgt	780
30							
35							
40							
45							
50							
55							
60							
65							

ES 2 525 545 T3

acacaatgtg atggagactg ttacgcagga gacagctcct ccagatgaaa tgaccacatc 840  
 atttcctcc agtgtcacca acacactcat gatgacatca aagactataa caatgacaac 900  
 5 ctccacagac tccactcttg gaaacacaga agagacatca acagcaggaa ctgaaagttc 960  
 taccacagtg acctcagcag tctcaataac agctggacag gaaggacaat cacgaacaac 1020  
 ttcttgagg acctctatcc aagacacatc agctttctt cagaaccact ggactcggag 1080  
 10 cacgcagacc accagggaa ctcaaaccag caccctaaca cacagaacca cttcaactcc 1140  
 ttctttctct ccaagtgtac acaatgtgac agggactggt tctcagaaga catctcctc 1200  
 aggtgaaaca gctacctcat ccctctgtag tgtcacaac acatccatga tgacatcaga 1260  
 15 gaagataaca gtgacaacct ccacaggctc cactcttggga aaccaggagg agacatcatc 1320  
 agtacctggt actggaagtc ttatgccagt cacctcagca gccttagtaa cagttgatcc 1380  
 20 agaaggaca tcaccagcaa ctttctcaag gacttctact caggacacaa cagctttttc 1440  
 taagaaccac cagactcaga gcgtggagac caccagagta tctcaaatca acacctcaa 1500  
 caccctcaca ccggttaca catcaactgt ttatcctca ccaagtggat tcaaccaag 1560  
 25 tggaacagtt tctcaggaga cattcccttc tgggaaaca accatctcat ccccttcag 1620  
 tgtcagcaat acattcttg taacatcaa ggtgttcaga atgccaatct ccagagactc 1680  
 tactcttggga aacacagagg agacatcact atctgtaagt ggaaccattt ctgcaatcac 1740  
 30 ttccaaagtt tcaaccatat ggtggtcaga cactctgtca acagcactct ccccagttc 1800  
 tctacctca aaaatatcca cagctttcca caccagcag agtgaagggtg cagagaccac 1860  
 35 aggacggcct catgagagga gctcattctc tccagggtgtg tctcaagaaa tatttactct 1920  
 acatgaaaca acaacatggc cttcctcatt ctccagcaaa ggccacacaa cttggtcaca 1980  
 aacagaactg ccctcaacat caacaggtgc tgccactagg cttgtcacag gaaatccatc 2040  
 40 tacaagggca gctggcacta ttccaaggtt cccctctaag gtctcagcaa taggggaacc 2100  
 aggagagccc accacatact cctcccacag cacaactctc ccaaaaacaa caggggcagg 2160  
 cgcccagaca caatggacac aagaaacggg gaccactgga gaggctcttc tcagcagccc 2220  
 45 aagctatagt gtgattcaga tgataaaaac ggccacatcc ccatcttctt cacctatgct 2280  
 ggatagacac acatcacaac aaattacaac ggcaccatca acaaatcatt caacaataca 2340  
 50 ttccacaagc acctctctc aggaatcacc agctgtttcc caaaggggtc aactcagac 2400  
 cccgcagacc acacaagaat cacaaccac gaggtccgtc tccccatga ctgacaccaa 2460  
 gacagtcacc accccaggtt cttccttcac agccagtggt cactcgccct cagaaattgt 2520  
 55 tcctcaggac gcaccacca taagtgcagc aacaacctt gcccagctc ccaccgggaa 2580  
 tggtcacaca acccaggccc cgaccacagc actgcaggca gcaccagca gccatgatgc 2640  
 caccctgggg ccctcaggag gcacgtcact ttccaaaaca ggtgccctta ctctggccaa 2700  
 60 ctctgtagtg tcaacaccag gggcccaga aggacaatgg acatcagcct ctgccagcac 2760  
 ctcacctgac acagcagcag ccatgacca taccaccag gctgagagca cagaggcctc 2820

65

ES 2 525 545 T3

	tggacaaaca	cagaccagcg	aaccggcctc	ctcaggggtca	cgaaccacct	cagcgggcac	2880
5	agctaccct	tcctcatccg	gggcgagtgg	cacaacacct	tcaggaagcg	aaggaatatc	2940
	cacctcagga	gagacgaaa	ggttttcatc	aaaccctcc	agggacagtc	acacaacca	3000
	gtcaacaacc	gaattgctgt	ccgcctcagc	cagtcatggt	gccatcccag	taagcacagg	3060
10	aatggcgtct	tcgatcgtcc	ccggcacctt	tcatcccacc	ctctctgagg	cctccactgc	3120
	agggagaccg	acaggacagt	caagcccaac	ttctcccagt	gcctctcctc	aggagacagc	3180
	cgccatttcc	cggatggccc	agactcagag	gacaggaacc	agcagagggg	ctgacactat	3240
15	cagcctggcg	tcccaggcaa	ccgacacctt	ctcaacagtc	ccaccacac	ctccatcgat	3300
	cacatccagt	gggcttacat	ctccacaaac	ccagaccac	actctgtcac	cttcagggtc	3360
20	tggtaaaacc	ttcaccacgg	ccctcatcag	caacgccacc	cctcttctg	tcaccagcac	3420
	ctcctcagcc	tccacaggtc	acgccacccc	tcttgctgtc	agcagtgcta	cctcagcttc	3480
	cacagtatcc	tcggactccc	ctctgaagat	ggaacatca	ggaatgaaa	caccgtcact	3540
25	gaagacagac	ggtgggagac	gcacagccac	atcaccaccc	cccacaacct	cccagaccat	3600
	catttccacc	attcccagca	ctgccatgca	caccgctcc	acagctgcc	ccatccccat	3660
	cctgcctgag	agaggagttt	ccctcttccc	ctatggggca	ggcgcgggg	acctggagtt	3720
30	cgtcaggagg	accgtggact	tcacctccc	actcttcaag	ccggcgactg	gcttccccct	3780
	tggctcctct	ctccgtgatt	ccctctactt	cacagacaat	ggccagatca	tcttcccaga	3840
	gtcagactac	cagattttct	cctaccccaa	cccactccca	acaggcttca	caggccggga	3900
35	ccctgtggcc	ctgggtggctc	cgttctggga	cgatgctgac	ttctccactg	gtcgggggac	3960
	cacattttat	caggaatacg	agacgttcta	tggtgaacac	agcctgctag	tccagcaggc	4020
40	cgagtcttgg	attagaaa	tgacaaaaca	cgggggctac	aaggccaggt	gggccctaaa	4080
	ggtcacgtgg	gtcaatgccc	acgcctatcc	tgcccagtgg	accctcggga	gcaacaccta	4140
	ccaagccatc	ctctccacgg	acgggagcag	gtcctatgcc	ctgtttctct	accagagcgg	4200
45	tgggatgcag	tgggacgtgg	cccagcgtc	aggcaaccg	gtgctcatgg	gcttctctag	4260
	tggagatggc	tatttcgaaa	acagcccact	gatgtcccag	ccagtgtggg	agaggtatcg	4320
	ccctgataga	ttcctgaatt	ccaactcagg	cctccaagg	ctgcagttct	acaggctaca	4380
50	ccgggaagaa	aggcccaact	accgtctcga	gtgcctgcag	tggtgaaga	gccagcctcg	4440
	gtggcccagc	tggggctgga	accaggtctc	ctgcccttgt	tcctggcagc	agggacgacg	4500
55	ggacttacga	ttccaaccg	tcagcatagg	tcgctggggc	ctcggcagta	ggcagctgtg	4560
	cagcttcacc	tcttggcgag	gaggcgtgtg	ctgcagctac	gggccctggg	gagagtttcg	4620
	tgaaggctgg	cacgtgcagc	gtccttggca	gttggcccag	gaactggagc	cacagagctg	4680
60	gtgctgccgc	tggaatgaca	agccctacct	ctgtgccctg	taccagcaga	ggcggcccca	4740
	cgtgggctgt	gctacataca	ggccccaca	gcccgcctgg	atgttcgggg	acccccacat	4800
	caccaccttg	gatggtgtca	gttacacctt	caatgggctg	ggggacttcc	tgctggctcg	4860
65	ggcccaagac	gggaactcct	ccttctgct	tcagggccgc	accgccca	ctggctcagc	4920

ES 2 525 545 T3

ccagggcacc aacttcatcg cctttgcggc tcagtaccgc tccagcagcc tgggccccgt 4980  
 cacggtccaa tggctccttg agcctcacga cgcaatccgt gtcctgctgg ataaccagac 5040  
 5 tgtgacattt cagcctgacc atgaagacgg cggaggccag gagacgttca acgccaccgg 5100  
 agtcctcctg agccgcaacg gctctgaggt ctcggccagc ttcgacggct gggccaccgt 5160  
 10 ctcggtgatc gcgctctcca acatcctcca cgctccgcc agcctccgc ccgagtacca 5220  
 gaaccgcacg gaggggctcc tgggggtctg gaataacaat ccagaggacg acttcaggat 5280  
 gcccaatggc tccaccattc ccccaggag ccctgaggag atgcttttcc actttggaat 5340  
 15 gacctggcag atcaacggga caggcctcct tggcaagagg aatgaccagc tgccttccaa 5400  
 cttcaccctt gtttttact cacaactgca aaaaaacagc tcctgggctg aacatttgat 5460  
 ctccaactgt gacggagata gctcatgcat ctatgacacc ctggccctgc gcaacgcaag 5520  
 20 catcggactt cacacgaggg aagtcagtaa aaactacgag caggcgaacg ccaccctcaa 5580  
 tcagtaccgg ccctccatca atggtggtcg tgtgattgaa gcctacaagg ggcagaccac 5640  
 gctgattcag tacaccagca atgctgagga tgccaacttc acgctcagag acagctgcac 5700  
 25 cgacttgag ctctttgaga atgggacgtt gctgtggaca cccaagtcgc tggagccatt 5760  
 cactctggag attctagcaa gaagtgcaa gattggcttg gcatctgcac tccagcccag 5820  
 30 gactgtggtc tgccattgca atgcagagag ccagtgtttg tacaatcaga ccagcaggg 5880  
 gggcaactcc tccctggagg tggctggctg caagtgtgac gggggcacct tcggccgcta 5940  
 ctgcgagggc tccgaggatg cctgtgagga gccgtgcttc ccgagtgtcc actgcttcc 6000  
 35 tgggaagggc tgcgaggcct gccctccaaa cctgactggg gatgggcggc actgtgcggc 6060  
 tctggggagc tctttcctgt gtcagaacca gtcctgcct gtgaattact gctacaatca 6120  
 aggccactgc tacatctccc agactctggg ctgtcagccc atgtgcacct gccccccagc 6180  
 40 cttactgac agccgctgct tcctggctgg gaacaacttc agtccaactg tcaacctaga 6240  
 acttccctta agagtcattc agctcttgct cagtgaagag gaaaatgcct ccatggcaga 6300  
 45 ggtcaacgcc tcggtggcat acagactggg gaccctggac atgcgggcct ttctccgcaa 6360  
 cagccaagtg gaacgaatcg attctgcagc accggcctcg ggaagcccca tccaactg 6420  
 gatggtcatc tcggagtcc agtaccgcc tcggggcccg gtcattgact tcctgaacaa 6480  
 50 ccagctgctg gccgcggtgg tggaggcgtt cttataccac gttccacgga ggagtgagga 6540  
 gcccaggaac gacgtggtct tccagcccat ctccggggaa gacgtgcgag atgtgacagc 6600  
 cctgaacgtg agcacgctga aggcttactt cagatgcgat ggctacaagg gctacgacct 6660  
 55 ggtctacagc ccccagagcg gcttcacctg cgtgtccccg tgcagtaggg gctactgtga 6720  
 ccatggaggc cagtgccagc acctgcccag tgggccccgc tgcagctgtg tgtccttctc 6780  
 60 catctacagc gcctggggcg agcactgtga gcacctgagc atgaaactcg acgcgttctt 6840  
 cggcatcttc tttggggccc tgggcggcct cttgctgctg ggggtcggga cgttcgtggt 6900  
 cctgcgcttc tggggttgct ccggggccag gttctcctat ttctgaact cagctgaggc 6960  
 65

ES 2 525 545 T3

	cttgccttga aggggcagct gtggcctagg ctacctcaag actcacctca tccttaccgc	7020
	acatttaagg cgccattgct tttgggagac tggaaaaggg aagggtgactg aaggctgtca	7080
5	ggattcttca aggagaatga atactgggaa tcaagacaag actatacctt atccataggc	7140
	gcaggtgcac agggggaggc cataaagatc aaacatgcat ggatgggtcc tcacgcagac	7200
10	acaccacag aaggacacta gcctgtgcac gcgcgcgtgc acacacacac acacacacac	7260
	gagttcataa tgtggtgatg gccctaagtt aagcaaatg cttctgcaca caaaactctc	7320
	tggtttactt caaattaact ctatttaaataa aaagtctctc tgactttttg tgtctccaaa	7380
15	a	7381
	<210> 103	
	<211> 2323	
20	<212> ADN	
	<213> Homo sapiens	
	<400> 103	
25		
30		
35		
40		
45		
50		
55		
60		
65		

