

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 494 843**

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 06814892 (3)**

97 Fecha y número de publicación de la concesión europea: **04.06.2014 EP 1934615**

54 Título: **Métodos y materiales 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:

16.09.2014

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 2 494 843 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 materiales para identificar el origen de un carcinoma de origen primario desconocido**Descripción**

5 CAMPO DE LA INVENCION

La presente invención proporciona métodos, usos de kits, etc. 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 neoplasias heterogéneas, confirmadas por biopsia, en las que la enfermedad metastásica se presenta sin un tejido de origen (ToO) o sitio del tumor primario identificable. Este problema representa aproximadamente el 3%-5% de todos los cánceres, por lo que es la séptima neoplasia más común. Ghosh *et al.* (2005); y Mintzer *et al.* (2004). El pronóstico y régimen terapéutico de los pacientes dependen del origen del tumor primario, lo que recalca la necesidad de identificar el sitio del tumor primario. Greco *et al.* (2004); Lembersky *et al.* (1996); y Schlag *et al.* (1994).

Actualmente se utilizan diversos métodos para resolver este problema. En las figuras 1-2 se esquematizan varios métodos seguidos. Pueden utilizarse Marcadores tumorales séricos para el diagnóstico diferencial. Aunque carecen de especificidad adecuada, pueden utilizarse en combinación con la información clínica y patológica. Ghosh *et al.* (2005). Pueden utilizarse métodos de inmunohistoquímica (IHC) para identificar el linaje del tumor, pero muy pocos Marcadores IHC son específicos al 100%. Por lo tanto, los patólogos suelen utilizar un panel de Marcadores IHC. Varios estudios han demostrado precisiones del 66%-88% utilizando de cuatro a 14 Marcadores IHC. Brown *et al.* (1997); DeYoung *et al.* (2000); y Dennis *et al.* (2005a). Las pruebas diagnósticas más caras incluyen métodos de formación de imágenes tales como la radiografía de tórax, la tomografía computarizada (CT) y la tomografía 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 para resolver los 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 determinar perfiles 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 el documento 20060094035. Estos estudios demostraron la viabilidad de la identificación del tejido de origen en base al perfil de expresión génica. Para que estas tecnologías de determinación de perfiles de expresión sean útiles en la práctica clínica, deben superarse dos obstáculos principales. En primer lugar, puesto que la determinación de perfiles de expresión génica se lleva a cabo totalmente en tejidos primarios, los marcadores génicos candidatos deben validarse en los tejidos metastásicos para confirmar que su expresión específica de tejido se conserva en la metástasis. En segundo lugar, la tecnología de determinación de perfiles de expresión génica debe ser capaz de utilizar tejido fijado en formalina e incluido en parafina (FFPE), ya que las muestras de tejidos fijados son el material convencional en la práctica actual. La fijación en formalina da como resultado la degradación del ARN (Lewis *et al.* (2001); y Masuda *et al.* (1999)) por lo que los protocolos de micromatrices existentes no funcionarán con la misma fiabilidad. Bibikova *et al.* (2004). Además, la tecnología de determinación de perfiles debe ser robusta, reproducible y fácilmente accesible.

Se ha demostrado que la RTPCR cuantitativa (qRTPCR) genera resultados fiables a partir 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 utilizar un método pangenómico como herramienta de descubrimiento y desarrollar un ensayo de diagnóstico en base a una tecnología más robusta. Ramaswamy (2004). Sin embargo, este paradigma requiere un conjunto de genes más pequeño para su desarrollo. Oien y colaboradores utilizaron el análisis en serie de la expresión génica (SAGE) para identificar 61 Marcadores tumorales a partir de los que desarrollaron un método RTPCR en base a once genes para cinco tipos de tumores. Dennis *et al.* (2002). Otro estudio que acoplaba SAGE y qRTPCR desarrolló un panel de cinco genes para cuatro tipos de tumores y consiguió una exactitud del 81%. Buckhaults *et al.* (2003). Un estudio más reciente acoplaba la determinación de perfiles mediante micromatrices con qRTPCR, pero utilizaba 79 Marcadores. Tohill *et al.* (2005).

55 RESUMEN DE LA INVENCION

La presente invención proporciona un método de identificación del origen de una metástasis de origen desconocido como se define en las reivindicaciones.

60 BREVE DESCRIPCION DE LOS DIBUJOS

Las figuras 1-2 representan los métodos de la técnica anterior de identificación del origen de una metástasis de origen desconocido.
La figura 3 representa el presente algoritmo de diagnóstico de CUP.

La figura 4 representa los datos de micromatriz que muestran las 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.

5 La figura 5 representa los electroferogramas obtenidos a partir de un Agilent Bioanalyzer. Se aisló el ARN a partir de tejido FFPE mediante digestión con proteinasa K durante tres horas (A) o dieciséis horas (B). La muestra C22 (rojo) era un bloque de un año, mientras que la muestra C23 (azul) era un bloque de cinco años. En verde se muestra un marcador de tamaño.

10 La figura 6 representa una comparación de los valores de Ct obtenidos a partir de tres métodos de qRT-PCR diferentes: cebado con hexámeros aleatorios en la transcripción inversa seguido de qPCR con el ADNc resultante (etapa RH 2), cebado específico de gen (cebador inverso) en la transcripción inversa seguido de qPCR con el ADNc resultante (etapa SPG 2) o cebado específico de gen y qRT-PCR en una reacción de una sola etapa (etapa GSP 1). Se dividió ARN de once muestras en los tres métodos y se midieron los niveles de ARN de tres genes: β -actina (A), HUMSPB (B) y TTF (C). La mediana del valor de Ct obtenida con cada método se indica mediante la línea continua.

15 La figura 7 representa los diagramas de placas del ensayo de CUP.

La figura 8 es una serie de gráficos que representan el rendimiento del ensayo a lo largo de un intervalo de concentraciones de ARN.

20 La figura 9 es un diagrama de flujo de trabajo experimental: designación y validación de Marcadores candidatos (9A); y optimización del ensayo y construcción y ensayo del algoritmo de predicción (9B).

La figura 10 representa la expresión de 10 Marcadores génicos candidatos específicos de tejido seleccionados en adenocarcinoma primario de próstata y carcinomas metastásicos FFPE. Para cada gráfico el eje X representa el valor de expresión de Marcador normalizado.

25 La figura 11 representa la optimización del ensayo. (A y B) Electroferogramas obtenidos de un Agilent Bioanalyzer. Se aisló el ARN a partir de tejido FFPE mediante digestión con proteinasa K durante tres horas (A) o dieciséis horas (B). La muestra C22 (rojo) era un bloque de un año, mientras que la muestra C23 (azul) era un bloque de cinco años. En verde se muestra un marcador de tamaño. (C y D) Comparación de los valores de Ct obtenidos a partir de tres métodos de qRT-PCR diferentes: cebado con hexámeros aleatorios en la transcripción inversa seguido de qPCR con el ADNc resultante (etapa RH 2), cebado específico de gen (cebador inverso) en la transcripción inversa seguido de qPCR con el ADNc resultante (etapa SPG 2) o cebado específico de gen y qRT-PCR en una reacción de una sola etapa (GSP 1 etapa). Se dividió ARN de once muestras en los tres métodos y se midieron los niveles de ARN para dos genes: β -actina (C), HUMSPB (D). La mediana del valor de Ct obtenida con cada método se indica mediante la línea continua.

30 La figura 12 es un mapa de calor que muestra los niveles relativos de expresión del panel de 10 Marcadores en 35 239 muestras. El color rojo indica mayor expresión.

DESCRIPCIÓN DETALLADA

40 La identificación del 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. Se validaron Marcadores candidatos mediante reacción en cadena de la polimerasa con transcriptasa inversa (RT-PCR) en 205 carcinomas metastásicos FFPE procedentes de estos seis tejidos, así como metástasis procedentes de otros tipos de cáncer para determinar la especificidad. Se seleccionó una firma genética de diez genes que predijo el tejido de origen de los carcinomas metastásicos para estos seis tipos de cáncer. A continuación, se optimizaron los métodos de qRT-PCR y aislamiento de ARN para estos diez Marcadores, y se aplicó el ensayo qRT-PCR a un conjunto de 260 tumores metastásicos, generando una exactitud global del 78%. Por último, se ensayó un conjunto independiente de 48 muestras metastásicas. Es importante señalar que treinta y siete muestras en este conjunto tenían un primario conocido o inicialmente presentado como CUP, pero se resolvieron posteriormente, y el ensayo demostró una precisión del 78%.

50 Un Biomarcador es cualquier indicio del nivel de expresión de un gen Marcador indicado. Los indicios pueden ser directos o indirectos y medir la sobreexpresión o subexpresión del gen dados los parámetros fisiológicos y en comparación con un control interno, el tejido normal u otro carcinoma. Los Biomarcadores incluyen, sin limitación, los ácidos nucleicos (sobreexpresión y subexpresión y directo e indirecto). El uso de ácidos nucleicos como Biomarcadores puede incluir cualquier método conocido en la técnica, incluidos, sin limitación, la medición de la amplificación de ADN, ARN, micro ARN, la pérdida de heterocigosidad (LOH), los polimorfismos de nucleótido único (SNPs, Brookes (1999)), ADN microsatélite, hipometilación o hipermetilación de ADN. El uso de proteínas como Biomarcadores incluye cualquier método conocido en la técnica, incluidos, sin limitación, la medición de la cantidad, la actividad, modificaciones tales como glicosilación, fosforilación, ribosilación de ADP, ubiquitinación, etc., o inmunohistoquímica (IHC). Otros Biomarcadores incluyen formación de imágenes, recuento celular y Marcadores de apoptosis.

60 Los genes indicados proporcionados en el presente documento son los asociados con un tipo de tejido o tumor concreto. Un gen Marcador puede asociarse con numerosos tipos de cáncer, pero siempre que la expresión del gen esté lo suficientemente asociada con un tipo de tejido o tumor a identificar utilizando el algoritmo descrito en el presente documento para que sea específico para un origen concreto, puede utilizarse el gen en la invención

reivindicada para determinar el tejido de origen para un carcinoma de origen primario desconocido (CUP). En la técnica se conocen numerosos genes asociados con uno o más tipos de cáncer. La presente invención proporciona genes Marcadores preferentes y combinaciones de genes Marcadores aún más preferentes tal como se define en las reivindicaciones. Estas se describen detalladamente en el presente documento.

5 "Origen" en la expresión "tejido de origen" se refiere al tipo de tejido (pulmón, colon, etc.) o al tipo histológico (adenocarcinoma, carcinoma de células escamosas, etc.) en función de las circunstancias médicas concretas y será entendido por cualquier experto en la materia.

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

20 Los métodos y usos de la invención incluyen uno o más genes Marcadores como se define en las reivindicaciones. "Marcador" o "gen Marcador" se utiliza a lo largo de la presente memoria para referirse a los genes y productos de expresión génica que se corresponden con cualquier gen cuya sobreexpresión o subexpresión esté asociada con un tipo de tejido o tumor. Los genes Marcadores preferentes se describen más detalladamente en la Tabla 1.

25

30

35

40

45

50

55

60

65

Tabla 1

SEQ ID NO:	Nombre	Denominación matriz	secuencia
1	SP-B	209810_at	gaaaaaccagccactgctttacaggacaggggggtgaagctgagccccgc ctcacacccacccccatgcaactcaaaagattggattttacagctacttgaatt caaaattcagaagaataaaaaatgggaacatacagaactctaaaagataga catcagaaattgtaagtaagcttttcaaaaaaicagcaattccccagcgta gtcaaggggggacactgcacgctctggcatgatgggatggcgaccggggc aagctttctctcagatgctctgctgcttgagagctattgctttgtaagatat aaaaaggggtttcttttctgttaagggtggacttccagatttgattgaaa gtcctagggtgattctatttctgctgtgattatctgctgaaagctcagctggg gttgtgcaagctagggaccattcctgtgtaatacaatgctgcaccaatgct
2	TTF1	211024_s_at	gtgattcaaatgggtttccacgctagggcggggcagagattggagagggc tctgtgctgacatggctctggactctaaagaccaaacttcaactctgggcaca ctctgccagcaaagaggactcgtttaaataccaggattttttttttttaa gggaggacgggagctggggagaggaagagcttcaacataaccactt gtcactgacacaaaggaagtgccccctccccggcaccctctggccgcta ggctcagcggcgaccgcctccgcgaaaatagtttgtaaatgtgaacttgt agctgtaaaacgctgtcaaaagttggactaaatgcctagttttagtaatctgt acattttgtgtaaaaagaaaaccactcccagtcccagccctcacattttt atgggcattgacaaatctgtgtatattttggcagtttggatttgcggcgta gtcttttctgttgaact
3	DSG3	205595 at	ccatcccatagaagtccagcagacaggattttaaagtccagactttgtca ggaagtcaaggagcttctgttctccgcctctgggtctgtccagccagctgt ttccatccctgaccctctgcagcatgtaactatttagtaacggagacttactc ggcttctggttccctcgtgcaaccttccactgcaggetttagccacttctcac acaaaatgtgatagtgacagaaagggatctgtcccattccagtggtctctg gcaacctagctggccaacgcagctacgaggggtcacatactatgctctgta cagaggatcctgtctcccgttaatatgaccagaatgagctggaataaccaca ctgaccaaactctgatctttggactaaagtattcaaaatagcatagcaaaagct cactgtattgggctaataattggcacttattagcttctctcataaactgateac gattataaattaaatgttgggttcataccccaaaagcaaatatgttgcactct aatttcaagtac

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

4	HPT1	209847_at	ctgcaccacactacttagatatttcatgtgctatagacattagagagattttca ttttccatgacattttctctctgcaaaggccttagctactgtgttttccctttt ggggcaagacagactcattaaatattctgtacattttttttatcaaggagata tatcagtggtgtctcatagaactgcctggattccatttatgtttttctgattccatc ctgtgtccccttcatectgactccttggatttactgaattcaaacattgtc
5	PSCA	205319_at	ttcctgaggcacatcctaacgcaagtttgaccatgtatgtttgacccctttcc ccnaacctgacctcccattggcctttccaggattccnaccnggcagatc agttttagtganacanatccgcntgcagatggccctccaacnttntgttg ntgttccatggcccagcattttccacctaacctgtgttcaggcactnttc cccaggaagccttccctgcccacccattatgaattgagccaggtttggt ccgtggtgtccccgcaccagcaggggacaggcaatcaggaggggccc agtaaaggctgagatgaagtggactgagtagaactggaggacaagagttg acgtgagttcctgggagttccagagatg
6	F5	204713_s_at	atcctctacagccagatgtcacagggatacgtctactttcacttgggtgctgga gaattcanaagtcaagaacatgctaagcnaagggacccaaggtagaaag agatcaagcagcaagcacaggttctcctggatgaaattactagcacataa agttgggagacacctaagccaagacactgggtctcctccggaatgaggcc ctgggaggacctcctagccaagacactgggtctcctccagaatgaggcc ctggaaggacctcctagtgatctgttactcttaaaacaagtaactcatctaa gattttggtgggagatggcattggcttctgagaaaggtagctatgaaataat ccaagatactgatgaagacacagctgtaacaattggctgatcagcccca gaatgcctcacgtgcttggggagaaagcaccctcttccaacaagcctgg aaag
7	MGB1	206378_at	gcagcagcctcaccatgaagttgctgatggctctcatgctggcggccctctc ccagcactgctacgcaggctctggctgccccttattggagaatgtgattcca agacaatcaatccacaagtgtctaagactgaatacaagaacttctcaaga gttcatagacgacaatgccactacaaatgccatagatgaattgaaggaatgt tttctaaccnaacggatgaaactctgagcaatgttgagggtttatgcaatta atatatgacagcagctcttggattattttaacttctgcaagaccttggctca cagaactgcagggtatggtgagaaaccaactacggattgctgcaaacac acctctcttcttatgtcttttact
8	PDEF	220192_x_at	gagtggggcccttaaactggattcaaaaaatgctctaaacataggaatggtt gaagaggcttgcagcttcagatgaaactaaatctctagaagaggcaciaa gaatggctaaagcaattcatccaagggccaccggaagtaattagagctttg aaaaaatctgtttgtcaggcagagagctatattggaggaagcattacagaa cgaaagagatcttttaggaacagttgggggtggcctgcaaatttagaggct attgctaagaaaggaaaaaataaaataattggttttcgtgtggatgtactcc aagtaaagctcagtgactaatatgtataaatgttaaatgatataaatatgaa catcagtaaaaaaaaaaattcttaaggctactataaatgcagacttacttta atcattgaaatctgaactcattfacctcatttcttccaattactcccttgggtat ttactgcgta

5	9	PSA	204582_s_at	<p>tggtgtaattttgctctctgtgtcctggggaactggccatgcctggagac atatcactcaatttctctgaggacacagataggatggggtgtctgtgtattgt ggggtacagagatgaaagaggggtgggatccactgagagagtggag agtgacatgtgtggacactgtccatgaagcactgagcagaagctggagg cacaacgcaccagacactcacagcaaggatggagctgaaaacataacc actctgtcc †</p>
10	10	WT1	206067_s_at	<p>atagatgtacatacctccttcacaaatggagggaattcatttcatcactgg gagtgtccttagtgataaaaaccatgctggtatattgctcaagttgtaaaa taaaagtactttaaagaaaatagggatggccaggatctccactgataa gactgttttaagtaacttaaggaccttgggtctacaagtatatgtaaaaa atgagacttactgggtgaggaaatccattgttaagatggctgtgtgtgt gt taccgttcttgaaactgtgtaaatatgtctgataatgattgctcttgac aactaaaattaggactgtataagtagatgcactgggtgtgtatcttac aagat</p>

25 La presente invención proporciona un método de identificación del origen de una metástasis de origen desconocido tal como se define en las reivindicaciones.

30 Los genes Marcadores se seleccionan de entre i) SP-B, TTF y DSG3, ii) F5 y PSCA o iii) CDH17. Preferentemente, los genes Marcadores son SP-B, TTF, DSG3, KRT6F, p73H y/o SFTPC. Más preferentemente, los genes Marcadores son SP-B, TTF y/o DSG3. Los genes Marcadores adicionalmente pueden incluir o ser reemplazados por KRT6F, p73H y/o SFTPC.

35 En una forma de realización, los genes Marcadores son F5, PSCA, ITGB6, KLK10, CLDN18, TR10 y/o FKBP10. Más preferentemente, los genes Marcadores son F5 y/o PSCA. Preferentemente, los genes Marcadores pueden incluir o ser reemplazados por ITGB6, KLK10, CLDN18, TR10 y/o FKBP10.

En otra forma de realización, los genes Marcadores son CDH17, CDX1 y/o FABP1, preferentemente, CDH17. Los genes Marcadores adicionalmente pueden incluir o ser reemplazados por CDX1 y/o FABP1.

40 En una forma de realización, la expresión génica se mide utilizando al menos una de las SEQ ID NO: 11-58.

La presente descripción también abarca métodos que miden la expresión génica obteniendo y midiendo la formación de al menos uno de los amplicones SEQ ID NO: 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54 y/o 58.

45 En una forma de realización, los genes Marcadores puede seleccionarse de un Marcador específico de género seleccionado de al menos uno de entre: i) en el caso de un paciente masculino KLK3, KLK2, NGEF o NPY; o ii) en el caso de una paciente femenina PDEF, MGB, PIP, B305D, B726 o GABA-Pi; y/o WT1, PAX8, STAR o EMX2. Preferentemente, el gen Marcador es KLK2 o KLK3. En esta forma de realización, los genes Marcadores pueden incluir o ser reemplazados por NGEF y/o NPY. En una forma de realización, los genes Marcadores son PDEF, MGB, PIP, B305D, B726 o GABA-Pi, preferentemente, PDEF y MGB. En esta forma de realización, los genes Marcadores pueden incluir o ser reemplazados por PIP, B305D, B726 o GABA-Pi. En una forma de realización, los genes Marcadores son WT1, PAX8, STAR o EMX2, preferentemente, WT1. En esta forma de realización, los genes Marcadores pueden incluir o ser reemplazados por PAX8, STAR o EMX2.

55 La presente descripción proporciona métodos para obtener información clínica adicional, incluido el sitio de metástasis para determinar el origen del carcinoma; obtener conjuntos de Biomarcadores óptimos para carcinomas que comprenden las etapas de utilizar metástasis de origen conocido, determinar los Biomarcadores para ello y comparar los Biomarcadores con los Biomarcadores de metástasis de origen desconocido; proporcionar orientación para el tratamiento determinando el origen de una metástasis de origen desconocido e identificar los tratamientos apropiados para ello; y proporcionar un pronóstico determinando el origen de una metástasis de origen desconocido e identificar el pronóstico correspondiente para ello.

60 La presente descripción proporciona adicionalmente métodos para encontrar Biomarcadores determinando el nivel de expresión de un gen Marcador en una metástasis concreta, medir un Biomarcador para el gen Marcador para determinar la expresión del mismo, analizar la expresión del gen Marcador según cualquiera de los métodos

proporcionados en el presente documento o conocidos en la técnica y determinar si el gen Marcador es realmente específico para el tumor de origen.

5 La presente invención proporciona adicionalmente el uso de una composición que contiene al menos una secuencia aislada seleccionada de entre las SEQ ID NO: 11-58 en los métodos de la invención. La presente invención proporciona adicionalmente el uso de un kit para llevar a cabo un ensayo según los métodos proporcionados en el presente documento que contiene ARN o ADNc que hibrida con los genes Marcadores utilizados en los métodos de la invención.

10 La presente invención proporciona adicionalmente el uso de micromatrices o matrices génicas para llevar a cabo los métodos descritos en el presente documento.

15 La presente descripción proporciona adicionalmente carteras de diagnóstico/pronóstico que contienen secuencias aisladas de ácidos nucleicos, sus complementos, o porciones de los mismos de una combinación de genes como se describe en el presente documento, en los que la combinación es suficiente para medir o caracterizar la expresión génica en una muestra biológica con células metastásicas con respecto a células de diferentes carcinomas o tejido normal.

20 Cualquier método descrito en la presente descripción puede incluir adicionalmente medir la expresión de al menos un gen expresado constitutivamente en la muestra.

25 Preferentemente, los Marcadores para el cáncer de páncreas son el factor de coagulación V (F5), el antígeno de células madre de próstata (PSCA), la integrina $\beta 6$ (ITGB6), la calicreína 10 (KLK10), claudina18 (CLDN18), la isoforma trio (TR10) y la proteína hipotética FLJ22041 similar a proteínas de unión a FK506 (FKBP10). Preferentemente, los Biomarcadores para F5 y PSCA se miden conjuntamente. Pueden medirse los Biomarcadores para ITGB6, KLK10, CLDN18, TR10 y FKBP10, además de o en lugar de F5 y/o PSCA. F5 se describe por ejemplo en los documentos 20040076955; 20040005563 y W02004031412. PSCA se describe por ejemplo en los documentos WO1998040403; 20030232350 y W02004063355. ITGB6 se describe, por ejemplo, en los documentos WO2004018999 y 6339148. KLK10 se describe, por ejemplo, en los documentos WO2004077060 y 20030235820. CLDN18 se describe, por ejemplo, en los documentos WO2004063355 y WO2005005601. TR10 se describe por ejemplo en el documento 20020055627. FKBP10 se describe, por ejemplo, en el documento WO2000055320.

35 Preferentemente, los genes Marcadores del cáncer de colon son el transportador asociado al péptido intestinal HPT-1 (CDH17), el factor de transcripción 1 de la caja homeótica de tipo caudal (CDX1) y la proteína 1 de unión a ácidos grasos (FABP1). Preferentemente, un Biomarcador para CDH17 se mide en solitario. Pueden medirse los Biomarcadores para CDX1 y FABP1, además de, o en lugar de un Biomarcador para CDH17. CDH17 es descrito por ejemplo por Takamura *et al.* (2004); y en el documento W02004063355. CDX1 es descrito por ejemplo por Pilozzi *et al.* (2004); en los documentos 20050059008 y 20010029020. FABP1 es descrito por ejemplo por Borchers *et al.* (1997); Chan *et al.* (1985); Chen *et al.* (1986); y Lowe *et al.* (1985).

40 Preferentemente, los genes Marcadores para el cáncer de pulmón son la proteína B surfactante (PS-B), el factor de transcripción tiroideo (TTF), la desmogleína 3 (DSG3), la queratina 6 isoforma 6F (KRT6F), el gen relacionado con p53 (p73H) y la proteína C surfactante (SFTPC). Preferentemente, los Biomarcadores para SP-B, TTF y DSG3 se miden conjuntamente. Pueden medirse los Biomarcadores para KRT6F, p73H y SFTPC, además de, o en lugar de cualquiera de los Biomarcadores para SP-B, TTF y/o DSG3. SP-B es descrito, por ejemplo, por Pilot-Mathias *et al.* (1989); en los documentos 20030219760 y 20030232350. TTF es descrito, por ejemplo, por Jones *et al.* (2005); en los documentos US20040219575; WO1998056953; WO2002073204; 20030138793 y WO2004063355. DSG3 es descrito, por ejemplo, por Wan *et al.* (2003); en los documentos 20030232350; aWO2004030615 y W02002101357. KRT6F es descrito, por ejemplo, por Takahashi *et al.* (1995); en los documentos 20040146862 y 20040219572. p73H es descrito, por ejemplo, por Senoo *et al.* (1998); y en el documento 20030138793. SFTPC es descrito, por ejemplo, por Glasser *et al.* (1988).

55 Los genes Marcadores pueden seleccionarse adicionalmente de un Marcador específico de género tal como, en el caso de un paciente masculino KLK3, KLK2, NGEF o NPY; o en el caso de una paciente femenina PDEF, MGB, PIP, B305D, B726 o GABA-Pi; y/o WT1, PAX8, STAR o EMX2.

60 Preferentemente, los genes Marcadores para el cáncer de mama son el factor epitelial derivado de próstata (PDEF), la mamaglobina (MG), la proteína inducible por prolactina (PIP), B305D, B726 y GABA- π . Preferentemente, los Biomarcadores para PDEF y MG se miden conjuntamente. Pueden medirse los Biomarcadores para PIP, B305D, B726 y GABA-Pi, además de, o en lugar de los Biomarcadores para PDEF y/o MG. PDEF es descrito, por ejemplo, en los documentos WO2004030615; WO2000006589; WO2001073032; por Wallace *et al.* (2005); Feldman *et al.* (2003); y Oettgen *et al.* (2000). MG es descrito, por ejemplo, en los documentos W02004030615; 20030124128; por Fleming *et al.* (2000); Watson *et al.* (1996 y 1998); y en el documento 5668267. PIP es descrito, por ejemplo, por Autiero *et al.* (2002); Clark *et al.* (1999); Myal *et al.* (1991) y Murphy *et al.* (1987). B305D, B726 y GABA-Pi son descritos por Reinholz *et al.* (2005). NGEF es descrito, por ejemplo, por Bera *et al.* (2004).

Preferentemente, los Marcadores para el cáncer de ovario son el tumor de Wilms 1 (WT1), PAX8, la proteína reguladora aguda esteroideogénica (STAR) y EMX2. Preferentemente, se miden los Biomarcadores para WT1. Pueden medirse los Biomarcadores para STAR y EMX2, además de o en lugar de los Biomarcadores para WT1. WT1 es descrito, por ejemplo en los documentos 5350840; 6232073; 6225051; 20040005563; y por Bentov *et al.* (2003). PAX8 es descrito, por ejemplo, en el documento 20050037010; por 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). STAR es descrito, por ejemplo, por Gradi *et al.* (1995); y Kim *et al.* (2003). EMX2 es descrito, por ejemplo, por Noonan *et al.* (2001).

Preferentemente, los Marcadores para el cáncer de próstata son KLK3, KLK2, NGEF y NPY. Preferentemente, se miden los Biomarcadores para KLK3. Pueden medirse los Biomarcadores para KLK2, NGEF y NPY, además de o en lugar de KLK3. KLK2 y KLK3 son descritos, por ejemplo, por Magklara *et al.* (2002). KLK2 se describe, por ejemplo, en los documentos 20030215835 y 5786148. KLK3 se describe, por ejemplo, en el documento 6261766.

El método también puede incluir obtener información clínica adicional, incluido el sitio de metástasis para determinar el origen del carcinoma. En la figura 3 se proporciona un diagrama de flujo.

La descripción proporciona adicionalmente un método para obtener conjuntos de Biomarcadores óptimos para carcinomas utilizando metástasis de origen conocido, determinar los Biomarcadores para ello y comparar los Biomarcadores con los Biomarcadores de metástasis de origen desconocido.

La descripción proporciona adicionalmente un método para proporcionar orientación para el tratamiento determinando el origen de una metástasis de origen desconocido según los métodos descritos en el presente documento e identificando el tratamiento apropiado para ello.

La descripción proporciona adicionalmente un método para proporcionar un pronóstico determinando el origen de una metástasis de origen desconocido según los métodos descritos en el presente documento e identificando el pronóstico correspondiente para ello.

La descripción proporciona adicionalmente un método para encontrar Biomarcadores que comprende determinar el nivel de expresión de un gen Marcador en una metástasis concreta, medir un Biomarcador para el gen Marcador para determinar la expresión del mismo, analizar la expresión del gen Marcador según los métodos descritos en el presente documento y determinar si el gen Marcador es realmente específico para el tumor de origen.

La invención proporciona adicionalmente el uso de composiciones que comprenden al menos una secuencia aislada seleccionada de entre las SEQ ID NO: 11-58 en los métodos de la invención.

La invención proporciona adicionalmente el uso de kits, micromatrices o matrices génicas para llevar a cabo los ensayos descritos en el presente documento.

Sólo en raras ocasiones se ha descubierto que la mera presencia o ausencia de secuencias concretas de ácidos nucleicos en una muestra de tejido tenga valor diagnóstico o pronóstico. Por otro lado, se considera cada vez más importante la información sobre la expresión de diversas proteínas, péptidos o ARNm. La mera presencia de secuencias de ácidos nucleicos que tengan el potencial de expresar proteínas, péptidos o ARNm (tales secuencias conocidas como "genes") dentro del genoma por sí misma no es determinante de si una proteína, péptido, o ARNm se expresa en una determinada célula. Que un determinado gen capaz de expresar proteínas, péptidos o ARNm, lo haga o no, y en qué medida se produzca tal expresión, en todo caso, viene determinado por diversos factores complejos. Independientemente de las dificultades en la comprensión y la evaluación de estos factores, el ensayo de la expresión génica puede proporcionar información útil sobre la existencia de eventos importantes, tales como la tumorigénesis, la metástasis, la apoptosis, y otros fenómenos clínicamente pertinentes. Los indicios relativos del grado en que los genes son activos o inactivos pueden encontrarse en los perfiles de expresión génica. Los perfiles de expresión génica de la presente invención se utilizan para proporcionar un diagnóstico y para tratar a los pacientes de CUP.

La preparación de la muestra requiere la recogida de muestras de pacientes. Las muestras de pacientes utilizadas en el método de la invención son las que se sospecha contienen células enfermas, tales como células tomadas de un ganglio en una biopsia por aspiración con aguja fina (FNA) de tejido. La preparación de la masa de tejido obtenida de una biopsia o una muestra quirúrgica y la microdissección de captura por láser también son adecuadas para su uso. La tecnología de microdissección de captura por láser (LCM) es una manera de seleccionar las células a estudiar, minimizando la variabilidad debida a la heterogeneidad del tipo de célula. Por consiguiente, pueden detectarse fácilmente cambios moderados o pequeños en la expresión del gen Marcador entre las células normales o benignas y las cancerosas. Las muestras también pueden comprender células epiteliales circulantes extraídas de sangre periférica. Éstas pueden obtenerse según varios métodos, pero el método más preferente es la técnica de separación magnética descrita en el documento 6136182. Una vez se ha obtenido la muestra que

contiene las células de interés, se obtiene un perfil de expresión génica utilizando un Biomarcador, para los genes en las carteras apropiadas.

5 Los métodos preferentes para establecer los perfiles de expresión génica incluyen determinar la cantidad de ARN que es producido por un gen que puede codificar una proteína o péptido. Esto se logra mediante PCR con transcriptasa inversa (RT-PCR), RT-PCR competitivo, RT-PCR en tiempo real, RT-PCR de presentación diferencial, análisis de Northern Blot y otros ensayos relacionados. Aunque es posible llevar a cabo estas técnicas mediante reacciones de PCR individuales, es mejor amplificar ADN complementario (ADNc) o ARN complementario (ARNc) producido a partir de ARNm y analizarlo mediante micromatrices. Los expertos en la materia conocen varios métodos y configuraciones de matriz diferentes para su producción y se describen, por ejemplo, en los documentos 10 5445934, 5532128, 5556752, 5242974, 5384261, 5405783, 5412087, 5424186, 5429807, 5436327, 5472672, 5527681, 5529756, 5545531, 5554501, 5561071, 5571639, 5593839, 5599695, 5624711, 5658734 y 5700637.

15 La tecnología de micromatrices permite medir a nivel del ARNm en estado estacionario de miles de genes simultáneamente, lo que proporciona una poderosa herramienta para la identificación de efectos tales como la aparición, detención o modulación de la proliferación celular incontrolada. Actualmente se encuentra muy extendido el uso de dos tecnologías de micromatrices, las matrices de oligonucleótidos y de ADNc. Aunque existen diferencias en la construcción de estas matrices, esencialmente todas las salidas y análisis de datos de bajada son iguales. El producto de estos análisis son por lo general mediciones de la intensidad de la señal recibida desde una sonda 20 marcada utilizada para detectar una secuencia de ADNc de la muestra que hibrida con una secuencia de ácido nucleico en una ubicación conocida en el micromatriz. Por lo general, la intensidad de la señal es proporcional a la cantidad de ADNc, y por lo tanto de ARNm, expresado en las células de la muestra. Se dispone de un gran número de tales técnicas y resultan útiles. Los métodos preferentes para determinar la expresión génica pueden encontrarse en los documentos 25 6271002, 6218122, 6218114 y 6004755.

El análisis de los niveles de expresión se lleva a cabo comparando tales intensidades de señal. Esto se realiza mejor generando una matriz de relación de las intensidades de expresión de los genes en una muestra de ensayo frente a las de una muestra de control. Por ejemplo, las intensidades de expresión génica de un tejido enfermo pueden compararse con las intensidades de expresión generadas a partir de tejido benigno o normal del mismo tipo. Una relación de estas intensidades de expresión indica el número de veces que cambia la expresión 30 génica entre las muestras de ensayo y de control.

La selección puede basarse en ensayos estadísticos que producen listas ordenadas relacionadas con la evidencia de significación para la expresión diferencial de cada gen entre los factores relacionados con el sitio de origen original del tumor. Los ejemplos de tales ensayos incluyen ANOVA y Kruskal-Wallis. Las clasificaciones pueden utilizarse como ponderaciones en un modelo diseñado para interpretar la adición total de tales ponderaciones, hasta un punto de corte, como la preponderancia de la evidencia a favor de una clase sobre otra. También puede utilizarse la evidencia anterior como se describe en la literatura para ajustar las ponderaciones. 35

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 un grupo *ad hoc* de ensayos estadísticos, optimización de la media-varianza y el conocimiento de expertos. En una forma de realización alternativa, los métodos de extracción de características podrían automatizarse para seleccionar y ensayar los Marcadores a través de enfoques de aprendizaje supervisado. A medida que crece la base de datos, puede repetirse la selección de 40 Marcadores con el fin de producir la mayor precisión diagnóstica posible en cualquier estado determinado de la base de datos.

Una forma de realización preferente es normalizar cada medición identificando un conjunto de control estable y escalando este conjunto a varianza cero en todas las muestras. Este conjunto de control se define como cualquier transcrito endógeno único o conjunto de transcritos endógenos influidos por el error sistemático en el ensayo, y que no se sepa que cambien independientemente de este error. Todos los Marcadores se ajustan mediante el factor específico de muestra que genera varianza cero para cualquier estadístico descriptivo del conjunto de control, tal como la media o la mediana, o para una medición directa. Como alternativa, si la premisa de variación de los controles relacionados solamente con el error sistemático no es cierta, y sin embargo el error de clasificación resultante es menor cuando se realiza la normalización, el conjunto de control seguirá siendo utilizado como se indica. También podrían ser útiles los controles de adición no endógenos, pero no resultan preferentes. 50

Después del Marcador de selección, se utilizan estas variables seleccionadas en un clasificador diseñado para producir la mayor precisión de clasificación posible. Puede utilizarse un algoritmo de aprendizaje supervisado diseñado para relacionar un conjunto de medidas de entrada con un conjunto de salida de predictores con el fin de 55 construir un modelo a partir de las 10 entradas para predecir el tejido de origen. El problema puede enunciarse como: los datos de entrenamiento dados $\{(x_1, y), \dots, (x_n, y)\}$ producen un clasificador $h: X \rightarrow Y$ que mapea una muestra $x \in X$ a su etiqueta de tejido de origen $y \in Y$. Las predicciones se basan en los casos resueltos con anterioridad que están contenidos en la base de datos y por lo tanto componen el conjunto de entrenamiento.

60 El algoritmo de aprendizaje supervisado debe encontrar parámetros basados en las relaciones entre las variables de entrada y las salidas conocidas que minimicen el error de clasificación esperado. A continuación,

65

pueden utilizarse estos parámetros para predecir el tejido de origen a partir de la entrada de una nueva muestra. Ejemplos de estos algoritmos incluyen modelos lineales de clasificación, clasificadores cuadráticos, métodos basados en árboles, redes neuronales y métodos de prototipo tales como un clasificador de k vecinos más cercanos o algoritmos de cuantificación de vectores de aprendizaje.

5 Una forma de realización específica para diseñar los 10 Marcadores normalizados es el método LDA, utilizando 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 $\vec{\mu}_{y=0}$, $\vec{\mu}_{y=1}$ y las covarianzas $\Sigma_{y=0}$, $\Sigma_{y=1}$ para las etiquetas de clase y 0 y 1, se busca una combinación lineal de $\vec{\mu}$ que tenga unas medias $\vec{\mu}$ y unas varianzas $\Sigma_{y=\mu}$ que maximicen la relación entre la varianza entre las clases y la varianza dentro de las clases:

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

20 LDA puede generalizarse a un análisis discriminante para múltiples clases, donde y tiene N posibles estados, en lugar de sólo dos. Se hace una estimación de las medias y varianzas de clase a partir de los valores contenidos en la base de datos para los Marcadores elegidos. En una forma de realización preferente, la matriz de covarianza se pondera por las probabilidades a priori iguales de cada tipo de tumor sujeta a la siguiente. La predicción en los pacientes masculinos se realiza mediante un modelo en el que las distribuciones de probabilidad a priori son cero para cada grupo de tumores de los órganos reproductores femeninos. Del mismo modo, la predicción en las pacientes femeninas se realiza mediante un modelo en el que las distribuciones de probabilidad a priori son cero para los órganos reproductores masculinos. En la presente invención, las distribuciones de probabilidad a priori son cero para los ensayos en mujeres para próstata y cero para hombres sometidos a ensayo para mama y ovario. Además, las muestras con un antecedente idéntico a una etiqueta de clase se someten a ensayo mediante un modelo en el que la probabilidad a priori es cero para esa etiqueta de clase concreta.

25 El problema anterior puede verse como una maximización del cociente de Rayleigh manejado como un problema de valor propio generalizado. El subespacio reducido se utiliza en la clasificación calculando la distancia de cada muestra al centroide en el subespacio elegido. El modelo puede ajustarse mediante la máxima verosimilitud, y las probabilidades posteriores se calculan mediante el teorema de Bayes.

30 Un método alternativo puede incluir encontrar un mapa del espacio de características de dimensión n, donde n es el número de variables utilizadas, a un conjunto de etiquetas de clasificación implicará el reparto del espacio de características en regiones, a continuación, asignar una clasificación a cada región. Las puntuaciones de estos algoritmos de tipo vecinos más cercanos están relacionados con la distancia entre límites de decisión y no se traducen necesariamente en probabilidades de clase.

35 Si hay demasiadas variables para elegir, y muchas de ellas son ruido aleatorio, la selección de variables y el modelo corren el riesgo de una sobrevaloración del problema. Por lo tanto, suele utilizarse una lista clasificada en diversos puntos de corte como entradas con el fin de limitar el número de variables. También pueden utilizarse algoritmos de búsqueda tal como un algoritmo genético para seleccionar un subconjunto de variables que ensayan una función de coste. Puede intentarse un apareamiento simulado para limitar el riesgo de alcanzar la función de coste en un mínimo local. Sin embargo, estos procedimientos deben validarse con muestras independientes al proceso de selección y modelización.

40 También pueden utilizarse enfoques de variable latente. Puede utilizarse cualquier algoritmo de aprendizaje no supervisado para hacer una estimación de variedades de baja dimensión a partir de un espacio de alta dimensión para descubrir asociaciones entre las variables de entrada y lo bien que pueden ajustarse a un conjunto menor de variables latentes. Aunque las estimaciones de la eficacia de las reducciones son subjetivas, puede aplicarse un algoritmo supervisado en el conjunto de variables reducido con el fin de hacer una estimación de la precisión de la clasificación. Por lo tanto, un clasificador, que puede construirse a partir de las variables latentes, también puede construirse a partir de un conjunto de variables correlacionadas significativamente con las variables latentes. Un ejemplo de esto podría incluir el uso de variables correlacionadas con los componentes principales, a partir de un análisis de componentes principales, como entradas para cualquier modelo de clasificación supervisada.

45 Estos algoritmos pueden implementarse en cualquier código de software que tenga métodos para introducir las variables, entrenar las muestras con una función, ensayar una muestra en base al modelo y devolver los resultados a una consola. R, Octave, C, C++, Fortran, Java, Perl y Python tienen bibliotecas disponibles bajo una

licencia de código abierto para realizar muchas de las funciones enumeradas anteriormente. Los paquetes comerciales tales como S+ y Matlab también contienen muchos de estos métodos.

5 El código realiza las siguientes etapas en el siguiente orden utilizando R versión 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.

- 10 1) Los valores de CT para 10 genes Marcadores y 2 controles se almacenan en un disco duro para todas las muestras de conjunto de entrenamiento disponibles.
- 10 2) Para cada muestra, la resta del promedio específico de la muestra de los controles de cada marcador normaliza los 10 valores de gen marcador.
- 15 3) El conjunto de datos de entrenamiento está compuesto por las metástasis con sitios de origen conocidos, donde cada muestra tiene al menos uno de sus marcadores diana específico para el tejido de origen marcado con un valor de CT normalizado inferior a 5.
- 15 4) LDA construye 4 conjuntos de 2 modelos LDA a partir de los datos de entrenamiento en (3). En cada conjunto, un modelo que es específico para los hombres, y tiene la probabilidad a priori para mama y ovario establecida en cero, así como la probabilidad a priori para próstata ajustada a las distribuciones de probabilidad a priori equivalentes de las demás etiquetas de clase. El otro modelo en cada par es específico para las mujeres con la probabilidad a priori de próstata establecida en cero, y con las distribuciones de probabilidad a priori para mama y ovario ajustadas a las distribuciones de probabilidad a priori equivalentes encontradas en las demás etiquetas de clase.

- 25 a. El primer conjunto se utiliza para ensayar muestras de CUP encontradas en el colon, las probabilidades a priori para colon se establecen en cero y todas las demás etiquetas de clase de no reproductivo se ajustan a las distribuciones de probabilidad a priori equivalentes.
- 25 b. Un segundo conjunto de modelos es específico para un CUP encontrado en el ovario, con la probabilidad a priori para ovario establecida en cero y las demás etiquetas de clase de no reproductivo ajustadas a las distribuciones de probabilidad a priori equivalentes.
- 30 c. Un tercer conjunto es un CUP encontrado en el pulmón, con probabilidad a priori para pulmón establecida en cero. Todas las demás etiquetas de clase de no reproductivo tienen distribuciones de probabilidad a priori equivalentes.
- 30 d. El modelo general utilizado para todos los demás tejidos de fondo. Todas las distribuciones de probabilidad a priori se ajustan de manera equivalente a excepción de las etiquetas de clase específicas de reproductivo que se ajustan tal como se ha definido en 4.

35 Con el fin de ensayar una muestra, se ejecuta un programa R que realiza lo siguiente.

- 40 1) Lee un conjunto de datos de ensayo.
- 40 2) Genera un promedio específico de la muestra de ambos controles.
- 40 3) Para cada muestra, utiliza el promedio específico de la muestra a restar de cada Marcador.
- 40 4) Reemplaza cualquier CT normalizado generado a partir de un CT sin procesar de 40 con 12.
- 40 5) Para cada muestra en el conjunto de ensayo se ensaya lo siguiente.
 - 45 a. Si el promedio de ambos controles es superior a 34, la muestra se etiqueta como 'CTR_FAILURE' con ceros para las probabilidades a posteriori.
 - 45 b. Se verifican los antecedentes para colon, ovario o pulmón. Si se encuentra una coincidencia, se marca también el género. A continuación se utilizan los antecedentes y un modelo específico de género para evaluar la muestra.
 - 50 c. Si se encuentra mama, páncreas, SCC de pulmón o próstata como etiqueta de antecedentes, a continuación se da a la muestra una etiqueta de 'FAILURE_ineligible_sample', y todas las probabilidades a posteriori se establecen en cero.
 - 50 d. Para todas las demás muestras se utiliza el modelo general para masculino o femenino.

55 Los resultados se formatean y se escriben en un archivo.

La presente descripción incluye carteras de expresión génica obtenidas mediante este proceso.

60 Los perfiles de expresión génica pueden presentarse de varias maneras. Lo más común es disponer una matriz de relación o las intensidades de fluorescencia sin procesar en un dendrograma en el que las columnas indican las muestras de ensayo y las filas indican los genes. Los datos se disponen de manera los genes con perfiles de expresión similares estén próximos entre sí. La relación de expresión para cada gen se visualiza en un color. Por ejemplo, una relación inferior a uno (disminución de la expresión) aparece en la porción azul del espectro, mientras que una relación superior a uno (aumento de la expresión) aparece en la porción roja del espectro. Se dispone en el mercado de programas de software para presentar este tipo de datos, incluido "GeneSpring" (Silicon Genetics, Inc.) y "Discovery" e "Infer" (Partek, Inc.)

Las mediciones de la abundancia de especies de ARN únicas se recogen de tumores primarios o de tumores metastásicos de primarios de origen conocido. Estas lecturas, junto con la historia clínica, incluida, pero no limitada a la edad del paciente, el sexo, el lugar de origen del tumor primario y el sitio de metástasis (en caso de ser aplicable) se utilizan para generar una base de datos relacional. La base de datos se utiliza para seleccionar los transcritos de ARN y los factores clínicos que pueden utilizarse como variables de Marcador para predecir el origen primario de un tumor metastásico.

En el caso de medir los niveles de proteína para determinar la expresión génica, cualquier método conocido en la técnica es adecuado siempre que dé como resultado una especificidad y sensibilidad adecuadas. Por ejemplo, pueden medirse los niveles de proteína mediante unión a un anticuerpo o fragmento de anticuerpo específico para la proteína y midiendo la cantidad de proteína unida al anticuerpo. Los anticuerpos pueden marcarse mediante reactivos radiactivos, fluorescentes u otros reactivos detectables para facilitar la detección. Los métodos de detección incluyen, sin limitación, el ensayo inmunoabsorbente ligado a enzimas (ELISA) y técnicas de inmunotransferencia.

Los genes modulados utilizados en los métodos de la invención se describen en los Ejemplos. Los genes que se expresan diferencialmente presentan un aumento o una disminución de la expresión en pacientes con carcinoma de un origen concreto con respecto a aquellos con carcinomas de diferentes orígenes. "Aumento de la expresión" y "disminución de la expresión" son expresiones relacionadas que se refieren a que se encuentra una diferencia detectable (más allá de la contribución del ruido en el sistema utilizado para medirlo) en la cantidad de expresión de los genes con respecto a una línea basal. En este caso, la línea basal se determina en base al algoritmo. Los genes de interés en las células enfermas presentan entonces un aumento o una disminución de la expresión con respecto al nivel de la línea basal utilizando el mismo método de medición. "Enfermo/a", en este contexto, se refiere a una alteración del estado de un cuerpo que interrumpe o altera, o que tiene el potencial de alterar, el rendimiento adecuado de las funciones corporales tal como ocurre con la proliferación incontrolada de células. Se da un diagnóstico de enfermedad cuando algún aspecto del genotipo o fenotipo de la persona concuerda con la presencia de la enfermedad. Sin embargo, el acto de llevar a cabo un diagnóstico o pronóstico puede incluir la determinación de cuestiones sobre la enfermedad/estado, tal como la determinación de la probabilidad de recaída, el tipo de tratamiento y el control del tratamiento. En el control del tratamiento, se emiten juicios clínicos sobre el efecto de un determinado ciclo terapéutico comparando la expresión de los genes a lo largo del tiempo para determinar si los perfiles de expresión génica han cambiado o están cambiando a patrones más conformes con el tejido normal.

Los genes pueden agruparse de manera que la información obtenida sobre el conjunto de genes en el grupo proporcione una base sólida para emitir un juicio clínicamente pertinente, tal como el diagnóstico, el pronóstico o la elección del tratamiento. Estos conjuntos de genes componen las carteras descritas en el presente documento. Al igual que con la mayoría de los Marcadores de diagnóstico, a menudo resulta deseable utilizar el menor número de Marcadores suficiente para emitir un juicio médico correcto. Esto evita un retraso en el tratamiento a la espera de un análisis adicional, así como un uso improductivo de tiempo y recursos.

Un método para establecer carteras de expresión génica es mediante algoritmos de optimización tal como el algoritmo de varianza media ampliamente utilizado en el establecimiento de carteras de acciones. Este método se describe detalladamente en el documento 20030194734. Esencialmente, el método requiere establecer un conjunto de entradas (acciones en las aplicaciones financieras, expresión medida mediante la intensidad en el presente documento) que optimizará el retorno (por ejemplo, la señal que se genera) que se recibe para utilizarlo minimizando al mismo tiempo la variabilidad del retorno. Se dispone de muchos programas comerciales de software para llevar a cabo tales operaciones. Resulta preferente "Wagner Associates Mean-Variance Optimization Application", denominada "Wagner Software" a lo largo de la presente memoria. Este software utiliza las funciones de la "Wagner Associates Mean-Variance Optimization Library" para determinar una frontera eficaz y resultan preferentes las carteras óptimas en el sentido de Markowitz. Markowitz (1952). El uso de este tipo de software requiere la transformación de los datos de micromatriz de manera que puedan tratarse como una entrada en la forma en que se utilizan la rentabilidad de las acciones y las medidas de riesgo cuando el software se utiliza para los fines previstos de análisis financiero.

El proceso de selección de una cartera también puede incluir la aplicación de reglas heurísticas. Preferentemente, tales reglas se formulan en base a la biología y la comprensión de la tecnología utilizada para producir resultados clínicos. Más preferentemente, se aplican a la salida del método de optimización. Por ejemplo, puede aplicarse el método de la varianza media de selección de la cartera a los datos de micromatriz para varios genes expresados diferencialmente en sujetos con cáncer. La salida del método sería un conjunto optimizado de genes que podrían incluir algunos genes que se expresan en sangre periférica así como en tejido enfermo. Si las muestras utilizadas en el método de ensayo se obtienen a partir de sangre periférica y determinados genes expresados diferencialmente en casos de cáncer también puedan expresarse diferencialmente en sangre periférica, entonces puede aplicarse una regla heurística en la que una cartera se selecciona de la frontera eficaz excluyendo las que se expresan diferencialmente en sangre periférica. Por supuesto, la regla puede aplicarse antes de la formación de la frontera eficaz, por ejemplo, aplicando la regla durante la preselección de datos.

Pueden aplicarse otras reglas heurísticas que no están necesariamente relacionadas con la biología en cuestión. Por ejemplo, puede aplicarse una regla de que sólo un porcentaje prescrito de la cartera pueda estar representado por un gen o grupo de genes concreto. El software disponible en el mercado tal como el Wagner Software se ajusta fácilmente a estos tipos de heurísticos. Esto puede ser útil, por ejemplo, cuando factores distintos de la exactitud y la precisión (por ejemplo, derechos de licencia previstos) tengan un impacto sobre la conveniencia de incluir uno o más genes.

Los perfiles de expresión génica descritos en el presente documento también pueden utilizarse junto con otros métodos de diagnóstico no genéticos útiles en el diagnóstico, el pronóstico o el control del tratamiento del cáncer. Por ejemplo, en algunas circunstancias es beneficioso combinar la potencia diagnóstica de los métodos basados en la expresión génica descritos anteriormente con datos procedentes de Marcadores convencionales tales como Marcadores de proteína sérica (por ejemplo, el Antígeno de Cáncer 27.29 ("CA 27.29")). Existe una variedad de tales Marcadores, incluidos analitos tales como CA 27.29. En uno de tales métodos, se extrae periódicamente sangre de un paciente tratado y a continuación se somete a un inmunoensayo enzimático para uno de los Marcadores séricos descritos anteriormente. Cuando la concentración del Marcador sugiere la reaparición de tumores o el fracaso del tratamiento, se obtiene una fuente de muestras susceptible de análisis de la expresión génica. Cuando exista una masa sospechosa, se hace una biopsia por aspiración con aguja fina (FNA) y a continuación se analizan los perfiles de expresión génica de las células obtenidas de la masa como se ha descrito anteriormente. Como alternativa, las muestras de tejido pueden obtenerse de las áreas adyacentes al tejido del cual se eliminó anteriormente un tumor. Este enfoque puede ser especialmente útil cuando otros ensayos producen resultados ambiguos.

Los kits preparados como se describe en el presente documento incluyen ensayos formateados para determinar los perfiles de expresión génica. Estos pueden incluir todos o algunos de los materiales necesarios para llevar a cabo los ensayos, tal como reactivos e instrucciones y un medio a través del cual se ensayan los Biomarcadores.

Los artículos descritos en el presente documento incluyen representaciones de los perfiles de expresión génica útiles para tratar, diagnosticar, pronosticar y evaluar de otro modo las enfermedades. Estas representaciones de perfiles se reducen a un medio que puede ser leído automáticamente por una máquina tal como medios legibles por ordenador (magnético, óptico, y similares). Los artículos también pueden incluir instrucciones para evaluar los perfiles de expresión génica en dichos medios. Por ejemplo, los artículos pueden comprender un CD ROM con instrucciones de ordenador para comparar los perfiles de expresión génica de las carteras de genes descritas anteriormente. Los artículos también pueden tener perfiles de expresión génica registrados digitalmente en los mismos de manera que puedan compararse con los datos de expresión génica de muestras de pacientes. Como alternativa, los perfiles pueden grabarse en un formato de representación diferente. Un registro gráfico es uno de tales formatos. Los algoritmos de agrupación como los incorporados en el software "DISCOVERY" e "INFER" de Partek, Inc. mencionados anteriormente pueden ayudar mejor a visualizar estos datos.

Los diferentes tipos de artículos de fabricación como se describen en el presente documento son ensayos formateados o medios utilizados para revelar los perfiles de expresión génica. Estos pueden comprender, por ejemplo, micromatrices en las que las sondas o complementos de secuencia están fijados a una matriz con la que se combinan las secuencias indicativas de los genes de interés creando un determinante legible de su presencia. Como alternativa, los artículos tal como se describen en el presente documento pueden fabricarse en forma de kits de reactivos para llevar a cabo la hibridación, la amplificación y la generación de señales indicativas del nivel de expresión de los genes de interés para la detección del cáncer.

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

Ejemplo 1

Materiales y Métodos

Descubrimiento genético de Marcadores de cáncer pancreático.

Se aisló ARN a partir de tejidos de tumor pancreático, de páncreas normal, de pulmón, de colon, de mama y de ovario utilizando Trizol. A continuación se utilizó el ARN para generar ARN marcado amplificado (Lipshutz *et al.* (1999)) que a continuación se hibridó en matrices Affymetrix U133A. A continuación se analizaron los datos de dos maneras.

En el primer método, se filtró este conjunto de datos para conservar sólo los genes con al menos dos lecturas actuales por todo el conjunto de datos. Este filtrado dejó 14.547 genes. Se determinó que 2.736 genes se sobreexpresaban en el cáncer de páncreas frente al páncreas normal con un valor de p inferior a 0,05. Cuarenta y cinco genes de los 2.736 también se sobreexpresaban al menos el doble en comparación con la intensidad máxima encontrada a partir de tejidos de pulmón y colon. Por último, se descubrieron seis conjuntos de sondas que se

sobreexpresaban al menos el doble en comparación con la intensidad máxima encontrada a partir de tejidos de pulmón, colon, mama y ovario.

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

10 *Muestras de tejido.*

Se obtuvieron un total de 260 tejidos primarios y metástasis FFPE de diversos proveedores comerciales. Las muestras ensayadas fueron: 30 metástasis de mama, 30 metástasis de colorrectal, 56 metástasis de pulmón, 49 metástasis de ovario, 43 metástasis de páncreas, 18 tumores 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 melanomas, 3 linfomas).

Extracción de ARN.

20 El aislamiento de ARN a partir de secciones de tejido en parafina se basó en los métodos y reactivos descritos en el manual High Pure RNA Paraffin Kit (Roche) con las siguientes modificaciones. Se seccionaron muestras de tejido incluidas en parafina según el tamaño de la metástasis incluida (2-5 mm = 9 X 10 µm, 6-8 mm = 6 X 10 µm, 8-≥ 10 mm = 3 X 10 µm), y se colocaron en tubos Eppendorf de 1,5 ml con ARNasa/ADNasa. Se desparafinaron las secciones mediante la incubación en 1 ml de xileno durante 2-5 minutos a temperatura ambiente después de una agitación en vórtex de 10-20 segundos. A continuación se centrifugaron los tubos y se eliminó el sobrenadante y se repitió la etapa de desparafinación. Después de eliminar el sobrenadante, se añadió 1 ml de etanol y se agitó en vórtex la muestra durante 1 minuto, se centrifugó y se eliminó el sobrenadante. Este proceso se repitió una vez más. Se eliminó el etanol residual y se secó el sedimento en un horno a 55°C durante 5-10 minutos y se resuspendió en 100 µl de tampón de lisis tisular, 16 µl de SDS al 10% y 80 µl de proteinasa K. Las muestras se agitaron en vórtex y se incubaron en un agitador-calentador puesto 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 a continuación se mezcló, se centrifugó y el sobrenadante se añadió a la columna de filtración. La columna de filtración junto con el tubo de recogida se centrifugó durante 1 minuto a 8.000 rpm y se descartó el flujo a través. Se realizó 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 la que se añadió cada solución a la columna, se centrifugó y se descartó el flujo a través. A continuación, se centrifugó la columna a velocidad máxima durante 2 minutos, se colocó en un tubo de 1,5 ml nuevo y se añadieron 90 µl de tampón de elución. Se obtuvo ARN después de una incubación de 1 minuto a temperatura ambiente, seguida de una centrifugación de 1 minuto a 8.000 rpm. La muestra se trató con ADNasa con adición de 10 µl de tampón de incubación ADNasa, 2 µl de ADNasa I y se incubó durante 30 minutos a 37°C. Se inactivó la ADNasa después de la adición de 20 µl de tampón de lisis tisular, 18 µl de SDS al 10% y 40 µl de proteinasa K. Nuevamente, se añadieron 325 µl de tampón de unión y 325 µl de etanol a cada muestra que a continuación se mezcló, se centrifugó y el sobrenadante se añadió a la columna de filtración. Los lavados secuenciales y la elución de ARN continuaron como se ha indicado anteriormente salvo que se utilizaron 50 µl de tampón de elución para eluir el ARN. Para eliminar la contaminación de la fibra de vidrio transferida desde la columna, se centrifugó el ARN durante 2 minutos a velocidad máxima y se eliminó el sobrenadante en un tubo Eppendorf de 1,5 ml. Las muestras se cuantificaron mediante lecturas de DO 260/280 obtenidas mediante un espectrofotómetro y se diluyeron las muestras hasta 50 ng/µl. El ARN aislado se almacenó en agua libre de ARNasa a -80°C hasta su uso.

50 *Cebador TaqMan y diseño de sondas.*

Se utilizaron los números de referencia de la secuencia de referencia de ARNm apropiada junto con Oligo 6.0 para desarrollar ensayos de CUP TaqMan® (Marcadores pulmonares: proteína B asociada a surfactante pulmonar humana (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 pancreáticos: antígeno de células madre de próstata (PSCA), factor de coagulación V (F5), Marcador prostático calicreína 3 (KLK3)) y los ensayos de mantenimiento β actina, hidroximetilbilano sintasa (PBGD). Los cebadores y las sondas de hidrólisis para cada ensayo se enumeran en la Tabla 2. Se excluyó la amplificación de ADN genómico diseñando ensayos alrededor de los sitios de corte y empalme de exón-intrón. Las sondas de hidrólisis se marcaron en el nucleótido 5' con FAM como colorante indicador y en el nucleótido 3' con BHQ1-TT como colorante de extinción interno.

Reacción en cadena de la polimerasa en tiempo real cuantitativa.

65 Se llevó a cabo la cuantificación de ARN específico de gen en una placa de 384 pocillos en el sistema de detección de secuencias ABI Prism 7900HT (Applied Biosystems). Para cada ciclo de agitador-calentador se amplificaron los patrones y las curvas patrón. Los patrones para cada Marcador consistieron en transcritos *in vitro* de

genes diana que se diluyeron en ARN portador de riñón de rata a 1×10^5 copias. Las curvas patrón para los Marcadores de mantenimiento consistieron en transcritos *in vitro* de genes diana que se sometieron a dilución en serie en ARN portador de riñón de rata a 1×10^7 , 1×10^5 y 1×10^3 copias. También se incluyeron en cada ciclo de ensayo controles no diana para asegurar que no hubiese contaminación ambiental. Todas las muestras y los controles se procesaron por duplicado. Se realizó la qRTPCR con reactivos de uso general en el laboratorio en una reacción de 10 μ l que contenía: tampón de RT-PCR (Bicina/KOH 50 nM pH 8,2, KAc 115 nM, glicerol al 8%, $MgCl_2$ 2,5 mM, $MnSO_4$ 3,5 mM, 0,5 mM de cada uno de dCTP, dATP, dGTP y dTTP), aditivos (Tris-Cl 2 mM pH 8, albúmina bovina 0,2 mM, trehalosa 150 mM, Tween 20 al 0,002%), mezcla de enzimas (2U de Tth (Roche), 0,4 mg/ μ l de Ac TP6-25), mezcla de sonda y cebador (0,2 μ M de sonda, 0,5 μ M de cebadores). Se siguieron los siguientes parámetros de ciclación: 1 ciclo a 95°C durante 1 minuto; 1 ciclo a 55°C durante 2 minutos; Rampa 5%; 1 ciclo a 70°C durante 2 minutos; y 40 ciclos de 95°C durante 15 segundos, 58°C durante 30 segundos. Una vez terminada la reacción de PCR, se establecieron los valores iniciales y umbral en el software ABI 7900HT Prism y los valores de Ct calculados se exportaron a Microsoft Excel.

15 Reacción de una sola etapa vs. de dos etapas.

Se llevó a cabo la síntesis de la primera hebra utilizando 100 ng de hexámeros aleatorios o cebadores específicos de gen por reacción. En la primera etapa, se calentaron 11,5 μ l de Mezcla-1 (cebadores y 1 μ g de ARN total) a 65°C durante 5 minutos y a continuación se enfriaron en hielo. Se añadieron a la Mezcla-1 8,5 μ l de Mezcla-2 (1x tampón, DTT 0,01 mM, 0,5 mM de cada uno de los dNTP, 0,25 U/ μ l de RNasin[®], 10 U/ μ l de Superscript III) y se incubaron a 50°C durante 60 minutos, seguido de 95°C durante 5 minutos. El ADNc se almacenó a -20°C hasta el momento de utilizarlo. La qRTPCR para la segunda etapa de la reacción en 2 etapas se realizó como se ha indicado 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. La qRTPCR para la reacción de una sola etapa se realizó exactamente como se ha indicado en el párrafo anterior. Las reacciones de una sola etapa y de dos etapas se llevaron a cabo en 100 ng de molde (ARN/ADNc). Una vez terminada la reacción de PCR, se establecieron los valores iniciales y umbral en el software ABI 7900HT Prism y los valores de Ct calculados se exportaron a Microsoft Excel.

30 Generación de un mapa de calor.

Para cada muestra, se calculó un ΔCt tomando el Ct medio de cada Marcador CUP y restando el Ct medio de un promedio de los Marcadores de mantenimiento (HK) ($\Delta Ct = Ct(\text{Marcador CUP}) - Ct(\text{Marcador HK med.})$). Se determinó para cada muestra el ΔCt mínimo para cada conjunto de Marcadores del tejido de origen (pulmón, mama, próstata, colon, ovario y páncreas). Se puntuó el tejido de origen con el ΔCt mínimo global y todos los demás tejidos de origen se puntuaron cero. Se ordenaron los datos según el diagnóstico patológico. Se llenó Partek Pro con los datos de viabilidad modificados y se generó un diagrama de intensidad.

Resultados.

40 Descubrimiento de Marcadores pancreáticos novedosos del estado del cáncer y del tumor de origen.

En primer lugar, se analizaron cinco Marcadores candidatos de páncreas: antígeno de células madre de próstata (PSCA), inhibidor de serin-proteínasa, clado A miembro 1 (SERPINA1), citoqueratina 7 (KRT7), metaloproteasa de matriz 11 (MMP11) y mucina 4 (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)) utilizando micromatrices de ADN y un panel de 13 adenocarcinomas ductales pancreáticos, cinco tejidos normales de páncreas y 98 muestras de tumores de mama, colorrectal, de pulmón y de ovario. Sólo el PSCA demostró sensibilidad moderada (se detectaron seis de trece o el 46% de los tumores pancreáticos) a una alta especificidad (91 de 98 o el 93% fueron identificados correctamente como no de origen pancreático) (figura 4A). Por el contrario, KRT7, SERPINA1, MMP11 y MUC4 demostraron una sensibilidad del 38%, 31%, 85% y 31%, respectivamente, a especificidades del 66%, 91%, 82% y 81%, respectivamente. Estos datos estaban en concordancia con la qRTPCR realizada en 27 metástasis de origen pancreático y 39 metástasis de origen no pancreático para todos los Marcadores excepto para MMP11 que presentó una sensibilidad y especificidad más escasas con qRTPCR y las metástasis. En conclusión, los datos de micromatriz en tejido primario congelado de forma instantánea, sirven como buen indicador de la capacidad del Marcador para identificar una metástasis FFPE como de origen pancreático utilizando qRTPCR aunque pueden resultar útiles Marcadores adicionales para un rendimiento óptimo.

Debido a que el adenocarcinoma ductal pancreático se desarrolla a partir de células del epitelio ductal que comprenden sólo un pequeño porcentaje de todas las células pancreáticas (comprendiendo las células acinares y las células de los islotes la mayoría) y a que los tejidos de adenocarcinoma pancreático contienen una cantidad significativa de tejido normal adyacente (Prasad *et al.* (2005); e Ishikawa *et al.* (2005)), ha resultado difícil identificar Marcadores de cáncer pancreático (es decir, con aumento de la expresión en cáncer) que también diferencien este órgano de los órganos. Para su uso en un panel CUP, es necesaria tal diferenciación. El primero método de consulta (véase Materiales y Métodos) devolvió seis conjuntos de sondas: el factor de coagulación V (F5), una proteína hipotética FLJ22041 similar a proteínas de unión a FK506 (FKBP10), integrina $\beta 6$ (ITGB6), transglutaminasa 2 (TGM2), ribonucleoproteína nuclear heterogénea A0 (HNRPO) y BAX delta (BAX). El segundo método de consulta

(véase Materiales y Métodos) devolvió ocho conjuntos de sondas: F5, TGM2, factor de transcripción 1 de homeodominio similar a paired (PITX1), ARNm isoforma trio (TRIO), ARNm para p73H (P73), una proteína desconocida para MGC: 10264 (SCD) y dos conjuntos de sondas para claudina18. F5 y TGM2 estaban presentes en ambos resultados de la consulta y, de los dos, F5 parecía el más prometedor (figura 4B).

5

Optimización de la preparación de muestras y qRT-PCR utilizando tejidos FFPE.

A continuación, se optimizaron los métodos de qRT-PCR y aislamiento de ARN utilizando tejidos fijados antes de examinar el rendimiento del panel de Marcadores. En primer lugar, se analizó el efecto de la reducción del tiempo de incubación con proteinasa K de dieciséis horas a 3 horas. No hubo ningún efecto sobre el rendimiento. Sin embargo, algunas muestras presentaron fragmentos más largos de ARN cuando se utilizó la etapa de proteinasa K más corta (figura 5). Por ejemplo, cuando se aisló el ARN a partir de un bloque de un año (C22), no se observó diferencia en los electroferogramas. Sin embargo, cuando se aisló el ARN a partir de un bloque de cinco años (C23), se observó una mayor fracción de ARN de alto peso molecular, según la evaluación de la joroba en el hombro, cuando se utilizó la digestión con proteinasa K más corta. Esta tendencia se mantuvo generalmente cuando se procesaron otras muestras, independientemente del órgano de origen para la metástasis FFPE. En conclusión, acortar el tiempo de digestión con proteinasa K no sacrifica los rendimientos de ARN y puede ayudar a aislar ARN menos degradado y más largo.

10

15

20

25

30

35

A continuación, se compararon tres métodos diferentes de transcripción inversa: transcripción inversa con hexámeros aleatorios seguida de qPCR (dos etapas), transcripción inversa con un cebador específico de gen seguida de qPCR (dos etapas) y qRT-PCR de una sola etapa utilizando cebadores específicos de gen. Se aisló ARN de once metástasis y se compararon los valores de Ct entre los tres métodos para la β -actina, la proteína B surfactante humana (HUMSPB) y el 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 con hexámeros aleatorios seguida de qPCR (reacción de dos etapas) dio los valores de Ct más altos mientras que la transcripción inversa con un cebador específico de gen seguida de qPCR (reacción de dos etapas) dio valores de Ct ligeramente más bajos (pero estadísticamente significativos) que la correspondiente reacción en una etapa. Sin embargo, la RT-PCR en dos etapas con cebadores específicos de gen tenía una etapa de transcripción inversa más larga. Cuando los valores de Ct de HUMSPB y TTF se normalizaron al valor de β -actina correspondiente para cada muestra, no hubo diferencias en los valores de Ct normalizados entre los tres métodos. En conclusión, la optimización de las condiciones de reacción de RT-PCR puede generar valores de Ct más bajos, lo que puede ayudar en el análisis de los bloques de parafina más viejos (Cronin *et al.* (2004)), y una reacción de RT-PCR de una sola etapa con cebadores específicos de gen puede generar valores de Ct comparables a los generados en la correspondiente reacción de dos etapas.

Rendimiento diagnóstico de un ensayo qRT-PCR de CUP.