ES 2 525 545 T3

agctatgata gcaacacctt ggtggccatc gtggtgggtg tggggcgccct catcactggc 60  
 atggaccgag gcctcatggg catgtgtgtc aacgagcggc gacgcctcat tgtgcctccc 120  
 5 cacctgggct atgggagcat cggcctggcg gggctcattc caccggatgc caccctctac 180  
 ttgatgtgg ttctgtgga tgtgtggaac aaggaagaca cctgtcaggt gagcacattg 240  
 ctgccccgc cccactgccc ccgcatggtc caggacggcg actttgtccg ctaccactac 300  
 10 aatggcacc tgctggacgg cacctccttc gacaccagct acagtaaggc cggcacttat 360  
 gacacctacg tcggctctgg ttggctgata aagggcatgg accaggggct gctgggcatg 420  
 15 tgtcctggag agagaaggaa gattatcatc cctccattcc tggcctatgg cgagaaaggc 480  
 tatgggacag tgatcccccc acaggcctcg ctggtctttc acgtcctcct gattgacgtg 540  
 cacaacccga aggacgctgt ccagctagag acgctggagc tccccccgg ctgtgtccgc 600  
 20 agagccgggg cgggggactt catgagctac cactacaatg gctccttgat ggacggcacc 660  
 ctcttcgatt ccagctactc ccgcaaccac acctacaata cctatatcgg gcagggttac 720  
 atcatccccg ggatggacca ggggctgcag ggtgcctgca tgggggaacg ccggagaatt 780  
 25 accatcccc cgcacctcgc ctatggggag aatggaactg gagacaagat ccctggctct 840  
 gccgtgctaa tcttcaacgt ccatgtcatt gacttcaca accctgcgga tgtggtggaa 900  
 30 atcaggacac tgtcccggcc atccgagacc tgcaatgaga ccaccaagct tggggacttt 960  
 gttcgatacc attacaactg ttctttgetg gacggcacc agctgttcac ctgcatgac 1020  
 tacggggccc cccaggaggc gactctcggg gccaacaagg tgatcgaagg cctggacacg 1080  
 35 ggcctgcagg gcatgtgtgt gggagagagg cggcagctca tcgtgcccc gcacctggcc 1140  
 cacggggaga gtggagcccg gggagtccca ggcagtgtg tgctgctggt tgaggtggag 1200  
 ctggtgtccc gggaggatgg gctgcccaca ggctacctgt ttgtgtggca caaggaccct 1260  
 40 cctgccaacc tgtttgaaga catggacctc aacaaggatg gcgaggctcc tccggaggag 1320  
 ttctccacct tcatcaaggc tcaagtgagt gagggcaaag gacgcctcat gcctgggcag 1380  
 45  
 50  
 55  
 60  
 65



ES 2 525 545 T3

gaccctgaga aaaccatagg agacatgttc cagaaccagg accgcaacca ggacggcaag 1440  
 atcacagtcg acgagctcaa gctgaagtca gatgaggacg aggagcgggt ccacgaggag 1500  
 5 ctctgagggg cagggagcct ggccaggcct gagacacaga gggccactgc gagggggaca 1560  
 gtggcgggtg gactgacctg ctgacagtca ccctccctct gctgggatga ggtccaggag 1620  
 ccaactaaaa caatggcaga ggagacatct ctgggtgttcc caccacccta gatgaaaatc 1680  
 10 cacagcacag acctctaccg tgtttctctt ccatccctaa accacttctt taaaatgttt 1740  
 ggatttgcaa agccaatttg gggcctgtgg agcctgggggt tggatagggc catggctggt 1800  
 15 cccccacat acctcccctc cacatcactg acacagctga gcttgttatc catctcccca 1860  
 aactttctct ttctttgtac ttcttgcctc cccactccc agcccctatt cctctatgtg 1920  
 acagctggct aggaccctc tgccttcctt cccaatcctg actggctcct agggaagggg 1980  
 20 aaggctcctg gagggcagcc ctacctctcc catgcccttt gccctcctcc ctgcctcca 2040  
 gtggaggctg agctgacctt gggctgctgg aggcagact gggctgtagt tagcttttca 2100  
 tccctaaaga aggctttccc taaggaacca tagaagagag gaagaaaaca aagggcatgt 2160  
 25 gtgaggggag ctgcttgggt ggggtgtagg gctatgaaat cttggatttg gggctgaggg 2220  
 gtgggagggg gggcagagct ctgcacactc aaaggctaaa ctgggtgtcag tccttttttc 2280  
 30 ctttgttcca aataaaagat taaacaaaa aaaaaaaaaaaa aaa 2323  
  
 <210> 104  
 <211> 741  
 <212> ADN  
 35 <213> Homo sapiens  
  
 <400> 104  
  
 40 tcacgtgacc cgggcgcgct gcggccgccc gcgcggaccc ggcgagaggc ggcggcggga 60  
 gcggcgggtga tggacgggtc cggggagcag cccagaggcg gggggcccac cagctctgag 120  
 cagatcatga agacaggggc ctttttgctt caggggatga ttgccgccgt ggacacagac 180  
 45 tcccccgag aggtcttttt ccgagtggca gctgacatgt tttctgacgg caacttcaac 240  
 tggggccggg ttgtcgccct tttctacttt gccagcaaac tggtgctcaa ggccctgtgc 300  
 accaaggtgc cggaactgat cagaaccatc atgggctgga cattggactt cctccgggag 360  
 cggctgttgg gctggatcca agaccagggt ggttgggacg gcctcctctc ctactttggg 420  
 acgcccacgt ggcagaccgt gaccatcttt gtggcgggag tgctcaccgc ctgcctcacc 480  
 55 atctggaaga agatgggctg agggccccag ctgccttggg ctgtgttttt cctccataaa 540  
 ttatggcatt tttctgggag ggggtgggat tgggggacat gggcattttt ctacttttg 600  
 taattattgg ggggtgtggg gaagagtggg cttgaggggg taataaacct ccttcgggac 660  
 60 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 720  
 aaaaaaaaaa aaaaaaaaaa a 741

65

<210> 105  
<211> 2373  
<212> ADN  
<213> Homo sapiens  
  
<400> 105

5

10

15

20

25

30

35

40

45

50

55

60

65

ES 2 525 545 T3

5      cccaggccca ccccacccag caccctggc gcagggactg ctggaacctg gctgtgcgcg      60  
        ctgtcgcttt aagacagact ctgccggcgc cgtccggagc ctagaaacc ggccccggat      120  
        cgcgagccgg agccggagcc ggagccgggg ccggccgggc tgctgaggcc cgagcggcag      180  
        gagcgcagcg cggagcgctg agccaggcgc ccagtcgcga gaagctgccg ccgcctctgc      240  
 10      ccgcccggcg ccgcagcccc gggcggtcca tggggcgggc acggcgtcgc tgcaggcgcc      300  
        ggcagccctg gagggcagcc gcttaggcgc tgcgctcttg tccccgagg tcgcagccag      360  
        gggggcgggg cgcgcccagc cccggcccct ggagcgcccg ccgcggtccc cacctccatg      420  
 15      gacgccttca aggggggcat gagcctggag cggctgccgg aggggctccg gccgccgccg      480  
        ccgccacccc atgacatggg gccgccttc cacctggccc ggcccgccga cccccgcgag      540  
        ccgctcgaga actccgccag cgagtcgtct gacacggagc tgccagagaa ggagcgcggc      600  
 20      ggggaaccca aggggcccga ggacagtggg gcgggaggca cgggctgcgg cggcgcagac      660  
        gaccagcca agaagaagaa gcagcggcgg caacgtacgc acttcacaag ccagcagttg      720  
        caagagctag aggccacggt ccagaggaac cgctaccccg acatgagcat gagggaggag      780  
 25      atcgccgtgt ggaccaacct caccgagccg cgcgtgcggg tctggttcaa gaaccggcga      840  
        gccaaagtggc gtaagcgcga gcgtaaccag cagctggacc tgtgcaaggg tggctacgtg      900  
 30      ccgcaattca gcggcctagt gcagccctac gaggacgtgt acgccgccgg ctactcctac      960  
        aacaactggg ccgccaagag cctggcgcca gcgccgtct ccaccaagag cttcaccttc      1020  
        ttcaactcca tgagcccgtg gtcgtcgcag tccatgttct cagcaccag ctccatctcc      1080  
 35      tccatgacca tgccgtccag catgggcccc ggcgccgtgc ctggcatgcc caactcgggc      1140  
        ctcaacaaca tcaacaacct caccggctcc tcgctcaact cggccatgtc gccgggcgct      1200  
        tgcccgtacg gactccccgc ctgcacctac agcgtctacc gggacacgtg caactcgagc      1260  
 40      ctagccagcc tgccggtcaa gtccaaacag cactcgtcgt ttggctacgg cgccctgcag      1320  
        ggcccggcct cgggcctcaa cgcgtgccag tacaacagct gaccgccccg ccgcaccacg      1380  
 45      cgggccggcg gccggagcgg ggaagggcgc gggcgcggag gacgcacgcg gggccccggc      1440  
        tcgcaagccc cagctcaccg cgccgcggac ctcacacctg cgcagcccc tctcccact      1500  
        tcccactccg ggttggtttt gtgtttgctt ttccggacct cactctgccc tccaaaaaga      1560  
 50      caaaaaaaaa aaaaaaaaaa aaagcaaaaa gacgtcggag aaaagtgccg cgaaaaaatg      1620  
        gatgagttgc aatttctctc gggatggcgc ggggtggtgtg tgtgtgttcc cacgggcccc      1680  
 55      ggaggccac tccgcggagg gcacgcggcg cggtaggcga gcgccgaggc ccagcggccg      1740  
        ggggaggacg acctcgtatc ccgcgtcccc gccgcgctgg atccggactg agtggccggg      1800  
        cctgcggact ggatgtgcgg ggcctggact tgcctaggat tccccgacct cgtacaaacc      1860  
 60      aagttgccct ctccgagcta ggcccggccg agagcgcctt agctcgagtc ggatccgtgt      1920  
        tggggcgggc gttgggtttg gggggacggt gccccagacc caggatcggg cactcagtgg      1980

65

ES 2 525 545 T3

agccgcacac ggccccggcg cgcctggtag agcctcgtcg gccccgcgcc ccggagccct 2040  
 atattaaggc cacggagcga cagcgggcag tgcgggcctg gcgggaggtg ggggaggtcc 2100  
 5 atctcagaac accccagcct tgagcttagc tgcaggccca ggccctctgc tctgctcccg 2160  
 ggctaggagg tggccctctg tctgggcgaa cagccccctc ctcaaccgcc gccgtgcaag 2220  
 agtcgagccg gcagagcaag gggcgcggcc ccagggccct gcgcccactt tgcacacccg 2280  
 10 ctctccggcc cgcgcccctg tttacagcgt ccctgtgtat gttggactga ctgtaataaa 2340  
 tctgtctata tcgactaaaa aaaaaaaaaa aaa 2373  
 15 <210> 106  
 <211> 1314  
 <212> ADN  
 <213> Homo sapiens  
 20 <400> 106  
 aattcccggc tcggggacct ccacgcaccg cggctagcgc cgacaaccag ctagcgtgca 60  
 aggcgccgcy gctcagcgcg taccggcggg cttcgaaacc gcagtcctcc ggcgaccccg 120  
 25 aactccgctc cggagcctca gccccctgga aagtgatccc ggcacccgag agccaagatg 180  
 ccggcccact tgctgcagga cgatatctct agctcctata ccaccaccac caccattaca 240  
 30 gcgcctccct ccagggctct gcagaatgga ggagataagt tggagacgat gccccctctac 300  
 ttggaagacg acattcggcc tgatataaaa gatgatatat atgacccac ctacaaggat 360  
 aaggaaggcc caagcccaaa ggttgaatat gtctggagaa acatcatcct tatgtctctg 420  
 35 ctacacttgg gagccctgta tgggatcact ttgattccta cctgcaagtt ctacacctgg 480  
 ctttgggggg tattctacta ttttgtcagt gccctgggca taacagcagg agctcatcgt 540  
 ctgtggagcc accgctctta caaagctcgg ctgcccctac ggctctttct gatcattgcc 600  
 40 aacacaatgg cattccagaa tgatgtctat gaatgggctc gtgaccaccg tgcccaccac 660  
 aagttttcag aaacacatgc tgatcctcat aattcccgac gtggcttttt cttctctcac 720  
 45 gtgggttggc tgcttgtgcy caaacaccca gctgtcaaag agaaggggag tacgctagac 780  
 ttgtctgacc tagaagctga gaaactgggt atgttccaga ggaggtacta caaacctggc 840  
 ttgctgatga tgtgcttcat cctgcccacg cttgtgccct ggtatttctg gggtgaaact 900  
 50 tttcaaaaaca gtgtgttctg tgccactttc ttgcgatatg ctgtggtgct taatgccacc 960  
 tggctggtga acagtgtgcy ccacctcttc ggatatcgtc cttatgacaa gaacattagc 1020  
 ccccgggaga atatcctggt ttcacttggga gctgtgggtg agggcttcca caactaccac 1080  
 55 cactcctttc cctatgacta ctctgccagt gagtaccgct ggcacatcaa cttcaccaca 1140  
 ttcttcattg attgcatggc cgccctcggg ctggcctatg accggaagaa agtctccaag 1200  
 60 gccgccatct tggccaggat taaaagaacc ggagatggaa actaaaaaaaa aaaaaaaaaa 1260  
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaataaaa aaaaaaaaaa aaaa 1314

65

ES 2 525 545 T3

<210> 107  
 <211> 476  
 <212> ADN  
 <213> Homo sapiens

5

<400> 107

	gaaaaaccag ccactgcttt acaggacagg gggttgaagc tgagccccgc ctcacacca	60
10	ccccatgca ctcaaagatt ggattttaca gctacttgca attcaaaatt cagaagaata	120
	aaaaatggga acatacagaa ctctaaaaga tagacatcag aaattgttaa gttaagcttt	180
15	ttcaaaaaat cagcaattcc ccagcgtagt caagggtgga cactgcacgc tctggcatga	240
	tgggatggcg accgggcaag ctttcttctc cgagatgctc tgctgcttga gagctattgc	300
	tttgtttaaga tataaaaagg ggtttctttt tgtctttctg taaggtggac ttccagattt	360
20	tgattgaaag tcctagggtg attctatttc tgctgtgatt tatctgctga aagctcagct	420
	ggggttgtgc aagctagggga cccattcctg tgtaatacaa tgtctgcacc aatgct	476

25

30

35

40

45

50

55

60

65

**Reivindicaciones**

1. Un método de identificar el origen de una metástasis de origen desconocido que comprende los pasos de
  - 5 a. medir los Biomarcadores asociados con al menos seis carcinomas diferentes en una muestra que contiene células metastásicas en donde los Biomarcadores son niveles de expresión de al menos los genes Marcadores siguientes: SP-B, TTF1, DSG3, PSCA, F5, HPT1, PSA, MGB/MG, WT1 y PDEF;
  - b. combinar los datos de los Biomarcadores en un algoritmo donde el algoritmo
    - 10 i. normaliza los Biomarcadores frente a la referencia; y
    - ii. impone un punto de corte que optimiza la sensibilidad y especificidad de cada Biomarcador, pondera la prevalencia de los carcinomas y selecciona un tejido de origen;
  - 15 c. determinar el origen en base a la probabilidad más alta determinada por el algoritmo o determinar que el carcinoma no se deriva de un conjunto particular de carcinomas; y
  - d. medir opcionalmente los Biomarcadores específicos para uno o más carcinomas diferentes adicionales, y repetir los pasos c) y d) para los Biomarcadores adicionales.
2. El método de acuerdo con la reivindicación 1 en donde los genes Marcadores comprenden además
  - 20 (a) KRT6F, p73H, y/o SFTPC; y/o
  - (b) ITGB6, KLK10, CLDN18, TR10 y/o FKBP10; y/o
  - (c) CDX1 y/o FABP1
3. El método de una de las reivindicaciones 1 ó 2 en donde la expresión génica se mide usando al menos una de las SEQ ID NOs: 11-58.
4. El método de la reivindicación 3 en donde los genes Marcadores comprenden además genes Marcadores específicos del género seleccionados de al menos uno de
  - 30 i. en el caso de un paciente masculino KLK2, NGEP o NPY; o
  - ii. en el caso de un paciente femenino PIP, B305D, B726 o GABA-Pi; y/o PAX8, STAR o EMX2.
5. El método de la reivindicación 4 en donde, en el caso de un paciente masculino, los genes Marcadores comprenden además KLK2.
6. El método de la reivindicación 5 en donde, en el caso de un paciente masculino, los genes Marcadores comprenden además NGEP y/o NPY.
7. El método de la reivindicación 4 en donde, en el caso de un paciente femenino, los genes Marcadores comprenden además, PIP, B305D, B726 o GABA-Pi.
8. El método de la reivindicación 4 en donde, en el caso de un paciente femenino, los genes Marcadores comprenden además PAX8, STAR o EMX2.
9. El método de la reivindicación 8 en donde, en el caso de un paciente femenino, los genes Marcadores comprenden además PIP, B305D, B726 o GABA-Pi.
10. El método de acuerdo con una de las reivindicaciones 1-2, o 4-9 que comprende además medir la expresión de al menos un gen expresado consecutivamente en la muestra.
11. El método de la reivindicación 1 que comprende además obtener información clínica adicional incluyendo el sitio de metástasis para determinar el origen del carcinoma.
12. Un método de proporcionar un pronóstico o dirección de terapia determinando el origen de una metástasis de origen desconocido de acuerdo con la reivindicación 1 e identificar el pronóstico correspondiente para el mismo o el tratamiento apropiado para el mismo.
13. El uso de un kit para realizar el método de la reivindicación 1, en donde el kit comprende: ARN o ADNc que hibrida con los genes marcadores de la reivindicación 1, una micromatriz o un chip génico.

65

Figura 1

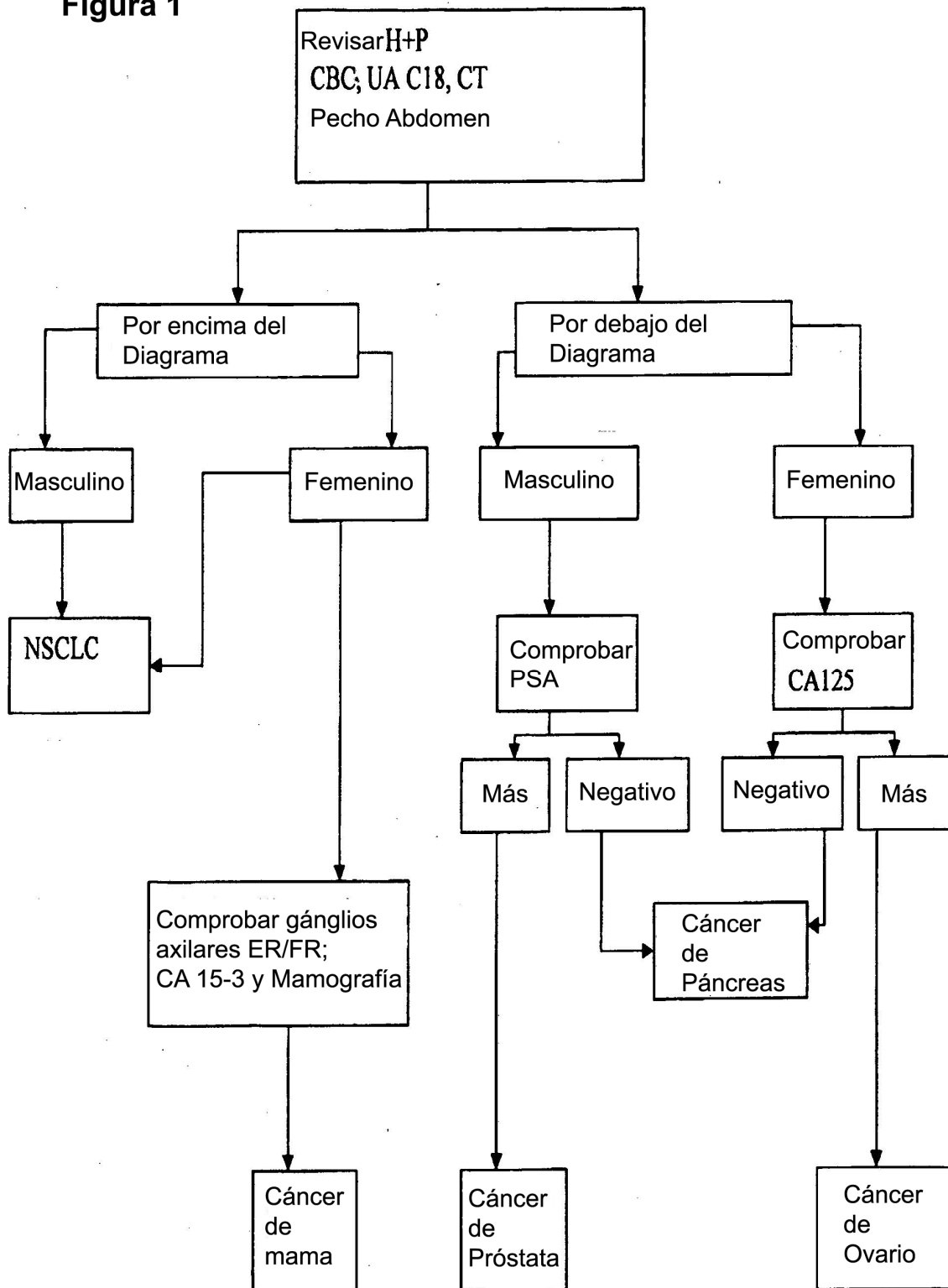
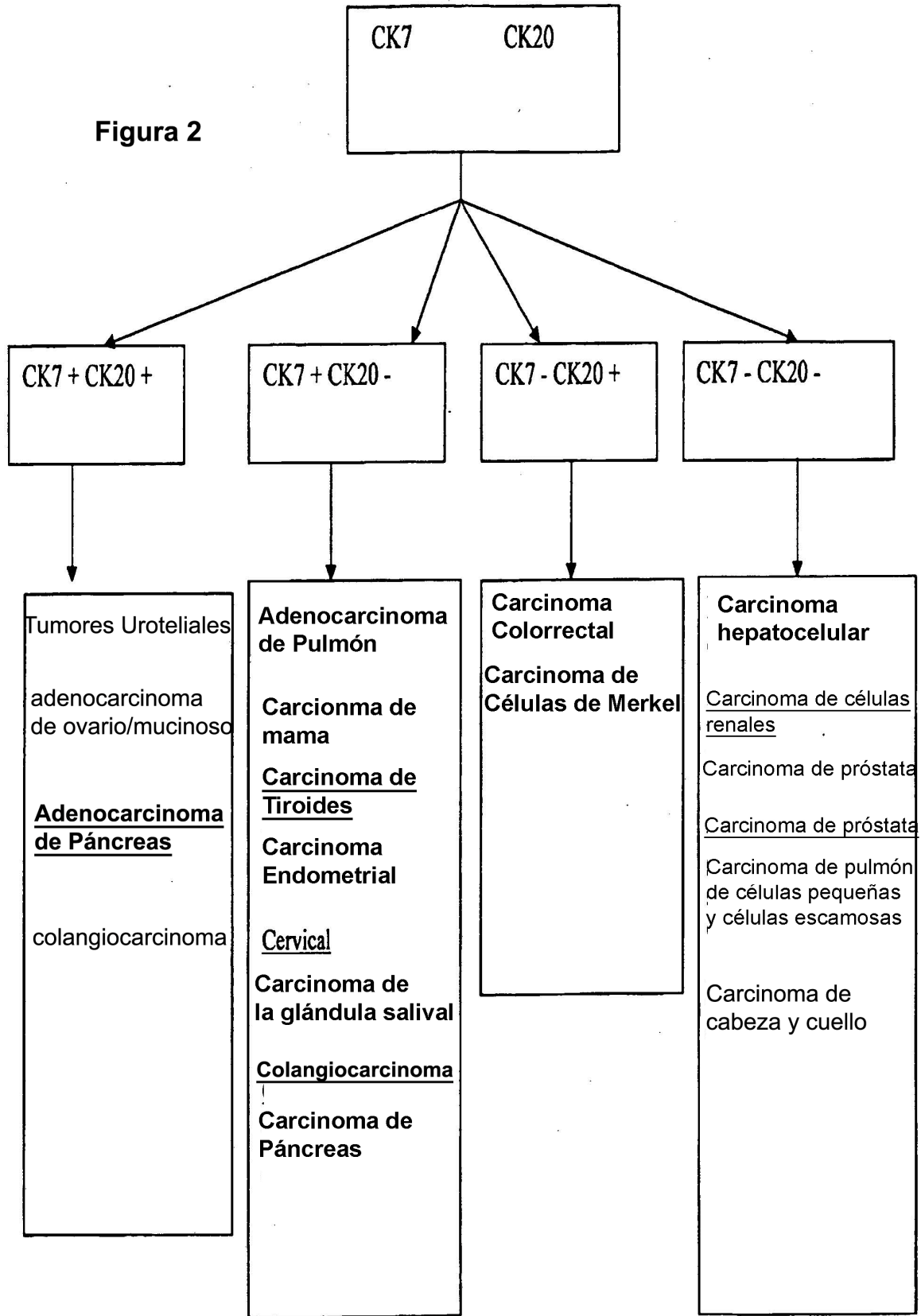