A continuación, se realizaron 12 reacciones de qRT-PCR (10 Marcadores y dos genes de mantenimiento) en 239 metástasis FFPE. Los Marcadores utilizados para el ensayo se muestran en la Tabla 2. Los Marcadores de pulmón fueron proteína B asociada a surfactante humana (HUMSPB), factor de transcripción tiroideo 1 (TTF1) y desmogleína 3 (DSG3). El Marcador colorrectal fue cadherina 17 (CDH17). Los Marcadores de mama fueron mamaglobina (MG) y el factor de transcripción Ets derivado de próstata (PDEF). El Marcador de ovario fue tumor de Wilms 1 (WT1). Los Marcadores pancreáticos fueron el antígeno de células madre de próstata (PSCA) y el factor de coagulación V (F5), y el Marcador prostático fue calcitreína 3 (KLK3). Para las descripciones de los genes, véase la Tabla 31.

40

45

Tabla 2. Secuencias de cebador y sonda, números de registro y longitudes de amplicón

50

Diana	SEQ ID NO	Secuencia (5'-3')	Descripción	SEQ ID NO
SP-B	59	cacagccccgaccttgatga	Cebador directo	11
		ggtcccagagcccgtctca	Cebador inverso	12
		agctgtccagctgcaaaggaaaagcc	Sonda*	13
		cacagccccgaccttgatgagaactcagctgtccagctgcaaaggaaaagc caagtgagacgggctctgggacc	Amplicón	14

55

60

65

ES 2 494 843 T3

5	TTF1	60	ccaacccagacccgcg	Cebador directo	15
			cgcccatgccgctcatgtca	Cebador inverso	16
			cccgccatctcccgtctatg	Sonda*	17
			ccaacccagacccgcgctcccgcctctcccgtctatggcccggcgagc ggcatgaacatgagcggcatggcg	Amplicón	18
10	DSG3	61	gcagagaaggagaagataactcaa	Cebador directo	19
			actccagagattcggtaggtga	Cebador inverso	20
			attgccaagattactcagattacca	Sonda*	21
			gcagagaaggagaagataactcaaaaagaacccaattgccaagattactc agattaccaagcaaccagaaaatcacctaccgaatctctggagt	Amplicón	22
15	CDH17	62	tcctcggcagtggaagctta	Cebador directo	23
			tcctcaactctgtgctctgga	Cebador inverso	24
			ccaaaatcaatggctactatgcccgactg	Sonda*	25
			tcctcggcagtggaagcttaaaaaacgactgggaattccaaaatcaatggt actatgcccgactgctaccaggcacacagagttgagga	Amplicón	26
20	MG	63	agttgctgatggtcctcatgc	Cebador directo	27
			cacttggtgattgattgcttga	Cebador inverso	28
			ccctctccagcactgctacgca	Sonda*	28
			agttgctgatggtcctcatgctggcgccctctccagcactgctacgcaggctct ggctgcccttattggagaatggtattccaagacaatcaatccacaagtg	Amplicón	30
25	PDEF	64	cgccacctggacatctgga	Cebador directo	31
			cactggtcgaggcacagtatga	Cebador inverso	32
			gtcagcggcctggatgaaagagcgg	Sonda*	33
			cgccacctggacatctggaagctcagcggcctggatgaaagagcggacttca cctgggctcgattactactgtgcctcgaccagt	Amplicón	34
30	WT1	65	gcgagcccaatacagaatacac	Cebador directo	35
			cggggctactccaggcaca	Cebador inverso	36
			tcagaggcattcaggatgctgagc	Sonda*	37
			gcgagcccaatacagaatacacacgcacggtgtcttcagaggcattcagga tgtgagcagtgctgctggagtagccccg	Amplicón	38
35	PSCA	66	ctgttgatggcaggctggc	Cebador directo	39
			tgctcacctggccttggca	Cebador inverso	40
			gcagccaggcactgcccctgct	Sonda*	41
			ctgttgatggcaggctggcctgagccaggcactgcccctgctgctactcct gcaaagcccaggtagca	Amplicón	42
40	F5	67	tgaagaaatctctgggattattca	Cebador directo	43
			tatgtgtatctctggaatcatca	Cebador inverso	44
			aaaagggaaacagatattgaagactc	Sonda*	45
			tgaagaaatctctgggattattcagaattgtcaaaagggaaacagatattgaa gactctgatgatattccagaagataccacata	Amplicón	46
45	KLK3	68	ccccagtggtcctcaca	Cebador directo	47
			aggatgaaacaagctgcccga	Cebador inverso	48
			caggaacaaaagcgtgatctgctgg	Sonda*	49
			ccccagtggtcctcacaagctgcccactgcatcaggaacaaaagcgtgatct tgctgggtcggcacagctgtttcatcct	Amplicón	50
50	β actina	69	gccctgaggcactctcca	Cebador directo	51
			cggtgtccacgtcacacttca	Cebador inverso	52
			cttctctctggcagtgagtcctg	Sonda*	53
			gccctgaggcactctccagccttctctctgggcatggagtcctgtggcatccac gaaactacctcaactccatcatgaagtgtgacgtggacatccg	Amplicón	54
55	PBGD	70	ccacacacagcctactttccaa	Cebador directo	55
			taccacgcgaatcactctca	Cebador inverso	56
			aacggcaatgctgctgaacggcgga	Sonda*	57
			ccacacacagcctactttccaaagcggagccatgtctggaacggcaatgctggc tgcaacggcggaagaaaacaacccaaagatgagagtgattcgcgtgggta	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, la especificidad moderada de los Marcadores de colon, pulmón y ovario, y la especificidad algo menor de los pancreáticos. La combinación de los datos normalizados de qRTPCR con el refinamiento computacional mejora el rendimiento del panel de Marcadores. Los resultados se obtuvieron a partir de los datos normalizados de qRTPCR combinados con el algoritmo y se determinó la precisión del ensayo qRTPCR.

Análisis.

5 En este ejemplo, se utilizó la determinación de perfiles de expresión basada en micromatrices en tumores primarios para identificar los Marcadores candidatos para su uso con las metástasis. El hecho de que los tumores primarios pueden utilizarse para descubrir Marcadores del tumor de origen para las metástasis es coherente con varios hallazgos recientes. Por ejemplo, Weigelt y colaboradores han demostrado que los perfiles de expresión génica de los tumores de mama primarios se mantienen en las metástasis a distancia. Weigelt *et al.* (2003). Italiano y colaboradores descubrieron que el estado de EGFR, según la evaluación de IHC, era similar en los 80 tumores colorrectales primarios y las 80 metástasis relacionadas. Italiano *et al.* (2005). Sólo cinco de los 80 presentaron discordancia en el estado de EGFR. Italiano *et al.* (2005). Backus y colaboradores identificaron supuestos Marcadores para detectar metástasis de cáncer de mama utilizando un análisis de la expresión génica de todo el genoma de mama y otros tejidos, y demostraron que la mamaglobina y CK19 detectaban metástasis clínicamente tratables en los ganglios linfáticos centinela de la mama con un 90% de sensibilidad y un 94% de especificidad. Backus *et al.* (2005).

20 Los estudios basados en micromatrices con tejido primario confirmaron la especificidad y la sensibilidad de los Marcadores conocidos. Como resultado, a excepción de F5, todos los Marcadores utilizados tienen alta especificidad para los tejidos estudiados en el presente documento. 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); Khoo y *et al.* (1997) Un estudio reciente determinó que, utilizando IHC, el PSCA se sobreexpresa en las metástasis de cáncer de próstata. Lam *et al.* (2005) Dennis *et al.* (2002) también demostraron que podría utilizarse PSCA como Marcador del tumor de origen para páncreas y próstata. Como se muestra en el presente documento, se descubrió una fuerte expresión de PSCA en algunos tejidos de próstata a nivel del ARN pero, debido a la inclusión del PSA en el ensayo, ya pueden separarse los cánceres de próstata y páncreas. Un hallazgo novedoso de este estudio fue el uso de F5 como Marcador complementario (de PSCA) para tejido de origen pancreático. En el conjunto de datos de micromatriz con tejido primario y el conjunto de datos de qRTPCR con metástasis FFPE, F5 complementaba a PSCA (figura 4 y Tabla 3).

30 Tabla 3. Datos de viabilidad

	Mama	Colon	Pulmón	Otro	Ovario	Páncreas	Próstata	Total
35 Total ensayados	30	30	56	32	49	43	20	260
#Correcto	22	27	45	16	43	31	20	204
#Otro / Ningún ensayo	1	1	3	n/a	1	4	0	10
#Incorrecto	7	2	8	16	5	8	0	46
% ensayado	96,67	96,67	94,64	100	97,96	90,70	100	96,15
% correcto de ensayados	75,86	93,10	84,91	0	89,58	79,49	100	81,60
40 Correcto del total (%)	73,33	90,00	80,36	50,00	87,76	72,09	100	78,46

45 Investigadores anteriores han generado ensayos de CUP utilizando IHC o micromatrices. Su *et al.* (2001); Ramaswamy *et al.* (2001); y Bloom *et al.* (2004). Más recientemente, se ha acoplado SAGE a un pequeño panel de Marcadores qRTPCR. Dennis (2002); y Buckhaults *et al.* (2003). Este estudio es el primero en combinar perfiles de expresión en base a micromatrices con un pequeño panel 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 anteriormente mediante estudios SAGE. Algunos estudios han demostrado que existe una concordancia moderada entre los datos de perfiles basados en micromatrices de ADN y basados en SAGE y que la correlación mejora para los genes con niveles de expresión más altos. Van Ruissen *et al.* (2005); y Kim (2003). Por ejemplo, Dennis y colaboradores identificaron PSA, MG, PSCA y HUMSPB mientras que Buckhaults y colaboradores (Dennis *et al.* (2002)) identificaron PDEF. Resulta preferente ejecutar el ensayo de CUP utilizando qRTPCR porque es una tecnología robusta y puede tener ventajas de rendimiento frente a IHC. Al-Mulla *et al.* (2005); y Haas *et al.* (2005). Como se muestra en el presente documento, se mejoró el protocolo qRTPCR mediante el uso de cebadores específicos de gen en una reacción de una sola etapa. Esta es la primera demostración del uso de cebadores específicos de gen en una reacción de qRTPCR de una sola etapa con tejido FFPE. Otros investigadores han realizado un qRTPCR de dos etapas (síntesis de ADNC en una reacción seguida de qPCR) o han utilizado hexámeros aleatorios o cebadores específicos de gen truncados. Abrahamsen *et al.* (2003); Specht *et al.* (2001); Godfrey *et al.* (2000); Cronin *et al.* (2004); y Mikhitarian *et al.* (2004).

60 **Ejemplo 2**

Protocolo de aislamiento de ARN total de CUP FFPE

65 (Kit Highpure Cat #3270289)

Objetivo:

Aislamiento de ARN total a partir de tejido FFPE

5 Procedimiento:

Preparación de las soluciones de trabajo

10 1. Proteinasa K (PK) en kit
 Disolver liofilizado en 4,5 ml de tampón de elución. Alicuotar y almacenar a -20°C, estable durante 12 meses. PK-4x250mg (cat #3115852)
 Disolver liofilizado en 12,5 ml de tampón de elución (1x tampón TE (pH 7,4-7). Alicuotar y almacenar a -20°C.

15 2. Tampón de lavado I
 Añadir 60 ml de etanol absoluto al tampón de lavado I, almacenar a temperatura ambiente.

20 3. Tampón de lavado II
 Añadir 200 ml de etanol absoluto al tampón de lavado II, almacenar a temperatura ambiente.

25 4. ADNasa I
 Disolver liofilizado en 400 µl de tampón de elución. Alicuotar y almacenar a -20°C, estable durante 12 meses.

Corte de bloques de parafina ~ 30-45 minutos para 12 bloques (12 bloques x 2 tubos = 24 tubos)

30 Las secciones cortadas del bloque deben procesarse de inmediato para la extracción de ARN
 1. Utilizar una hoja de afeitar limpia y afilada en el micrótopo para cortar 6 secciones de 10 micras de espesor de los bloques de tejido recortado (tamaño 3-4 x 5-10mm).

35 Comentario: Bloque nuevo - Descartar secciones de cera hasta la sección de tejido obtenida. Bloque utilizado – Descartar las 3 primeras secciones de tejido.

2. Colocar inmediatamente el tejido cortado en tubos de microcentrífuga de 1,5 ml y tapar herméticamente para minimizar la humedad.
 3. Se recomienda tomar el número de secciones en función del tamaño del tumor que se muestra en la Tabla 4.

Tabla 4

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

Desparafinación ~ 30-45 minutos

50 1. Añadir a cada muestra 1,0 ml de xileno y agitar en vórtex enérgicamente durante 10-20 segundos e incubar a temperatura ambiente 2-5 minutos. Centrifugar 2 minutos a velocidad máxima. Eliminar el sobrenadante cuidadosamente.

55 Comentario: si el tejido parece flotar, centrifugar durante 2 minutos más.

2. Repetir la etapa 1.
 3. Centrifugar 2 minutos a velocidad máxima. Eliminar el sobrenadante.
 4. Añadir 1 ml de etanol abs. y agitar en vórtex enérgicamente 1 minuto. Centrifugar 2 minutos a velocidad máxima. Eliminar el sobrenadante.

60 5. Repetir la etapa 4.
 6. Secar el tubo brevemente sobre una toalla de papel para deshacerse de los residuos de etanol.
 7. Secar en horno el sedimento de tejido durante 5-10 minutos a 55°C.

Comentario: es muy importante eliminar completamente el etanol y secar bien los sedimentos, el etanol residual puede inhibir la digestión con PK.

65 Comentario: si la PK está a -20C, calentar a temperatura ambiente 20-30 minutos.

Extracción de ARN ~ 2,5-3 horas

1. Añadir a un sedimento de tejido 100 µl de tampón de lisis tisular, 16 µl de SDS al 10% y **80 µl** de solución de trabajo de proteinasa K, agitar en vórtex brevemente en varios intervalos e incubar **2 horas** a 55°C agitando a 400 rpm.
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 velocidad máxima durante 2 minutos.
4. Combinar el tubo de filtración y el tubo de recogida (12 tubos), y pipetear el sobrenadante de lisado en el filtro.
5. Centrifugar durante 30 segundos a 8.000 rpm y descartar el flujo a través.
- 10 *Comentario: puede repetirse la etapa 4-5, si necesita combinarse el ARN con 2 preparaciones más de sedimentos de tejido.*
6. Repetir la centrifugación a 8.000 rpm durante 30 segundos para secar el filtro.
7. Añadir a la columna 500 µl de solución de trabajo de tampón de lavado I y centrifugar durante 15-30 segundos a 8.000 rpm, descartar el flujo a través.
- 15 8. Añadir 500 µl de solución de trabajo de tampón de lavado II. Centrifugar durante 15-30 segundos a 8.000 rpm, descartar el flujo a través.
9. Añadir 300 µl de solución de trabajo de tampón de lavado II, centrifugar durante 15-30 segundos a 8.000 rpm, descartar el flujo a través.
- 20 10. Centrifugar el filtro High Pure durante 2 minutos a velocidad máxima.
11. Colocar el tubo de filtración High Pure en un nuevo tubo de 1,5 ml y añadir 90 µl de tampón de elución. Incubar durante 1-2 minutos a temperatura ambiente. Centrifugar 1 minuto a 8.000 rpm. **Tratamiento con ADNasa I ~ 1,5 horas.**
12. Añadir al eluato 10 µl de 10x tampón de incubación de ADNasa y 1,0 µl de solución de trabajo de ADNasa I y mezclar. Incubar durante 45 minutos a 37°C **(o 2,0 µl de ADNasa I durante 30 minutos).**
- 25 13. Añadir 20 µl de tampón de lisis tisular, 18 µl de SDS al 10% y 40 µl de solución de trabajo de proteinasa K. Agitar en vórtex brevemente. Incubar durante **30 minutos** (30-60 minutos) a 55°C.
14. Añadir 325 µl de tampón de unión y 325 µl de etanol abs. Mezclar y pipetear en un nuevo tubo de filtración High Pure con el tubo de recogida (12 tubos).
- 30 15. Centrifugar durante 30 segundos a 8.000 rpm y descartar el flujo a través.
16. Repetir la centrifugación a 8.000 rpm durante 30 segundos para secar el filtro.
17. Añadir 500 µl de solución de trabajo de tampón de lavado I a la columna. Centrifugar durante 15 segundos a 8.000 rpm, descartar el flujo a través.
18. Añadir 500 µl de solución de trabajo de tampón de lavado II. Centrifugar durante 15 segundos a 8.000 rpm, descartar el flujo a través.
- 35 19. Añadir 300 µl de solución de trabajo de tampón de lavado II. Centrifugar durante 15 segundos a 8.000 rpm, descartar el flujo a través.
20. Centrifugar el filtro High Pure durante 2 minutos a velocidad máxima.
21. Colocar el tubo de filtración High Pure en un nuevo tubo de 1,5 ml. Añadir 50 µl de tampón de elución; incubar durante 1-2 minutos a temperatura ambiente. Centrifugar durante 1 minuto a 8.000 rpm para recoger el ARN eluido.
- 40 22. Centrifugar el eluato durante 2 minutos a velocidad máxima y transferir el sobrenadante a un nuevo tubo sin agitar las fibras de vidrio del fondo.
23. Tomar una lectura de DO a 260/280 y diluir a 50 ng/µl. Almacenar a -80°C.
- Protocolo del Ensayo ASR de CUP (ABI 7900)**
- Objetivo:** Utilizar qRT-PCR para determinar el tejido de origen de una muestra de CUP
- 45 **Preparación del control:**
- 1. Controles positivos (Consultar la Tabla 5 y la Placa C en Preparación de Placas, figura 7)**

Tabla 5. Diluciones en serie de IVT – 5 µl 1x10⁸ en 470 µl de H₂O + 25 µl de 10.000 ARNr

Control IVT	CE/µl	Muestra	Agua	ARNr residual
BACTIN	100E+05	50	425	25
CDH17	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 - 5 µl 50.000 CE/µl + 495 µl de H₂O
Alicuotar 10 µl por tira de tubos (2 placas); Colocar la mezcla a -80°C hasta el momento de utilizarla.

2. Curvas patrón (Consúltense la Tabla 6 y la Placa C en Preparación de Placas, figura 7)

Etapla 1: La curva patrón se preparó exactamente como se muestra en la Tabla 6.

5

Tabla 7. Solución de reserva - 1×10^8 IVT.
Diluir 50.000 CE/ μ l de ARNr a 500 CE/ μ l – 5 μ l 50.000 CE/ μ l + 495 μ l de H₂O

10

Control IVT	CE/ μ l	Muestra	Agua	ARNr residual
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

15

20

Alicuotar 10 μ l por tira de tubos (2 placas); Colocar la mezcla a -80°C hasta el momento de utilizarla.

Mezcla de enzimas:

25

1. **Mezcla maestra:** Enzima (Tth) / Anticuerpo (TP6-25), véase la tabla 7.

Tabla 7

30

Reactivo	2x
Enzima Tth (5U/ μ l)	600,00
Anticuerpo: TP6-25 (1mg/ml)	600,00
Agua	300,00
Total	1.500,00

35

Alicuotar 500 μ l/tubo y congelar a -20°C.

Mezcla maestra de CUP:

40

1. **2,5X Mezcla maestra de CUP (Tablas 8-11):**

Tabla 8

45

ml	5x Aditivos	2,5x Conc.
0,50	1M Tris-Cl pH 8	5mM
1,25	40mg/ml de Albúmina, bovina	500 μ g/ml
37,50	Trehalosa de reserva 1M	375mM
2,5	Tween 20 al 20% v	0,50%
7,00	ddH ₂ O	
48,75		

50

Dejar que el reactivo se mezcle completamente > 15 minutos

55

Tabla 9

60

ml	5x Aditivos	2,5 x Conc
12,50	Bicina/hidróxido de potasio 1M pH 8,2	125mM
5,75	Acetato de potasio 5M	287,5mM
20,00	Glicerol (V x D = M -> 19,6 x 1,26 = 24,6 g)	20%
1,25	Cloruro de magnesio 500mM	6,25mM
1,75	Cloruro de magnesio 500mM	8,75mM
5,00	ddH ₂ O	
46,25		

65

ES 2 494 843 T3

Dejar que el reactivo se mezcle completamente > 15 minutos; Combinar las mezclas anteriores en un recipiente estéril - añadir lo siguiente

5

Tabla 10

ml	5x Aditivos	2,5x Conc.
1,25	dATP 100mM	1,25mM
1,25	dCTP 100mM	1,25mM
1,25	dTTP 100mM	1,25mM
1,25	dGTP 100mM	1,25mM
100,00		

10

Dejar que el reactivo se mezcle completamente > 15 minutos; alicuotar 1,8 ml/tubo y congelar a -20°C

15

Tabla 11

Cebador/sonda	Reserva (µM)	FC (µM)	µl
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			1.000,0

20

25

Mezcla de cebador y sonda:

Alicuotar 250 µl/tubo y congelar a -20°C

30

Mezcla de reacción:

1. Mezcla maestra de CUP (CMM): (Consúltense las tablas 12-14 y la Placa A en Preparación de Placas, figura 7)

35

Tabla 12

Reactivo	FC	X1 (10µl)	450
2,5 x Mezcla maestra de CUP	1X	4,00	1800
ROX	1x	0,20	90
2x mezcla TthAb	2U	1,00	450
Agua		2,3	1035
Total		7,50	3375

40

45

Preferentemente, cada ciclo/placa no tendrá más de 356 reacciones: 12 muestras con 12 Marcadores (288 reacciones con 2 repeticiones para cada una) + 10 controles de curva patrón por duplicado (20) + 2 controles positivos y 2 controles negativos para cada Marcador. (4x12 = 48)

Ajustar el agua para el volumen de muestra - 4,3 µl de muestra MAX; Mezclar bien

50

Tabla 13

Reactivo	FC	X1 (10µl)	34
Cebadores 10µM/Sonda 4µM	0,5µM/0,2µM	0,50	17
CMM	1x	7,50	255
Total		8,00	272

55

2. Marcadores ToO: Mezclar bien

60

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

65

3. Marcadores β -actina y PBGD: Mezclar bien

Preparación de la muestra:

5

Tabla 15

10

15

20

Muestra	ID de la muestra	Conc	Agua añadida = 50ng/ μ l
A1			
A2			
A3			
A4			
A5			
A6			
A7			
A8			
A9			
A10			
A11			
A12			

25

1. Muestras de CUP: 12 muestras en placas de 96 pocillos: A1-A12 (Consúltense la Tabla 16 y la Placa B en Preparación de Placas, figura 7);

Alicuotar 50 μ l de 50 ng/ μ l (2 μ l/rxn)

Cargar la placa:

30

1. Preparación de placa de 384 pocillos: (Consúltense la Placa de D en Preparación de Placas, figura 7)

Se cargan sobre la placa 2 μ l de muestra y 8 μ l de CMM. (muestra = 50 ng/ μ l)

Se cargan sobre la placa 4 μ l de muestra y 6 μ l de CMM (muestra = 25 ng/ μ l)

Se sella y se etiqueta la placa. Se centrifuga a 2.000 rpm durante 1 minuto. **Preparación de ABI 7900HT:** Colocar en el ABI 7900. Seleccionar el programa "CUP 384" y pulsar "inicio".

35

Tabla 16

40

45

Condiciones de termociclación
95C x 60s
55C x 2m
RAMPA 5 %
70C x 2m
40 ciclos de
95C x 15s
58C x 30s

50

ROX conectado

Los datos se analizan, se extraen los Ct y se insertan en el algoritmo

Ejemplo 3

Algoritmo CUP

55

Los valores Δ Ct normalizados de actina para HPT, MGB, PDEF, PSA, SP-B, TFF, DSG, WT1, PSCA y F5 se colocan en 6 conjuntos en base al tejido de origen del que se seleccionaron originalmente. Se restan las constantes 9,00, 11,00, 7,50, 5,00, 10,00, 9,50, 6,50, 8,00, 9,00 y 800 de cada Δ Ct respectivamente. A continuación, 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, y min (PSCA o F5)) como variable representativa para el grupo.

60

Estas variables, y el sitio de metástasis se utilizan para clasificar la muestra utilizando discriminantes lineales. Deben construirse dos modelos diferentes, uno para hombres y otro para mujeres, a partir de los datos de entrenamiento que utilizan la función de biblioteca MASS "lda" (Venables *et al.* (2002) en R (versión 2.0.1). A continuación se calcula una probabilidad a posteriori para cada ToO utilizando la función "predict" para el modelo masculino o femenino.

65

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

5

Ejemplo de Código R:

```

#Entrenar el modelo masculino
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(0,0.09,0.23,0.43,0,0.16,0.02)/sum(c(0, 0.09, 0.23, 0.43, 0, 0.16, 0.02)))

#Entrenar el modelo femenino
dat.f<-CUP2.MIN.NORM[, c
('HPT', 'MFB.PDEF', 'SP.B.TTF.DSG3', 'WT1', 'PSCA.F5', 'Class', 'background')]
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)))

#si la muestra desconocida (i) es de hombre
predict(CUP.Ida.m, CUP2.MIN.NORM.TEST[i,])

#si la muestra desconocida (i) es de mujer
predict(CUP.Ida.f, CUP2.MIN.NORM.TEST[i,])

```

10

15

20

25

Para ejecutar este código, una trama de datos llamada CUP2.MIN.NORM debe contener los datos de entrenamiento con el valor mínimo calculado para cada tejido de origen establecido como se ha descrito anteriormente.

30

"Class" (clase) se corresponde con el tejido de origen, y "background" (antecedentes) se corresponde con los sitios de metástasis descritos anteriormente.

35

Los datos de ensayo pueden estar contenidos en CUP2.MIN.NORM.TEST, y puede ensayarse una muestra específica en la fila i utilizando la función "predict". Una vez más, los datos de ensayo deben estar en el mismo formato que el conjunto de entrenamiento y tener los ajustes de valor mínimo aplicados a él también.

Ejemplo 4

Muestras resueltas de CUP

40

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

45

Tabla 17

Categoría de muestra	Muestra #	Correcto	Incorrecto	Ningún ensayo	% 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

50

Ejemplo 5

55

Límites del ensayo de CUP

La figura 8 representa los resultados obtenidos, utilizando los métodos descritos en los Ejemplos 1-3, para determinar los límites de los ensayos de CUP. Se puso a prueba el rendimiento del ensayo a lo largo de un intervalo de concentraciones de ARN y se descubrió que los ensayos de CUP son eficaces en el intervalo de 100 ng - 12,5 ng de ARN.

60

Ejemplo 6

Ensayo qRTPCR

65

Materiales y métodos. Muestras de tejido congeladas para análisis de micromatrices. Se utilizó un total de 700 tejidos humanos primarios congelados para los perfiles de micromatrices de expresión génica. Las muestras se obtuvieron de diversas instituciones académicas, como la Universidad de Washington (St. Louis, MO), el Centro Médico Erasmus (Rotterdam, Países Bajos) y empresas de banco de tejidos comerciales, incluidas Genomics Collaborative, Inc (Cambridge, MA), Asterand (Detroit, MI), Oncomatrix (La Jolla, CA) y Clinomics Biosciences (Pittsfield, MA). Para cada muestra, se recogió también información clínica, patológica y demográfica del paciente. Se revisaron las características histopatológicas de cada muestra para confirmar el diagnóstico, y para hacer una estimación de la conservación de la muestra y del contenido del tumor.

Extracción de ARN e hibridación con Affymetrix Gene Chip. Se diseccionaron muestras de cáncer congeladas con más de un 70% de células tumorales, benignas y muestras normales, y se homogeneizaron con un homogeneizador mecánico (UltraTurrex T8, Alemania) en reactivo Trizol (Invitrogen, Carlsbad, CA). Se homogeneizó el tejido en reactivo Trizol siguiendo el protocolo estándar de Trizol para el aislamiento de ARN a partir de tejidos congelados (Invitrogen, Carlsbad, CA). Después de la centrifugación se recogió la fase líquida superior y se hizo precipitar el ARN total con alcohol isopropílico a -20°C . Los sedimentos de ARN se lavaron con etanol al 75%, se resolvieron en agua y se almacenaron a -80°C hasta su uso.

Se examinó la calidad del ARN con un equipo Agilent 2100 Bioanalyzer RNA 6000 Nano Assay (Agilent Technologies, Palo Alto, CA). Se preparó el ARNc marcado y se hibridó con la matriz de oligonucleótidos de alta densidad Hu133A Gene Chip (Affymetrix, Santa Clara, CA), que contenía un total de 22.000 conjuntos de sondas, según el protocolo estándar del fabricante. Se exploraron las matrices utilizando los protocolos y escáneres de Affymetrix. Para su posterior análisis, cada conjunto de sonda se consideró un gen distinto. Los valores de expresión para cada gen se calcularon utilizando el software de análisis Affymetrix Gene Chip MAS 5.0. Todas las matrices cumplían tres estándares de control de calidad: el porcentaje de lectura "actual" para la matriz fue superior al 35%, el factor de escala fue inferior a 12 cuando se escalaba a una intensidad diana global de 600, y el nivel basal medio fue inferior a 150.

Selección de Marcadores candidatos. Para la selección de los Marcadores candidatos de tejido de origen (ToO) para tejidos de pulmón, colon, mama, ovario y próstata, se midieron los niveles de expresión de los conjuntos de sonda en las muestras de ARN incluyendo un total de 682 tejidos normales, benignos y cancerosos de mama, colon, pulmón, ovario, próstata. Los Marcadores candidatos específicos de tejido se seleccionaron en base al número de consultas estadísticas.

Con el fin de generar candidatos pancreáticos, se utilizaron perfiles de expresión génica de 13 muestras de adenocarcinoma ductal pancreático primario, 5 muestras normales de páncreas y 98 muestras de cáncer de pulmón, colon, mama y ovario para seleccionar los Marcadores de adenocarcinoma pancreático. Se realizaron dos consultas. En la primera consulta, se creó un conjunto de datos contenía 14.547 genes con al menos 2 lecturas "actuales" en muestras de páncreas. Se identificó un total de 2.736 genes que se sobreexpresaban en el cáncer de páncreas en comparación con la normalidad mediante la *t* de Student ($p < 0,05$). Se seleccionaron los genes cuya mínima expresión en el 11º percentil de cáncer de páncreas era al menos 2 veces superior al máximo en el cáncer de colon y de pulmón, generando 45 conjuntos de sondas. Como etapa final, se seleccionaron 6 genes con una expresión máxima al menos 2 veces superior a la máxima expresión en cánceres de colon, pulmón, mama y ovario. En una segunda consulta, se creó un conjunto de datos de 4.654 conjuntos de sondas con un máximo de 2 lecturas "actuales" en las muestras de mama, colon, pulmón y ovario. Se seleccionó un total de 160 genes con al menos 2 lecturas "actuales" en muestras normales y cancerosas de páncreas. De los 160 genes, se seleccionaron 10 genes después de comparar su nivel de expresión entre los tejidos normales y de páncreas. Se combinaron los resultados de ambas consultas de páncreas.

Además del análisis de los perfiles de expresión génica, se seleccionaron algunos Marcadores de la literatura. Se combinaron los resultados de todas las consultas para crear una lista corta de Marcadores candidatos ToO para cada tipo de tejido. Se hizo una estimación de la sensibilidad y la especificidad de cada Marcador. Los Marcadores que demostraron la mayor capacidad para diferenciar los tejidos por su origen se designaron para el ensayo de RT-PCR en base a la redundancia y complementariedad de los Marcadores.

Tejidos de CUP y de carcinoma metastásico FFPE de origen conocido. Se adquirió un total de 386 carcinomas metastásicos FFPE (estadio III-IV) de origen conocido y 24 adenocarcinomas primarios de próstata FFPE de diversos proveedores comerciales, incluidos Proteogenex (Los Angeles, CA), Genómica Collaborative, Inc. (Cambridge, MA), Asterand (Detroit, MI), Ardáis (Lexington, MA) y Oncomatrix (La Jolla, CA). Se obtuvo un conjunto independiente de 48 tejidos de CUP y de carcinoma metastásico de primario conocido, de Albany Medical College (Albany, NY). Para cada muestra, se recogió información clínica, patológica y demográfica del paciente. Se revisaron las características histopatológicas de cada muestra para confirmar el diagnóstico, y para hacer una estimación de la conservación de la muestra y del contenido del tumor. Para las muestras metastásicas, se establecieron de manera inequívoca los diagnósticos de carcinoma metastásico y ToO en base a la historia clínica del paciente y la evaluación histológica del carcinoma metastásico en comparación con los correspondientes primarios.

Aislamiento de ARN a partir de muestras FFPE. El aislamiento de ARN a partir de secciones de tejido en parafina fue tal como se describe en el manual del kit High Pure RNA Paraffin (Roche) con las siguientes modificaciones. Las muestras de tejido incluidas en parafina se seccionaron según el tamaño de la metástasis incluida (2-5 mm = 9 X 10 μ m, 6-8 mm = 6 X 10 μ m, 8- \geq 10 mm = 3 X 10 μ m). Las secciones se desparafinaron como se describe en el manual del kit, el sedimento de tejido se secó en un horno a 55°C durante 5-10 minutos y se resuspendió en 100 μ l de tampón de lisis tisular, 16 μ l de SDS al 10% y 80 μ l de proteinasa K. Las muestras se agitaron en vórtex y se incubaron en un agitador-calentador puesto a 400 rpm durante 2 horas a 55°C. El posterior procesamiento de la muestra se realizó según el manual del kit High Pure RNA Paraffin. Las muestras se cuantificaron mediante lecturas de OD 260/280 obtenidas mediante un espectrofotómetro y las muestras se diluyeron a 50 ng/ μ l. El ARN aislado se almacenó en agua libre de ARNasa a -80°C hasta su uso.

qRTPCR para el cribado previo de Marcadores candidatos. Se sometió a transcripción inversa con hexámeros aleatorios un μ g de ARN total de cada muestra, utilizando la transcriptasa inversa Superscript II según las instrucciones del fabricante (Invitrogen, Carlsbad, CA). Los cebadores y las sondas MGB para los Marcadores génicos candidatos ensayados y el gen de control ACTB se diseñaron utilizando el software Primer Express (Applied Biosystems, Foster City, CA), o bien se utilizó ABI Assay-on-Demand (Applied Biosystems, Foster City, CA). Se ensayaron todos los cebadores y sondas de diseño propio para una eficacia de amplificación óptima superior al 90%. La amplificación por RT-PCR se llevó a cabo en una mezcla de reacción de 20 μ l que contenía 200 ng de ADNc molde, 2 x de mezcla TaqMan[®] Universal PCR Master Mix (10 μ l) (Applied Biosystems, Foster City, CA), 500 nM de cebadores directo e inverso y 250 nM de sonda. Las reacciones se llevaron a cabo 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 50°C, 10 minutos de activación de la polimerasa a 95°C y 50 ciclos a 95°C durante 15 segundos y temperatura de hibridación (60°C) durante 60 segundos. En cada ensayo, se incluyó por duplicado el control "no molde" junto con ADNc molde para el gen de interés y para el gen de control. La expresión relativa de cada gen diana se representó como Δ Ct, que es igual al Ct del gen diana menos el Ct del gen de control (ACTB).

qRTPCR de una sola etapa optimizada. Se utilizaron los números de referencia de la secuencia de referencia del ARNm apropiada junto con Oligo 6.0 para desarrollar los ensayos de CUP TaqMan[®] (Marcadores de pulmón: proteína B asociada a surfactante pulmonar humana (HUMPSBPA), 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 pancreáticos: antígeno de células madre de próstata (PSCA), factor de coagulación V (F5), Marcador prostático: calicleína 3 (KLK3)) y ensayos de mantenimiento beta actina (β -actina), hidroximetilbilano sintasa (PBGD). Los cebadores específicos de gen y las sondas de hidrólisis para el ensayo de qRT-PCR de una sola etapa optimizado se enumeran en la Tabla 2 (SEQ ID NO: 11-58). Se excluyó la amplificación de ADN genómico diseñando los ensayos alrededor de los sitios de corte y empalme de exón-intrón. Se marcaron las sondas de hidrólisis en el nucleótido 5' con FAM como colorante indicador y en el nucleótido 3' con BHQ1-TT como colorante de extinción interno.