Figura 2





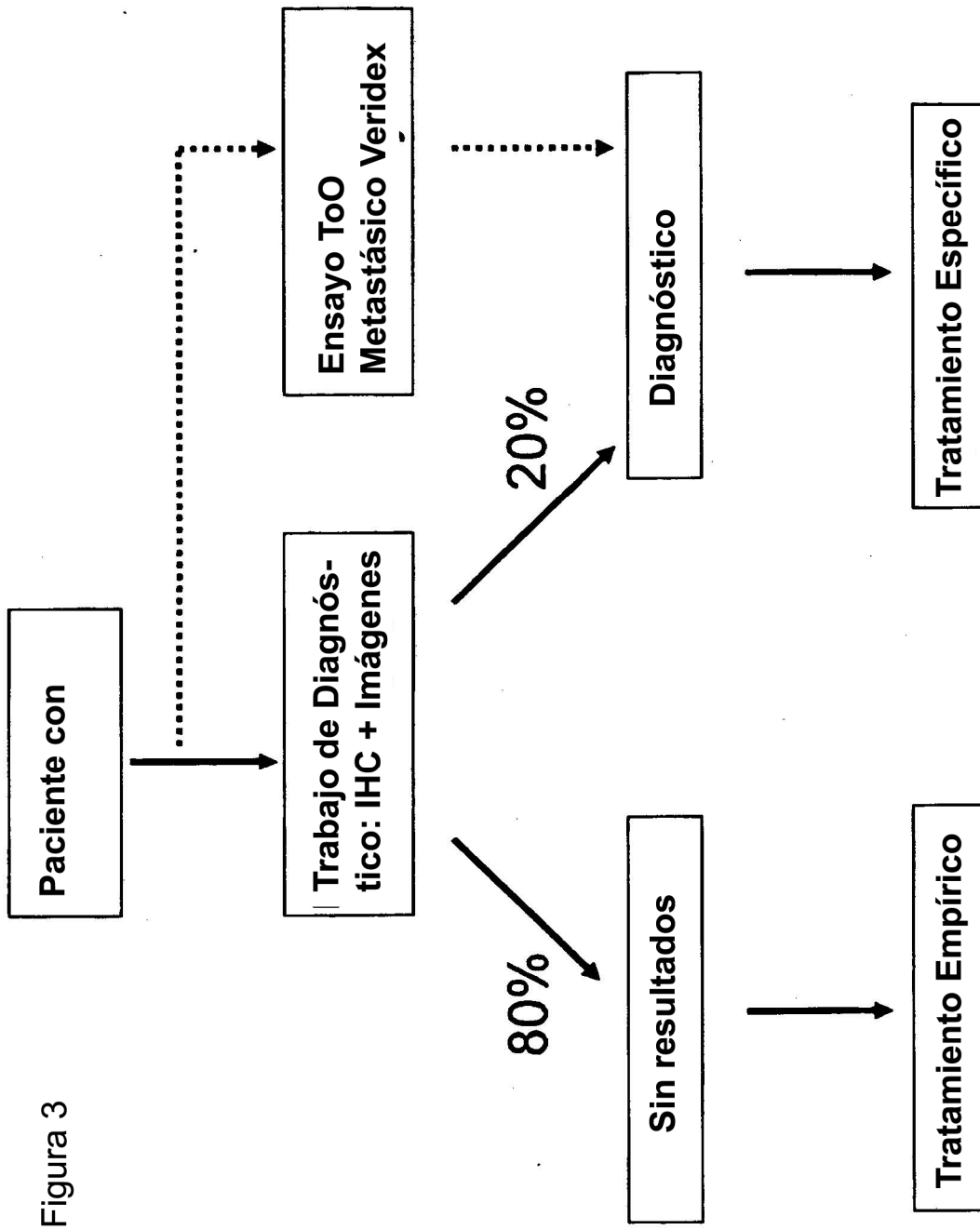


Figura 3

Figura 4(A)

205319\_at PSCA

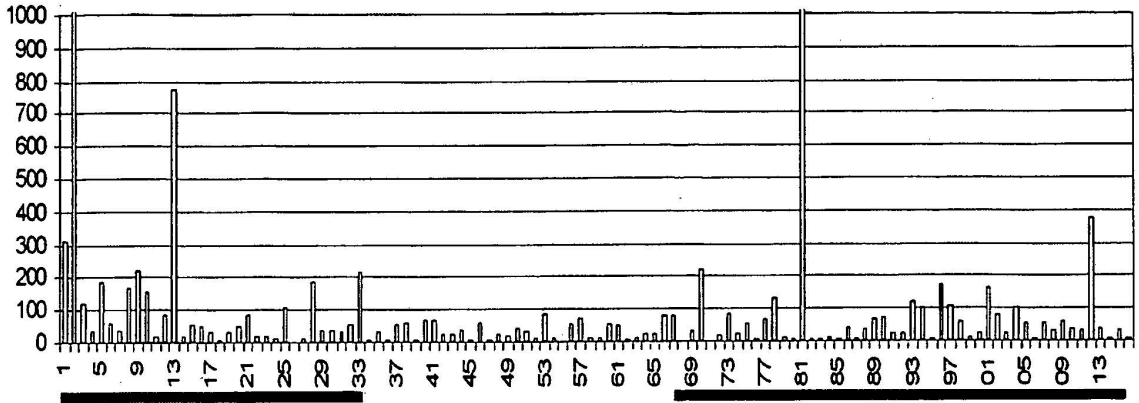


Figura 4(B)

204713\_s\_at F5

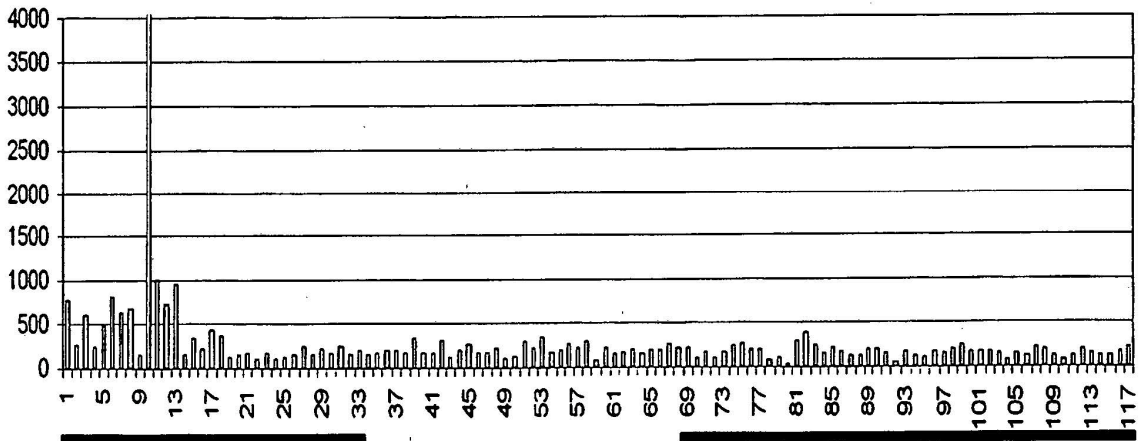


Figura 5(A)

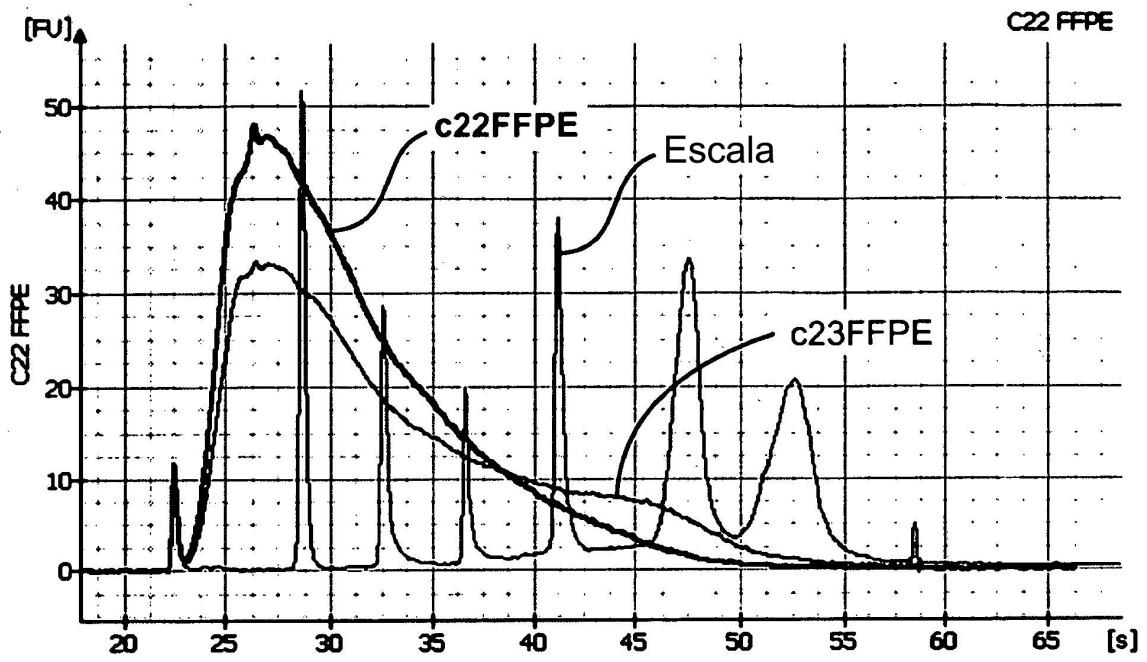


Figura 5(B)

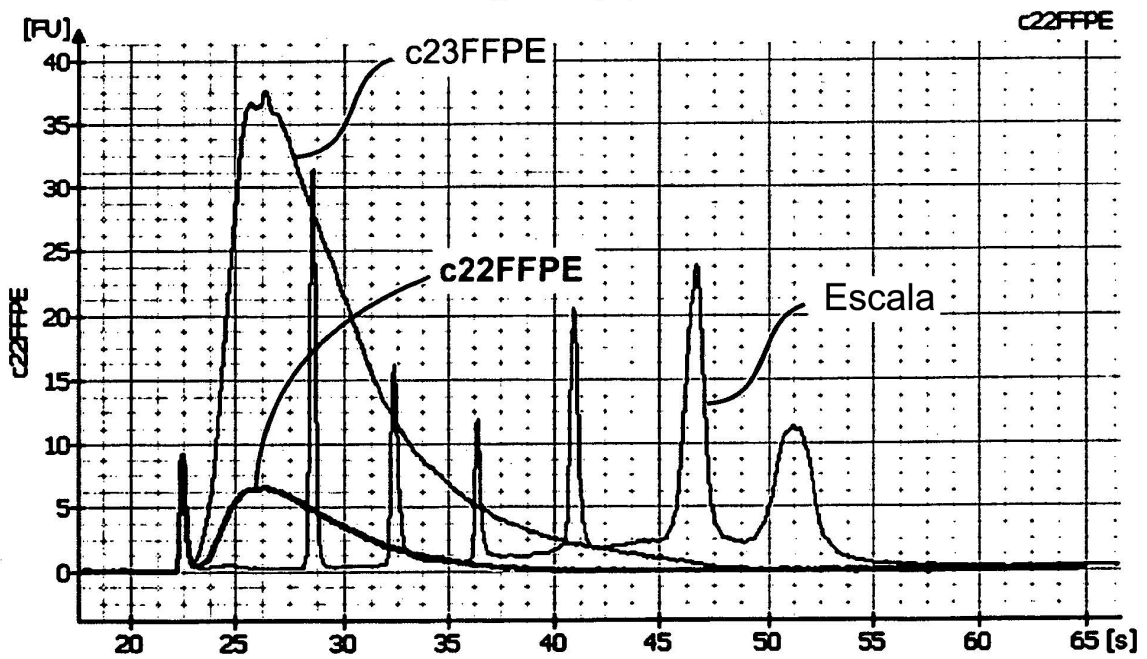


Figure 6(A)

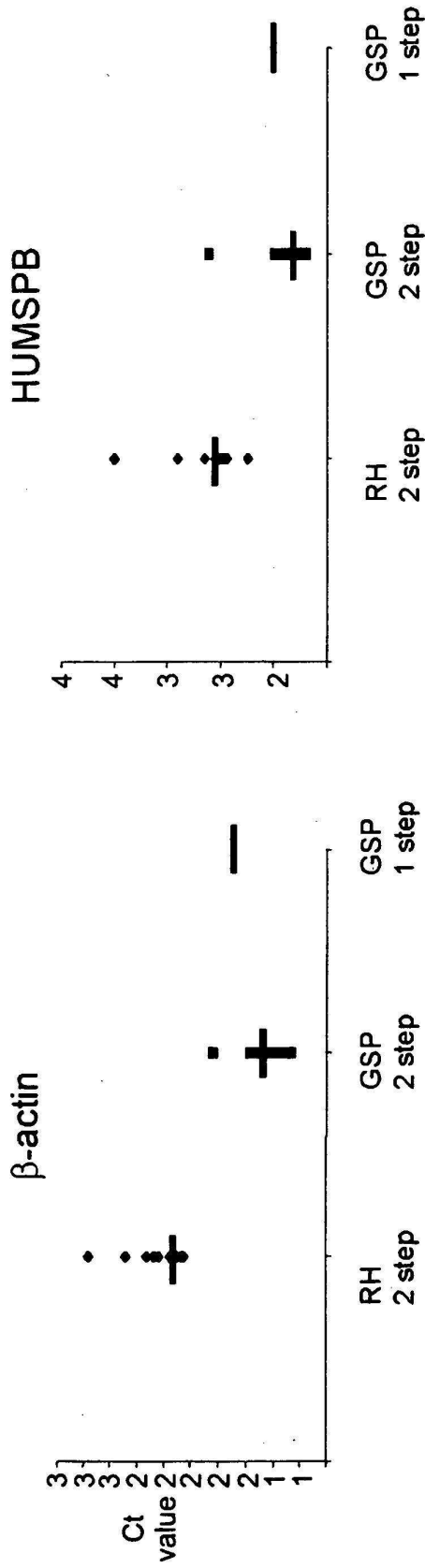
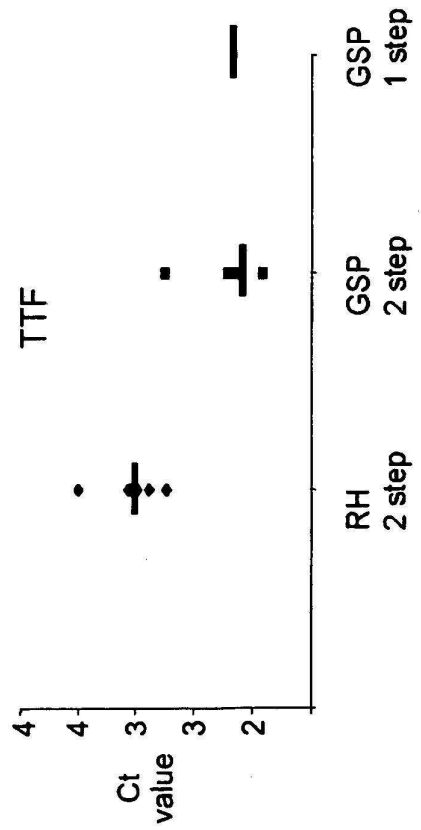