La cuantificación de ARN específico de 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 ciclo del agitador-calentador, se amplificaron los patrones y las curvas patrón. Los patrones para cada Marcador consistieron en transcritos *in vitro* de genes diana que se diluyeron en ARN portador de riñón de rata a 1×10^5 copias. Las curvas patrón para los Marcadores de mantenimiento consistieron en transcritos *in vitro* de genes diana que se sometieron a dilución en serie en ARN portador de riñón de rata a 1×10^7 , 1×10^5 y 1×10^3 copias. También se incluyeron en cada ciclo de ensayo controles no diana para asegurar que no hubiese contaminación ambiental. Todas las muestras y los controles se procesaron por duplicado. La qRTPCR se realizó con reactivos de uso general en el laboratorio en una reacción de 10 μ l que contenía: tampón de RT-PCR (Bicina/KOH 50 nM pH 8,2, KAc 115 nM, glicerol al 8%, MgCl₂ 2,5 mM, MnSO₄ 3,5 mM, 0,5 mM de cada uno de dCTP, dATP, dGTP y dTTP), aditivos (Tris-Cl 2 mM pH 8, albúmina bovina 0,2 mM, trehalosa 150 mM, Tween 20 al 0,002%), mezcla de enzimas (2U de Tth (Roche), 0,4 mg/ μ l de Ac TP6-25), mezcla de cebador y sonda (0,2 μ M de sonda, 0,5 μ M de cebadores). Se siguieron los siguientes parámetros de ciclación: 1 ciclo a 95°C durante 1 minuto; 1 ciclo a 55°C durante 2 minutos; rampa 5%; 1 ciclo a 70°C durante 2 minutos; y 40 ciclos de 95°C durante 15 segundos, 58°C durante 30 segundos. Una vez terminada la reacción de PCR, se establecieron los valores iniciales y umbral en el software ABI 7900HT Prism y los valores de Ct calculados se exportaron a Microsoft Excel.

Reacción de una sola etapa vs. de dos etapas. Para comparar las reacciones de RT-PCR de dos etapas y de una sola etapa, se llevó a cabo la síntesis de la primera hebra de la reacción de dos etapas utilizando 100 ng de hexámeros aleatorios o cebadores específicos de gen por reacción. En la primera etapa, se calentaron 11,5 μ l de Mezcla-1 (cebadores y 1 μ g de ARN total) a 65°C durante 5 minutos y a continuación se enfriaron en hielo. Se añadieron a la Mezcla-1, 8,5 μ l de Mezcla-2 (1x de tampón, DTT 0,01 mM, 0,5 mM de cada dNTP, 0,25 U/ μ l de RNasin[®], 10 U/ μ l de Superscript III) y se incubó a 50°C durante 60 minutos, seguido de 95°C durante 5 minutos. El ADNc se almacenó a -20°C hasta el momento de utilizarlo. La qRTPCR para la segunda etapa de la reacción de dos etapas 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. La qRTPCR para la reacción

de una sola etapa se realizó exactamente como se ha indicado en el párrafo anterior. La reacción de una sola etapa y la de dos etapas se llevaron a cabo en 100 ng de molde (ARN/ADNc). Una vez terminada la reacción de PCR, se establecieron los valores inicial y umbral en el software ABI 7900HT Prism y los valores de Ct calculados se exportaron a Microsoft Excel.

5

Desarrollo del algoritmo. Se construyeron discriminadores lineales utilizando la función de biblioteca MASS (Venables y Ripley) 'lda' en el lenguaje R (versión 2.1.1). El modelo utilizado depende 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, se establece en cero la clase distribución de probabilidad a priori para la clase que es equivalente al sitio de metástasis. Además, se establece en cero la probabilidad a priori para la clase mama y ovario en los pacientes masculinos, mientras que en pacientes femeninas, la distribución de probabilidad a priori de la clase próstata se establece en cero. Todas las demás probabilidades a priori utilizadas en los modelos son equivalentes. Además, la clasificación para cada muestra se basa en la probabilidad a posteriori más alta determinada por el modelo para cada clase. Para hacer una estimación del rendimiento de los modelos, se realizó la validación cruzada dejando uno fuera. Además de esto, los conjuntos de datos se dividieron al azar en mitades, preservando al mismo tiempo la relación proporcional entre cada clase, en conjuntos de entrenamiento y de ensayo. Esta división aleatoria se repitió tres veces.

10

15

20

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

25

Características de la muestra. Se utilizó ARN de un total de 700 muestras de tejido primario congelado para la determinación del perfil de expresión génica y la identificación de genes específicos del tipo de tejido. Las muestras incluyeron 545 carcinomas primarios (29 de pulmón, 13 de páncreas, 315 de mama, 128 de colorrectal, 38 de próstata, 22 de ovario), 37 lesiones benignas (1 de pulmón, 4 de colorrectal, 6 de mama, 26 de próstata) y 118 tejidos normales (36 de pulmón, 5 de páncreas, 36 de colorrectal, 14 de mama, 3 de próstata, 24 de ovario).

30

35

En el estudio se utilizó un total de 375 carcinomas metastásicos de origen conocido (estadio III-IV) y 26 muestras de adenocarcinoma primario de próstata. Los carcinomas metastásicos de origen pulmonar, pancreático, colorrectal, ovárico, prostático, así como otros tipos de cáncer. La categoría de muestra "otro" consistió en metástasis derivadas de tejidos distintos de pulmón, páncreas, colon, mama, ovario y próstata. Las características de los pacientes se resumen en la Tabla 18.

40

45

50

55

60

65

Tabla 18

		CUP metastatico	Conjunto de Muestras
5	Número total	401	48
	Promedio de edad	57,8±11*	62,13±11,7
	Genero		
	Femenino	241	20
	Masculino	160	28
10	Tejido de origen		
	Pulmón	65	9
	Páncreas	63	2
	Colorrectal	61	4
	Mama	63	5
15	Ovario	82	2
	Próstata	27	2
	Riñón	8	8
	Estómago	7	0
	Otro**	25	5
20	Carcinoma de primario desconocido		11
	Diagnóstico histopatológico		
	Adenocarcinoma, moderadamente/bien diferenciado	306	27
	Adenocarcinoma, poco diferenciado	49	4
	Carcinoma de células escamosas	16	5
25	Carcinoma poco diferenciado	16	10
	Carcinoma de células pequeñas	3	
	Melanoma	5	
	Linfoma	3	
	Carcinoma hepatocelular	2	
30	Mesotelioma	1	
	Otro***	14	2
	Sitio de metástasis		
	Nódulos linfáticos	73	1
	Cerebro	17	14
35	Pulmón	20	7
	Hígado	75	11
	Región pélvica (ovario, vejiga, trompas de falopio)	53	2
	Abdomen (epiplón (epiplón, mesenterio, colon, peritoneo)	91	5
40	Otro (piel, tiroides, pared torácica, zona umbilical)	44	8
	Desconocido	2	
	Primario (próstata)	26	
45	* Se desconoce la edad de 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 se separaron en dos conjuntos: el conjunto de validación (205 muestras), que se utilizó para validar la expresión diferencial específica de tejido de los Marcadores candidatos y el conjunto de entrenamiento (260 muestras), que se utilizó para ensayar el procedimiento qRTPCR de una sola etapa optimizado y entrenar un algoritmo de predicción. El primer conjunto de 205 muestras incluyó 25 metástasis de cáncer de pulmón, 41 de páncreas, 31 de colorrectal, 33 de mama, 33 de ovario, 1 de próstata, 23 de otras metástasis de cáncer y 18 cánceres primarios de próstata. El segundo conjunto consistió en 260 muestras que incluyeron 56 metástasis de cáncer de pulmón, 43 de páncreas, 30 de colorrectal, 30 de mama, 49 de ovario, 32 de otras metástasis de cáncer y 20 cánceres primarios de próstata. Sesenta y cuatro muestras, incluidas 16 metástasis de pulmón, 21 de páncreas, 15 de otras metástasis, y 12 carcinomas primarios de próstata fueron del mismo paciente en ambos conjuntos.

El conjunto independiente de muestras obtenido de Albany Medical College estaba compuesto por 33 muestras de CUP con un primario sugerido para 22 de ellos, y 15 carcinomas metastásicos de origen conocido. Para los CUP con un primario sugerido, el diagnóstico se hizo en base a las características morfológicas y/o a los resultados de ensayo con un panel de Marcadores IHC. Las características demográficas, clínicas y patológicas de los pacientes se presentan en la Tabla 18.

Selección de Marcadores candidatos. El análisis de los perfiles de expresión génica de 5 tipos de tejidos primarios (pulmón, colon, mama, ovario, próstata) dio como resultado la designación de los 13 Marcadores candidatos específicos de tejido para el ensayo qRTPCR. Se han identificado los mejores candidatos en estudios

anteriores 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 Khor et al. (1997). Además del análisis de los datos de micromatriz, se seleccionaron dos Marcadores de la literatura, incluido un Marcador DSG3 complementario de carcinoma de células escamosas de pulmón y el Marcador de mama PDEF. Backus *et al.* (2005). Los datos de micromatriz confirmaron la elevada sensibilidad y especificidad de estos Marcadores.

Se utilizó un enfoque especial para identificar los Marcadores específicos de páncreas. En primer lugar, se analizaron cinco Marcadores candidatos de páncreas: antígeno de células madre de próstata (PSCA), inhibidor de serin-proteinasa, clado A miembro 1 (SERPINA1), citoqueratina 7 (KRT7), metaloproteasa 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)) utilizando micromatrices de ADN y un panel de 13 adenocarcinomas ductales pancreáticos, cinco tejidos normales de páncreas y 98 muestras de tumores de mama, colorrectal, de pulmón y de ovario. Sólo PSCA demostró sensibilidad moderada (se detectaron seis de trece o el 46% de los tumores pancreáticos) con elevada especificidad (91 de 98 o el 93% se identificaron correctamente como no de origen pancreático). Por el contrario, KRT7, SERPINA1, MMP11 y MUC4 demostraron una sensibilidad del 38%, 31%, 85% y 31%, respectivamente, con especificidades del 66%, 91%, 82% y 81%, respectivamente. Estos datos estaban en concordancia con la qRT-PCR realizada en 27 metástasis de origen pancreático y 39 metástasis de origen no pancreático para todos los Marcadores excepto para MMP11 que presentó la sensibilidad y especificidad más escasas con qRT-PCR y las metástasis. En conclusión, los datos de micromatriz en tejido primario congelado de forma instantánea, sirve como buen indicador de la capacidad del Marcador para identificar una metástasis FFPE como de origen pancreático utilizando qRT-PCR, aunque pueden resultar útiles Marcadores adicionales para un rendimiento óptimo.

El adenocarcinoma ductal pancreático se desarrolla a partir de células del epitelio ductal que comprenden sólo un pequeño porcentaje de todas las células pancreáticas (comprendiendo las células acinares y de los islotes la mayoría) en el páncreas normal. Además, los tejidos de adenocarcinoma pancreático contienen una cantidad significativa de tejido normal adyacente. Prasad *et al.* (2005); e Ishikawa *et al.* (2005). Debido a esto, los Marcadores pancreáticos candidatos se enriquecieron para los genes elevados en el adenocarcinoma de páncreas con respecto a las células pancreáticas normales. El primer método de consulta devolvió seis conjuntos de sondas: factor de coagulación V (F5), una proteína hipotética FLJ22041 similar a proteínas de unión a FK506 (FKBP10), integrina beta 6 (ITGB6), transglutaminasa 2 (TGM2), ribonucleoproteína nuclear heterogénea A0 (HNRPO) y BAX delta (BAX). El segundo método de consulta (véase la sección de Materiales y Métodos para más detalles) devolvieron ocho conjuntos de sondas: F5, TGM2, factor de transcripción 1 de homeodominio similar a paired (PITX1), ARNm isoforma trio (TRIO), ARNm para p73H (p73), una proteína desconocida para MGC:10264 (SCD) y dos conjuntos de sondas para claudina18.

Se seleccionó un total de 23 Marcadores candidatos específicos de tejido para su posterior validación mediante RT-PCR en tejidos FFPE de carcinoma metastásico mediante qRT-PCR. Los Marcadores candidatos se ensayaron en 205 carcinomas metastásicos FFPE, de pulmón, páncreas, colon, mama, ovario, próstata y carcinomas primarios de próstata. La Tabla 19 proporciona los símbolos de genes de los Marcadores específicos de tejido seleccionados para la validación por RT-PCR y también resume los resultados de los ensayos realizados con estos Marcadores.

Tabla 19

5	Tipo de tejido	SEQ ID NO	Método ID		Filtros de selección de Marcadores			¿Marcador adecuado?
			Micromatriz	Lit.	Tejido met. corres. baja expr.	Redundancia de Marcador	Reactividad cruzada de tejido	
	Pulmón	1/59	X	X				X
		60	x	X				X
10		61		X		X		X
	Páncreas	66		X				X
		67	X					X
		71	X			X		
15		72	X		X			
		73		X				
		74		X				
		75		X				
		76		X				
20	Colon	4/85	X	X				X
		77	X	X				
		78	X	X		X		
		79	X	X		X		
25	Próstata	9/86	X	X				X
		80	X	X		X		
	Mama	63	X	X				X
		81	X	X			X	
		64		X				X
30	Ovario	82	X	X			X	
		83	X	X			X	
		65	X	X				X

35 De 23 Marcadores ensayados, se rechazaron trece en base a su reactividad cruzada, bajo nivel de expresión en los tejidos metastásicos correspondientes o redundancia. Se seleccionaron diez Marcadores para la versión final del ensayo. Los Marcadores pulmonares fueron proteína B asociada a surfactante pulmonar humana (HUMSPB), factor de transcripción tiroideo 1 (TTF1) y desmogleína 3 (DSG3). Los Marcadores pancreáticos fueron el antígeno de células madre de próstata (PSCA) y el factor de coagulación V (F5), y el Marcador prostático fue calicreína 3 (KLK3). El Marcador colorrectal fue cadherina 17 (CDH17). Los Marcadores de mama fueron mamaglobina (MG) y factor de transcripción Ets derivado de próstata (PDEF). El Marcador de ovario fue tumor de Wilms 1 (WT1). En la figura 10 se presentan los valores de expresión relativos normalizados medios de los Marcadores seleccionados en diferentes tejidos metastásicos.

45 *Optimización de la preparación de muestras y qRT-PCR utilizando tejidos FFPE.* A continuación, se optimizaron los métodos de qRT-PCR y aislamiento de ARN utilizando tejidos fijados antes de examinar el rendimiento del panel de Marcadores. En primer lugar, se analizó el efecto de la reducción del tiempo de incubación con proteinasa K de dieciséis horas a 3 horas. No hubo ningún efecto sobre el rendimiento. Sin embargo, algunas muestras presentaron fragmentos más largos de ARN cuando se utilizó la etapa de proteinasa K más corta (figura 11A, B). Por ejemplo, cuando el ARN se aisló a partir de un bloque de un año (C22), no se observó ninguna diferencia en los electroferogramas. Sin embargo, cuando el ARN se aisló a partir de un bloque de cinco años (C23), se observó una fracción mayor de ARN de mayor peso molecular, según lo evaluado por la joroba en el hombro, cuando se utilizó la digestión con proteinasa K más corta. Esta tendencia se mantuvo generalmente cuando se procesaron otras muestras, independientemente del órgano de origen para la metástasis FFPE. En conclusión, acortar el tiempo de digestión con proteinasa K no sacrifica el rendimiento de ARN y puede ayudar a aislar ARN menos degradado y más largo.

60 A continuación, se compararon tres métodos diferentes de transcripción inversa: transcripción inversa con hexámeros aleatorios seguida de qPCR (dos etapas), transcripción inversa con un cebador específico de gen seguida de qPCR (dos etapas) y qRT-PCR de una sola etapa utilizando cebadores específicos de gen. Se aisló ARN de once metástasis y se compararon los valores de Ct entre los tres métodos para β -actina, HUMSPB (figura 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 de qPCR (reacción de dos etapas) dio los valores más altos de Ct mientras que la transcripción inversa con un cebador específico de gen seguida de qPCR (reacción de dos etapas) dio valores de Ct ligeramente más bajos (pero estadísticamente

significativos) que la reacción de 1 etapa correspondiente. Sin embargo, la RT-PCR de dos etapas con cebadores específicos de gen tenía una etapa de transcripción inversa más larga. Cuando los valores de Ct de HUMSPB se normalizaron al valor de β -actina correspondiente para cada muestra, no hubo diferencias en los valores de Ct normalizados entre los tres métodos. En conclusión, la optimización de las condiciones de reacción de RT-PCR puede generar valores de Ct más bajos, lo que ayuda a analizar los bloques de parafina más viejos (Cronin *et al.* (2004)), y una reacción de RTPCR de una sola etapa con cebadores específicos de gen puede generar valores de Ct comparables a los generados en la reacción de dos etapas correspondiente.

Rendimiento diagnóstico del ensayo qRTPCR optimizado. Se realizaron 12 reacciones de qRTPCR (10 Marcadores y 2 genes de mantenimiento) en un nuevo conjunto de 260 metástasis FFPE. Veintiún muestras dieron valores de Ct altos para los genes de mantenimiento por lo que sólo se utilizaron 239 en un análisis de mapa de calor. El análisis de los valores de Ct normalizados en un mapa de calor reveló la alta especificidad de los Marcadores de mama y próstata, una especificidad moderada de los de colon, pulmón y ovario, y una especificidad algo más baja de los Marcadores pancreáticos (figura 12). La combinación de los datos de qRTPCR normalizados con refinamiento computacional mejora el rendimiento del panel de Marcadores.

Utilizando los valores de expresión, normalizados a la media de la expresión de dos genes de mantenimiento, se desarrolló un algoritmo para predecir el tejido metastásico de origen combinando los datos de qRTPCR normalizados con el algoritmo y se determinó la precisión del ensayo qRTPCR realizando una validación cruzada dejando uno fuera (LOOCV). Para los seis tipos de tejidos incluidos en el ensayo, se estimó por separado que tanto el número de lecturas falsos positivos, cuando una muestra se predijo erróneamente como otro tipo de tumor incluido en el ensayo (páncreas como de colon, por ejemplo), como el número de veces que una muestra no se predijo como aquellos incluidos en los tipos de tejidos de ensayo (otros). Los resultados de la 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	
#Correcto	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	

El tejido de origen se predijo correctamente para 204 de 260 muestras analizadas, con una precisión global del 78%. Una proporción significativa de las lecturas falsos positivos se debieron a la reactividad cruzada de los Marcadores en tejidos histológicamente similares. Por ejemplo, tres carcinomas metastásicos de células escamosas de origen faríngeo, laríngeo y esofágico se predijeron erróneamente como pulmonar debido a la expresión de DSG3 en estos tejidos. La expresión positiva de CDH17 en carcinomas distintos del GI de colon, incluidos estómago y páncreas, hicieron que se clasificaran erróneamente como de colon 4 de 6 metástasis de cáncer de estómago ensayadas y 3 de 43 metástasis de cáncer de páncreas ensayadas.

Además de un ensayo LOOCV, los datos se dividieron aleatoriamente en 3 pares de conjuntos de ensayo y de entrenamiento separados. Cada división contenía aproximadamente un 50% de las muestras de cada clase. A divisiones 50/50 en tres pares de conjuntos de ensayo y de entrenamiento separados, las precisiones de clasificación globales del ensayo fueron 77%, 71% y 75%, lo que confirma la estabilidad de rendimiento del ensayo.

Por último, se ensayó otro conjunto independiente de 48 carcinomas metastásicos FFPE que incluía carcinoma metastásico de origen primario conocido, muestras de CUP con un diagnóstico de tejido de origen emitido por la evaluación patológica incluida IHC y muestras de CUP que siguieron siendo CUP después del ensayo IHC. Se hizo una estimación por separado de la precisión de la predicción del tejido de origen para cada categoría de muestras. La Tabla 21 resume los resultados de ensayo.

Tabla 21

	Ensayado	Correcto	Precisión
Met. conocida	15	11	73,3
CUP resuelto	22	17	77,3
CUP no resuelto	11		

La predicción del tejido de origen concordaba, con sólo unas pocas excepciones, con el diagnóstico del tejido de origen o primario conocido evaluado mediante evaluación clínica/patológica incluida IHC. Al igual que en el conjunto de entrenamiento, el ensayo no fue capaz de diferenciar los carcinomas de células escamosas procedentes de diferentes fuentes y se predijeron falsamente como de pulmón.

El ensayo también realizó diagnósticos de supuesto tejido de origen para ocho de once muestras que siguieron siendo CUP después de ensayos de diagnóstico convencionales. Uno de los casos CUP fue especialmente interesante. Se diagnosticó carcinoma metastásico en pulmón y pleura a un paciente masculino con antecedentes de cáncer de próstata. Los ensayos de PSA en suero e IHC con anticuerpos para PSA en el tejido metastásico dieron negativo, por lo que el diagnóstico del patólogo fue de CUP con una inclinación hacia los tumores gastrointestinales. El ensayo predijo claramente (probabilidad a posteriori del 0,99) que el tejido de origen era colon.

Análisis. En este estudio, se utilizó la determinación de perfiles de expresión en base a micromatrices de tumores primarios para identificar Marcadores candidatos para su uso con metástasis. El hecho de que los tumores primarios pueden utilizarse para descubrir Marcadores del tumor de origen para las metástasis concuerda con varios hallazgos recientes. Por ejemplo, Weigelt y colaboradores han demostrado que los perfiles de expresión génica de tumores primarios de mama se mantienen en las metástasis distantes. Weigelt *et al.* (2003). Backus y colaboradores identificaron supuestos Marcadores para detectar metástasis de cáncer de mama utilizando un análisis de expresión génica de todo el genoma de mama y otros tejidos y demostraron que la mamaglobina y CK19 detectaban metástasis clínicamente tratables en los ganglios linfáticos centinela de mama con un 90% de sensibilidad y un 94% de especificidad. Backus *et al.* (2005).

Durante el desarrollo del ensayo, la selección se centró en seis tipos de cáncer, incluido pulmón, páncreas y colon que se encuentran entre los más prevalentes en el CUP (Ghosh *et al.* (2005); y Pavlidis *et al.* 2005)) y mama, ovario y próstata para los cuales el tratamiento podría ser potencialmente más beneficioso para los pacientes. Ghosh *et al.* (2005). Sin embargo, pueden añadirse al panel Marcadores y tipos de tejido adicionales, siempre y cuando no se comprometa la precisión global del ensayo y, en su caso, no se afecte a la logística de las reacciones de RT-PCR.

Los estudios basados en micromatrices con tejido primario confirmaron la especificidad y la sensibilidad de los Marcadores conocidos. Como resultado, la mayoría de los Marcadores específicos de tejido tiene una alta especificidad para los tejidos estudiados en el presente documento. Un estudio reciente descubrió que, mediante IHC, PSCA se sobreexpresa en las metástasis de cáncer de próstata. Lam *et al.* (2005). Dennis *et al.* (2002) también demostraron que podría utilizarse PSCA como Marcador del tumor de origen para páncreas y próstata. Hubo una fuerte expresión de PSCA en algunos tejidos de próstata a nivel del ARN pero, debido a la inclusión del PSA en el ensayo, ya pueden separarse los cánceres de próstata y de páncreas. Un nuevo hallazgo de este estudio fue el uso de F5 como Marcador complementario (para PSCA) para tejido de origen pancreático. Tanto en el conjunto de datos de micromatriz con el tejido primario como en los datos de qRTPCR establecidos con metástasis FFPE, se descubrió que F5 complementaba a PSCA.

Investigadores anteriores han generado ensayos de CUP mediante 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, se ha acoplado SAGE a un pequeño panel de Marcadores qRTPCR. Dennis *et al.* (2002); y Buckhaults *et al.* (2003). Este estudio es el primero en combinar la determinación de perfiles de expresión basada en micromatrices con un pequeño panel 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 anteriormente mediante estudios SAGE. Este hallazgo no resulta sorprendente dados los estudios que han demostrado que existe una concordancia moderada entre los datos de perfiles basados en micromatrices de ADN y basados en 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 colaboradores identificaron PSA, MG, PSCA y HUMSPB mientras que Buckhaults y colaboradores (Buckhaults *et al.* (2003)) identificaron PDEF.

La ejecución del ensayo de CUP se realiza preferentemente mediante qRTPCR debido a que es una tecnología robusta y puede tener ventajas de rendimiento frente a IHC. Al-Mulla *et al.* (2005); y Haas *et al.* (2005). Además, como se muestra en el presente documento, se ha mejorado el protocolo qRTPCR mediante el uso de cebadores específicos de gen en una reacción de una sola etapa. Esta es la primera demostración de la utilización de cebadores específicos de gen en una reacción qRTPCR de una sola etapa con tejido FFPE. Otros investigadores han realizado una qRTPCR de dos etapas (síntesis de ADNc en una reacción seguida de qPCR) o han utilizado

hexámeros aleatorios o cebadores específicos de gen truncados. Abrahamsen *et al.* (2003); Specht *et al.* (2001); Godfrey *et al.* (2000); Cronin *et al.* (2004); y Mikhitarian *et al.* (2004).

5 En resumen, la precisión global del 78% del ensayo para seis tipos de tejido supera a 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

10 En este estudio se construyeron carteras de marcadores génicos que utilizaban un clasificador, eligiendo de MVO y utilizando este clasificador para predecir el origen del tejido y el estado del cáncer para cinco tipos principales de cáncer incluidos de mama, colon, pulmón, ovario y próstata. Se analizaron 103 muestra de tejido humano normal congelado de forma instantánea, 23 de lesiones epiteliales proliferativas benignas y trescientas setenta y ocho de cáncer primario, mediante Affymetrix Human U133A GeneChip. También se analizaron muestras de leucocitos con el fin de restar la expresión génica potencialmente enmascarada por la co-expresión en células leucocitarias de fondo. Se desarrolló un nuevo método bioinformático basado en MVO para seleccionar carteras de marcadores génicos para los tejidos de origen y el estado del cáncer. Los datos demostraron que podría utilizarse un panel de 26 genes como clasificador para predecir con precisión el tejido de origen y el estado del cáncer de entre los 5 tipos de cáncer. Por lo tanto, puede obtenerse un método de clasificación de múltiples cánceres determinando los perfiles de expresión génica de un número razonablemente pequeño de marcadores génicos.

20 La Tabla 22 muestra los Marcadores identificados para los orígenes de tejido indicados. Para las descripciones de los genes, véase la Tabla 31.

25

Tabla 22

30

35

40

45

50

55

60

Tejido	SEQ ID NO:	Nombre
Pulmón	59	SP-B
	60	TTF1
	61	DSG3
Páncreas	66	PSCA
	67	F5
	71	ITGB6
	72	TGM2
	84	HNRPA0
Colon	85	HPT1
	77	FABP1
	78	CDX1
	79	GUCY2C
Próstata	86	PSA
	80	hKLK2
	63	MGB1
Mama	81	PIP
	64	PDEF
	82	HE4
Ovario	83	PAX8
	65	WT1

65

El conjunto de muestras incluía un total de 299 carcinomas metastásicos de colon, mama, páncreas, ovario, próstata, pulmón y otros carcinomas y muestras de cáncer primario de próstata. Se implementó una QC basada en la evaluación histológica, el rendimiento de ARN y la expresión del gen de control beta-actina. Otras categorías de muestras incluían metástasis originadas a partir de carcinomas de estómago (5), de riñón (6) colangiocarcinoma/de

ES 2 494 843 T3

vesícula biliar (4), de hígado (2), de cabeza y cuello (4), de íleon (1) y un mesotelioma. Los resultados se resumen en la Tabla 23.

Tabla 23

Tipo de tejido	Recogido	QC histología	QC aislamiento de ARN	QC pto. corte ACTB
Pulmón	41	37	36	25
Páncreas	63	57	49	41
Colon	45	42	42	31
Mama	40	35	35	34
Ovario	37	36	35	33
Próstata	27	27	25	19
Otro	46	34	29	23
Total	299	268	251	205

El ensayo de las muestras anteriormente indicadas dio como resultado la reducción del conjunto de Marcadores a los de la Tabla 24, con los resultados que se observan en la Tabla 25.

Tabla 24

Tabla final de Marcadores		
Pulmón	proteína asociada a surfactante	SP-B
	factor de transcripción tiroideo 1	TTF1
	desmogleína 3	DSG3
Páncreas	antígeno de células madre de próstata	PSCA
	factor de coagulación 5	F5
Colon	transportador asociado a péptido intestinal	HPT1
Próstata	antígeno específico de próstata	PSA
Mama	mamaglobina	MGB
	Factor de transcripción Ets	PDEF
Ovario	Tumor de Wilms	WT1 1

Tabla 25

Cáncer	Muestras #	Macador	Correcto	Sensi. %	Incorrecto	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	16/33	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 demostraron que de 205 tumores metastásicos incluidos en parafina; 166 muestras (81%) presentaron resultados de ensayo concluyentes, Tabla 26.

Tabla 26

	Candidato	Correcto	Incorrecto	Sin	Precisión (%)	
5	Pulmón	SP-B + TFF+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
10	Ovario	WT1	23	2	8	70
	Otro		20	3		87
	Global		155	11	39	76

15 De los resultados falsos positivos, muchos falsos provenían de tejidos histológicamente y embriológicamente similares, Tabla 27.

Tabla 27

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

35 Se consideraron los siguientes parámetros para el desarrollo de modelos:

40 Separar los Marcadores en conjuntos femeninos y masculinos y calcular la probabilidad de CUP por separado para los 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. Se excluyó la expresión de fondo de los resultados de ensayo: Pulmón: SP_B, TTF1, DSG3; Ovario: WT1; y Colón: HPT1.

45 El modelo CUP se ajustó a la prevalencia de CUP (%): pulmón 23, páncreas 16, colorrectal 9, mama 3, ovario 4, próstata 2, otro 43. La prevalencia para mama y ovario se ajustó al 0% para los pacientes masculinos y la próstata se ajustó al 0% para las pacientes femeninas.

50 Se tomaron las siguientes etapas: colocar los Marcadores a 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 a partir de las muestras restantes; ensayar la muestra dejada fuera; repetir hasta haber ensayado el 100% de las muestras; dejar fuera al azar ~50% de las muestras (~50% por tejido); construir el modelo a partir de las muestras restantes; ensayar ~50% de las muestras; y repetir para 3 divisiones aleatorias diferentes.

55 Se ajustó la precisión de la clasificación a la prevalencia del tipo de cáncer para producir los resultados que se resumen en la Tabla 28 con los datos no tratados que se muestran en la Tabla 29.

60

65

Tabla 28

	Mama	Colon	Pulmón	Otro	Ovario	Pánc.	Próstata	Gobal	Ajustado
5	Correcto	23	29	22	19	24	35	19	171
	Sin ensayo	3	2	2		2	3	0	12
	Incorrecto	7	0	1	4	7	3	0	22
	Prevalencia	0,03	0,09	0,23	0,43	0,04	0,16	0,02	
10	Ensayados/total %	91	94	92	100	94	93	100	94
	Correctos/total %	70	94	88	83	73	85	100	89
	Sin ensayo %	9	6	8	n/a	6	7	0	6
15	Correcto	23	25	19	20	20	24	19	150
	Sin ensayo %	7	6	5		10	15	0	43
	Incorrecto	3	0	1	3	3	2	0	12
	Prevalencia	0,03	0,09	0,23	0,43	0,04	0,16	0,02	
20	Ensayados/total %	79	81	80	100	70	63	100	79
	Correctos/total %	70	81	76	87	61	59	100	73
	Correctos/ensayados %	88	100	95	87	87	92	100	93
	Sin ensayo %	21	19	20	n/a	30	37	0	21

25

30

35

40

45

50

55

60

65

Tabla 29
Muestra

ID	Género	Origen	Ant. Predicción	BACTIN	PBGD	Ave	CDH17	DSG3	F5	HUMP	KLK3	MG	PDEF	PSCA	TTF1	WT1
128	f	mama	pulmón	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	desc. 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	desc. 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	desc. 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	desc. 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	desc. 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	desc. 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	desc. 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	desc. 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	desc. 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	desc. 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	desc. pulmón	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	desc. 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	desc. 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	desc. pulmón	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	desc. pulmón	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	desc. 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	desc. pulmón	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	desc. 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	desc. 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	desc. 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	desc. 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	desc. 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	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	desc. 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	desc. páncreas	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	desc. 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	desc. 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	desc. páncreas	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	desc. 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	desc. 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	desc. 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	desc. páncreas	23.55	29.88	26.72	40.00	40.00	30.60	38.84	40.00	32.12	31.76	40.00	38.59
417	f	páncreas	desc. páncreas	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	próstata	pulmón próstata	22.37	26.95	24.66	40.00	40.00	29.47	23.69	21.38	24.70	24.28	30.89	31.16
407	m	próstata	pulmón próstata	28.20	31.87	30.04	40.00	40.00	40.00	27.70	25.98	27.65	40.00	39.13	38.76
116	f	CUP	desc. pulmón SCC	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	pulmón 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	desc. páncreas	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	desc. páncreas	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	desc. pulmón	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	desc. 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	desc. 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	desc. 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 85 Estudio prospectivo de firma genética del cáncer metastásico de sitio primario desconocido CUP para predecir el tejido de origen

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

Objetivo principal: Confirmar la viabilidad de realizar el análisis genético a partir de muestras de biopsia core en pacientes consecutivos con CUP.

15 Objetivo secundario: correlacionar los resultados del ensayo de RT-PCR de firma genética de 10 genes con la evaluación diagnóstica realizada en el M.D. Anderson Cancer Center (MDACC). Tercer objetivo: Correlacionar la prevalencia 6 tipos de cáncer predichos por el ensayo con la prevalencia proveniente de la literatura y de la experiencia del MDACC.

20 Se utilizó el método descrito en el presente documento para realizar un análisis de expresión génica con micromatriz de 700 carcinomas primarios congelados, y muestras benignas y normales y marcadores génicos candidatos identificados, específicos para carcinomas de pulmón, páncreas, colon, mama, próstata y ovario. Se ensayaron los marcadores génicos candidatos mediante RT-PCR en 205 muestras fijadas en formalina e incluidas en parafina (FFPE) de carcinoma metastásico (estadio III-IV) originado a partir de pulmón, páncreas, colon, mama, ovario y próstata, así como metástasis originadas a partir de otros tipos de cáncer para el control de la especificidad. Los demás tipos de cáncer metastásico incluían carcinomas gástrico, de células renales, hepatocelular, colangiocarcinoma/de vesícula biliar y de cabeza y cuello. Los resultados permitieron seleccionar una firma genética de 10 genes que predijo el tejido de origen del carcinoma metastásico y dio una precisión global del 76%. El CV medio para mediciones repetidas en los experimentos de RT-PCR es del 1,5%, calculado en base a 4 puntos de datos duplicados. Se utilizó la beta-actina (ACTB) como gen de mantenimiento y la mediana de su expresión fue similar en muestras metastásicas de origen diferente (CV = 5,6%).