Figure 6(B)

Figure 6(C)





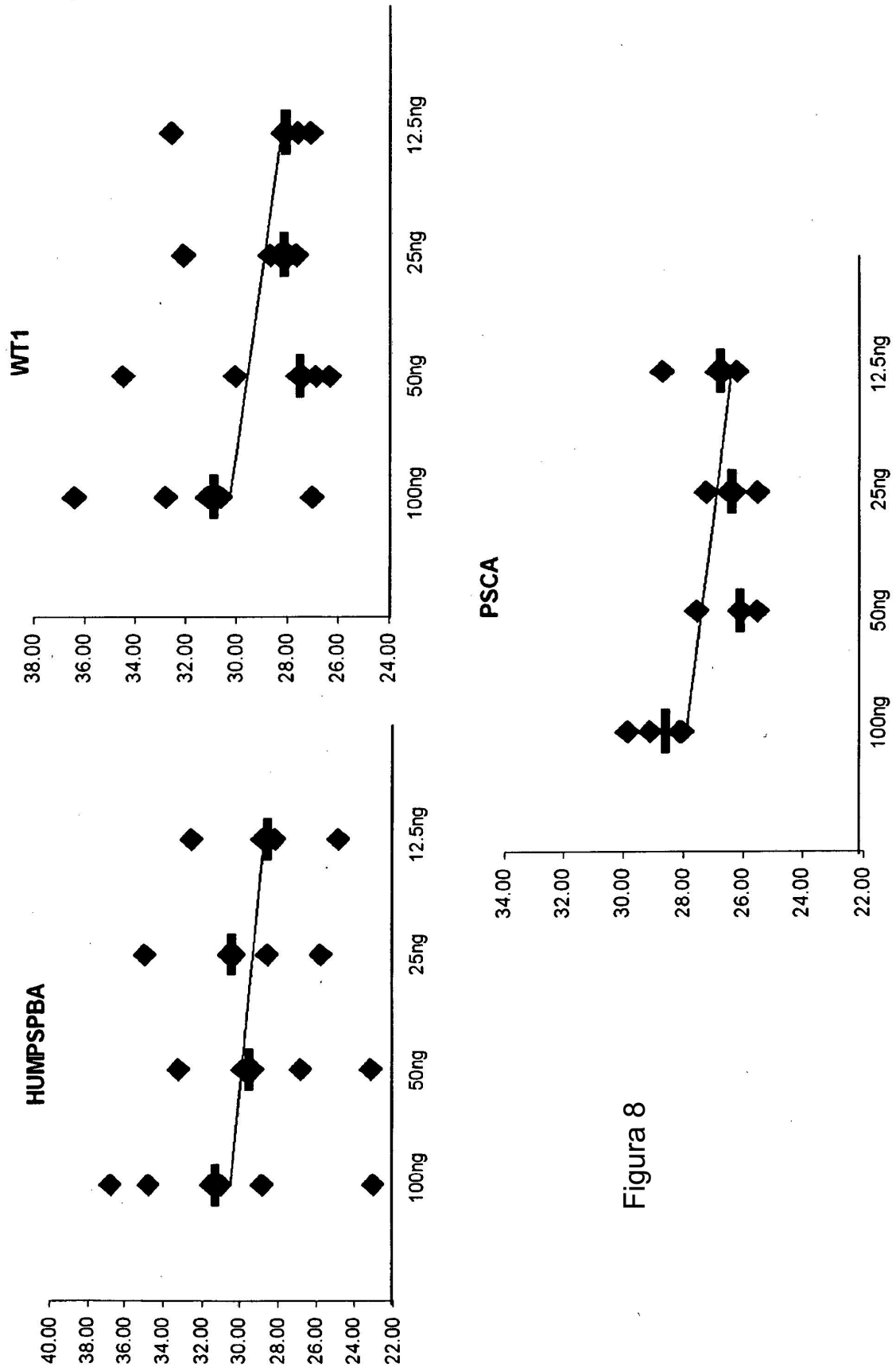


Figura 8

Figura 9A

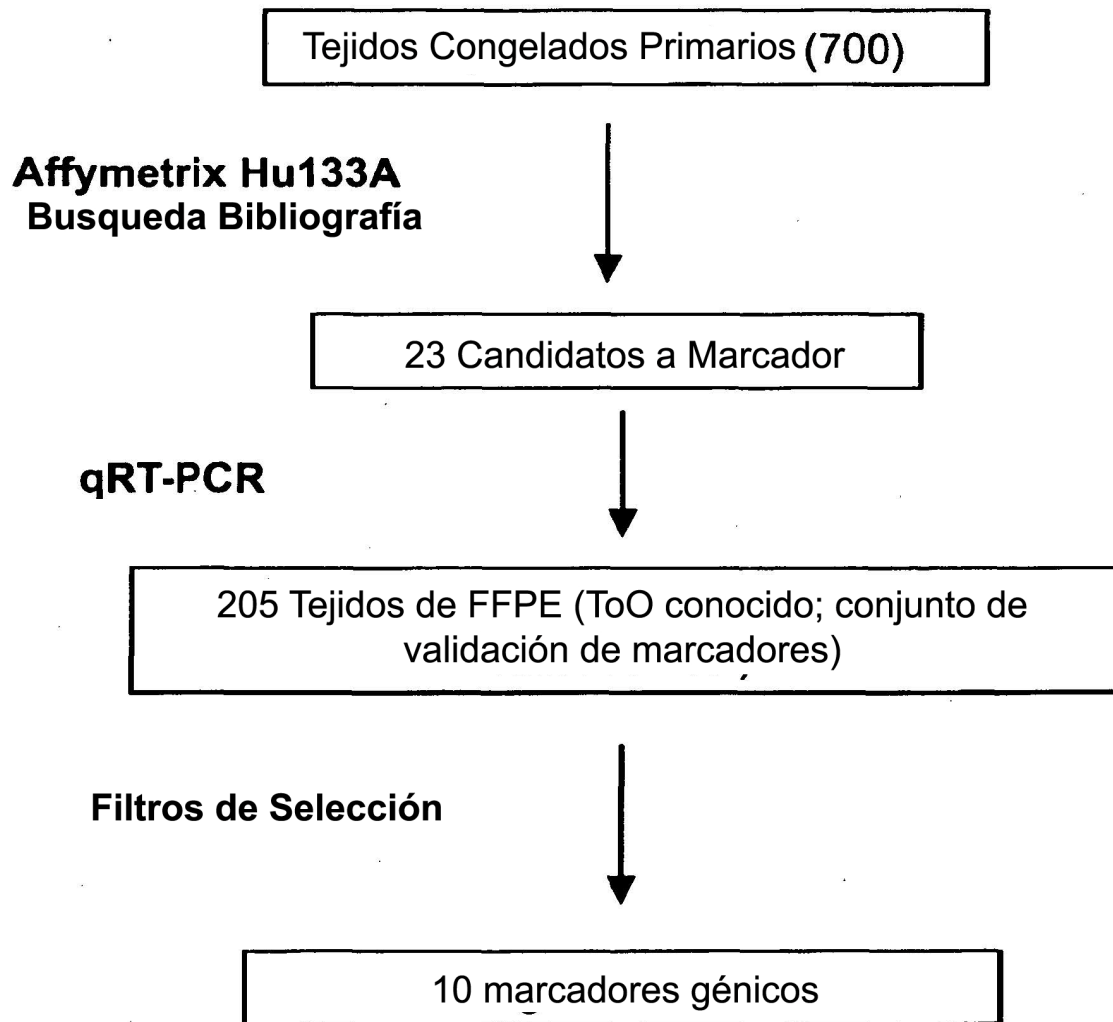


Figura 9B

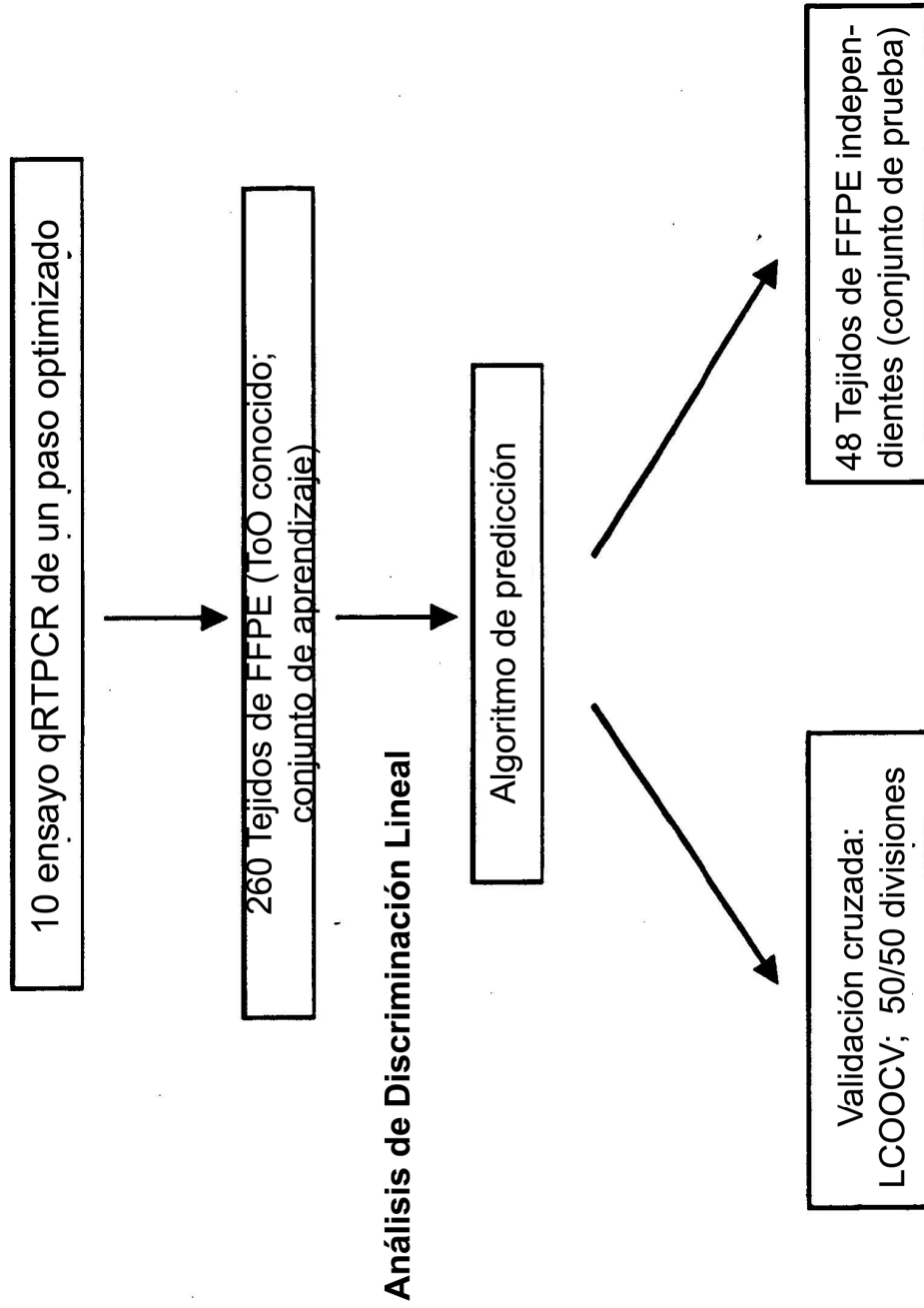




Figura 10