25 El objetivo específico de este estudio era validar la capacidad de la firma genética de 10 genes para predecir el tejido de origen del carcinoma metastásico en los pacientes con CUP en comparación con la evaluación diagnóstica exhaustiva.

35 Elegibilidad de los pacientes

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

40 Pacientes han cumplido los criterios para CUP: ningún primario detectado después de una evaluación completa que se define como la historia completa y un examen físico, un examen detallado de laboratorio, estudios de formación de imágenes y estudios invasivos dirigidos a los síntomas o signos. Sólo se permitió participar en el estudio a los pacientes no tratados.

45 Si un paciente ha sido tratado con quimioterapia o radiación, se permite su participación en el estudio si se dispone de tejido anterior (al tratamiento) en forma de bloques archivados dentro del período de 10 años.

50 Los pacientes dieron su consentimiento/autorización por escrito para participar en este estudio.

Diseño del estudio

55 Se permitió participar en el estudio a pacientes con diagnóstico de CUP que se habían sometido a una biopsia core o biopsia por escisión de la lesión metastásica más accesible. Sólo los pacientes con biopsia FNA no fueron elegibles. Se admitió a los primeros 60 pacientes consecutivos que se presentaron que cumplieran con los criterios de inclusión y el consentimiento para el estudio. Si el MDACC solicitaba una biopsia repetida para fines de diagnóstico para su tratamiento, se obtenía tejido adicional para el estudio si el paciente daba su consentimiento. Se registró a todos los participantes en el protocolo en el sistema de gestión de datos de pacientes (PDMS) institucional.

60 Se realizó una evaluación diagnóstica completa, que incluía evaluaciones clínicas y patológicas, en todos los pacientes incluidos según las normas del MDACC. La parte de patología de la evaluación diagnóstica puede haber incluido ensayos de inmunohistoquímica (IHC) con Marcadores que incluían CK-7, CK-20, TTF-1 y otros que el patólogo consideró indicados. Esto es parte del trabajo de rutina de todos los pacientes que se presentan con CUP.

65

Recogida de muestras de tejido

El estudio incluía muestras de carcinoma metastásico fijadas en formalina e incluidas en parafina recogidas de pacientes con CUP.

Se utilizaron seis secciones de 10 µm para el aislamiento de ARN, las muestras de tejido más pequeñas requieren nueve secciones de 10 µm. Se confirmaron el diagnóstico histopatológico y el contenido del tumor para cada muestra utilizada para el aislamiento de ARN en una sección adicional teñida con hematoxilina y eosina (HE). La muestra de tumor debía tener más de un 30% de contenido del tumor en la sección HE.

Los datos clínicos fueron suministrados de forma anónima a Veridex e incluían la edad del paciente, el sexo, la histología del tumor mediante microscopía óptica, el grado de malignidad del tumor (diferenciación), el sitio de metástasis, la fecha de obtención de la muestra, la descripción de la evaluación diagnóstica realizada para cada paciente.

Procesamiento de tejidos y experimentos de RT-PCR

Se extrajo el ARN total de cada muestra de tejido utilizando el protocolo descrito anteriormente. Sólo se utilizaron las muestras que produjeron más de 1 µm de ARN total de la cantidad estándar de tejido para el posterior ensayo de RT-PCR. Las muestras con menos rendimiento de ARN se consideran degradadas y se excluyeron de los posteriores experimentos. Se llevó a cabo control de la integridad del ARN en base a la expresión de mantenimiento con el fin de excluir las muestras con ARN degradado, según el procedimiento estándar de Veridex.

Se utilizó un ensayo de RT-PCR que incluye un panel de 10 genes y 1-2 genes de control para el análisis de las muestras de ARN. El ensayo de PCR y la transcripción inversa se llevan a cabo utilizando los protocolos descritos anteriormente.

Se calculó el valor de expresión relativa para cada gen ensayado presentado como ΔCt , que es igual al Ct del gen diana menos el Ct de los genes de control, y se utilizó para la predicción del tejido de origen.

Tamaño de muestra e interpretación de datos

Se estudió un tamaño de muestra limitado de 60 pacientes debido a la naturaleza exploratoria del estudio piloto. Hasta la fecha, se han sometido a ensayo 22 pacientes. Las muestras de un paciente no consiguieron producir suficiente ARN para la RT-PCR y 3 no lograron pasar el control QC evaluado mediante RT-PCR con genes de control. Se utilizó un total de 18 pacientes para determinada la probabilidad de lesión metastásica del paciente.

Se utilizó el modelo estadístico para determinar la probabilidad del tejido de origen del carcinoma metastásico de las siguientes siete categorías: pulmón, páncreas, colon, mama, próstata, ovario y ningún ensayo (otro). Para cada muestra, la probabilidad para cada categoría se calcula a partir de un modelo de clasificación lineal. Los resultados de ensayo se resumen en la Tabla 30.

La probabilidad de lesión metastásica de un paciente (con primarios conocidos) que viene de uno de estos 6 sitios (colon, páncreas, pulmón, próstata, ovario, mama) es de aproximadamente un 76%. Este número proviene de la literatura dada la incidencia de diversos tipos de cáncer y el potencial de propagación y datos no publicados generados en el M.D. Anderson a partir de registro de tumores. Para las muestras ensayadas, la prevalencia de 6 sitios fue del 67% (12 de 18 muestras ensayadas), en estrecha concordancia con las observaciones anteriores.

Tabla 30

Datos del paciente			Probabilidad a posteriori ToO (%)								
ID	M/F	predicción	Mama	Colon	Pulmón	SCC pulmón	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	SCC pulmón	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	
15	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	
20	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	SCC pulmón	0,00	0,03	0,00	65,64	7,96	0,00	26,38	0,00	
3	F	no pasó el ctrl.	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
8	M	no pasó el ctrl.	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
25	20	F	no pasó el ctrl.	0,00	0,00	0,00	0,00	0,00	0,00	0,00	

Aunque la invención precedente se ha descrito con cierto detalle a modo de ilustración y ejemplo con fines de claridad de comprensión, las descripciones y los ejemplos no deben interpretarse como limitativos del alcance de la invención.

Tabla 31

Nombre	SEQ ID NO	Registro	Descripción
CDH17	62	NM_004063	cadherina 17
CDX1	78	NM_001804	Factor de transcripción 1 de la caja homeótica
DSG3	61/3	NM_001944	desmogleína 3
F5	67/6	NM_000130	factor de coagulación V
FABP1	71	NM_001443	proteína de unión a ácidos grasos 1, hígado
GUCY2C	79	NM_004963	guanilato ciclasa 2C
HE4	82	NM_006103	supuesto marcador de carcinoma de ovario
KLK2	80	BC005196	calicreína 2, de próstata
HNRPA0	84	NM_006805	ribonucleoproteína nuclear heterogénea A0
HPT1	85/4	U07969	Transportador asociado a péptido intestinal
ITGB6	71	NM_000888	Integrina, beta 6
KLK3	68	NM_001648	calicreína 3
MGB1	63/7	NM_002411	mamaglobina 1
PAX8	83	BC001060	Gen paired box 8
PBGD	70	NM_000190	hidroximetilbilano sintasa
PDEF	64/8	NM_012391	Factor de transcripción Ets que contiene dominio
PIP	81	NM_002652	Proteína inducida por prolactina
PSA	86/9	U17040	Precursor de antígeno específico de próstata
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
TTF1	60/2	NM_003317	similar a factor de transcripción tiroideo 1
WT1	65/10	NM_024426	Tumor de Wilms 1
β -actin	69	NM_001101	β -actina
KRT6F	87	L42612	queratina 6 isoforma K6f
p73H	88	AB010153	proteína relacionada con p53
SFTPC	89	NM_003018	proteína C asociada a surfactante pulmonar
KLK10	90	NM_002776	calicreína 10
CLDN18	91	NM_016369	Claudina 18
TR10	92	BD280579	Receptor del factor de necrosis tumoral
B305D	93		
B726	94		
GABA-pi	95	BC109105	receptor de ácido gamma-aminobutírico A, pi
StAR	96	NM_01007243	reguladora esteroideogénica aguda
EMX2	97	NM_004098	homólogo 2 de espiráculos vacíos (<i>Drosophila</i>)
NGEP	98	AY617079	variante larga de NGEP
NPY	99	NM_000905	neuropéptido Y
SERPINA1	100	NM_000295	inhibidor de serpin peptidasa, clado A miembro 1
KRT7	101	NM_005556	Queratina 7
MMP11	102	NM_005940	metalopeptidasa de matriz 11 (estromelina 3)
MUC4	103	NM_018406	asociada a mucina 4 de superficie celular
FLJ22041	104	AK025694	
BAX	105	NM_138763	Variante Δ de transcrito proteína asociada a BCL2
PITX1	106	NM_002653	factor 1 de trans de homeodominio similar a paired
MGC:10264	107	BC005807	estearoil-CoA desaturasa (Δ -9-desaturasa)

REFERENCIAS

Patentes y publicaciones de solicitud de patentes de los Estados Unidos

5242974	5700637	20030194733
5350840	5786148	20030198970
5384261	6004755	20030215803
5405783	6136182	20030215835
5412087	6218114	20030219760
5424186	6218122	20030219767

5429807	6225051	20030232350
5436327	6232073	20030235820
5445934	6261766	20040005563
5472672	6271002	20040009154
5527681	6339148	20040009489
5529756	20010029020	20040018969
5532128	20020055627	20040029114
5545531	20020068288	20040076955
5554501	20020168647	20040126808
5556752	20030044859	20040146862
5561071	20030087818	20040219572
5571639	20030104448	20040219575
5593839	20030124128	20050037010
5599695	20030124579	20050059008
5624711	20030138793	20060094035
5658734	20030190656	

Patentes y publicaciones de patentes extranjeras

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

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

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 G1:4320-4324

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

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

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

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

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
- 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
- 5 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
- Dennis *et al.* (2005b) Markers of adenocarcinoma characteristic of the site of origin: development of a diagnostic algorithm *Clin Can Res* 11:3766-3772
- 10 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
- 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
- 15 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
- 60 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
- 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 genespecific 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 (PIPGDFP-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
- 20 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
- 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
- 30 Pillozzi *et al.* (2004) CDX1 expression is reduced in colorectal carcinoma and is associated with promoter hypermethylation *J Pathol* 204:289-295
- Poleev *et al.* (1992) PAX8, a human paired box gene: isolation and expression in developing thyroid, kidney and Wilms' tumors *Development* 116:611-623
- 35 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
- 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
- 40 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
- 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
- 45 Senoo *et al.* (1998) A second p53-related protein, p73L, with high homology to p73 *Biochem Biophys Res Comm* 248:603-607
- 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
- 50 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
- 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
- Tothill *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
- 55 van Ruissen *et al.* (2005) Evaluation of the similarity of gene expression data estimated with SAGE and Affymetrix GeneChips *BMC Genomics* 6:91
- 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
- 60 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
- 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
- 65 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

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

LISTADO DE SECUENCIAS

10 <110> Veridex, LLC Wang, Yixin Mazumder, Abhijit Talantov, Dmitri Jatkoe, Timothy Baden, Jonathan

<120> Métodos y materiales para identificar el origen de un carcinoma de origen primario desconocido

15 <130> VDX5007WOPCT

<150> 60/718.501

<151> 19-09-2005

20 <150> 60/725.680

<151> 12-10-2005

<160> 107

25 <170> PatentIn versión 3.3

<210> 1

<211> 476

<212> ADN

<213> humano

30 <400> 1

```

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

50 <210> 2
 <211> 493
 <212> ADN
 <213> humano

55 <400> 2

60

65

ES 2 494 843 T3

5 ctgcacccac ctacttagat atttcatgtg ctatagacat tagagagatt tttcattttt 60
 ccatgacatt tttcctctct gcaaatggct tagctacttg tgtttttccc ttttggggca 120
 agacagactc attaaatatt ctgtacattt tttctttatc aaggagatat atcagtgttg 180
 10 tctcatagaa ctgcctggat tccatttatg ttttttctga ttccatcctg tgtccccctc 240
 atccttgact cctttggtat ttcactgaat ttcaaacatt tgtc 284

15 <210> 5
 <211> 394
 <212> ADN
 <213> humano

20 <220>
 <221> misc_feature
 <222> (58)..(58)
 <223> n es a, c, g, o t

25 <220>
 <221> misc_feature
 <222> (95)..(95)
 <223> n es a, c, g, o t

30 <220>
 <221> misc_feature
 <222> (99)..(99)
 <223> n es a, c, g, o t

35 <220>
 <221> misc_feature
 <222> (119)..(119)
 <223> n es a, c, g, o t

40 <220>
 <221> misc_feature
 <222> (123)..(123)
 <223> n es a, c, g, o t

45 <220>
 <221> misc_feature.
 <222> (130)..(130)
 <223> n es a, c, g, o t

50 <220>
 <221> misc_feature
 <222> (151)..(151)
 <223> n es a, c, g, o t

55 <220>
 <221> misc_feature
 <222> (155)..(155)
 <223> n es a, c, g, o t

60 <220>
 <221> misc_feature
 <222> (161)..(161)

65 <223> n es a, c, g, o t
 <220>

ES 2 494 843 T3

<221> misc_feature
 <222> (212)..(212)
 <223> n es a, c, g, o t

5
 <400> 5

10	ttcctgaggc acatcctaac gcaagtttga ccatgtatgt ttgcaccctt tttcccnaa	60
	ccctgacctt cccatgggcc tttccagga ttccnaccng gcagatcagt tttagtana	120
	canatccgcn tgcagatggc cctccaacc nttntgttg ntgtttccat ggccagcat	180
15	tttccacctt taacctgtg ttcaggcact tnttcccca ggaagcctt cctgcccacc	240
	ccatttatga attgagccag gtttgggccg tgggtgcccc cgcaccagc aggggacagg	300
	caatcaggag ggcccagtaa aggctgagat gaagtggact gagtagaact ggaggacaag	360
20	agttgacgtg agttcctggg agtttccaga gatg	394

25
 <210> 6
 <211> 470
 <212> ADN
 <213> humano

30
 <220>
 <221> misc_feature
 <222> (61)..(61)
 <223> n es a, c, g, o t

35
 <220>
 <221> misc_feature
 <222> (82) .. (82)
 <223> n es a, c, g, o t

40
 <400> 6

	atcctctaca gccagatgtc acagggatac gtctactttc acttgggtgct ggagaattca	60
45	naagtcaaga acatgctaag cntaagggac ccaaggtaga aagagatcaa gcagcaaagc	120
	acaggttctc ctggatgaaa ttactagcac ataaagttgg gagacaccta agccaagaca	180
	ctggttctcc ttccggaatg aggccctggg aggaccttc tagccaagac actggttctc	240
50	cttccagaat gaggccctgg aaggacctc ctagtgatct gttactctta aaacaagta	300
	actcatctaa gattttggtt gggagatggc atttggcttc tgagaaagg agctatgaaa	360
	taatccaaga tactgatgaa gacacagctg ttaacaattg gctgatcagc cccagaatg	420
55	cctcacgtgc ttggggagaa agcaccctc ttgccaacaa gcctggaaag	470

60
 <210> 7
 <211> 396
 <212> ADN
 <213> humano

65
 <400> 7

ES 2 494 843 T3

5 gcagcagcct caccatgaag ttgctgatgg tcctcatgct ggcggccctc tcccagcact 60
 gctacgcagg ctctggctgc cccttattgg agaatgtgat ttccaagaca atcaatccac 120
 aagtgtctaa gactgaatac aaagaacttc ttcaagagtt catagacgac aatgccacta 180
 caaatgccat agatgaattg aaggaatgtt ttcttaacca aacggatgaa actctgagca 240
 10 atgttgagggt gtttatgcaa ttaatatatg acagcagtct ttgtgattta ttttaacttt 300
 ctgcaagacc tttggctcac agaactgcag ggtatgggtga gaaaccaact acggattgct 360
 15 gcaaaccaca ccttctcttt ctatgtctt tttact 396

20 <210> 8
 <211> 491
 <212> ADN
 <213> humano

<400> 8

25 gagtggggcc cttaaactgg attcaaaaaa tgctctaaac ataggaatgg ttgaagaggt 60
 cttgcagtct tcagatgaaa ctaaattctt agaagaggca caagaatggc taaagcaatt 120
 30 catccaaggg ccaccggaag taattagagc ttgaaaaaa tctgtttgtt caggcagaga 180
 gctatatattg gaggaagcat tacagaacga aagagatctt ttaggaacag ttgggggtgg 240
 35 gcctgcaaat ttagaggcta ttgctaagaa aggaaaatth aataaataat tggtttttcg 300
 tgtggatgta ctccaagtaa agctccagtg actaatatgt ataaatgta aatgatatta 360
 40 aatatgaaca tcagttaaaa aaaaaattct ttaaggctac tattaatatg cagacttact 420
 tttaatcatt tgaaatctga actcatttac ctcatttctt gcccaattact cccttgggta 480
 tttactgcgt a 491

45 <210> 9
 <211> 265
 <212> ADN
 <213> humano

50 <400> 9

55 tgggtgtaatt ttgtcctctc tgtgtcctgg ggaatactgg ccatgcctgg agacatatca 60
 ctcaatttct ctgaggacac agataggatg ggggtgtctgt gttattttgtg gggtagagag 120
 atgaaagagg ggtgggatcc aactgagag agtggagagt gacatgtgct ggacactgtc 180
 60 catgaagcac tgagcagaag ctggaggcac aacgcaccag aactcacag caaggatgga 240
 gctgaaaaca taaccactc tgtcc 265

65 <210> 10
 <211> 441

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

ES 2 494 843 T3

ccaaccaga cccgcg 17

5 <210> 16
<211> 21
<212> ADN
<213> humano

10 <400> 16
cgccatgcc gctcatgtc a 21

15 <210> 17
<211> 21
<212> ADN
<213> humano

20 <400> 17
ccgcatct ccgctcat g 21

25 <210> 18
<211> 78
<212> ADN
<213> humano

30 <400> 18

ccaaccaga cccgcgcttc ccgcatct ccgctcat gggccggcg agcggcatga 60

30 acatgagcgg catggcg 78

35 <210> 19
<211> 23
<212> ADN
<213> humano

40 <400> 19
gagagaagga gaagataact caa 23

45 <210> 20
<211> 22
<212> ADN
<213> humano

50 <400> 20
actcagaga ttcgtaggt ga 22

55 <210> 21
<211> 26
<212> ADN
<213> humano

60 <400> 21
attgcaaga ttactcaga ttacca 26

65 <210> 22
<211> 97
<212> ADN
<213> humano

<400> 22

gcagagaagg agaagataac tcaaaaagaa acccaattgc caagattact tcagattacc 60
5 aagcaaccca gaaaatcacc taccgaatct ctggagt 97

<210> 23
<211> 21
10 <212> ADN
<213> humano

<400> 23
15 tccctcggca gtggaagctt a 21

<210> 24
<211> 24
<212> ADN
20 <213> humano

<400> 24
tctcaaact ctgtgtcct ggta 24

<210> 25
25 <211> 29
<212> ADN
<213> humano

<400> 25
30 ccaaaatcaa tggactcat gcccgactg 29

<210> 26
<211> 95
35 <212> ADN
<213> humano

<400> 26

40 tccctcggca gtggaagctt acaaaacgac tgggaagttt ccaaaatcaa tggactcat 60
gcccgactgt ctaccaggca cacagagttt gagga 95

45 <210> 27
<211> 21
<212> ADN
50 <213> humano

<400> 27
agttgctgat ggtcctcatg c 21

<210> 28
55 <211> 24
<212> ADN
<213> humano

<400> 28
60 cactgtgga tggattgtct tgga 24

<210> 29
<211> 23
65 <212> ADN
<213> humano

ES 2 494 843 T3

<400> 29
 ccctctccca gcaactgctac gca 23

5 <210> 30
 <211> 107
 <212> ADN
 <213> humano

10 <400> 30

agttgctgat ggtcctcatg ctggcggccc tctcccagca ctgctacgca ggctctggct 60
 gccccttatt ggagaatgtg atttccaaga caatcaatcc acaagtg 107

15

<210> 31
 <211> 20
 <212> ADN
 <213> humano

20

<400> 31
 cgcccacctg gacatctgga 20

25

<210> 32
 <211> 23
 <212> ADN
 <213> humano

30

<400> 32
 cactggtcga ggcacagtag tga 23

35

<210> 33
 <211> 25
 <212> ADN
 <213> humano

40

<400> 33
 gtcagcggcc tggatgaaag agcgg 25

45

<210> 34
 <211> 86
 <212> ADN
 <213> humano

<400> 34

50 cgcccacctg gacatctgga agtcagcggc ctggatgaaa gagcggactt cacctggggc 60
 gattcactac tgtgcctcga ccagtg 86

55

<210> 35
 <211> 23
 <212> ADN
 <213> humano

60

<400> 35
 gcggagccca atacagaata cac 23

65

<210> 36
 <211> 19
 <212> ADN

ES 2 494 843 T3

<213> humano
 <400> 36
 5 cggggctact ccaggcaca 19
 <210> 37
 <211> 25
 <212> ADN
 <213> humano
 10
 <400> 37
 tcagaggcat tcaggatg cgacg 25
 <210> 38
 <211> 80
 <212> ADN
 <213> humano
 15
 <400> 38
 20
gcggagccca atacagaata cacacgcacg gtgtcttcag aggcattcag gatgtgacgac 60
 25 **gtgtgcctgg agtagccccg 80**
 <210> 39
 <211> 20
 <212> ADN
 <213> humano
 30
 <400> 39
 35 ctggtgatgg caggcttggc 20
 <210> 40
 <211> 20
 <212> ADN
 <213> humano
 40
 <400> 40
 ttgctcacct gggctttgca 20
 <210> 41
 <211> 21
 <212> ADN
 <213> humano
 45
 <400> 41
 50 gcagccaggc actgccctgc t 21
 <210> 42
 <211> 74
 <212> ADN
 <213> humano
 55
 <400> 42
 60 **ctggtgatgg caggcttggc cctgcagcca ggcactgccc tgctgtgcta ctctgcaaa 60**
gcccagggtga gcaa 74
 65

5	<210> 43 <211> 25 <212> ADN <213> humano		
	<400> 43 tgaagaaata tcctgggatt attca	25	
10	<210> 44 <211> 27 <212> ADN <213> humano		
15	<400> 44 tatgtggtat cttctggaat atcatca	27	
20	<210> 45 <211> 27 <212> ADN <213> humano		
25	<400> 45 acaaagggaa acagatattg aagactc	27	
30	<210> 46 <211> 87 <212> ADN <213> humano		
	<400> 46		
35	tgaagaaata tcctgggatt attcagaatt tgtacaaagg gaaacagata ttgaagactc		60
	tgatgatatt ccagaagata ccacata		87
40	<210> 47 <211> 19 <212> ADN <213> humano		
45	<400> 47 cccccagtgg gtcctcaca	19	
50	<210> 48 <211> 22 <212> ADN <213> humano		
55	<400> 48 aggatgaaac aagctgtgcc ga	22	
60	<210> 49 <211> 26 <212> ADN <213> humano		
65	<400> 49 caggaacaaa agcgtgatct tgctgg	26	
	<210> 50 <211> 82 <212> ADN		

ES 2 494 843 T3

<213> humano
 <400> 50

5
 ccccagtg gtcctcacag ctgccactg catcaggaac aaaagcgtga tcttgctggg 60
 tccgcacagc ttgtttcatc ct 82

10
 <210> 51
 <211> 19
 <212> ADN
 <213> humano

15
 <400> 51
 gccctgaggc actcttcca 19

20
 <210> 52
 <211> 22
 <212> ADN
 <213> humano

25
 <400> 52
 cggatgtcca cgtcacactt ca 22

30
 <210> 53
 <211> 25
 <212> ADN
 <213> humano

35
 <400> 53
 ctctctcttct gggcatggag tcctg 25

40
 <210> 54
 <211> 100
 <212> ADN
 <213> humano

45
 <400> 54

45
 gccctgaggc actcttccag ccttccttcc tgggcatgga gtcctgtggc atccacgaaa 60
 ctacccttcaa ctccatcatg aagtgtgacg tggacatccg 100

50
 <210> 55
 <211> 22
 <212> ADN
 <213> humano

55
 <400> 55
 ccacacacag cctactttcc aa 22

60
 <210> 56
 <211> 21
 <212> ADN
 <213> humano

65
 <400> 56
 taccacgcg aatcactctc a 21

<210> 57

ES 2 494 843 T3

<211> 27
<212> ADN
<213> humano

5 <400> 57
aacggcaatg cggctgcaac ggcggaa 27

10 <210> 58
<211> 103
<212> ADN
<213> humano

<400> 58

15 ccacacacag cctactttcc aagcggagcc atgtctggta acggcaatgc ggctgcaacg 60
gcggaagaaa acagcccaaâ gatgagagtg attcgcgtgg gta 103

20

<210> 59
<211> 2724
<212> ADN
<213> humano

<400> 59

30

35

40

45

50

55

60

65

5 ggtgccatgg ctgagtcaca cctgctgcag tggctgctgc tgctgctgcc cacgctctgt 60
 ggcccaggca ctgctgcctg gaccacctca tccttggcct gtgcccaggg ccctgagttc 120
 tggtgccaaa gcctggagca agcattgcag tgcagagccc tagggcattg cctacaggaa 180
 10 gtctggggac atgtgggagc cgatgaccta tgccaagagt gtgaggacat cgtccacatc 240
 cttacaaga tggccaagga ggccattttc caggacacga tgaggaagt cctggagcag 300
 gagtgcaacg tcctccccct gaagctgctc atgccccagt gcaaccaagt gcttgacgac 360
 15 tacttcccc tggtcatcga ctacttccag aaccagactg actcaaacgg catctgtatg 420
 cacctggggc tgtgcaaatc ccggcagcca gagccagagc aggagccagg gatgtcagac 480
 cccctgcccc aacctctgcg ggaccctctg ccagaccctc tgctggacia gctcgtcctc 540
 20 cctgtgctgc ccggggccct ccaggcgagg cctgggcctc acacacagga tctctccgag 600
 cagcaattcc ccattcctct cccctattgc tggctctgca gggctctgat caagcggatc 660
 caagccatga ttcccaaggg tgcgctagct gtggcagtg cccaggttg ccgctggtgta 720
 25 cctctggtgg cgggcggcat ctgccagtgc ctggctgagc gctactccgt catcctgctc 780
 gacacgctgc tgggccgcat gctgccccag ctggtctgcc gcctcgtcct ccggtgctcc 840
 atggatgaca gcgctggccc aaggtcgccg acaggagaat ggctgccgcg agactctgag 900
 30 tgccacctct gcatgtccgt gaccaccag gccgggaaca gcagcgagca ggccatacca 960
 caggcaatgc tccaggcctg tgttggetcc tggctggaca gggaaaagt caagcaattt 1020
 35 gtggagcagc acacgcccc gctgctgacc ctggtgcccc ggggctggga tgcccacacc 1080
 acctgccagg ccctcgggggt gtgtgggacc atgtccagcc ctctccagt tatccacagc 1140
 cccgaccttt gatgagaact cagctgtcca gaaaaagaca ccgtccttta aagtgtgca 1200
 40 gtatggccag acgtggtggc tcacacctgc aatcccagca ccttaggagg ccgaggcagg 1260
 aggatccttg aggtcaggag ttcgagacca gcctcgccaa catggtgaaa ccccatttct 1320
 45 actaaaaata caaaaaatta gccaaagtgt gtggcatatg cctgtaatcc caactactca 1380
 gaaggccgag gcaggagaat tacttgaacg caggagaatc actgcagccc aggaggcaga 1440
 50 ggttgcatg agccgagatt gcaccactgc actccagcct gggtgacaga gcaagactcc 1500

50

55

60

65

5 atctcagtaa ataaataaat aaataaaaag cgctgcagta gctgtggcct caccctgaag 1560
 tcagcgggcc caggcctacc tcaactctctc ccttggcaga gaagcagacg tccatagctc 1620
 ctctccctca caagegctcc cagcctgccc tccagctgct gctctccctt cccagtctct 1680
 10 actcactggg atgagggttag gtcattgagga caccaaaaac ctaaaaataa acaaaaagcc 1740
 aaacaagcct tagcttttct taaagactga aatgcctgga agtgtccctt tatttataaa 1800
 ataacttttg tcatatttct tatacatggt tcttgaaga aattcagaaa ctacagacaa 1860
 agagagtgga aattaccac tgctcaggct ctgagcccaa gctaagccat catatcccct 1920
 15 gtgccctgca cgtatacacc cagatggcct gaagcaactg aagatccaca aaagaagtga 1980
 aaatagccag ttctgctt aactgatgac attccaccat tgtgatttgt tctgccccca 2040
 ccctaactga tcaattgacc ttgtgacaat acaccttccc cacccttgag aagggtgcttt 2100
 20 gtaatattct ccccaaccac cccacgcccg ccccccgca cccttaagaa ggtattttgt 2160
 aatattctct ccgccattga gaatgtgctt tgtaagatcc accccctgcc cacaaaaaat 2220
 tgctcctaac tccaccgctt atcccaaacc tacaagaact aatgataatc ccaccacct 2280
 25 ttgctgactc tttttggact cagcccacct gcaccagggt gattaaaaag ctttattggt 2340
 cacacaaagc ctgtttgga gtctcttcac agggaagcat gtgacacca caatcccacc 2400
 tagcccagga gagagctacg gcaggggtggt tgttttgaca ctgagcttgg ggctttttcc 2460
 30 atcttctccc cacagctctt ggctccacac ctccaccggt caagcgccag aaagagctgt 2520
 ctatgcagcc tgctcttggg cctggggatg agacacacaa ttcatggct cctggatttt 2580
 aagtagacat ttgtaaactt atagctaact actgtcctta aagccattgt ttccattaca 2640
 35 aaatccaact ctctgagaga aaaggggtgt ttaaatttaa aaaaataaaa acaaaaaagt 2700
 ttgattgaga aaaaaaaaaa aaaa 2724

40
 <210> 60
 <211> 2352
 <212> ADN
 45 <213> humano
 <400> 60

50 gaaacttaaa ggtgtttacc ttgtcatcag catgtaagct aattatctcg ggcaagatgt 60
 aggcttctat tgtcttgttg ctttagcgtt tacgccccgc ctctggtggc tgcctaaaac 120
 55 ctggcgccgg gctaaaacaa acgcgaggca gccccgagc ctccactcaa gccaatgaag 180
 gaggactcgg tccactccgt tacgtgtaca tccaacaaga tcggcgtaa ggtaacacca 240
 gaatatttgg caaagggaga aaaaaaaaag agcgaggctt cgccttcccc ctctcccttt 300
 60 ttttctctcc tcttcttccc tctccagcc gccgccaat catgtcgatg agtccaaagc 360
 acaagactcc gttctcagtg tctgacatct tgagtcccct ggaggaaagc tacaagaag 420

65

5 tgggcatgga gggcggcggc ctcggggctc cgctggcggc gtacaggcag ggccaggcgg 480
 caccgccaac agcggccatg cagcagcacg ccgtggggca ccacggcgcc gtcaccgccg 540
 cctaccacat gacggcggcg ggggtgcccc agctctcgca ctccgccgtg gggggctact 600
 10 gcaacggcaa cctgggcaac atgagcgagc tgccgccgta ccaggacacc atgaggaaca 660
 gcgcctctgg ccccggatgg tacggcgcca acccagaccc gcgcttcccc gccatctccc 720
 gcttcatggg cccggcgagc ggcatgaaca tgagcggcat gggcggcctg ggctcgtctg 780
 15 gggacgtgag caagaacatg gccccgctgc caagcgcgcc gcgcaggaag cgccgggtgc 840
 tcttctcgca ggcgcaggtg tacgagctgg agcgacgctt caagcaacag aagtacctgt 900
 cggcgcggga gcgcgagcac ctggccagca tgatccacct gacgcccacg caggtcaaga 960
 20 tctggttcca gaaccaccgc taaaaatga agcgcaggc caaggacaag gcggcgcagc 1020
 agcaactgca gcaggacagc ggcggcggcg ggggcggcgg gggcaccggg tgcccgcagc 1080
 agcaacagge tcagcagcag tcgccgcgac gcgtggcggg gccggctctg gtgaaagacg 1140
 25 gcaaaccgtg ccaggcgggt gccccgcgc cgggcgccgc cagcctacaa ggccacgcgc 1200
 agcagcagge gcagcaccag gcgcagggcg cgcaggcggc ggagcggcc atctccgtgg 1260
 30 gcagcgggtg cgccggcctt ggcgcacacc cgggccacca gccaggcagc gcaggccagt 1320
 ctccggacct ggcgcaccac gccgccagcc ccgcggcgct gcagggccag gtatccagcc 1380
 tgtccacct gaactcctcg ggctcggact acggcaccat gtcctgctcc accttctat 1440
 35 acggtcggac ctggtgagag gacgccggc cggccctagc ccagcgtct gcctcaccgc 1500
 ttccctcctg cccgccacac agaccaccat ccaccgctgc tccacgcgct tcgacttttc 1560
 ttaacaacct ggccgcgttt agaccaagga acaaaaaaac cacaaggcc aaactgctgg 1620
 40 acgtctttct tttttcccc ccctaaaatt tgtgggtttt ttttttaaa aaaagaaaat 1680
 gaaaaacaac caagcgcac caatctcaag gaatctttaa gcagagaagg gcataaaaca 1740
 45 gctttggggg gtctttttt ggtgattcaa atgggtttc cacgctaggg cggggcacag 1800
 attggagagg gctctgtgct gacatggctc tggactctaa agaccaaact tcaactctggg 1860
 cacactctgc cagcaaagag gactcgttg taaataccag gattttttt ttttttgaa 1920
 50 gggaggacgg gagctgggga gaggaagag tcttcaacat aaccacttg tcaactgacac 1980
 aaaggaagtg cccctcccc ggcacctct ggccgcctag gctcagcggc gaccgccctc 2040
 cgcgaaaata gtttgtttaa tgtgaacttg tagctgtaa acgctgtcaa aagttggact 2100
 55 aaatgcctag ttttagtaa tctgtacatt ttgtgtaa aagaaaaacc actcccagtc 2160
 cccagccctt cacatTTTTT atgggcattg acaaatctgt gtatattatt tggcagttg 2220
 gtatttgcg cgtcagtctt tttctgtgt aacttatgta gatatttggc ttaaataag 2280
 60 ttcctaagaa gcttctaata aattatacaa attaaaaaga ttcttttct gattaaaaaa 2340
 aaaaaaaaa aa 2352
 65

ES 2 494 843 T3

<210> 61
 <211> 3336
 <212> ADN
 <213> humano

5

<400> 61

10

ttttcttaga cattaactgc agacggctgg caggatagaa gcagcggctc acttggactt 60

tttcaccagg gaaatcagag acaatgatgg ggctcttccc cagaactaca ggggctctgg 120

15

ccatcttcgt ggtggtcata ttggttcatt gagaattgag aatagagact aaaggtcaat 180

atgatgaaga agagatgact atgcaacaag ctaaaagaag gcaaaaacgt gaatgggtga 240

aatttgccaa accctgcaga gaaggagaag ataactcaa aagaaacca attgccaaga 300

20

ttacttcaga ttaccaagca acccagaaaa tcacctaccg aatctctgga gtgggaatcg 360

atcagccgcc ttttggaaatc tttgttgttg acaaaaacac tggagatatt aacataacag 420

ctatagtcga ccgggaggaa actccaagct tcctgatcac atgtcgggct ctaaagcccc 480

25

aaggactaga tntagagaaa ccacttatac taacgggtta aattttggat attaattgata 540

atcctccagt attttcacaa caaattttca tgggtgaaat tgaagaaaat agtgcctcaa 600

actcactggt gatgatacta aatgccacag atgcagatga accaaaccac ttgaattcta 660

30

aaattgcctt caaaattgtc tctcaggaac cagcaggcac acccatgttc ctccaaagca 720

gaaacactgg ggaagtccgt actttgacca attctcttga ccgagagcaa gctagcagct 780

atcgtctggt tgtgagtggg gcagacaaag atggagaagg actatcaact caatgtgaat 840

gtaatatata agtgaaagat gtcaacgata acttcccaat gtttagagac tctcagtatt 900

cagcacgtat tgaagaaaat attttaagtt ctgaattact tcgatttcaa gtaacagatt 960

40

tggatgaaga gtacacagat aattggcttg cagtatattt ctttacctct ggggaatgaag 1020

gaaattgggt tgaatacaa actgatccta gaactaatga aggcattctg aaagtgggtga 1080

aggctctaga ttatgaacaa ctacaaagcg tgaacttag tattgctgtc aaaaacaaag 1140

45

ctgaatttca ccaatcagtt atctctcgat accgagttca gtcaaccca gtcacaattc 1200

aggtaataaa tgtaagagaa ggaattgcat tccgtcctgc ttccaagaca ttactgtgc 1260

aaaaaggcat aagtagcaaa aaattgggtg attatattct ggggaacatat caagccatcg 1320

50

atgaggacac taacaaagct gcctcaaatg tcaaatatgt catgggacgt aacgatggtg 1380

gatacctaat gattgattca aaaactgctg aaatcaaatt tgtcaaaaat atgaaccgag 1440

attctacttt catagttaac aaaacaatca cagctgaggt tctggccata gatgaatata 1500

55

cgggtaaaac ttctacaggc acggtatatg ttagagtacc cgatttcaat gacaattgtc 1560

60

65

caacagctgt cctcgaaaaa gatgcagttt gcagttcttc accttccgtg gttgtctccg 1620
 5 ctagaacact gaataataga tacactggcc cctatacatt tgcactggaa gatcaacctg 1680
 taaagttgcc tgccgatgg agtatcaciaa cctcaatgc tacctcggcc ctctcagag 1740
 10 cccaggaaca gatacctcct ggagtatacc acatctccct ggtacttaca gacagtcaga 1800
 acaatcggtg tgagatgcca cgcagcttga cactggaagt ctgtcagtgt gacaacaggg 1860
 gcatctgtgg aacttcttac ccaaccaciaa gccctgggac caggtatggc aggccgcaact 1920
 15 cagggaggct ggggcctgcc gccatcggcc tgctgctcct tggctcctg ctgctgctgt 1980
 tggccccct tctgctgtg acctgtgact gtggggcagg ttctactggg ggagtgcag 2040
 gtggttttat cccagttcct gatggctcag aaggaacaat tcatcagtgg ggaattgaag 2100
 20 gagcccatcc tgaagacaag gaaatcaciaa atatttgtgt gcctcctgta acagccaatg 2160
 gagccgattt catggaaagt tctgaagttt gtacaaatac gtatgccaga ggcacagcgg 2220
 25 tggaaaggcac ttcaggaatg gaaatgacca ctaagcttgg agcagccact gaatctggag 2280
 gtgctgcagg cttgcaaca gggacagtgt caggagctgc ttcaggattc ggagcagcca 2340
 ctggagttgg catctgttcc tcagggcagt ctggaaccat gagaacaagg cattccactg 2400
 30 gaggaaccaa taaggactac gctgatgggg cgataagcat gaattttctg gactcctact 2460
 tttctcagaa agcatttgcc tgtgcggagg aagacgatgg ccaggaagca aatgactgct 2520
 tgttgatcta tgataatgaa ggcgcagatg cactggttc tcctgtgggc tccgtgggtt 2580
 35 gttgcagtti tattgctgat gacctggatg acagcttctt ggactcactt ggacccaat 2640
 ttaaaaaact tgcagagata agccttggtg ttgatggtga aggcaaagaa gttcagccac 2700
 40 cctctaaaga cagcggttat gggattgaat cctgtggcca tcccatagaa gtccagcaga 2760
 caggatttgt taagtgccag actttgtcag gaagtcaagg agcttctgct ttgtccgct 2820
 ctgggtctgt ccagccagct gtttccatcc ctgacctct gcagcatggt aactatttag 2880
 45 taacggagac ttactcggct tctggttccc tcgtgcaacc ttccactgca ggctttgatc 2940
 cacttctcac acaaaatgtg atagtgcag aaagggatg ctgtccatt tccagtgttc 3000
 50 ctggcaacct agctggccca acgcagctac gagggtcaca tactatgctc tgtacagagg 3060
 atccttgctc ccgtctaata tgaccagaat gagctggaat accacactga ccaaatctgg 3120
 atctttggac taaagtattc aaaatagcat agcaaagctc actgtattgg gctaataatt 3180
 55 tggcacttat tagcttctct cataaactga tcacgattat aaattaaatg tttgggttca 3240
 taccceaaa gcaatatgtt gtcactccta attctcaagt actattcaaa ttgtagtaaa 3300
 60 tcttaaagtt tttcaaaacc ctaaaatcat attcgc 3330

<210> 62
 <211> 3697
 65 <212> ADN
 <213> humano

<400> 62

5 agggagtggt cccgggggag atactccagt cgtagcaaga gtctcgacca ctgaatggaa 60
 gaaaaggact ttaaccacc attttgtagc ttacagaaag gaatttgaat aaagaaaact 120
 10 atgatacttc aggcccatct tcaactccctg tgtcttctta tgctttatth ggcaactgga 180
 tatggccaag aggggaagtt tagtggaccc ctgaaaccca tgacatthtc tathatgaa 240
 ggccaagaac cgagtcaaat tatattccag ttaaggcca atcctcctgc tgtgacttht 300
 15 gaactaactg gggagacaga caacatath gtgatagaac gggagggact tctgtattac 360
 aacagagcct tggacagga aacaagatct actcacaatc tccaggtgc agccctggac 420
 gctaathgaa ttatagtga ggtccagtc cctatcacca tagaagtga ggacatcaac 480
 20 gacaatcgac ccacgttht ccagtcaag tacgaaggct cagtaaggca gaactctgc 540
 ccaggaaagc cthcttgta tgtcaatgc acagacctg atgatccggc cactccaat 600
 25 ggccagctth attaccagat tgtcatccag thcccatga tcaacaatgt catgtactth 660
 cagatcaaca acaaacggg agccatctct thaccagg agggatctca ggaathgat 720
 cctgctaaga atccttcta taatctggt atctcagta aggacatggg agccagagt 780
 30 gagaathcct tcagtgatac cacatctgt gatatcatag tgacagagaa tathgaaa 840
 gcacaaaaac ctgtggagat ggtgaaaac tcaactgatc thaccccat caaathact 900
 caggtgcggt ggaathgatc cggtgcaca thctcttag ttgacaaaga gaagctgcca 960
 35 agathccat thcaathga ccaggaagga gatathtag tgactcagcc thtgaccga 1020
 gaagaaaagg atgathatgt ththtagca gttgaaagg atgagtagg aaaaccact 1080
 40 tcatatccgc tggaaathca tgtaaaagtt aaagathata atgathatc acctacatgt 1140
 ccgtcaccag taaccgtath tgaggthcag gaaathgaa gactgggtaa cagthcggg 1200
 accthactg cacatgacag ggatgaagaa aathctgcca acagththt aaactacagg 1260
 45 athgtggagc aaactccca actthccatg gatggactct thctaatca aactatgct 1320
 ggaathgtac agthtagta acagthctg aagaagcaag athctctca gtacaacta 1380
 acgathagag tgtctgaca agaththca accctthgt thtgcaaat cacgthath 1440
 50 gatathaat atcagathcc catctthgaa aathcagath atggaaacct gactctgct 1500
 gaagacacaa acathgggtc caccathth accathcagg cactgatgc tgathgacca 1560
 55 ththctggga gthctaaaat thctgathcat athataaagg gagacagta gggacgctg 1620
 ggggthgaca cagathccca thcaacacc ggathatgca thathaaaa gcctctgath 1680
 ththgaaacag cagctgtht caathgtg thcaagcag aaathctga gcctctagth 1740
 60 ththgtgta agthacaatgc aagthctth gcaagthca cgctthathg athcagathg 1800

65

	aatgaagcac	ctcaattttc	ccaacacgta	ttccaagcga	aagtcagtga	ggatgtagct	1860
5	ataggcacta	aagtgggcaa	tgtgactgcc	aaggatccag	aaggctctga	cataagctat	1920
	tcactgaggg	gagacacaag	aggttggctt	aaaattgacc	acgtgactgg	tgagatcttt	1980
	agtgtggctc	cattggacag	agaagccgga	agtccatata	gggtacaagt	ggtggccaca	2040
10	gaagtagggg	ggtcttcctt	gagctctgtg	tcagagttcc	acctgatcct	tatggatgtg	2100
	aatgacaacc	ctcccaggct	agccaaggac	tacacgggct	tgttcttctg	ccatcccctc	2160
15	agtgcacctg	gaagtctcat	tttcgaggct	actgatgatg	atcagcactt	atctcggggt	2220
	ccccatttta	cattttccct	cggcagtgga	agcttacaaa	acgactggga	agtttcctca	2280
	atcaatggta	ctcatgccc	actgtctacc	aggcacacag	agtttgagga	gagggagtat	2340
20	gtcgtcttga	tccgcatcaa	tgatgggggt	cggccaccct	tggaaggcat	tgtttcttta	2400
	ccagttacat	tctgcagttg	tgtggaagga	agttgtttcc	ggccagcagg	tcaccagact	2460
	gggataccca	ctgtgggcat	ggcagttggt	atactgctga	ccacccttct	ggtgattggt	2520
25	ataattttag	cagttgtggt	tatccgcata	aagaaggata	aaggcaaaga	taatgttgaa	2580
	agtgtctcaag	catctgaagt	caaacctctg	agaagctgaa	tttgaaaagg	aatgtttgaa	2640
30	tttatatagc	aagtgttatt	tcagcaacaa	ccatctcatc	ctattacttt	tcacttaacg	2700
	tgattataaa	ttttttaaac	agatattccc	tcttgtcctt	taatatttgc	taaataattc	2760
	tttttgagg	tggagtcttg	ctctgtcgcc	caggctggag	tacagtgggtg	tgatcccagc	2820
35	tcactgcaac	ctccgectcc	tgggttcaca	tgattctcct	gcctcagctt	cctaagtagc	2880
	tgggtttaca	ggcaccacc	accatgccc	gctaattttt	gtatttttaa	tagagacggg	2940
	gtttcgccat	ttggccaggc	tggctcttga	ctcctgacgt	caagtgatct	gcctgccttg	3000
40	gtctcccaat	acaggcatga	accactgcac	ccacctactt	agatatttca	tgtgctatag	3060
	acattagaga	gatttttcat	ttttccatga	catttttctt	ctctgcaaat	ggcttagcta	3120
45	cttgtgtttt	tcccttttgg	ggcaagacag	actcattaaa	tattctgtac	attttttctt	3180
	tatcaaggag	atatatcagt	gttgtctcat	agaactgcct	ggattccatt	tatgtttttt	3240
	ctgattccat	cctgtgtccc	cttcatcctt	gactcctttg	gtatttctact	gaatttcaaa	3300
50	catttgtcag	agaagaaaaa	cgtgaggact	caggaaaaat	aaataaataa	aagaacagcc	3360
	ttttccctta	gtattaacag	aatgttttct	gtgtcattaa	ccatctttta	tcaatgtgac	3420
	atgttgctct	ttggctgaaa	ttcttcaact	tggaaatgac	acagaccac	agaagggtgt	3480
55	caaacacaac	ctactctgca	aaccttggtg	aaggaaccag	tcagctggcc	agatttctct	3540
	actacctgcc	atgcatacat	gctgcgcatg	ttttcttcat	tcgtatgtta	gtaaagtttt	3600
60	ggttattata	tatttaacat	gtggaagaaa	acaagacatg	aaaagagtgg	tgacaaatca	3660
	agaataaaca	ctggttgtag	tcagttttgt	ttgttaa			3697

65

<210> 63
 <211> 503
 <212> ADN
 <213> humano

5

<400> 63

10	gacagcggct tccttgatcc ttgccacccg cgactgaaca ccgacagcag cagcctcacc	60
	atgaagttgc tgatggctct catgctggcg gccctctccc agcactgcta cgcaggctct	120
15	ggctgcccc tattggagaa tgtgatttcc aagacaatca atccacaagt gtctaagact	180
	gaatacaaag aacttcttca agagttcata gacgacaatg ccaactacaa tgccatagat	240
	gaattgaagg aatgttttct taaccaaacg gatgaaactc tgagcaatgt tgagggtgtt	300
20	atgcaattaa tatatgacag cagtctttgt gatttatttt aactttctgc aagacctttg	360
	gctcacagaa ctgcagggta tggtgagaaa ccaactacgg attgctgcaa accacacctt	420
	ctctttctta tgtcttttta ctacaaacta caagacaatt gttgaaacct gctatacatg	480
25	tttattttaa taaattgatg gca	503

<210> 64
 <211> 1894
 <212> ADN
 <213> humano

30

<400> 64

35	gtctgacttc ctcccagcac attcctgcac tctgccgtgt ccacactgcc ccacagaccc	60
	agtcctccaa gcctgctgcc agctccctgc aagcccctca ggttgggctt tgccacggtg	120
40	ccagcaggca gccctgggct gggggtaggg gactccctac aggcacgcag ccctgagacc	180
	tcagagggcc accccttgag ggtggccagg cccccagtgg ccaacctgag tgctgcctct	240
	gccaccagcc ctgctggccc ctggttccgc tgcccccca gatgcctggc tgagacacgc	300
45	cagtggcctc agctgcccac acctcttccc ggcccctgaa gttggcactg cagcagacag	360
	ctccctgggc accaggcagc taacagacac agccgccagc tcaaacagca gcggcatggg	420
50	cagcgcagc ccgggtctga gcagcgtatc cccagccac ctctgctgc cccccgacac	480
	ggtgtcgcgg acaggcttg agaaggcggc agcgggggca gtgggtctcg agagacggga	540
	ctggagtccc agtccacccg ccacgcccga gcagggcctg tccgcttct acctctcta	600
55	ctttgacatg ctgtaccctg aggacagcag ctgggcagcc aaggccccctg gggccagcag	660
	tcgggaggag ccacctgagg agcctgagca gtgcccggtc attgacagcc aagccccagc	720
	gggcagcctg gacttgggtgc ccggcgggct gaccttggag gagcactcgc tggagcaggt	780
60	gcagtccatg gtgggtggcg aagtgtctaa ggacatcgag acggcctgca agctgtctaa	840
	catcaccgca gatcccatgg actggagccc cagcaatgtg cagaagtggc tcctgtggac	900

65

agagcaccaa taccggctgc ccccatggg caaggccttc caggagctgg cgggcaagga 960
 5 gctgtgcgcc atgtcggagg agcagttccg ccagcgcctc cccctgggtg gggatgtgct 1020
 gcacgcccac ctggacatct ggaagtcagc ggccctggatg aaagagcggg cttcacctgg 1080
 ggcgattcac tactgtgcct cgaccagtga ggagagctgg accgacagcg aggtggactc 1140
 10 atcatgctcc gggcagccca tccacctgtg gcagttcctc aaggagttgc tactcaagcc 1200
 ccacagctat ggccgcttca ttaggtggct caacaaggag aagggcatct tcaaaattga 1260
 ggactcagcc caggtggccc ggctgtgggg catccgcaag aaccgtcccc ccatgaacta 1320
 15 cgacaagctg agccgctcca tccgccagta ttacaagaag ggcacatcc ggaagccaga 1380
 catctcccag cgcctcgtct accagtctgt gcaccccatc tgagtgcctg gccagggcc 1440
 tgaaaccgc cctcaggggc ctctctctg cctgcctgc ctcagccagg ccctgagatg 1500
 20 ggggaaaacg ggcagtctgc tctgctgctc tgacctcca gagcccaagg tcagggaggg 1560
 gcaaccaact gccccagggg gatatgggtc ctctggggcc ttcgggacca tggggcaggg 1620
 25 gtgcttctc ctcaggecca gctgctcccc tggaggacag agggagacag ggctgctccc 1680
 caacacctgc ctctgacccc agcatttcca gagcagagcc tacagaaggg cagtgactcg 1740
 acaaaggcca caggcagctc aggcctctct ctgctccatc cccctgcctc ccattctgca 1800
 30 ccacacctgg catggtgcag ggagacatct gcaccctga gttgggcagc caggagtgcc 1860
 cccgggaatg gataataaag atactagaga actg 1894

35
 <210> 65
 <211> 3029
 <212> ADN
 40 <213> humano
 <400> 65

45 ccaggcagct ggggtaagga gttcaaggca ggcgccacac ccgggggctc tccgcaacct 60
 gaccgctgt ccgctcccc acttcccgc ctccctcca cctactcatt caccaccca 120
 50 cccaccaga gccgggacgg cagcccaggc gccggggccc cgcctctcc tcgccgcat 180
 cctggacttc ctcttgcctc aggaccggc ttccacgtgt gtcccggagc cggcgtctca 240
 gcacacgctc cgtccgggc ctgggtgctt acagcagcca gagcagcagg gagtccggga 300
 55 cccgggcggc atctgggcca agttaggcgc cgcagaggcc agcgtgaac gtctccaggg 360
 ccggaggagc cgcggggcgt ccgggtctga gccgcagcaa atgggctccg acgtgcggga 420
 cctgaacgcg ctgctgccc ccgtcccctc cctgggtggc ggcggcggct gtgccctgcc 480
 60 tgtgagcggc gcggcgcagt gggcgcgggt gctggacttt gcgccccgg gcgcttcggc 540
 ttacgggtcg ttgggcggcc ccgcgccgc accggctccg ccgccacccc cgcctccgcc 600
 65 gcctcactcc ttcacaaac aggagccgag ctggggcggc gcggagccgc acgaggagca 660

5 gtgcctgagc gccttcaactg tccacttttc cggccagttc actggcacag ccggagcctg 720
 tcgctacggg cccttcggtc ctctccgcc cagccaggcg tcatccggcc aggccaggat 780
 gtttcctaac gcgccctacc tgcccagctg cctcgagagc cagcccgcta ttcgcaatca 840
 10 gggttacagc acggtcacct tcgacgggac gccagctac ggtcacacgc cctcgacca 900
 tgcggcgag tcccccaacc actcattcaa gcatgaggat cccatgggccc agcagggtc 960
 gctgggtgag cagcagtact cggtgccgcc cccggtctat ggctgccaca cccccaccga 1020
 15 cagctgcacc ggcagccagg ctttgctgct gaggacgcc tacagcagtg acaatttata 1080
 ccaaatgaca tcccagcttg aatgcatgac ctggaatcag atgaacttag gagccacctt 1140
 aaagggagtt gctgctggga gctccagctc agtgaaatgg acagaagggc agagcaacca 1200
 20 cagcacaggg tacgagagcg ataaccacac aacgccatc ctctgaggag cccaatacag 1260
 aatacacacg cacgggtgtct tcagaggcat tcaggatgtg cgacgtgtgc ctggagtagc 1320
 25 cccgactctt gtacggtcgg catctgagac cagtgagaaa cgccccttca tgtgtgctta 1380
 cccaggctgc aataagagat attttaagct gtcccactta cagatgcaca gcaggaagca 1440
 cactggtgag aaaccatacc agtgtgactt caaggactgt gaacgaaggt tttctcgttc 1500
 30 agaccagctc aaaagacacc aaaggagaca tacaggtgtg aaaccattcc agtgtaaaac 1560
 ttgtcagcga aagttctccc ggtccgacca cctgaagacc cacaccagga ctcatcagg 1620
 taaaacaagt gaaaagccct tcagctgtcg gtggccaagt tgcagaaaa agtttgcccg 1680
 35 gtcagatgaa ttagtccgcc atcacaacat gcatcagaga aacatgacca aactccagct 1740
 ggcgctttga ggggtctccc tcggggaccg ttcagtgtcc caggcagcac agtgtgtgaa 1800
 40 ctgctttcaa gtctgactct ccactcctcc tactaaaaa ggaaacttca gttgatcttc 1860
 ttcattcaac ttccaagaca agataccggt gcttctggaa actaccaggt gtgcctggaa 1920
 gagttggtct ctgccctgcc tacttttagt tgactcacag gccctggaga agcagctaac 1980
 45 aatgtctggt tagttaaag cccattgcca tttggtgtgg attttctact gtaagaagag 2040
 ccatagctga tcatgtcccc ctgacccttc ctttctttt ttatgctcgt tttcgtggg 2100
 gatggaatta ttgtaccatt ttctatcatg gaatatttat aggccagggc atgtgtatgt 2160
 50 gtctgctaata gtaaactttg tcatggtttc cattactaa cagcaacagc aagaaataaa 2220
 tcagagagca aggcacggg ggtgaatctt gtctaacatt cccgaggtca gccaggctgc 2280
 55 taacctggaa agcaggatgt agttctgcca ggcaactttt aaagctcatg catttcaagc 2340
 agctgaagaa aaaatcagaa ctaaccagta cctctgtata gaaatctaaa agaattttac 2400
 cattcagtta attcaatgtg aacactggca cactgctctt aagaaactat gaagatctga 2460
 60 gatttttttg tgtatgtttt tgactctttt gagtggaat catatgtgtc tttatagatg 2520

65

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

25 <210> 66
 <211> 1064
 <212> ADN
 <213> humano

30 <400> 66

tttgaggcca tataaagtca cctgaggccc tctccaccac agcccaccag tgaccatgaa 60
 35 ggctgtgctg cttgccctgt tgatggcagg cttggccctg cagccaggca ctgccctgct 120
 gtgctactcc tgcaaagccc aggtgagcaa cgaggactgc ctgcaggtgg agaactgcac 180
 ccagctgggg gagcagtgct ggaccgcgcg catccgcgca gttggcctcc tgaccgtcat 240
 40 cagcaaaggc tgcagcttga actgcgtgga tgactcacag gactactacg tgggcaagaa 300
 gaacatcacg tgctgtgaca ccgacttgtg caacgccagc ggggcccattg ccctgcagcc 360
 ggctgctgcc atccttgccg tgctccctgc actcggcctg ctgctctggg gaccggcca 420
 45 gctctaggct ctggggggcc ccgctgcagc ccacactggg tgtggtgccc caggcctctg 480
 tgccactcct cacacacccg gccagtgagg agcctgtcct ggttcctgag gcacatccta 540
 acgcaagtct gaccatgtat gtctgcgccc ctgtcccca ccctgacct cccatggccc 600
 50 tctccaggac tcccaccccg cagatcggct ctattgacac agatccgcct gcagatggcc 660
 cctccaacc tctctgctgc tgtttccatg gccagcatt ctccaccctt aaccctgtgc 720
 tcaggcacct cttccccag gaagccttc ctgccacc cactatgac ttgagccagg 780
 55 tctggtccgt ggtgtcccc gcaccagca ggggacaggc actcaggagg gcccggtaaa 840
 ggctgagatg aagtggactg agtagaactg gaggacagga gtcgacgtga gttcctggga 900
 60 gtctccagag atggggcctg gaggcctgga ggaaggggccc aggcctcaca ttcgtggggc 960
 tcctgaatg gcagcctcag cacagcgtag gcccttaata aacacctgtt ggataagcca 1020
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaa 1064

65

ES 2 494 843 T3

<210> 67
 <211> 6962
 <212> ADN
 <213> humano

5

<400> 67

10 gcaagaactg caggggagga ggacgctgcc acccaCagcc tctagagctc attgcagctg 60
 ggacagcccg gagtgtgggt agcagctcgg caagcgtgc ccaggtcctg ggggtggggc 120
 agccagcggg agcaggaaag gaagcatggt cccaggctgc ccacgcctct gggtcctggt 180
 15 ggtcttgggc accagctggg taggctgggg gagccaaggg acagaagcgg cacagctaag 240
 gcagttctac gtggctgctc agggcatcag ttggagctac cgacctgagc ccacaaactc 300
 aagtttgaat ctttctgtaa cttcctttaa gaaaattgtc tacagagagt atgaaccata 360
 20 ttttaagaaa gaaaaaccac aatctacat ttcaggactt cttgggccta ctttatatgc 420
 tgaagtcgga gacatcataa aagttcactt taaaaataag gcagataagc ccttgagcat 480
 ccatcctcaa ggaattaggt acagtaaatt atcagaaggt gcttcttacc ttgaccacac 540
 25 attccctgcg gagaagatgg acgacgctgt ggctccaggg cgagaataca cctatgaatg 600
 gagtatcagt gaggacagtg gaccaccca tgatgacct ccatgcctca cacacatcta 660
 ttactcccat gaaaatctga tCGaggattt caactcgggg ctgattgggc ccctgcttat 720
 30 ctgtaaaaaa gggaccctaa ctgaggggtg gacacagaag acgtttgaca agcaaatcgt 780
 gctactatit gctgtgtttg atgaaagcaa gagctggagc cagtcacat ccctaattgta 840
 cacagtcaat ggatatgtga atgggacaat gccagatata acagtttgtg cccatgacca 900
 35 catcagctgg catctgctgg gaatgagctc ggggccagaa ttattctcca ttcatttcaa 960
 cggccaggtc ctggagcaga accatcataa ggtctcagcc atcaccttg tcagtgctac 1020
 atccactacc gcaaatatga ctgtgggccc agagggaaag tggatcatat cttctctcac 1080
 40 cccaaaacat ttgcaagctg ggatgcaggg ttacattgac attaaaaact gccc aaagaa 1140
 aaccaggaat cttagaaaa taactcgtga gcagagggcg cacatgaaga ggtgggaata 1200
 cttcattgct gcagaggaag tcatttggga ctatgcacct gtaataccag cgaatatgga 1260
 45 caaaaaatac aggtctcagc atttggataa tttctcaaac caaattggaa aacattataa 1320
 gaaagtatg tacacacagt acgaagatga gtccttcacc aaacatacag tgaatcccaa 1380
 50 tatgaaagaa gatgggattt tgggtcctat tatcagagcc caggtcagag acacactcaa 1440
 aatcgtgttc aaaaatatgg ccagccgcc ctatagcatt taccctcatg gagtgacctt 1500
 ctgccttat gaagatgaag tcaactctt cttcacctca ggcaggaaca acaccatgat 1560
 55 cagagcagtt caaccagggg aaacctatac ttataagtgg aacatcttag agtttgatga 1620
 acccacagaa aatgatgcc agtgcttaac aagaccatac tacagtgacg tggacatcat 1680
 gagagacatc gcctctgggc taataggact acttctaate tgtaagagca gatccctgga 1740

60

65

5 caggcgagga atacagaggg cagcagacat cgaacagcag gctgtgtttg ctgtgtttga 1800
 tgagaacaaa agctggtacc ttgaggacaa catcaacaag ttttgtgaaa atcctgatga 1860
 ggtgaaacgt gatgacccca agttttatga atcaaacatc atgagcacta tcaatggcta 1920
 10 tgtgcctgag agcataacta ctcttggatt ctgctttgat gacactgtcc agtggcactt 1980
 ctgtagtgtg gggaccaga atgaaattht gaccatccac ttcactgggc actcattcat 2040
 ctatggaaag aggcatgagg acaccttgac cctcttcccc atgctgtggag aatctgtgac 2100
 15 ggtcacaatg gataatggtg gaacttggat gttacttcc atgaattcta gtccaagaag 2160
 caaaaagctg aggctgaaat tcagggatgt taaatgtatc ccagatgatg atgaagactc 2220
 atatgagatt tttgaacctc cagaatctac agtcatggct acacggaaaa tgcattgatcg 2280
 20 tttagaacct gaagatgaag agagtgatgc tgactatgat taccagaaca gactggctgc 2340
 agcattagga atcaggatcat tccgaaactc atcattgaat caggaagaag aagagttaa 2400
 tcttactgcc ctactctgg agaatggcac tgaattcgtt tcttcaaca cagatataat 2460
 25 tgttggttca aattattctt cccaagtaa tattagtaag ttcactgtca ataacctgc 2520
 agaacctcag aaagcccctt ctaccaaca agccaccaca gctggttccc cactgagaca 2580
 30 cctcattggc aagaactcag ttctcaattc ttccacagca gagcattcca gcccatattc 2640
 tgaagacctt atagaggatc ctctacagcc agatgtcaca gggatacgtc tactttact 2700
 tgggtgctgga gaattcaaaa gtcaagaaca tgctaagcat aagggacca aggtagaaag 2760
 35 agatcaagca gcaaagcaca ggttctcctg gatgaaatta ctacacata aagttgggag 2820
 acacctaagc caagacactg gttctcttc cggaatgagg ccctgggagg accttctag 2880
 ccaagacact ggttctctt ccagaatgag gccctggaag gaccctcta gtgatctgtt 2940
 40 actcttaaaa caaagtaact catctaagat tttggttggg agatggcatt tggcttctga 3000
 gaaaggtagc tatgaaataa tccaagatac tgatgaagac acagctgta acaattggct 3060
 gatcagcccc cagaatgcct cacgtgcttg gggagaaagc accctcttg ccaacaagcc 3120
 45 tggaaagcag agtggccacc caaagtttc tagagttaga cataaatctc tacaagtaag 3180
 acaggatgga ggaaagagta gactgaagaa aagccagttt ctattaaga cacgaaaaaa 3240
 50 gaaaaaagag aagcacacac accatgctcc tttatctccg aggaccttc accttctaag 3300
 aagtgaagcc tacaacacat tttcagaaag aagacttaag cattcgttgg tgcttcataa 3360
 atccaatgaa acatctctc ccacagacct caatcagaca ttgccctcta tggattttgg 3420
 55 ctggatagcc tcaattctg accataatca gaattcctca aatgacactg gtcaggcaag 3480
 ctgtcctcca ggtctttatc agacagtgcc cccagaggaa cactatcaaa cattccccat 3540
 60 tcaagacctt gatcaaatgc actctacttc agacccagc cacagatcct cttctccaga 3600
 gctcagtga atgcttgagt atgaccgaag tcacaagtcc tccccacag atataagtca 3660

65

	aatgtcccct tcctcagaac atgaagtctg gcagacagtc atctctccag acctcagcca	3720
5	ggtgaccctc tctccagaac tcagccagac aaacctctct ccagacctca gccacacgac	3780
	tctctctcca gaactcattc agagaaacct ttccccagcc ctcggtcaga tgcccatttc	3840
10	tccagacctc agccatacaa ccctttctcc agacctcagc catacaacct tttctttaga	3900
	cctcagccag acaaacctct ctccagaact cagtcagaca aacctttctc cagccctcgg	3960
	tcagatgccc ctttctccag acctcagcca tacaacctt tctctagact tcagccagac	4020
15	aaacctctct ccagaactca gccatagac tctctctcca gaactcagtc agacaaacct	4080
	ttccccagcc ctcggtcaga tgcccatttc tccagacctc agccatacaa ccctttctct	4140
	agacttcagc cagacaaacc tctctccaga actcagtcaa acaaaccttt ccccagccct	4200
20	cggtcagatg cccctttctc cagaccccag ccatacaacc ctttctctag acctcagcca	4260
	gacaaacctc tctccagaac tcagtcagac aaacctttcc ccagacctca gtgagatgcc	4320
	cctctttgca gatctcagtc aaattcccct taccacagac ctcgaccaga tgacactttc	4380
25	tccagacctt ggtgagacag atctttcccc aaactttggt cagatgtccc tttccccaga	4440
	cctcagccag gtgactctct ctccagacat cagtgacacc acccttctcc cggatctcag	4500
30	ccagatatca cctcctccag accttgatca gatattctac ctttctgaat ctagtctcagc	4560
	attgcttctt caagaattta atgagtcttt tccttatcca gaccttggtc agatgccatc	4620
	tccttcatct cctactctca atgatacttt tctatcaaag gaatttaatc cactggttat	4680
35	agtgggcctc agtaaagatg gtacagatta cattgagatc attccaaagg aagaggcca	4740
	gagcagtgaa gatgactatg ctgaaattga ttatgtgccc tatgatgacc cctacaaaac	4800
	tgatgttagg acaaacatca actcctccag agatcctgac aacattgcag catggtacct	4860
40	ccgcagcaac aatggaaaca gaagaaatta ttacattgct gctgaagaaa tatcctggga	4920
	ttattcagaa tttgtacaaa gggaaacaga tattgaagac tctgatgata ttccagaaga	4980
45	taccacatat aagaaagtag tttttcgaaa gtacctcgac agcactttta ccaaactgta	5040
	tcctcgaggg gagtatgaag agcatctcgg aattcttggt cctattatca gagctgaagt	5100
	ggatgatggt atccaagttc gttttaaaaa tttagcatcc agaccgtatt ctctacatgc	5160
50	ccatggactt tcctatgaaa aatcatcaga gggaaagact tatgaagatg actctcctga	5220
	atggtttaag gaagataatg ctgttcagcc aaatagcagt tataacctacg tatggcatgc	5280
	cactgagcga tcagggccag aaagtcctgg ctctgcctgt cgggcttggg cctactactc	5340
55	agctgtgaac ccagaaaaag atattcactc aggcttgata ggtcccctcc taatctgcca	5400
	aaaaggaata ctacataagg acagcaacat gcctatggac atgagagaat ttgtcttact	5460
60	atztatgacc tttgatgaaa agaagagctg gtactatgaa aagaagtccc gaagttcttg	5520
65		

5 gagactcaca tcctcagaaa tgaaaaaatc ccatgagttt cacgccatta atgggatgat 5580
ctacagcttg cctggcctga aaatgtatga gcaagagtgg gtgaggttac acctgctgaa 5640
cataggcggc tccaagaca ttcacgtggt tcactttcac ggccagacct tgctggaaaa 5700
10 tggcaataaa cagcaccagt taggggtctg gcccttctg cctggttcat ttaaaactct 5760
tgaaatgaag gcatcaaac ctggctggtg gctcctaac acagaggttg gagaaaacca 5820
gagagcaggg atgcaaacgc catttcttat catggacaga gactgtagga tgccaatggg 5880
15 actaagcact ggtatcatat ctgattcaca gatcaaggct tcagagtttc tgggttactg 5940
ggagcccaga ttagcaagat taaacaatgg tggatcttat aatgcttggg gtgtagaaaa 6000
acttgcagca gaatttgcct ctaaaccttg gatccagggtg gacatgcaa aggaagtcac 6060
20 aatcacaggg atccagacc aaggtgcaa aactacctg aagtcctgct ataccacaga 6120
gttctatgta gcttacagtt ccaaccagat caactggcag atcttcaaag ggaacagcac 6180
25 aaggaatgtg atgtatttta atggcaattc agatgcctct acaataaaag agaatcagtt 6240
tgaaccacct attgtggcta gatatattag gatctctcca actcgagcct ataacagacc 6300
taccttctga ttggaactgc aaggttctga ggtaaatgga tgtccacac ccctgggtat 6360
30 ggaaaatgga aagatagaaa acaagcaaat cacagcttct tcgtttaaga aatcttgggtg 6420
gggagattac tgggaaccct tccgtgcccg tctgaatgcc cagggacgtg tgaatgcctg 6480
35 gcaagccaag gcaacaaca ataagcagtg gctagaaatt gatctactca agatcaagaa 6540
gataacggca attataacac agggctgcaa gtctctgtcc tctgaaatgt atgtaaagag 6600
ctataccatc cactacagtg agcagggagt ggaatggaaa ccatacaggc tgaaatctc 6660
40 catggtggac aagatttttg aaggaaatac taataccaaa ggacatgtga agaacttttt 6720
caacccccca atcatttcca ggtttatccg tgtcattcct aaaacatgga atcaaagtat 6780
45 tgcacttctc ctggaactct ttggctgtga tatttactag aattgaacat tcaaaaacc 6840
ctggaagaga ctctttaaga cctcaaacca tttagaatgg gcaatgtatt ttacgctgtg 6900
ttaaagtta acagttttcc actatttctc tttctttct attagtgaat aaaattttat 6960
50 ac 6962

<210> 68
<211> 1464
<212> ADN
55 <213> humano
<400> 68

60 agccccaagc ttaccacctg caccgggaga gctgtgtcac catgtgggtc ccggttctct 60
tcctcaccct gtccgtgacg tggattgggt ctgcaccct catcctgtct cggattgtgg 120
65 gaggctggga gtgcgagaag cattccaac cctggcaggt gcttgtggcc tctcgtggca 180

5 ggcagctctg cggcgggtgtt ctgggtgcacc cccagtgggg cctcacagct gcccaactgca 240
 tcaggaacaa aagcgtgata ttgctgggtc ggcacagcct gtttcatcct gaagacacag 300
 gccaggattt tcaggtcagc cacagcttcc cacaccgct ctacgatatg agcctcctga 360
 10 agaatcgatt cctcaggcca ggtgatgact ccagccacga cctcatgctg ctccgctgt 420
 cagagcctgc cgagctcacg gatgctgtga aggtcatgga cctgcccacc caggagccag 480
 cactggggac cacctgctac gcctcaggct ggggcagcat tgaaccagag gagttcttga 540
 ccccaaagaa acttcagtgt gtggacctcc atgttatttc caatgacgtg tgtgcaag 600
 15 ttcaccctca gaaggtgacc aagttcatgc tgtgtgctgg acgctggaca gggggcaaaa 660
 gcacctgctc ggggtgattct gggggccac ttgtctgtaa tgggtgtgctt caaggtatca 720
 cgtcatgggg cagtgaacca tgtgccctgc ccgaaaggcc ttcctgtac accaagggtg 780
 20 tgcattaccg gaagtggatc aaggacacca tcgtggccaa cccctgagca cccctatcaa 840
 ccccctattg tagtaaactt ggaaccttgg aatgaccag gccaaagactc aagcctcccc 900
 agttctactg acctttgtcc ttaggtgtga ggtccagggt tgctaggaaa agaaatcagc 960
 25 agacacaggt gtagaccaga gtgtttctta aatggtgtaa tttgtcctc tctgtgtcct 1020
 ggggaatact ggccatgcct ggagacatat cactcaattt ctctgaggac acagatagga 1080
 30 tgggggtgtct gtgttatttg tggggtacag agatgaaaga ggggtgggat ccacactgag 1140
 agagtggaga gtgacatgtg ctggacactg tccatgaagc actgagcaga agctggaggc 1200
 acaacgcacc agacactcac agcaaggatg gagctgaaaa cataaccac tctgtcctgg 1260
 35 aggcactggg aagcctagag aaggctgtga gccaaaggag gagggcttc ctttggcatg 1320
 ggatggggat gaagtaagga gagggactgg accccctgga agctgattca ctatgggggg 1380
 aggtgtattg aagtcctcca gacaaccctc agatttgatg atttcctagt agaactcaca 1440
 40 gaaataaaga gctgttatac tgtg 1464

45 <210> 69
 <211> 1793
 <212> ADN
 <213> humano

50 <400> 69

55 cgcgtccgcc ccgcgagcac agagcctcgc ctttgccgat ccgccgccg tccacaccgg 60
 ccgccagctc accatggatg atgatatcgc cgcgctcgtc gtcgacaacg gctccggcat 120
 gtgcaaggcc ggcttcgcgg gcgacgatgc cccccgggcc gtcttcccct ccatcgtggg 180
 gcgccccagg caccagggcg tgatgggtgg catgggtcag aaggattcct atgtgggcga 240
 60 cgaggcccag agcaagagag gcatectcac cctgaagtac cccatcgagc acggcatcgt 300
 caccaactgg gacgacatgg agaaaatctg gcaccacacc ttctacaatg agctgcgtgt 360

65

5 ggctcccgag gagcaccgag tgctgctgac cgaggccccc ctgaacccca aggccaaccg 420
 cgagaagatg acccagatca tgtttgagac cttcaacacc ccagccatgt acgttgctat 480
 ccaggctgtg ctatccctgt acgcctctgg ccgtaccact ggcacgtga tggactccgg 540
 tgacggggtc acccacactg tgcccatcta cgaggggtat gccctcccc atgccatcct 600
 10 gcgtctggac ctggctggcc gggacctgac tgactacctc atgaagatcc tcaccgagcg 660
 cggctacagc ttcaccacca cggccgagcg ggaaatcgtg cgtgacatta aggagaagct 720
 gtgctacgtc gccctggact tcgagcaaga gatggccacg gctgcttcca gtcctccct 780
 15 ggagaagagc tacgagctgc ctgacggcca ggtcatcacc attggcaatg agcggttccg 840
 ctgccctgag gcactcttcc agccttcctt cctgggcatg gagtctgtg gcatccacga 900
 aactaccttc aactccatca tgaagtgtga cgtggacatc cgcaaagacc tgtacgcaa 960
 20 cacagtgtg tctggcggca ccaccatgta ccctggcatt gccgacagga tgcagaagga 1020
 gatcactgcc ctggcaccca gcacaatgaa gatcaagatc attgctcctc ctgagcgcaa 1080
 25 gtactccgtg tggatcggcg gctccatcct ggcctcgtg tccaccttcc agcagatgtg 1140
 gatcagcaag caggagtatg acgagtcggg cccctccatc gtccaccgca aatgcttcta 1200
 ggcggactat gacttagttg cgttacacc tttcttgaca aaacctaact tgcgcagaaa 1260
 30 acaagatgag attggcatgg ctttatttgt tttttttgtt ttgttttggg tttttttttt 1320
 tttttggctt gactcaggat ttaaaaactg gaacggtgaa ggtgacagca gtcggttggg 1380
 gcgagcatcc cccaaagttc acaatgtggc cgaggacttt gattgcacat tgttgttttt 1440
 35 ttaatagtca ttccaaatat gagatgcatt gttacaggaa gtcccttgcc atcctaaaag 1500
 ccacccact tctctctaag gagaatggcc cagtcctctc ccaagtccac acaggggagg 1560
 40 tgatagcatt gctttcgtgt aaattatgta atgcaaaatt tttttaatct tcgccttaat 1620
 acttttttat tttgttttat tttgaatgat gagccttctg gccccctt ccccctttt 1680
 gtcccccaac ttgagatgta tgaaggctt tggctcctt gggagtgggt ggaggcagcc 1740
 45 agggcttacc tgtacactga cttgagacca gttgaataaa agtgcacacc tta 1793

50 <210> 70
 <211> 1526
 <212> ADN
 <213> humano

<400> 70

55 ccggaagtga cgcgaggctc tgcggagacc aggagtcaga ctgtaggacg acctcgggtc 60
 ccacgtgtcc ccggtactcg ccggccggag cccccggctt cccggggccg ggggacctta 120
 60 gcggcaccca cacacagcct actttccaag cggagccatg tctggtaacg gcaatgcggc 180
 tgcaacggcg gaagaaaaca gcccaaagat gagagtgatt cgcgtgggta cccgcaagag 240

65

5 ccagcttgct cgcatacaga cggacagtgt ggtggcaaca ttgaaagcct cgtaccctgg 300
 cctgcagttt gaaatcattg ctatgtccac cacaggggac aagattcttg atactgcact 360
 ctctaagatt ggagagaaaa gcctgtttac caaggagctt gaacatgccc tggagaagaa 420
 tgaagtggac ctggttgctt actccttgaa ggacctgccc actgtgcttc ctctggctt 480
 10 caccatcgga gccatctgca agcgggaaaa ccctcatgat gctgttgtct ttcacccaaa 540
 atttgttggg aagaccctag aaaccctgcc agagaagagt gtggtgggaa ccagctccct 600
 gcgaagagca gcccagctgc agagaaagt cccgcactctg gagttcagga gtattcgggg 660
 15 aaacctcaac acccgcttc ggaagctgga cgagcagcag gagttcagtg ccatcatcct 720
 ggcaacagct ggctgcagc gcatgggctg gcacaaccgg gtggggcaga tctgcaccc 780
 tgaggaatgc atgtatgctg tgggccaggg ggccttgggc gtggaagtgc gagccaagga 840
 20 ccaggacatc ttggatctgg tgggtgtgct gcacgatccc gagactctgc ttcgctgcat 900
 cgctgaaagg gccttctga ggcacctgga aggaggctgc agtgtgccag tagccgtgca 960
 tacagctatg aaggatggc aactgtacct gactggagga gtctggagtc tagacggctc 1020
 25 agatagcata caagagacca tgcaggctac catccatgtc cctgcccagc atgaagatgg 1080
 ccctgaggat gaccacagt tggtaggcat cactgctcgt aacattccac gagggcccca 1140
 gttggctgcc cagaacttg gcatcagcct ggccaacttg ttgctgagca aaggagccaa 1200
 30 aacatcctg gatgttgac ggcagcttaa cgatgccat taactggtt gtggggcaca 1260
 gatgcctggg ttgctgctgt ccagtgccta catcccggc ctcagtgcc cattctcact 1320
 35 gctatctggg gagtgattac cccgggagac tgaactgcag ggttcaagcc ttccagggat 1380
 ttgcctcacc ttggggcctt gatgactgcc ttgcctctc agtatgtggg ggcttcatct 1440
 ctttagagaa gtccaagcaa cagccttga atgtaaccaa tctactaat aaaccagttc 1500
 40 tgaaggtgta aaaaaaaaa aaaaaa 1526

45 <210> 71
 <211> 2397
 <212> ADN
 <213> humano

<400> 71

50 gcaagaactg aaacgaatgg ggattgaact gctttgcctg ttctttctat ttctaggaag 60
 gaatgatcac gtacaagggtg gctgtgccct gggagggtgca gaaacctgtg aagactgcct 120
 55 gcttattgga cctcagtgtg cctggtgtgc tcaggagaat tttactcatc catctggagt 180
 tggcgaaagg tgtgataccc cagcaaact ttagctaaa ggatgtcaat taaacttcat 240
 cgaaaacctt gtctccaag tagaaatact taaaaataag cctctcagtg taggcagaca 300
 60 gaaaaatagt tctgacattg ttcagattgc gcctcaaagc ttgatcctta agttgagacc 360

65

aggtggtgcg cagactctgc aggtgcatgt ccgccagact gaggactacc cgggtggattt 420
 5 gtattacctc atggacctct ccgcctccat ggatgacgac ctcaacacaa taaaggagct 480
 gggctccccg ctttccaaag agatgtctaa attaaccagc aactttagac tgggcttcgg 540
 atcttttgtg gaaaaacctg tatccccctt cgtgaaaaca acaccagaag aaattgccaa 600
 10 cccttgacgt agtattccat acttctgttt acctacattt ggattcaagc acattttgcc 660
 attgacaaat gatgctgaaa gattcaatga aattgtgaag aatcagaaaa tttctgctaa 720
 tattgacaca cccgaagggt gatttgatgc aattatgcaa gctgctgtgt gtaaggaaaa 780
 15 aattggctgg cggaatgact ccctccacct cctggctctt gtgagtgatg ctgattctca 840
 ttttggaatg gacagcaaac tagcaggcat cgtcattcct aatgacgggc tctgtcactt 900
 ggacagcaag aatgaatact ccatgtcaac tgtcttggaa tatccaacaa ttggacaact 960
 20 cattgataaa ctggtacaaa acaactgttt attgatcttc gctgtaacc aagaacaagt 1020
 tcatttatat gagaattacg caaaacttat tcctggagct acagtaggtc tacttcagaa 1080
 ggactccgga aacattctcc agctgatcat ctacagcttat gaagaactgc ggtctgaggt 1140
 25 ggaactggaa gtattaggag aactgaagg actcaacttg tcatttacag ccatctgtaa 1200
 caacggtacc ctcttccaac accaaaagaa atgctctcac atgaaagtgg gagacacagc 1260
 ttccttcagc gtgactgtga atatcccaca ctgcgagaga agaagcaggc acattatcat 1320
 30 aaagcctgtg gggctggggg atgccctgga attacttgtc agcccagaat gcaactgca 1380
 ctgtcagaaa gaagtggaag tgaacagctc caaatgtcac cacgggaacg gctctttcca 1440
 35 gtgtgggggtg tgtgctgcc accctggcca catggggcct cgctgtgagt gtggcgagga 1500
 catgctgagc acagattcct gcaaggaggc cccagatcat cctcctgca gcggaagggg 1560
 tgactgctac tgtgggcagt gtatctgcca cttgtctccc tatggaaaca tttatgggcc 1620
 40 ttattgccag tgtgacaatt tctcctgctg gagacacaaa gggctgctct gcggaggtaa 1680
 cggcgactgt gactgtgggtg aatgtgtgtg caggagcggc tggactggcg agtactgcaa 1740
 ctgcaccacc agcacggact cctgctctc tgaagatgga gtgctctgca gcgggcgagg 1800
 45 ggactgtgtt tgtggcaagt gtgtttgcac aaaccctgga gcctcaggac caacctgtga 1860
 acgatgtcct acctgtgggtg acccctgtaa ctctaaacgg agctgcattg agtgccacct 1920
 gtcagcagct ggccaagccc gagaagaatg tgtggacaag tgcaaaactag ctggtgagc 1980
 50 catcagtgaa gaagaagatt tctcaaagga tggttctgtt tcctgctctc tgcaaggaga 2040
 aatgaatgt cttattacat tcctaataac tacagataat gaggggaaaa ccatcattca 2100
 cagcatcaat gaaaaagatt gtccgaagcc tccaacatt ccatgatca tgtaggggt 2160
 55 ttcctggct attcttctca tcggggtgt cctactgtgc atctggaagc tactggtgtc 2220
 atttcatgat cgtaaagaag ttgccaaatt tgaagcagaa cgatcaaaag ccaagtggca 2280

60

65

ES 2 494 843 T3

	aacgggaacc aatccactct acagaggatc cacaagtact tttaaaaatg taacttataa	2340
5	acacagggaa aaacaaaagg tagaccttc cacagattgc tagaactact ttatgca	2397
10	<210> 72 <211> 2118 <212> ADN <213> humano	
15	<400> 72	
15	tggggagccc aagcagaaac gcaagctggt ggctgagggtg tccctgcaga acccgctccc	60
20	tgtggccctg gaaggctgca ctttactgt ggagggggcc ggctgactg aggagcagaa	120
20	gacggtggag atcccagacc cctggaggc aggggaggaa gttaagggtga gaatggacct	180
25	gctgccgctc cacatgggccc tccacaagct ggtggtgaac ttcgagagcg acaagctgaa	240
25	ggctgtgaag ggcttccgga atgtcatcat tggccccgccc-taagggacct ctgctcccag	300
25	cctgctgaga gccccacct tgatcccaat ctttatccca agctagttag caaaatatgc	360
30	cccttcttgg gccccagacc ccagggcagg gtgggcagcc tatgggggct ctcggaatg	420
30	gaatgtgccc ctggcccatc tcagcctcct gagcctgtgg gtccccactc acccccttgg	480
35	ctgtgaggaa tgctctgtgc cagaaacagt gggagccctg accttggctg actggggctg	540
35	gggtgagaga ggaaagacct acattccctc tctgcccag atgcccttg gaaagccatt	600
35	gaccaccac catattgttt gatctacttc atagctcctt ggagcaggca aaaaaggac	660
40	agcatgcccc ttggctggat cagggaatcc agctccctag actgcatccc gtacctctc	720
40	ccatgactgc acccagctcc agggggcctt gggacagcca gagctgggtg gggacagtga	780
40	taggcccag gtcccccca catcccagca gcccaagctt aatagccctc cccctcaacc	840
45	tcaccattgt gaagcaccta ctatgtgctg ggtgcctccc acacttgctg gggctcacgg	900
45	ggcctccaac ccatttaatc accatgggaa actgtttgtg gcgctgcttc caggataagg	960
45	agactgaggc ttagagagag gaggcagccc cctccacacc agtggcctcg tggttattag	1020
50	caaggctggg taatgtgaag gcccaagagc agagtctggg cctctgactc tgagtccact	1080
50	gtccattta taaccaccag ctgacctgag actgtcggag aggctgtctg gggcctttat	1140
55	caaaaaaaga ctgagccaag acaaggagggt agagagggga ctgggggact gggagtcaga	1200
55	gccctggctg ggttcaggtc ccacgtctgg ccaggcactg ccttctcctc tctgggcctt	1260
55	tgtttccttg ttggtcagag gagtgattga accagctcat ctccaaggat cctctccact	1320
60	ccatgtttgc aatgctttta tatggcccag ccttgtaaata aaccacaagg tccactcct	1380
60	gctccacgaa gccttaagcc ataggcccag gatatttctg agagtgaaac catgactgtg	1440
60	accaccttct gtccccagcc ctgtcctggt tccttcctat gccaggtac cacccttcag	1500
65	acccagttc taggggagaa gagccctgga caccctgct ctaccatga gcctgcccgc	1560

5 tgcaatgcct agacttccca acagccttag ctgccagtgc tggctactaa ccaacaaggt 1620
 tggcacccca gctaccctt ctttgcaggg ctaaggcccc caaacatagc ccctgccccg 1680
 gaggaagctt ggggaacca tgagttgtca gctttgactt tatctcctgc tctttctaca 1740
 10 tgactgggcc tcccttgggc tggaagaatt ggggattctc tattggaggt gagatcacag 1800
 cctccagggc cccccaatc ccaggggaagg acttggagag aatcatgctg ttgcatttag 1860
 aactttctgc ttgacacagg aaagagtcac acaattaatc aacatgtata ttttctctat 1920
 15 acatagagct ctatttctct acggttttat aaaagccttg ggttccaacc aggcagtaga 1980
 tgtgcttctg aaccgcaagg agcaaacact gaaataaaat agtttatttt tcacactcaa 2040
 aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2100
 20 aaaaaaaaa aaaaaaaaaa 2118

25 <210> 73
 <211> 2832
 <212> ADN
 <213> humano

30 <400> 73

 aaagctcaaa ccgacaccct cacgcagatg atgacatcaa ctcttttttc ttccccaagt 60
 35 gtacacaatg tgatggagac tgttacgcag gagacagctc ctccagatga aatgaccaca 120
 tcatttccct ccagtgctac caacacactc atgatgacat caaagactat aacaatgaca 180
 acctccacag actccactct tggaaacaca gaagagacat caacagcagg aactgaaagt 240
 40 tctaccccag tgacctcagc agtctcaata acagctggac aggaaggaca atcacgaaca 300
 acttcttggg ggaaccttat ccaagacaca tcagcttctt ctcagaacca ctggactcgg 360
 agcacgcaga ccaccaggga atctcaaacc agcacctaa cacacagaac cacttcaact 420
 45 ccttctttct ctccaagtgt acacaatgtg acagggactg tttctcagaa gacatctcct 480
 tcaggtgaaa cagctacctc atccctctgt agtgtcacia acacatccat gatgacatca 540
 gagaagataa cagtgacaac ctccacaggc tccactcttg gaaaccagg ggagacatca 600
 50 tcagtacctg ttactggaag tcttatgcca gtcacctcag cagccttagt aacagttgat 660
 ccagaaggac aatcaccagc aactttctca aggacttcta ctcaggacac aacagctttt 720
 55 tctaagaacc accagactca gagcgtggag accaccagag tatctcaaat caacaccctc 780
 aacaccctca caccggttac aacatcaact gttttatcct caccaagtgg attcaacca 840
 agtggaacag tttctcagga gacattcctt tctggtgaaa caaccatctc atccccttcc 900
 60 agtgtcagca atacattcct ggtaacatca aaggtgttca gaatgccaat ctccagagac 960
 tctactcttg gaaacacaga ggagacatca ctatctgtaa gtggaacat ttctgcaatc 1020
 65 acttccaaag tttcaacat atggtggtca gacactctgt caacagcact ctccccagt 1080

5 tctctacctc caaaaatatc cacagctttc cacacccagc agagtgaagg tgcagagacc 1140
 acaggacggc ctcatgagag gagctcattc tctccagggtg tgtctcaaga aatatttact 1200
 ctacatgaaa caacaacatg gcccttctca ttctccagca aaggccacac aacttgggtca 1260
 caaacagAAC tGCCctcaac atcaacaggt gctgcccacta ggcttgtcac aggaaatcca 1320
 10 tctacaggggA cagctggcac tattccaagg gtccccctcta aggtctcagc aataggggaa 1380
 ccaggagagc ccaccacata ctctccccac agcacaactc tccccaaaac aacaggggca 1440
 ggcgcccaga cacaatggac acaagaaacg gggaccactg gagaggctct tctcagcagc 1500
 15 ccaagctaca gtgtgactca gatgataaaa acggccacat ccccatcttc ttcacctatg 1560
 ctggatagac acacatcaca acaaattaca acggcaccat caacaaatca ttcaacaata 1620
 cattccacaa gcacctctcc tcaggaatca ccagctgttt cccaaagggg tcacactcaa 1680
 20 gccccgcaga ccacacaaga atcacaaacc acgagggtccg tctcccccat gactgacacc 1740
 aagacagtca ccaccccagg ttcttccttc acagccagtg ggcactcgcc ctcagaaatt 1800
 25 gttcctcagg acgcaccac cataagtga gcaacaacct ttgccccagc tcccaccggg 1860
 gatggtcaca caaccaggc cccgaccaca gactgcagg cagcaccag cagccatgat 1920
 gccaccctgg ggcctcagg aggcacgtca ctttccaaaa caggtgcctt tactctggcc 1980
 30 aactctgtag tgtcaacacc agggggccca gaaggacaat ggacatcagc ctctgccagc 2040
 acctcacctg acacagcagc agccatgacc catacccacc aggctgagag cacagaggcc 2100
 tctggacaaa cacagaccag cgaaccggcc tcctcagggt cacgaaccac ctcagcgggc 2160
 35 acagctaccc cttcctcacc cggggcgagt ggcacaacac cttcaggaag cgaaggaata 2220
 tccacctcag gagagacgac aaggttttca tcaaaccctt ccagggatag tcacacaacc 2280
 cagtcaacaa ccgaattgct gtccgcctca gccagtcatg gtgcatccc agtaagcaca 2340
 40 ggaatggcgt cttcgtatgt ccccggcacc tttcatcca cctctctga ggcctccact 2400
 gcagggagac cgacaggaca gtcaagccca acttctcca gtgcctctcc tcaggagaca 2460
 gccgccattt cccggatggc ccagactcag aggacaagaa ccagcagagg gtctgacact 2520
 45 atcagcctgg cgtcccaggc aaccgacacc ttctcaacag tcccaccac acctccatcg 2580
 atcacatcca ctgggcttac atctccacaa acccagacc aactctgtc accttcaggg 2640
 50 tctggtaaaa cttcaccac ggccctcacc agcaacgcca cccctcttc tgtcacctac 2700
 gttcctcgg catccacagg tcacaccacc cctcttcatg tcaccgatgc ttctcagta 2760
 tccacaggtc acgccacccc tcttctgtc accagccett cctcagtatc cacaggtcac 2820
 55 accaccctc tt 2832

60 <210> 74
 <211> 1607
 <212> ADN
 <213> humano

65 <400> 74

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

55 <210> 75
 <211> 1753
 <212> ADN
 <213> humano
 60 <400> 75

65

5 cagccccgcc cctacctgtg gaagcccagc cgcccgtcc cgccgataaa aggcgcggag 60
 tgtccccgag gtcagcgagt gcgcgctcct cctcgcccgc cgctaggtcc atccccggcc 120
 agccaccatg tccatccact tcagctcccc ggtattcacc tcgcgctcag ccgccttctc 180
 gggccgcggc gcccagggtgc gcctgagctc cgctcgcccc ggcggccttg gcagcagcag 240
 10 cctctacggc ctcggcgctt cacggccgcg cgtggccgtg cgctctgcct atggggggcc 300
 ggtgggcgcc ggcatccgcg aggtcaccaat taaccagagc ctgctggccc cgctgcggct 360
 ggacgccgac ccctccctcc agcgggtgcg ccaggaggag agcgagcaga tcaagacct 420
 caacaacaag tttgcctcct tcatcgacaa ggtgcggttt ctggagcagc agaacaagct 480
 15 gctggagacc aagtggagcg tgctgcagga gcagaagtcg gccaaagaca gccgcctccc 540
 agacatcttt gaggcccaga ttgctggcct tcggggtcag cttgaggcac tgcaaggcca 600
 tgggggcccgc ctggaggcgg agctgcggag catgcaggat gtggtggagg acttcaagaa 660
 20 taagtacgaa gatgaaatta accaccgac agctgctgag aatgagtttg tgggtctgaa 720
 gaaggatgtg gatgctgcct acatgagcaa ggtggagctg gaggccaagg tggatgcctt 780
 gaatgatgag atcaacttcc tcaggacct caatgagacg gagttgacag agctgcagtc 840
 25 ccagatctcc gacacatctg tgggtctgtc catggacaac agtcgctccc tggacctgga 900
 cggcatcatc gctgaggta aggcgcagta tgaggagatg gccaaatgca gccgggctga 960
 ggctgaagcc tggtagcaga ccaagttga gaccctccag gccaggctg ggaagcatgg 1020
 30 ggacgacctc cggaataccc ggaatgagat ttcagagatg aaccgggcca tccagaggct 1080
 gcaggctgag atcgacaaca tcaagaacca gcgtgccaag ttggaggccg ccattgccga 1140
 ggctgaggag cgtggggagc tggcgtcaa ggatgctctg gccaaagcagg aggagctgga 1200
 35 agccgccctg cagcggggca agcaggatat ggcacggcag ctgctgtgag accaggaact 1260
 catgagcgtg aagctggccc tggacatcga gatgccacc taccgcaagc tgctggaggg 1320
 cgaggagagc cggttggctg gagatggagt gggagccgtg aatatctctg tgatgaattc 1380
 40 cactggtggc agtagcagtg gcggtggcat tgggctgacc ctcgggggaa ccatgggagc 1440
 caatgccctg agcttctcca gcagtgcggg tcctgggctc ctgaaggctt attccatccg 1500
 gaccgcatcc gccagtcgca ggagtgcccg cgactgagcc gcctcccacc actccactcc 1560
 45 tccagccacc acccacaatc acaagaagat tcccaccctt gcctcccatg cctggctcca 1620
 agacagtgag acagtctgga aagtgatgtc agaatagctt ccaataaagc agcctcattc 1680
 50 tgaggcctga gtgatccacg tgaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1740
 aaaaaaaaaa aaa 1753

55 <210> 76
 <211> 2255
 <212> ADN
 <213> humano

60 <400> 76

65

gatggctccg gccgcctggc tccgcagcgc ggccgcgcgc gccctcctgc ccccgatgct 60
 gctgctgctg ctccagccgc cgccgctgct ggcccgggct ctgccgccgg acgcccacca 120
 5 cctccatgcc gagaggaggg ggccacagcc ctggcatgca gccctgccca gtagcccggc 180
 acctgcccct gccacgcagg aagccccccg gcctgccagc agcctcaggc ctccccgctg 240
 10 tggcgtgccc gaccatctg atgggctgag tgcccgcaac cgacagaaga ggttcgtgct 300
 ttctggcggg cgctgggaga agacggacct cacctacagg atccttcggt tcccatggca 360
 gttggtgcag gagcaggtgc ggcagacgat ggcagaggcc ctaaaggatg ggagcgatgt 420
 15 gacgccactc acctttactg aggtgcacga gggccgtgct gacatcatga tcgacttcgc 480
 caggtactgg catggggacg acctgccgtt tgatgggcct gggggcatcc tggcccatgc 540
 20 cttctcccc aagactcacc gagaagggga tgtccacttc gactatgatg agacctggac 600
 tatcggggat gaccagggca cagacctgct gcaggtggca gcccatgaat ttggccacgt 660
 gctggggctg cagcacacaa cagcagccaa ggccctgatg tccgccttct acacctttcg 720
 25 ctaccactg agtctcagcc cagatgactg caggggcggt caacacctat atggccagcc 780
 ctggcccact gtcacctca ggaccccagc cctgggcccc caggctggga tagacaccaa 840
 tgagattgca ccgctggagc cagacgcccc gccagatgcc tgtgaggcct cctttgacgc 900
 30 ggtctccacc atccgaggcg agctcttttt cttcaaagcg ggcttttgtg ggcgcctccg 960
 tgggggcccag ctgcagcccg gctaccacgc attggcctct cgccactggc agggactgcc 1020
 cagccctgtg gacgctgcct tcgaggatgc ccagggccac atttggttct tccaagggtgc 1080
 35 tcagtactgg gtgtacgacg gtgaaaagcc agtcctgggc cccgcacccc tcaccgagct 1140
 gggcctggtg aggttcccgg tccatgctgc cttggtctgg ggtcccgaga agaacaagat 1200
 ctactttctc cgaggcaggg actactggcg tttccacccc agcaccggc gtgtagacag 1260
 40 tcccgtgccc cgcagggcca ctgactggag aggggtgccc tctgagatcg acgctgcett 1320
 ccaggatgct gatggctatg cctacttctt gcgcggccgc ctctactgga agtttgacct 1380
 45 tgtgaagggtg aaggctctgg aaggcttccc ccgtctcgtg ggtcctgact tctttggctg 1440
 tgccgagcct gccaacactt tcctctgacc atggcttggg tgccctcagg ggtgctgacc 1500
 cctgccaggc cacgaatata aggctagaga ccatggcca tctttgtggc tgtgggcacc 1560
 50 aggcattggga ctgagcccat gtctctcag ggggatgggg tggggtacaa ccaccatgac 1620
 aattgccggg agggccacgc aggtcgtggt cacctgccag cgactgtctc agactgggca 1680

55

60

65

5
 10
 15
 20