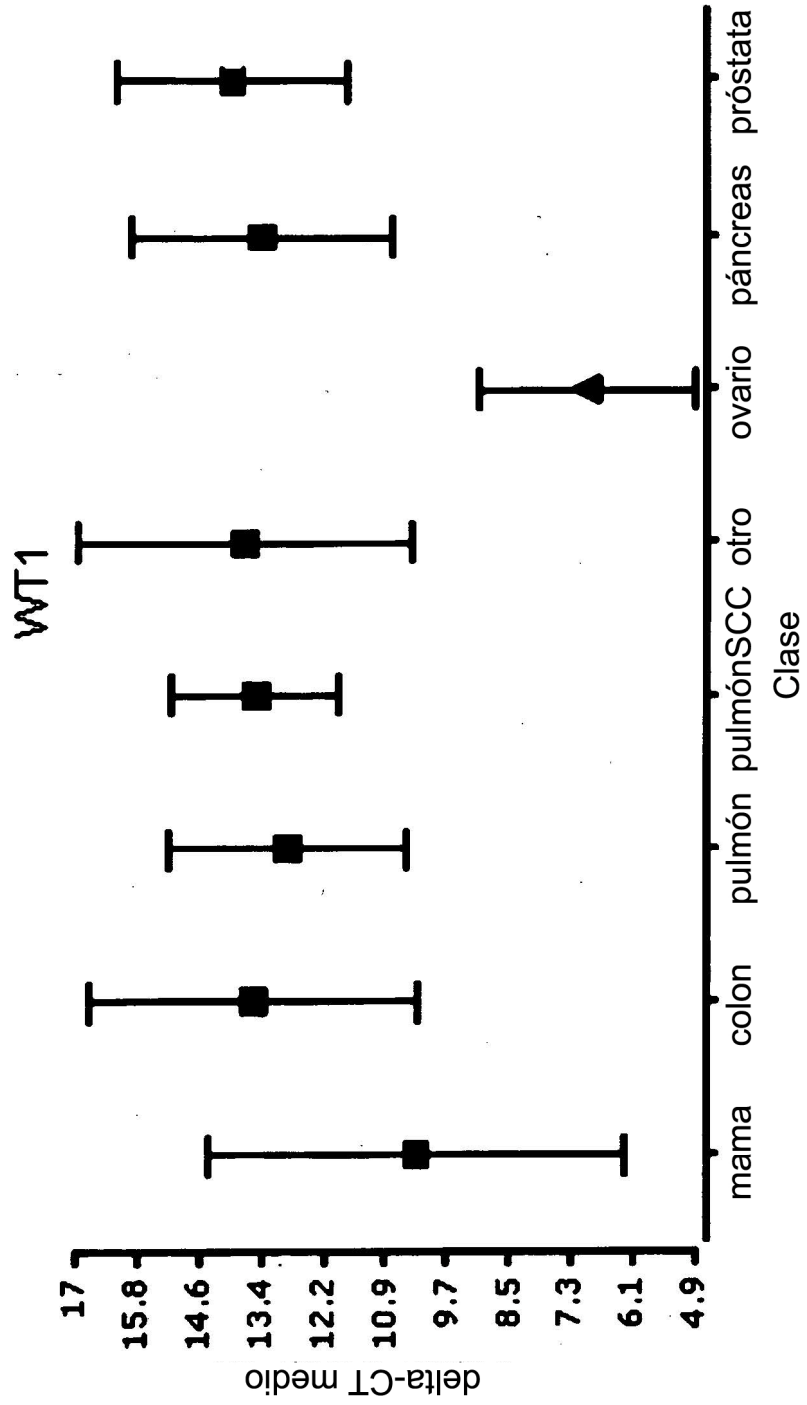




Figura 10 (cont)

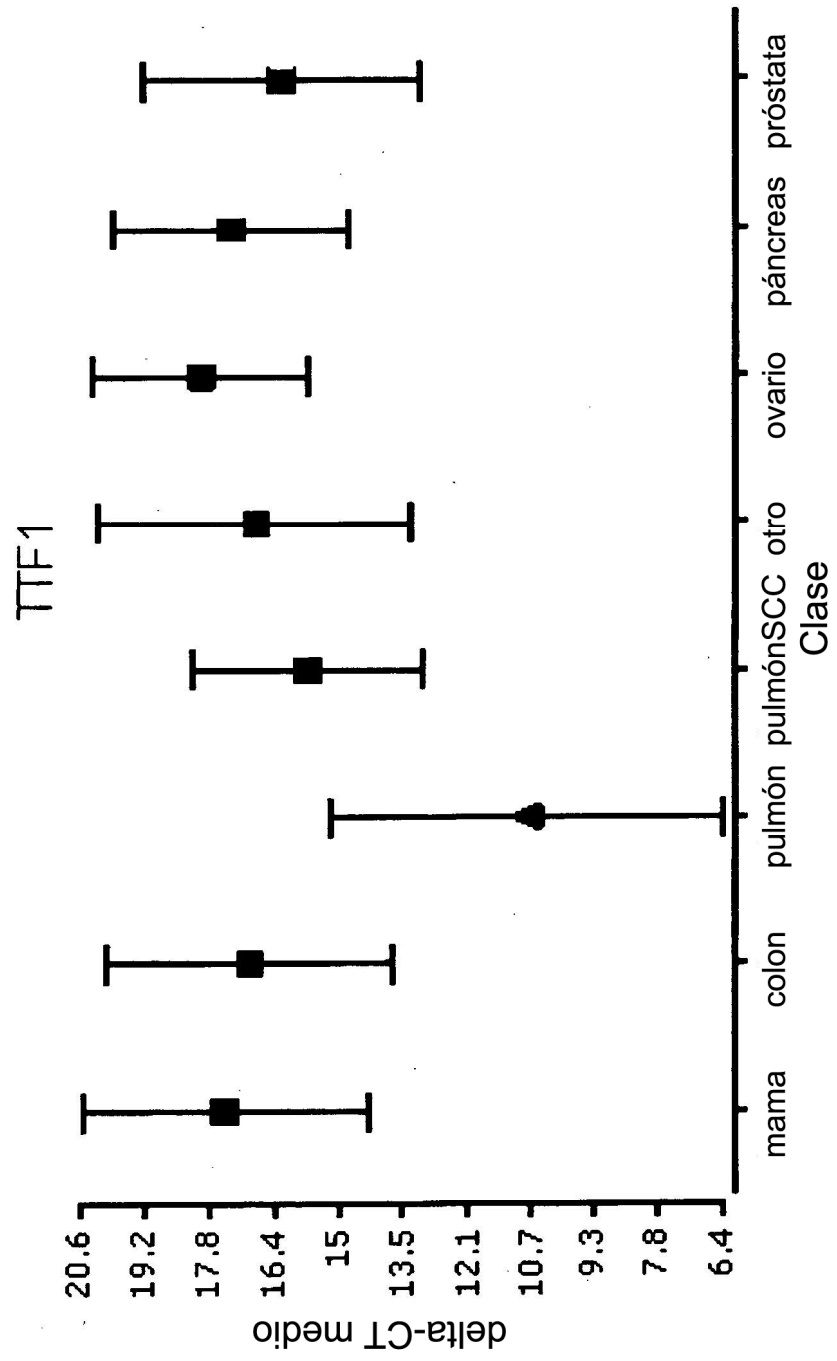


Figura 10 (cont.)

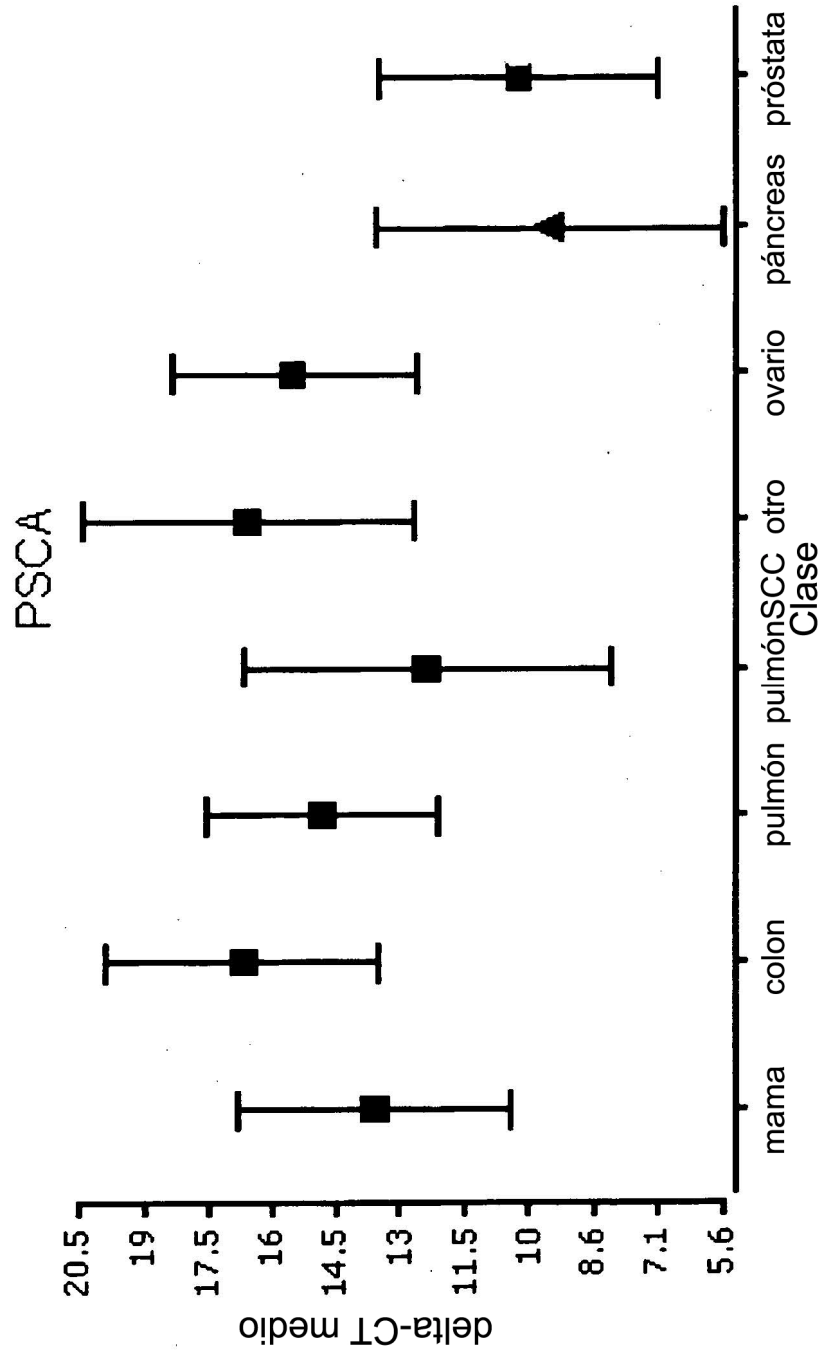


Figura 10 (cont.)

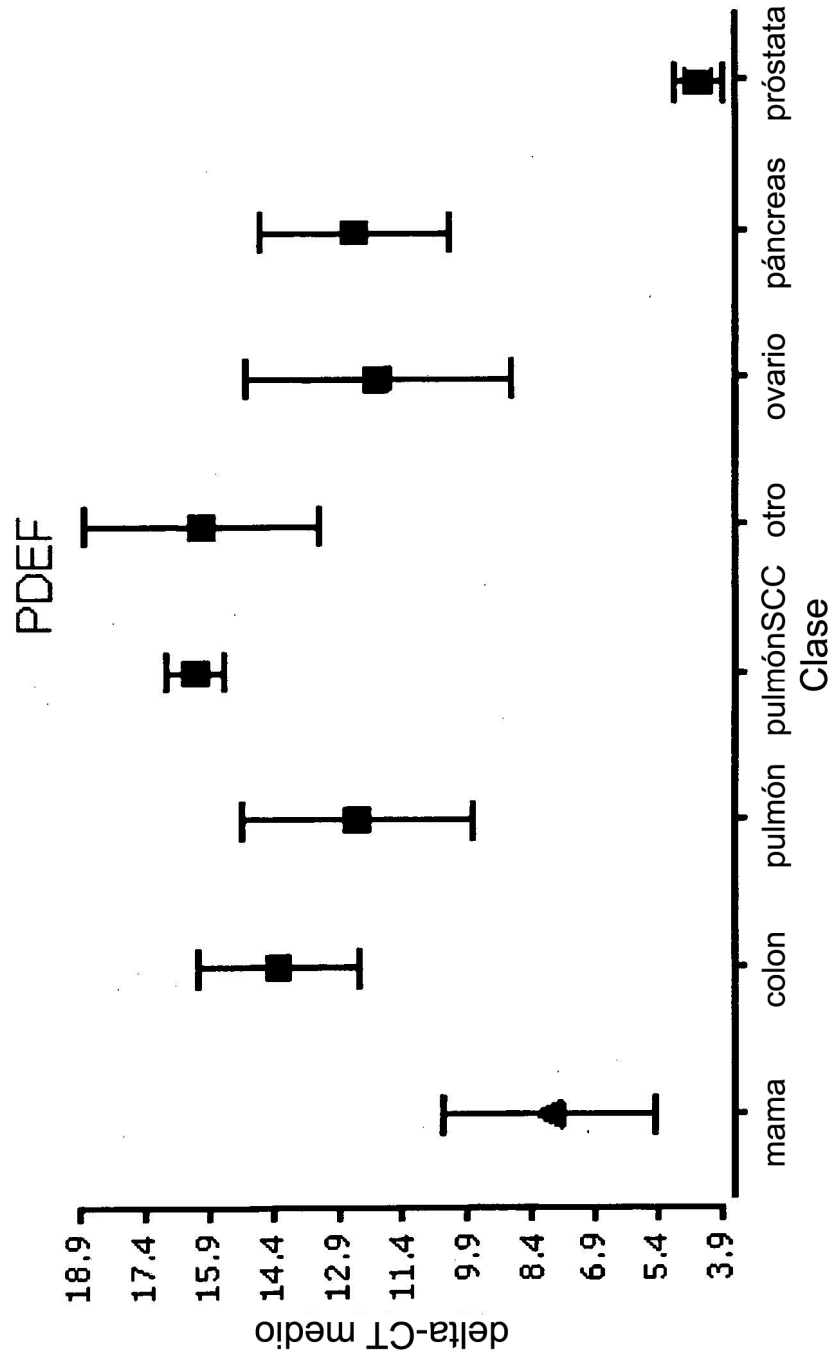


Figura 10 (cont.)