```

gggaggcttt ggcattgactt aagaggaagg gcagtcttgg gcccgcctatg caggctctgg 1740
caaacctggc tgccctgtct ccatccctgt cctcagggt agcaccatgg caggactggg 1800
ggaactggag tgtccttgct gtatccctgt tgtgagggtt cttccagggg ctggcactga 1860
agcaaggggtg ctggggcccc atggccttca gccctggctg agcaactggg ctgtagggca 1920
gggccacttc ctgaggtcag gtcttggtag gtgcctgcat ctgtctgcct tctggctgac 1980
aatcctggaa atctgttttc cagaatccag gccaaaaagt tcacagtcaa atggggaggg 2040
gtattcttca tgcaggagac cccaggcctt ggaggctgca acatacctca atcctgtccc 2100
aggccggatc ctctgaagc ccttttcgca gcactgctat cctccaaagc cattgtaa at 2160
gtgtgtacag tgtgtataaa cctttcttct cttttttttt ttttaaactg aggattgtca 2220
ttaaacacag ttgttttcta aaaaaaaaaa aaaaa 2255
  
```

25
 <210> 77
 <211> 462
 <212> ADN
 <213> humano

30
 35
 40
 45

```

agctctattg ccaccatgag tttctccggc aagtaccaac tgcagagcca ggaaaacttt 60
gaagccttca tgaaggcaat cggctctgcc gaagagctca tccagaaggg gaaggatc 120
aaggggggtg cggaaatcgt gcagaatggg aagcacttca agttcaccat caccgctggg 180
tccaaagtga tccaaaacga attcacgggt ggggaggaat gtgagctgga gacaatgaca 240
ggggagaaaag tcaagacagt ggttcagttg gaaggtgaca ataaactggt gacaactttc 300
aaaaacatca agtctgtgac cgaactcaac ggcgacataa tcaccaatac catgacattg 360
ggtgacattg tcttcaagag aatcagcaag agaatttaa caagtctgca tttcatatta 420
ttttagtgtg taaaattaat gtaataaagt gaactttggt tt 462
  
```

50
 <210> 78
 <211> 2108
 <212> ADN
 <213> humano

55
 60
 65