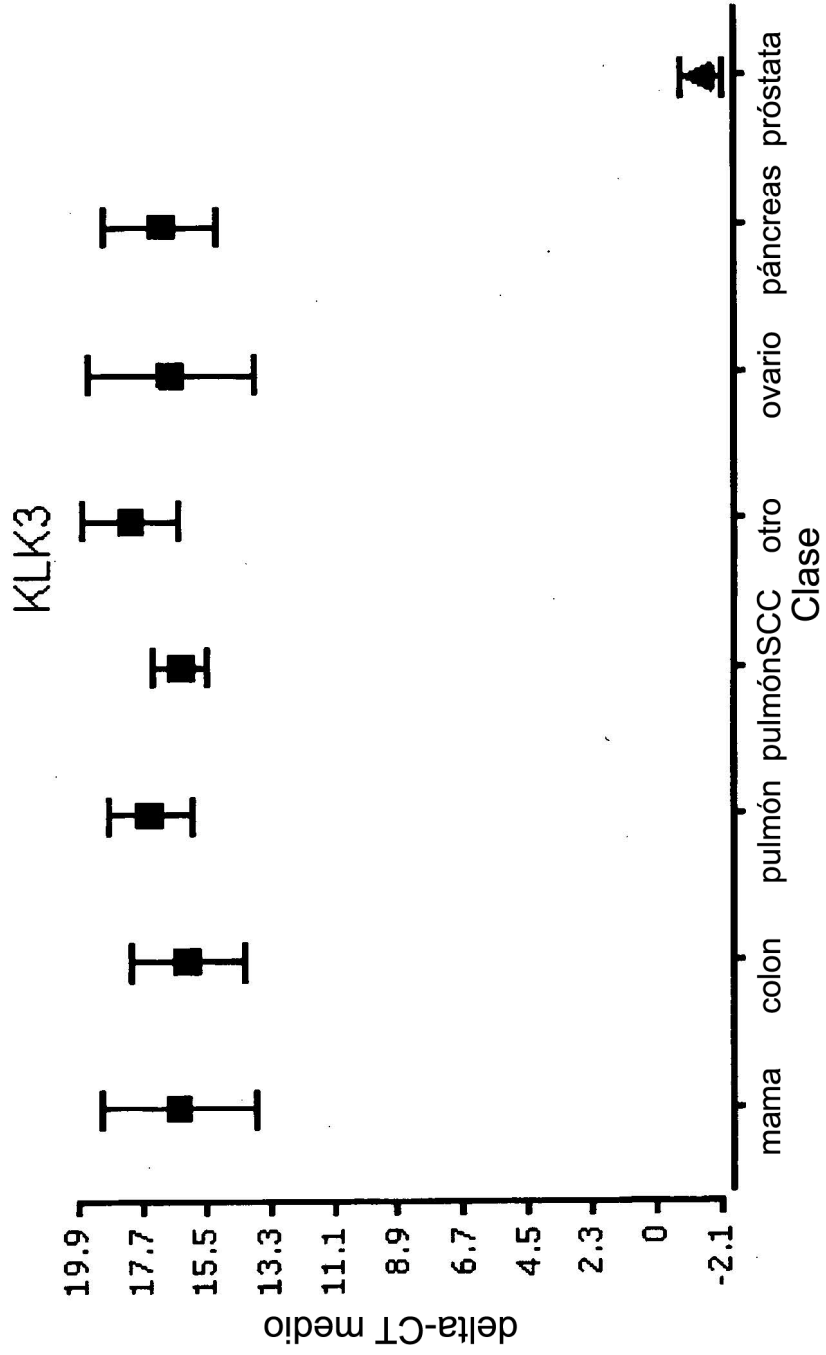


Figura 10 (cont.)

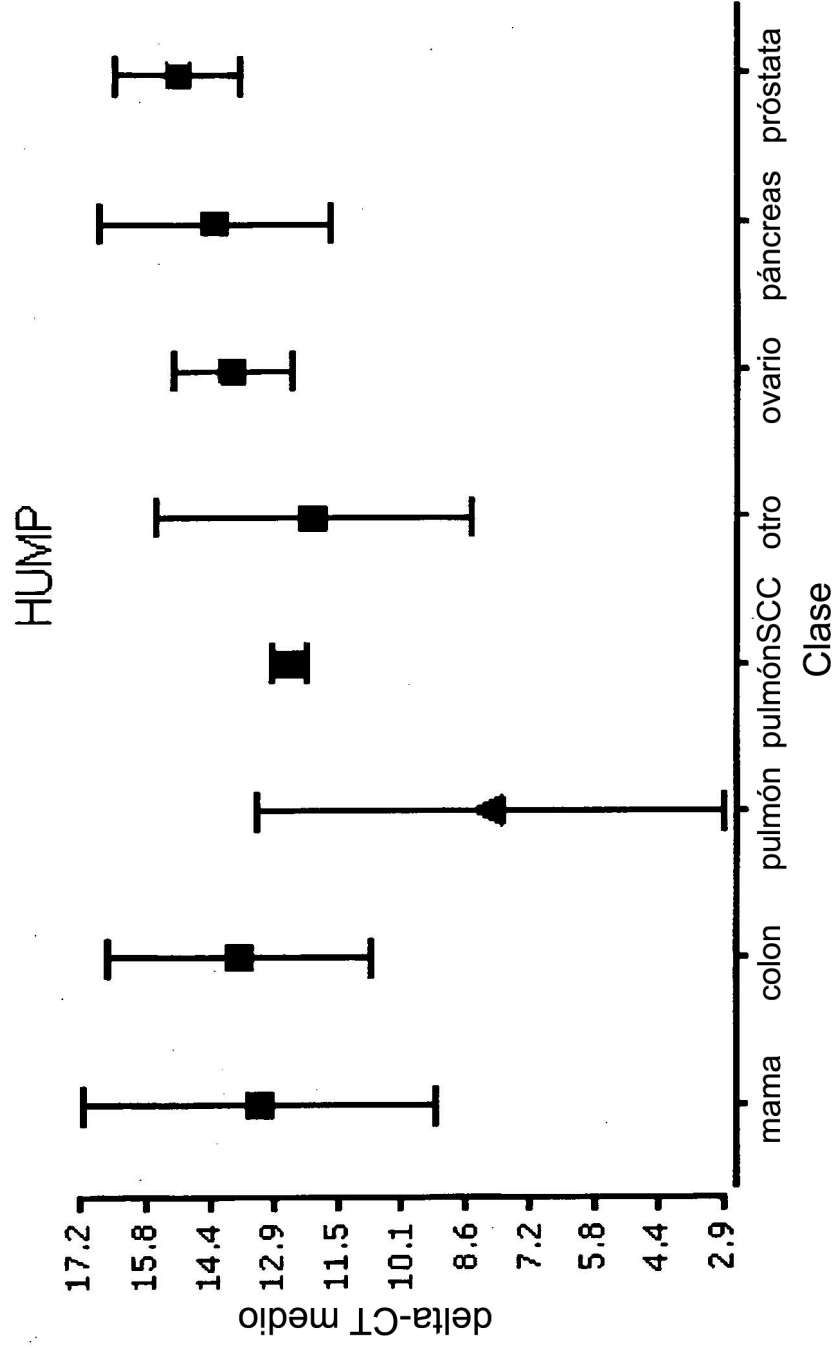


Figura 10 (cont.)

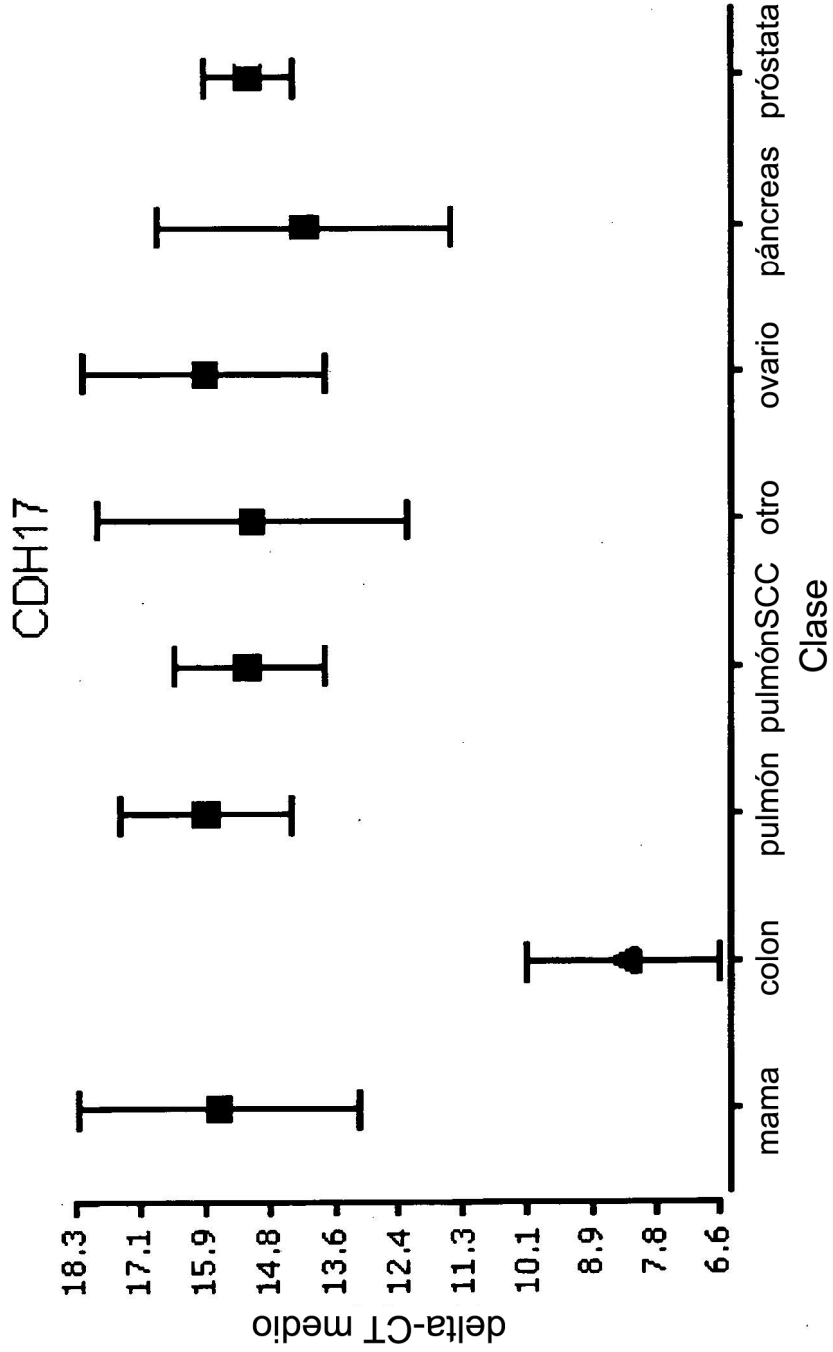




Figura 10 (cont.)

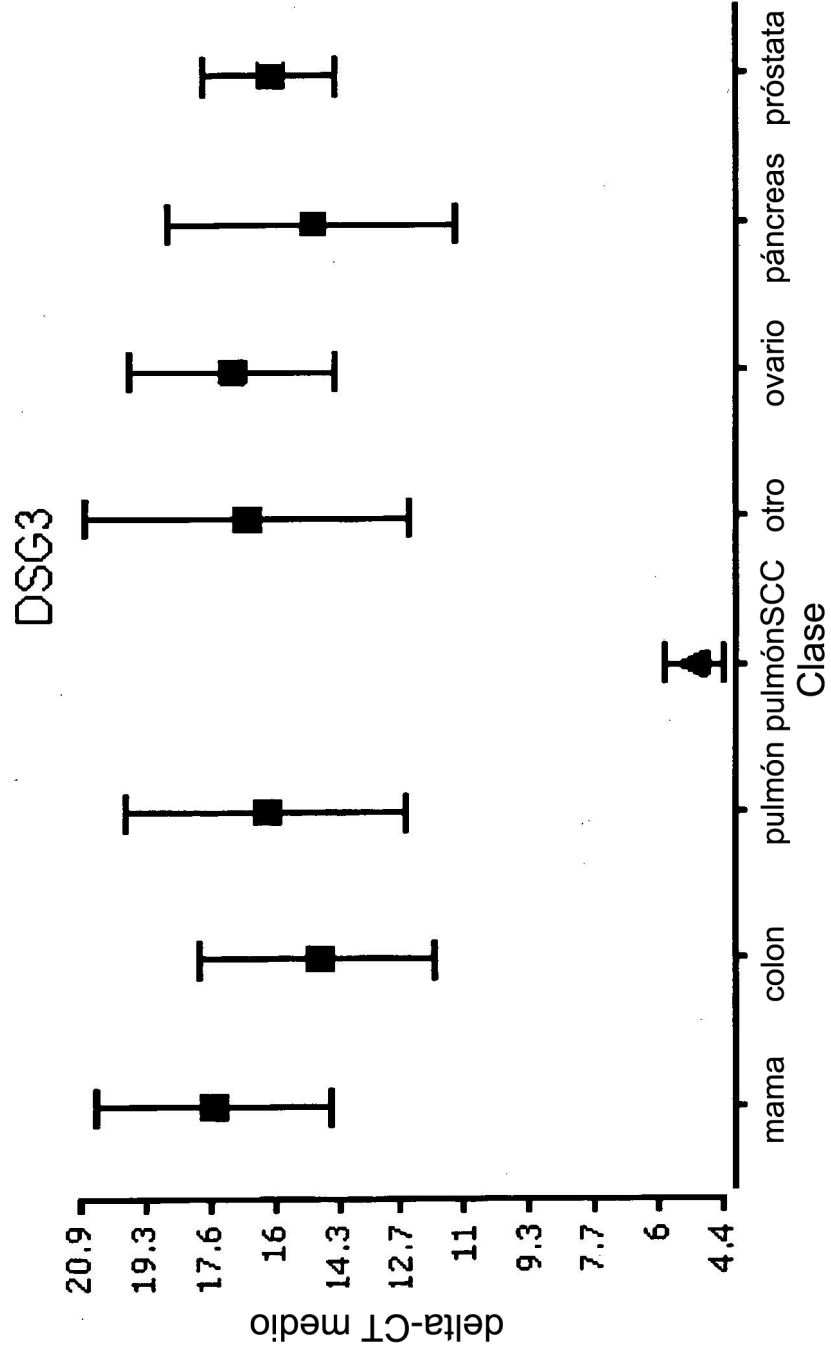


Figura 10 (cont.)

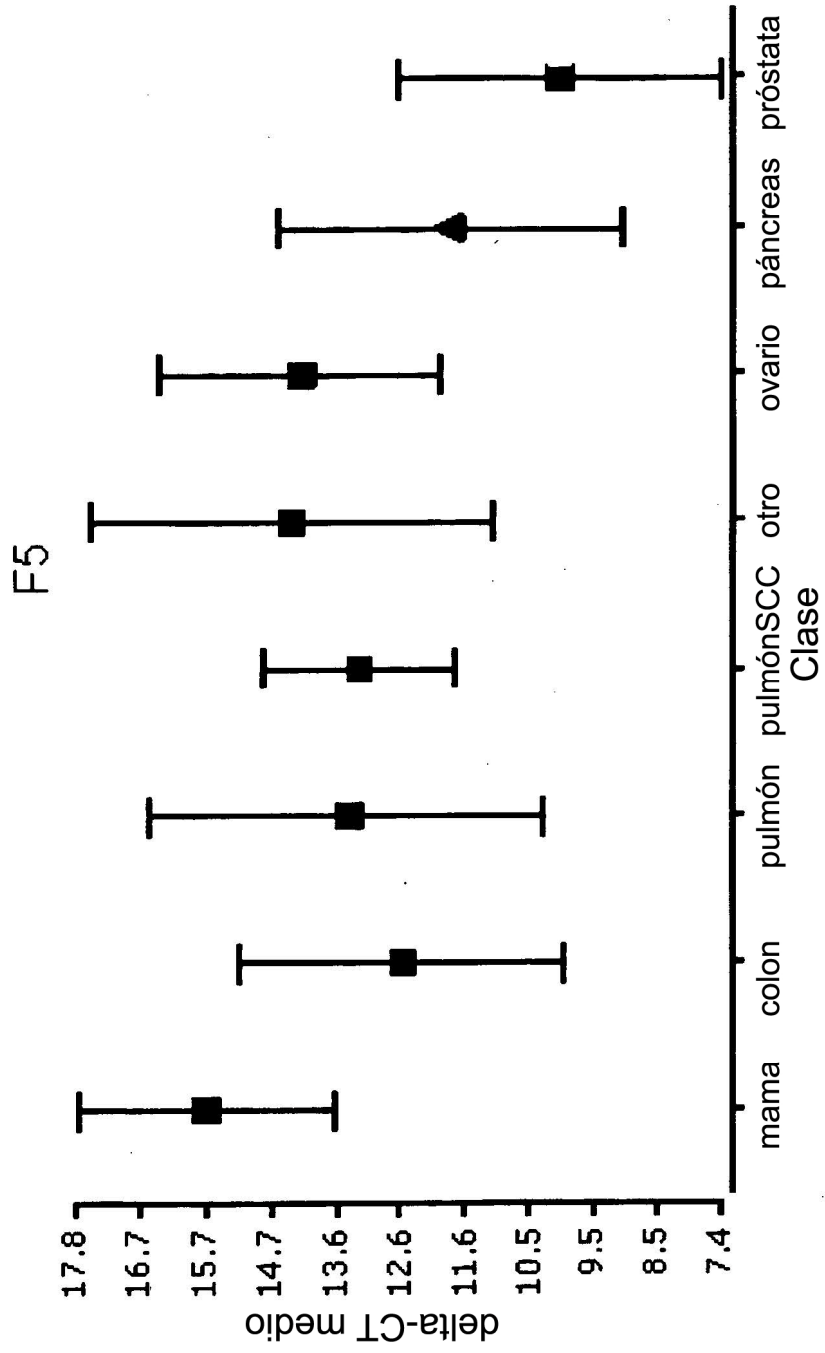


Figura 11(A)

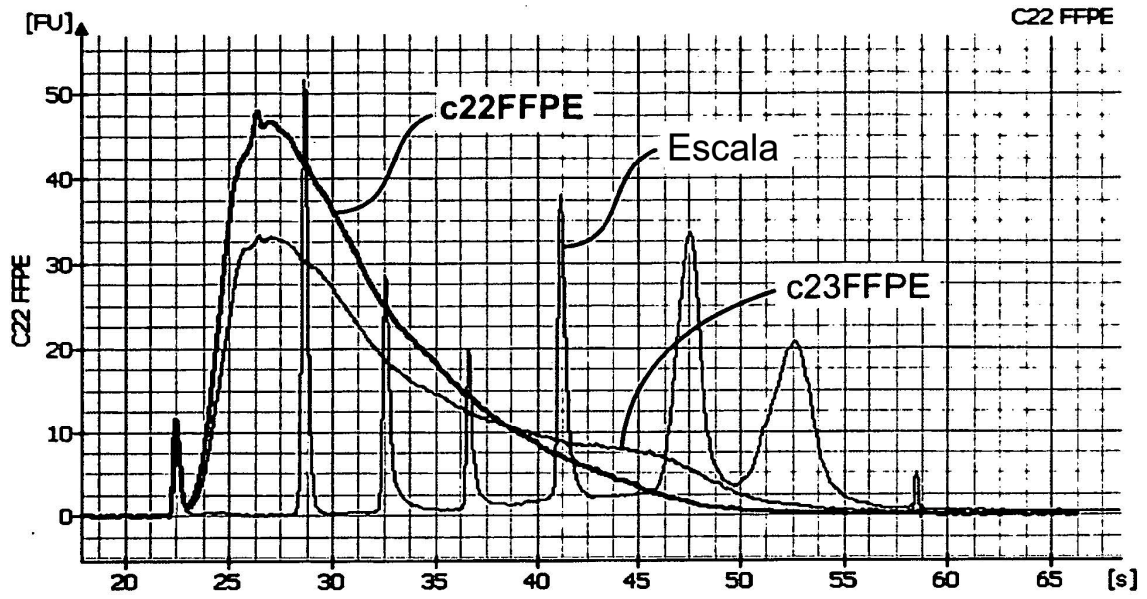


Figura 11(B)

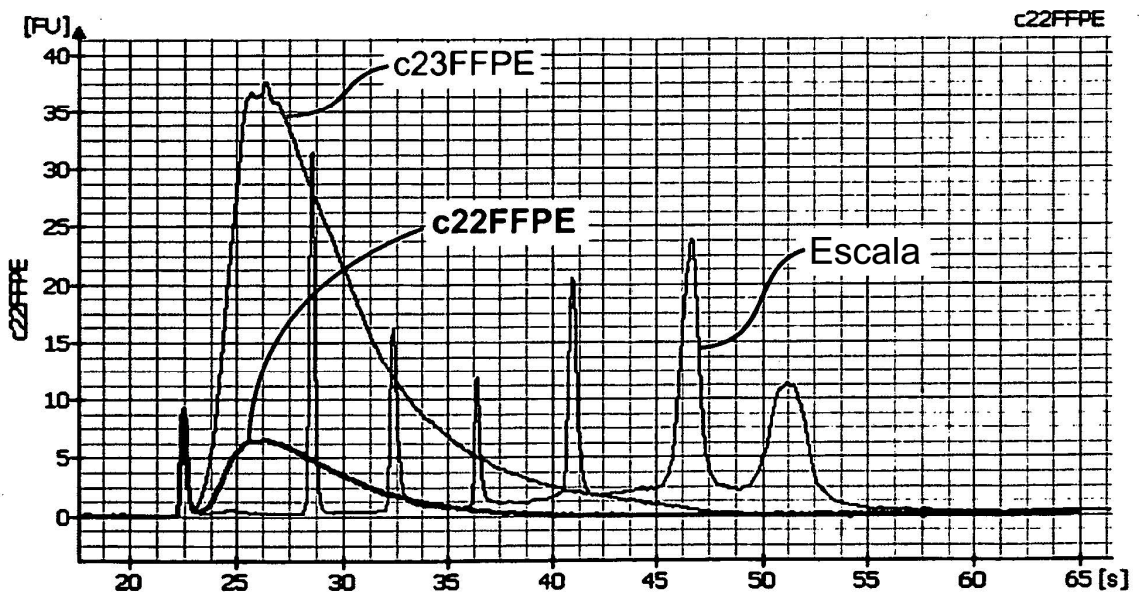


Figura 11(C)  
 $\beta$ -actina

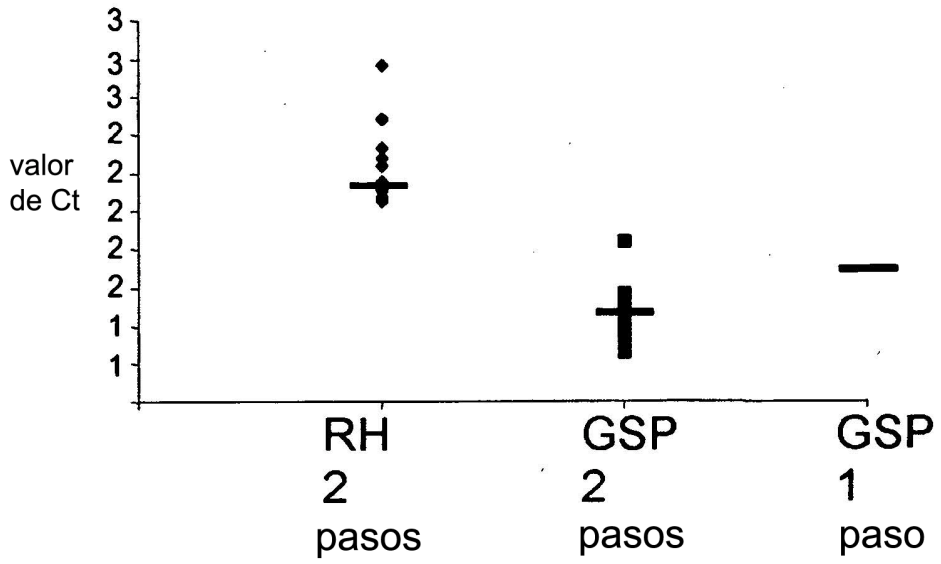


Figura 11(D)  
HUMSPB

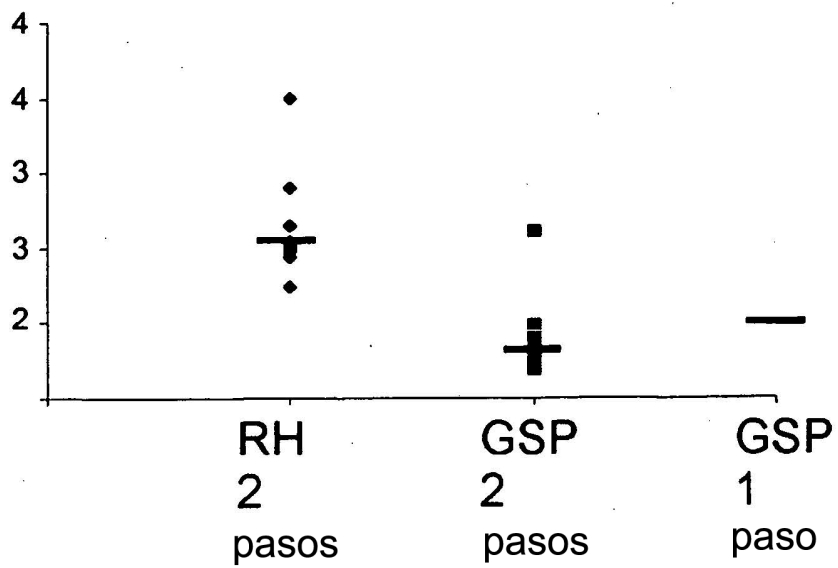


Figura 12