```

gggaccgcct cggaggcaga agagccgcga ggagccagcg gagcaccgcg ggctggggcg 60
cagccaccg ccgctcctcg agtcccctcg cccctttccc ttcgtgcccc ccggcagcct 120
ccagcgtcgg tccccaggca gcatgggtgag gtctgtctcc ggaccctcgc caccatgtac 180
gtgagctacc tcctggacaa ggacgtgagc atgtacccta gctccgtgcg cactctggc 240
ggcctcaacc tggcgccgca gaacttcgtc agccccccgc agtaccgga ctacggcggt 300
taccacgtgg cggccgcgac tgcagcggca gcgaacttgg acagcgcgca gtccccggg 360
ccatcctggc cggcagcgtg tggcgcccca ctccgggagg actggaatgg ctacgcgccc 420
  
```

5 ggagggcgccg cggccgcccgc caacgccgtg gctcacggcc tcaacgggtgg ctccccggcc 480
 gcagccatgg gctacagcag ccccgcagac taccatccgc accaccaccc gcatcaccac 540
 ccgcaccacc cggccgcccgc gccttctctg gcttctgggc tgctgcaaac gctcaacccc 600
 ggccctcctg ggccccggcg caccgtgccc gccgagcagc tgtctcccgg cggccagcgg 660
 10 cggaacctgt gcgagtggat gcggaagccg gcgcagcagt ccctcggcag ccaagtgaaa 720
 accaggacga aagacaaata tcgagtgggtg tacacggacc accagcggct ggagctggag 780
 aaggagtttc actacagtcg ctacatcacc atccggagga aagccgagct agccgccacg 840
 15 ctggggctct ctgagaggca ggttaaaatc tggtttcaga accgcagagc aaaggagagg 900
 aaaatcaaca agaagaagt gcagcagcaa cagcagcagc agccaccaca gccgcctccg 960
 ccgccaccac agcctcccca gcctcagcca ggtcctctga gaagtgtccc agagcccttg 1020
 20 agtccggtgt ctccctgca agcctcagtg tctggctctg tcctgggggt tctggggcca 1080
 actggggggg tgctaaaccc caccgtcacc cagtgacca ccggggctctg cagcggcaga 1140
 25 gcaattccag gctgagccat gaggagcgtg gactctgcta gactcctcag gagagacccc 1200
 tcccctcca cccacagcca tagacctaca gacctggctc tcagaggaaa aatgggagcc 1260
 aggagtaaga caagtgggat ttggggcctc aagaaatata ctctcccaga tttttacttt 1320
 30 ttcccatctg gctttttctg ccaactgagga gacagaaagc ctccgctggg cttcattccg 1380
 gactggcaga agcattgcct ggactgacca caccaaccag gccttcatcc tcctccccag 1440
 ctcttctctt cctagatctg caggctacac ctctggctag agccgagggg agagagggac 1500
 35 tcaagggaaa ggcaagcttg aggccaagat ggctgctgcc tgctcatggc cctcggaggt 1560
 ccagctgggc ctctgcctc cgggcaggca aggtttacac tgcggaagcc aaaggcagct 1620
 40 aagatagaaa gctggactga ccaaagactg cagaaccccc aggtggcctg cgtctttttt 1680
 ctcttcctt cccagaccag gaaaggcttg gctgggtgat gcacaggggtg tggtagagg 1740
 ggggtggttat tggactccag gcctgaccag ggggcccga cagggacttg ttagagagc 1800
 45 ctgtcaccag agcttctctg ggctgaatgt atgtcagtc tataaatgcc agagccaacc 1860
 tggacttct gtcattttca caatcttggg gctgatgaag aaggggggtg ggggagtttg 1920
 50 tgttgttgtt gctgctgttt gggttgttg tctgtgtaac atccaagcca gagtttttaa 1980
 agccttctgg atccatggg ggagaagtga tatggtgaag ggaagtggg agtatttgaa 2040
 cacagttgaa ttttttctaa aaagaaaaag agataaatga gctttccaga aaaaaaaaaa 2100
 55 aaaaaaaaaa 2108

60 <210> 79
 <211> 3745
 <212> ADN
 <213> humano

65 <400> 79

cgcaaagcaa gtgggcacaa ggagtatggt tctaactgta ttggggtcac gaagacgttg 60
 5 ctggttgact tggctttgtg gtcactgctc ttccagcccc ggtggctgtc ctttagttcc 120
 caggtgagtc agaactgcca caatggcagc tatgaaatca gcgtcctgat gatgggcaac 180
 tcagcctttg cagagcccct gaaaaacttg gaagatgcgg tgaatgaggg gctggaaata 240
 10 gtgagaggac gtctgcaaaa tgctggccta aatgtgactg tgaacgctac tttcatgtat 300
 tcggatggtc tgattcataa ctcaggcgac tgccggagta gcacctgtga aggcctcgac 360
 ctactcagga aaatttcaaa tgcacaacgg atgggctgtg tcctcatagg gccctcatgt 420
 15 acatactcca ccttccagat gtaccttgac acagaattga gctaccccat gatctcagct 480
 ggaagttttg gattgtcatg tgactataaa gaaaccttaa ccaggctgat gtctccagct 540
 20 agaaagtga tgtacttctt ggtaactttt tggaaaacca acgatctgcc cttcaaaact 600
 tattcctgga gcacttcgta tgttacaag aatggtacag aaactgagga ctgtttctgg 660
 taccttaatg ctctggaggc tagcgttcc tatttctccc acgaactcgg ctttaagggtg 720
 25 gtgtaagac aagataagga gtttcaggat atcttaatgg accacaacag gaaaagcaat 780
 gtgattatta tgtgtggtgg tccagagttc ctctacaagc tgaagggtga ccgagcagtg 840
 gctgaagaca ttgtcattat tctagtggat cttttcaatg accagtactt ggaggacaat 900
 30 gtcacagccc ctgactatat gaaaaatgct cttgttctga cgctgtctcc tgggaattcc 960
 cttctaataa gctctttctc caggaatcta tcaccaacaa aacgagactt tgctcttgcc 1020
 tatttgaatg gaatcctgct ctttggacat atgctgaaga tatttctga aaatggagaa 1080
 35 aatattacca ccccaaat tgctcatgct ttcaggaatc tcacttttga agggatgac 1140
 ggtccagtg ccttggatga ctggggggat gttgacagta ccatggtgct tctgtatacc 1200
 40 tctgtggaca ccaagaaata caaggttctt ttgacctatg ataccacgt aaataagacc 1260
 tatcctgtgg atatgagccc cacattcact tggagaact ctaaacttc taatgatatt 1320
 acaggccggg gccctcagat cctgatgatt gcagtcttca ccctcactgg agctgtggtg 1380
 45 ctgctcctgc tcgtcgtct cctgatgctc agaaaatata gaaaagatta tgaactcgt 1440
 cagaaaaaat ggtcccacat tcctcctgaa aatatcttcc ctctggagac caatgagacc 1500
 aatcatgta gcctcaagat cgatgatgac aaaagacgag atacaatcca gagactacga 1560
 50 cagtcaaat acgacaaaa gcgagtgatt ctcaaagatc tcaagcaca tgatggtaat 1620
 ttactgaaa aacagaagat agaattgaac aagttgcttc agattgacta ttacaacctg 1680
 55 accaagttct acggcacagt gaaacttgat accatgatct tcggggtgat agaatactgt 1740
 gagagaggat ccctccggga agttttaaat gacacaattt cctaccctga tggcacattc 1800

60

65

atggattggg agtttaagat ctctgtcttg tatgacattg ctaagggaat gtcatatctg 1860
 5 cactccagta agacagaagt ccatggctgt ctgaaatcta ccaactgcgt agtggacagt 1920
 agaatggtgg tgaagatcac tgattttggc tgcaattcca ttttacctcc aaaaaaggac 1980
 ctgtggacag ctccagagca cctccgcca gccaacatct ctcagaaagg agatgtgtac 2040
 10 agctatggga tcatcgaca ggagatcatt ctgcggaag aaaccttcta cactttgagc 2100
 tgtcgggacc ggaatgagaa gattttcaga gtggaaaatt ccaatggaat gaaacccttc 2160
 cgcccagatt tattcttggg aacagcagag gaaaaagagc tagaagtgta cctacttgta 2220
 15 aaaaactgtt gggaggaaga tccagaaaag agaccagatt tcaaaaaaat tgagactaca 2280
 cttgccaaga tatttggact ttttcatgac caaaaaaatg aaagctatat ggataccttg 2340
 atccgacgtc tacagctata ttctcgaac ctggaacatc tggtagagga aaggacacag 2400
 20 ctgtacaagg cagagagggg cagggtgac agacttaact ttatggtgct tccaaggcta 2460
 gtggtaaagt ctctgaagga gaaaggcttt gtggagccgg aactatatga ggaagttaca 2520
 25 atctacttca gtgacattgt aggtttcact actatctgca aatacagcac ccccatggaa 2580
 gtggtggaca tgcttaatga catctataag agttttgacc acattggtga tcatcatgat 2640
 gtctacaagg tggaaacat cggtgatgcg tacatggtgg ctagtggttt gcctaagaga 2700
 30 aatggcaatc ggcatgcaat agacattgcc aagatggcct tggaaatcct cagcttcatg 2760
 gggacctttg agctggagca tcttctggc ctccaatat ggattcgcat tggagttcac 2820
 tctggtccct gtgctgctgg agttgtggga atcaagatgc ctcgttattg tctatttggg 2880
 35 gataggtca acacagctc taggatggaa tccactggcc tccctttgag aattcacgtg 2940
 agtggctcca ccatagccat cctgaagaga actgagtgcc agttcctta tgaagtgaga 3000
 40 ggagaaacat acttaaaggg aagaggaaat gagactacct actggctgac tgggatgaag 3060
 gaccagaaat tcaacctgcc aaccctctct actgtggaga atcaacagcg tttgcaagca 3120
 gaattttcag acatgattgc caactctta cagaaaagac aggcagcagg gataagaagc 3180
 45 caaaaacca gacgggtagc cagctataaa aaaggcactc tggataactt gcagctgaat 3240
 accacagaca aggagagcac ctatttttaa acctaaatga ggtataagga ctcacacaaa 3300
 ttaaaataca gctgactga ggcagcgacc tcaagtgtcc tgaaagctta cattttcctg 3360
 50 agacctcaat gaagcagaaa tgtacttagg cttggctgcc ctgtctggaa catggacttt 3420
 cttgcatgaa tcagatgtgt gttctcagtg aaataactac cttccactet ggaaccttat 3480
 tccagcagtt gttccagggg gcttctacct ggaaaagaaa agaaatgaat agactatcta 3540
 55 gaacttgaga agattttatt cttatttcat ttattttttg tttgtttatt tttatcgttt 3600
 ttgtttactg gctttccttc tgtattcata agatttttta aattgtcata attatatttt 3660
 60 aaatacccat cttcattaaa gtatatttaa ctcataattt ttgcagaaaa tatgctatat 3720
 attaggcaag aataaaagct aaagg 3745

65

ES 2 494 843 T3

<210> 80
 <211> 901
 <212> ADN
 <213> humano

5

<400> 80

10 agccccaac tcaccacctg gccgtggaca cctgtgtcag catgtgggac ctggttctct 60
 ccatcgctt gtctgtggg tgcactggtg ccgtgccct catccagtct cggattgtgg 120
 gaggtgga gtgtgagaag cattccaac cctggcaggt ggctgtgtac agtcatggat 180
 15 gggcacactg tgggggtgtc ctggtgcacc cccagtgggt gctcacagct gccattgcc 240
 taagaagaa tagccaggtc tggctgggtc ggcacaacct gtttgagcct gaagacacag 300
 gccagaggggt ccctgtcagc cacagcttcc cacacccgct ctacaatatg agccttctga 360
 20 agcatcaaag ccttagacca gatgaagact ccagccatga cctcatgctg cttcgcctgt 420
 cagagcctgc caagatcaca gatgttgtga aggtcctggg cctgcccacc caggagccag 480
 25 cactggggac cacctgctac gcctcaggct ggggcagcat cgaaccagag gagttcttgc 540
 gccccaggag tcttcagtgt gtgagcctcc atctcctgtc caatgacatg tgtgctagag 600
 ctactctga gaaggtgaca gagttcatgt tgtgtgctgg gctctggaca ggtggtaaag 660
 30 acacttgtgg ggggtattct gggggtccac ttgtctgtaa tgggtgtgctt caaggtatca 720
 catcatgggg ccctgagcca tgtgcctgc ctgaaaagcc tgctgtgtac accaaggtgg 780
 35 tgcattaccg gaagtggatc aaggacacca tgcagccaa cccctgagtg cccctgtccc 840
 acccctacct ctagtaaatt taagtccacc tcaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 900
 a 901

40

<210> 81
 <211> 618
 <212> ADN
 <213> humano

45

<400> 81

50 ggggaccact tctctgggac acattgcctt ctgttttctc cagcatgccc ttgctccagc 60
 tcctgttcag ggccagcctt gccaccctgc tcctggttct ctgcctgcag ttgggggcca 120
 acaaagctca ggacaacact cggagatca taataaagaa ttttgacatt cccaagtcag 180
 55 tacgtccaaa tgacgaagtc actgcagtgc ttgcagttca aacagaattg aaagaatgca 240
 tgggtggttaa aacttacctc attagcagca tccctctaca aggtgcattt aactataagt 300
 atactgcctg cctatgtgac gacaatccaa aaaccttcta ctgggacttt tacaccaaca 360
 60 gaactgtgca aattgcagcc gtcggtgatg ttattcggga attagcatc tgcctgatg 420

65

5 atgctgctgt aatccccatc aaaaacaacc ggttttatac tattgaaatc ctaaaggtag 480
 aataatggaa gccctgtctg ttgcccacac ccaggtgatt tcctctaaag aaacttggct 540
 ggaatttctg ctgtggctca taaaataaac ttcttaacat gcttctacaa aaaaaaaaaa 600
 aaaaaaaaaa aaaaaaaa 618

10 <210> 82
 <211> 594
 <212> ADN
 <213> humano
 15 <400> 82

20 gtcggtttag gactttctgc ctccactatt gctatcggta ctggaatagc aggcatttca 60
 acatctgtca cgaccttcca tagcctatat aatgacttat ctgctagcat cacagacata 120
 tcacaaactt tatcagtcct ccaggcccaa gttgaatctt tagctgcagt tgtcctccaa 180
 25 aaccgccgag gccttgactt acttactgct taaagaggag gactctgcat attcttaaat 240
 gaggagtgtt gtttttacct aatcaatct ggcctgggtg atgacaacat aaaaaaattc 300
 aaggatagag cccaaaaact taccaacca gcaagtaatt tcaactgaacc cccttgggca 360
 30 ctccctaatt ggggtgcctg ggtcctccca attcttagtc cttaataacc catttttctc 420
 ctctttttat tcagacctg tatcttctgt ttagcttctc aattcatcca aaaccatatac 480
 caggccatca ccaatcattc tatacgacaa atgtttctta taacatcccc acaatatacac 540
 35 cccttaccac aagacctccc ttcaacttaa tctctcccga tataggttcc caca 594

40 <210> 83
 <211> 1372
 <212> ADN
 <213> humano
 45 <400> 83

50 gaattcggcg atgcctcaca actccatcag atctggccat ggagggctga accagctggg 60
 aggggacctt gtgaatggca gacctctgcc ggaagtggc cgccagcgca tcgtagacct 120
 ggcccaccag ggtgtaaggc cctgcgacat ctctcgccag ctccgctca gccatggttg 180
 cgtcagcaag atccttggca ggtactacga gactggcagc atccggcctg gagtgatagg 240
 55 gggctccaag cccaaggtgg ccacccccaa ggtggtggag aagattgggg actacaaacg 300
 ccagaacctt accatgtttg cctgggagat ccgagaccgg ctcttggctg agggcgctctg 360
 tgacaatgac actgtgcca gtgtcagctc cattaataga atcatccgga ccaaagtgca 420
 gcaaccattc aacctcccta tggacagctg cgtggccacc aagtccctga gtcccggaca 480
 60 cacgctgatc ccagctcag ctgtaactcc cccggagtca cccagctcg attccctggg 540
 ctccacctac tccatcaatg ggctcctggg catcgtcag cctggcagcg acaagaggaa 600
 65 aatggatgac agtgatcagg atagctgccg actaagcatt gactcacaga gcagcagcag 660

5 cggacccccga aagcaccttc gcacggatgc cttcagccag caccacctcg agccgctcga 720
 gtgcccattt gagcggcagc actaccaga ggcctatgcc tccccagcc acaccaaagg 780
 cgagcagggc ctctaccegc tgcccttget caacagcacc ctggacgacg ggaaggccac 840
 10 cctgaccctt tccaacacgc cactggggcg caacctctcg actcaccaga cctaccccgt 900
 ggtggcagat cctcactcac ctttcgccat aaagcaggaa acccccgagg tgtccagttc 960
 tagctccacc ctttctctt tatctagctc cgcctttttg gatctgcagc aagtcggctc 1020
 15 cggggctccg cccttcaatg cttttccca tgctgcctcc gtgtacgggc agttcacggg 1080
 ccaggccctc ctctcagggc gagagatggt ggggccacg ctgcccgat acccacccca 1140
 catccccacc agcggacagg gcagctatgc ctctctgcc atcgaggca tggtagcagg 1200
 20 aagtgaatac tctggcaatg cctatggcca caccctctac tcctctaca gcgaggcctg 1260
 gcgcttccc aactccagct tgctgagttc cccatattat tacagttcca catcaaggcc 1320
 25 gagtgcaccg cccaccactg ccacggcctt tgaccatctg tagttgaagc tt 1372

30 <210> 84
 <211> 2983
 <212> ADN
 <213> humano
 <400> 84

35 gcccagatag gggagcggag gtggcggcgg cggcggtagc ggtggccttg gttgtcttcc 60
 agtctcctcg gctcgcctt tagccggcac cgtcctcctt cctccccct tcctctcttc 120
 40 cttcttccc tcccctccc ttttccctt ccccgctcggg gagcggcggg ggtggctcca 180
 gcaacggctg ggcccaagct gtgtagaggc ctaaccaac gataacggcg gcgacggcga 240
 aacctcggag ctgcagggc gggggcaagg cccgggcctt ggagatggag aattctcagt 300
 45 tgtgtaagct gttcatcggc ggcctcaatg tgacagcag tgagtcgggc ctgcgcggcc 360
 actttgaggc ctttgggact ctgacggact gcgtggtggt ggtgaatccc cagaccaagc 420
 gctcccgttg ctttggcttc gtgacctact ccaatgtgga ggaggcggac gccgccatgg 480
 50 ccgctcggc ccatgccgtg gacggcaaca ctgtggagct gaagcggggc gtgtcccggg 540
 aggattcggc gcggcccggg gccacgcca aggttaagaa gctctttgtc ggaggcctta 600
 aaggagacgt ggctgagggc gacctgatcg agcacttctc gcagtttggc accgtggaaa 660
 55 aggccgagat tattgccgac aagcagtcg gcaagaagcg tggattcggc ttcgtgtatt 720
 tccagaatca cgacgcggca gacaaggccg cgggtgtcaa gttccatccg attcagggcc 780
 60 atcgcgtgga ggtgaagaaa gcagtcccca aggaggatat ctactccggg ggggggtggag 840
 gcggctcccg atcctcccgg ggcggccgag gcggccgggg gcgcggcggg ggtcgagacc 900
 65 agaacggcct ttccaagggc ggcggcggcg gttacaacag ctacgggtggt tacggcggcg 960

5 gcgaggcgg cggtacaat gcctacggag gcgggcggcg cggttcgtcc tacggtggga 1020
 gcgactacgg taacggcttc ggcggcttcg gcagctacag ccagcatcag tcctcctatg 1080
 ggcccatgaa gagcggcggc ggcggcggcg gtggaggcag tagctggggc ggtcgcagta 1140
 atagtggacc ttacagaggc ggctatggcg gtgggggtgg ctatggaggc agctccttct 1200
 10 aaaagaaaat ttaaaatgcc tgggagtggc tataggggta gctctttcca acagcccaag 1260
 tggggtaaac tcctaagccc caccctca cacacaccgc cttccctggt ttgcccttgg 1320
 gggagccact tctaaggctg cttacccttg ggggtgttcc tctatttgcc tgccacctct 1380
 15 cttgtctctc cctctgaaga tggactcggc cccacataca catttttggg ttacagtcat 1440
 tgatggactc tttttttta ttattacttg gaccttggtc gtttttatac tagcaaaatg 1500
 tcttgtttta atttgtgttt tttgggggga gggaggaggt gaacttgctg attctgtagc 1560
 20 aaaacctggg tgggggttgg ggtggggggt agtttacttt gttgtaagga cttgataacc 1620
 tggctacagc gttttctatg aaatctactt ggatcccatg cctgaaattt ggaagcatat 1680
 gtacaaaaat catttttacg ttttttttt aataaatcat tgtgtttgac cgtacatgtc 1740
 25 taacattttt tttctaggat ccattccgta ccgtttttta agggatattt gtttaagact 1800
 ttacgtgtta attctttatt cttgatgtgt acttagagaa acttaagagg tcctgtgggt 1860
 30 tttttccct ctctgttgc cctgctagtt gcgtgttgaa ttatatccct tacaggcaaa 1920
 acttttgaag tgggtgatgt ggctttttta actcttaagt ttctgtgcat ccatctcttg 1980
 tactaagcga attgtttatc atcttgacat ggttggctcat ttctatgaca atttacttca 2040
 35 aactgtgtac tgtgtagttc tatatagttt gtgtaagca tgtcattcat ataaactggt 2100
 taaaattttt cagatggcct agtttcatcc ctcttactgg tttgtctgta atgaatgggt 2160
 aaaaataagg gttatatttt accctcaaat gcgtttttgt actttcagag caggtttaaa 2220
 40 cgtttttttt ttttttttcc tatatccgaa ctgttggcct catggaaatc cttttccgga 2280
 tctttgtagc accatctact ggcagaatgg cagagtagct gcgaaacaat ttgtttaaaa 2340
 acttgcttaa gacaattgca tcagatttgg aagttttgcc atcaaaattc tttgcagaat 2400
 45 tggaggttaa cacatttgct tgtaactgag atgggcttca caggaatgta gttgccagtt 2460
 catatcacia tagccctttc tatatgaggt ttgaaaatgt aaactgctat gcatagcttg 2520
 ggcaatagcc ctaaattgct atgacaacta atgaaccagc tacgtatact ggtatttttag 2580
 50 gtgcaagttg taaagcaaaa tatctgtgta ttctgcttgg ttaacaaatg tatatttgta 2640
 gccctttcct gcaatagcat tcaagttggt gtttataaga gaagaacaaa agtgataata 2700
 55 ggtgaaaatt gcctttctgg atagaaatag agaatagcaa cgtttatgga tatcacaaat 2760
 aaagaattca attctttaca tgattgagtg agagtatgta taacctggtg ggtgggttca 2820
 60 gagtaccttt taatctagta tgcttaactt gatgtaata ttttaactta atatttgact 2880
 tacatgttga cgttgaaggc tcaaagctat actaagaagc tttctgaaag attgggcttt 2940
 aaaataaaat aatattttta tattgaaaa aaaaaaaaaa aaa 2983

65

ES 2 494 843 T3

<210> 85
 <211> 3345
 <212> ADN
 <213> humano

5

<400> 85

10 gaattccgtc tcgaccactg aatggaagaa aaggactttt aaccaccatt ttgtgactta 60
 cagaaaggaa tttgaataaa gaaaactatg atacttcagg cccatcttca ctccctgtgt 120
 cttcttatgc tttatttggc aactggatat ggccaagagg ggaagtttag tggaccctcg 180
 15 aaacccatga cattttctat ttatgaaggc caagaaccga gtcaaattat attccagttt 240
 aaggccaatc ctccctgctgt gacttttgaa ctaactgggg agacagacaa catatttgtg 300
 atagaacggg agggacttct gtattacaac agagccttgg acagggaaac aagatctact 360
 20 cacaatctcc aggttgacgc cctggacgct aatggaatta tagtggaggg tccagtcctt 420
 atcaccatag aagtgaagga catcaacgac aatcgacca cgtttctcca gtcaaagtac 480
 gaaggctcag taaggcagaa ctctcgcca ggaaagccct tcttgatgt caatgccaca 540
 25 gacctggatg atccggccac tcccaatggc cagctttatt accagattgt catccagctt 600
 cccagatca acaatgtcat gtactttcag atcaacaaca aaacgggagc catctctctt 660
 acccgagagg gatctcagga attgaatcct gctaagaatc cttcctataa tctggtgatc 720
 30 tcagtgaagg acatgggagg ccagagtgag aattccttca gtgataccac atctgtggat 780
 atcatagtga cagagaatat ttggaaagca ccaaaacctg tggagatggt ggaaaactca 840
 actgatcctc accccatcaa aatcactcag gtgcggtgga atgatcccg tgacacaatat 900
 35 tccttagttg acaaagagaa gctgccaaga tccccatttt caattgacca ggaaggagat 960
 atttacgtga ctccagccctt ggaccgagaa gaaaaggatg catatgtttt ttatgcagtt 1020
 gcaaaggatg agtacggaaa accactttca tatccgctgg aaattcatgt aaaagttaaa 1080
 40 gatattaatg ataatccacc tacatgtccg tcaccagtaa ccgtatttga ggtccaggag 1140
 aatgaacgac tgggtaacag tatcgggacc cttactgcac atgacagga tgaagaaaat 1200
 actgccaaca gttttctaaa ctacaggatt gtggagcaaa ctcccaaact tcccatggat 1260
 45 ggactcttcc taatccaaac ctatgctgga atgttacagt tagctaaaca gtccttgaag 1320
 aagcaagata ctctcagta caacttaacg atagaggtgt ctgacaaaga tttcaagacc 1380
 ctttgttttg tgcaaatcaa cgttattgat atcaatgatc agatccccat ctttgaaaaa 1440
 50 tcagattatg gaaacctgac tcttctgtaa gacacaaaca ttgggtccac catcttaacc 1500

55

60

65

atccaggcca ctgatgctga tgagccattt actgggagtt ctaaaattct gtatcatatc 1560
 5 ataaagggag acagtgaggg acgcctgggg gttgacacag atccccatac caacaccgga 1620
 tatgtcataa ttaaaaagcc tcttgatfff gaaacagcag ctgtttccaa cattgtgttc 1680
 aaagcagaaa atcctgagcc tctagtgttt ggtgtgaagt acaatgcaag ttcttttgcc 1740
 10 aagttcacgc ttattgtgac agatgtgaat gaagcacctc aattttccca acacgtattc 1800
 caagcgaaag tcagtgagga tgtagctata ggcactaaag tgggcaatgt gactgccaag 1860
 gatccagaag gtctggacat aagctattca ctgaggggag acacaagagg ttggcttaaa 1920
 15 attgaccacg tgactggtga gatctttagt gtggctccat tggacagaga agccggaagt 1980
 ccatatcggg tacaagtggg ggccacagaa gtaggggggt cttccttaag ctctgtgtca 2040
 gagttccacc tgatccttat ggatgtgaat gacaaccctc ccaggctagc caaggactac 2100
 20 acgggcttgt tcttctgcca tcccctcagt gcacctggaa gtctcatttt cgaggctact 2160
 gatgatgatc agcacttatt tcgggggtccc cattttacat tttccctcgg cagtgggaagc 2220
 25 ttacaaaacg actgggaagt ttccaaaatc aatggtactc atgcccgact gtctaccagg 2280
 cacacagact ttgaggagag ggcgtatgtc gtcttgatcc gcatcaatga tgggggtcgg 2340
 ccacccttgg aaggcattgt ttctttacca gttacattct gcagttgtgt ggaaggaagt 2400
 30 tgtttccggc cagcaggtca ccagactggg ataccactg tgggcatggc agttggtata 2460
 ctgctgacca cccttctggg gattggtata attttagcag ttgtgtttat ccgcataaag 2520
 aaggataaag gcaaagataa tgttgaaagt gctcaagcat ctgaagtcaa acctctgaga 2580
 35 agctgaattt gaaaaggaat gtttgaattt atatagcaag tgctatttca gcaacaacca 2640
 tctcatccta ttacttttca tctaactgtc attataattt tttaaacaga tattccctct 2700
 40 tgtcctttta tatttgctaa atatttcttt ttgaggtgg agtcttgctc tgtcgcccag 2760
 gctggagtac agtgggtgtga tcccagctca ctgcaacctc cgcctcctgg gttcacatga 2820
 ttctcctgcc tcagcttctt aagtagctgg gtttacaggc acccaccacc atgcccagct 2880
 45 aatTTTTgta tttttaatag agacgggggt tgcctatttg gccaggctgg tcttgaactc 2940
 ctgacgtcaa gtgatctgcc tgccttggtc tcccaataca ggcatagaacc actgcacca 3000
 cctacttaga tatttcatgt gctatagaca ttagagagat ttttcatttt tccatgacat 3060
 50 tttcctctc tgcaaatggc ttagctactt gtgtttttcc cttttggggc aagacagact 3120
 cattaatat tctgtacatt ttttctttat caaggagata tatcagtggt gtctcataga 3180
 actgcctgga ttccatttat gtttttctg attccatcct gtgtcccctt catccttgac 3240
 55 tcctttggta tttcactgaa tttcaaacat ttgtcagaga agaaaaaagt gaggactcag 3300
 gaaaaataaa taaataaaag aacagccttt tgcggccgcy aattc 3345

60 <210> 86
 <211> 990
 <212> ADN
 <213> humano
 65 <400> 86

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

35 <210> 87
 <211> 1805
 <212> ADN
 <213> humano
 40 <400> 87

45 gcgcacactc tcctaagccc tctcatctcc tggaaaccatg gccagcacat ccaccacat 60
 caggagccac agcagcagcc gccggggttt cagtgccaac tcagccaggc tccctgggggt 120
 cagccgctct ggcttcagca gcattctcgt gtcccgtctc aggggcagtg gtggcctggg 180
 50 tggcgcgatgt ggaggagctg gctttggcag ccgcagtctg tatggcctgg ggggctccaa 240
 gaggatctcc attggagggg gcagctgtgc catcagtggc ggctatggca gcagagccgg 300
 aggcagctat ggctttgggt gcgccgggag tggatttggg ttcgggtggg gagccggcat 360
 55 tggctttggg ctgggtgggt gagccggcct tgctggtggc tttgggggccc ctggcttccc 420
 tgtgtgcccc cctggaggca tccaagaggt cactgtcaac cagagtctcc tgactcccct 480

60
 65

5 caacctgcaa attgacccccg ccatccagcg ggtgcggggcc gaggagcgtg agcagatcaa 540
gaccctcaac aacaagtttg cctccttcat cgacaagggtg cggttcctag agcagcagaa 600
caaggttctg gacaccaagt ggaccctgct gcaggagcag ggcaccaaga ctgtgaggca 660
10 gaacctggag ccgttgttcg agcagtacat caacaacctc aggaggcagc tggacaacat 720
cgtgggggaa cggggtcgtc tggactcggg gctgagaaac atgcaggacc tgggtggagga 780
cctcaagaac aaatatgagg atgaaatcaa caagcgaca gcagcagaga atgaatttgt 840
15 gactctgaag aaggatgtgg atgctgccta catgaacaag gttgaaactgc aagccaaggc 900
agacactctt acagatgaga tcaacttctt gagagccttg tatgatgcag agctgtccca 960
gatgcagacc cacatctcag acacatccgt ggtgctatcc atggacaaca accgcaacct 1020
20 ggacctggac agcatcatcg ctgagggtcaa ggccaatat gaggagattg ctgagaggag 1080
cagggctgag gctgagtcct ggtaccagac aaagtacgag gagctgcaga tcacagcagg 1140
cagacatggg gacgacctgc gcaacaccaa gcaggagatt gctgagatca accgcatgat 1200
25 ccagaggctg agatctgaga tcgaccatgt caagaagcag tgtccaacc tacaggccgc 1260
cattgctgat gctgagcagc gtggggagat ggccctcaag gatgctaaga acaagctgga 1320
agggctggag gatgccctgc agaaggccaa gcaggacctg gcccgctgc tgaaggagta 1380
30 ccaggagctg atgaacgtca agctggccct ggacgtggag atcgccacct accgcaagct 1440
gctggagggc gaggagtgcg ggctgaatgg cgaaggcgtt ggacaagtca acatctctgt 1500
35 agtgcagtc accgtctcca gtggctatgg cgggtccagc ggtgtcggca gtggcttagg 1560
cctgggtgga ggaagcagct actcctatgg cagtggctt ggcgttggag gcggcttag 1620
ttccagcagc ggcagagcca ctgggggtgg cctcagctct gttggaggcg gcagttccac 1680
40 catcaagtac accaccacct cctcctccag caggaagagc tacaagcact gaagtcgtgc 1740
cgccagctct cagtcccaca gctctcaggc ccctctctgg cagcagagcc ctctcctcag 1800
gttgc 1805

45
<210> 88
<211> 2820
<212> ADN
50 <213> humano
<400> 88

55 tggcaaaatc ctggagccag aagaaaggac agcagcattg atcaatctta cagctaacat 60
gttgtaacctg gaaaacaatg cccagactca atttagtgag ccacagtaca cgaacctggg 120
60 gctcctgaac agcatggacc agcagattcg gaacggctcc tcgtccacca gtcctataa 180
cacagaccac gcgcagaaca gcgtcacggc gccctcggcc tacgcacagc ccagccccac 240
cttcgatgct ctctctccat caccggccat cccctccaac accgactacc caggccccga 300

65

5 cagttccgac gtgtccttcc agcagtcgag caccgccaag tcggccacct ggacgtattc 360
 cactgaactg aagaaactct actgccaaat tgcaaagaca tgccccatcc agatcaaggt 420
 gatgacccca cctcctcagg gagctgttat ccgcgccatg cctgtctaca aaaaagctga 480
 gcacgtcacg gaggtggtga agcggtgccc caaccatgag ctgagccgtg agttcaacga 540
 10 gggacagatt gcccctccta gtcatttgat tcgagtagag gggaacagcc atgcccagta 600
 tgtagaagat cccatcacag gaagacagag tgtgctggta cttatgagc caccacaggt 660
 tggcactgaa ttcacgacag tcttgtacaa tttcatgtgt aacagcagtt gtggtgagg 720
 15 gatgaaccgc cgtccaattt taatcattgt tactctggaa accagagatg ggcaagtcct 780
 gggccgacgc tgctttgagg cccggatctg tgcttgccca ggaagagaca ggaaggcgga 840
 tgaagatagc atcagaaagc agcaagtttc ggacagtaca aagaacggtg atggtacgaa 900
 20 gcgcccgttt cgtcagaaca cacatggtat ccagatgaca tccatcaaga aacgaagatc 960
 cccagatgat gaactgttat acttaccagt gaggggccgt gagacttatg aaatgctgtt 1020
 gaagatcaaa gagtccctgg aactcatgca gtaccttctt cagcacacaa ttgaaacgta 1080
 25 caggcaacag caacagcagc agcaccagca cttacttcag aaacagacct caatacagtc 1140
 tccatcttca tatggtaaaca gctccccacc tctgaacaaa atgaacagca tgaacaaget 1200
 gccttctgtg agccagctta tcaaccctca gcagcgcaac gccctcactc ctacaacat 1260
 30 tcctgatggc atgggagcca acattcccat gatgggcacc cacatgcca tggctggaga 1320
 catgaatgga ctcagcccca cccaggcact cctccccca ctctccatgc catccacctc 1380
 35 ccaactgcaca cccccactc cgtatccac agattgcagc attgtcagtt tcttagcgag 1440
 gttgggctgt tcatcatgtc tggactatth cacgaccag gggctgacca ccatctatca 1500
 gattgagcat tactccatgg atgatctggc aagtctgaaa atccctgagc aatttcgaca 1560
 40 tgcgatctgg aagggcatcc tggaccaccg gcagctccac gaattctcct ccccttctca 1620
 tctcctgcgg accccaagca gtgcctctac agtcagtggt ggctccagtg agaccggggg 1680
 tgagcgtggt attgatgctg tgcgattcac cctccgccag accatctctt tcccaccccg 1740
 45 agatgagtggt aatgacttca actttgacat ggatgctcgc cgcaataagc aacagcgcag 1800
 caaagaggag ggggagtgag cctcaccatg tgagctcttc ctatccctct cctaactgcc 1860
 agccccctaa aagcactcct gcttaatctt caaagccttc tcctagctc ctccccctcc 1920
 50 tcttgtctga tttcttaggg gaaggagaag taagaggcta cctcttacct aacatctgac 1980
 ctggcatcta attctgattc tggctttaag ccttcaaac tatagcttgc agaactgtag 2040
 ctgccatggc taggtagaag tgagcaaaaa agagttgggt gtctccttaa gctgcagaga 2100
 55 tttctcattg acttttataa agcatgttca cccttatagt ctaagactat atatataaat 2160
 gtataaatat acagtataga tttttgggtg gggggcattg agtattgttt aaaatgtaat 2220

60

65

5 ttaaatgaaa gaaaattgag ttgcacttat tgaccatttt ttaatttact tgttttggat 2280
 ggcttgtcta tactccttcc ctttaaggggt atcatgtatg gtgataggta tctagagctt 2340
 aatgctacat gtgagtgcga tgatgtacag attctttcag ttctttggat tctaaataca 2400
 10 tgccacatca aacctttgag tagatccatt tccattgctt attatgtagg taagactgta 2460
 gatatgtatt cttttctcag tgttgggtata ttttatatta ctgacatttc ttctagtgat 2520
 gatggttcac gttgggggtga tttaatccag ttataagaag aagtccatgt ccaaacggtc 2580
 15 ctcttttagtt tttggttggg aatgaggaaa attcttaaaa ggcccatagc agccagttca 2640
 aaaacacccg acgtcatgta tttgagcata tcagtaacct ccttaaattt aatacccaga 2700
 taccttatct tacaatggtg attgggaaaa catttgctgc ccattacaga ggtattaaaa 2760
 20 ctaaatttca ctactagatt gactaactca aatacacatt tgctactggt gtaagaattc 2820

25 <210> 89
 <211> 991
 <212> ADN
 <213> humano

<400> 89

30 cttatctcgg cttcgtttct ggagggccag gaacaaacag gcttcaaagc caagggcttg 60
 gctggcacac agggggcttg gtccttcacc tctgtcccct ctccctacgg acacatataa 120
 35 gaccctggtc acacctggga gaggaggaga ggagagcata gcacctgcag caagatggat 180
 gtgggcagca aagaggtcct gatggagagc cgcgccgact actccgcagc tccccggggc 240
 cgatttggca ttccctgctg cccagtgcac ctgaaacgcc ttcttatcgt ggtggtggtg 300
 40 gtggtcctca tcgtcgtggt gattgtggga gccctgctca tgggtctcca catgagccag 360
 aaacacacgg agatggttct ggagatgagc attggggcgc cggaagccca gcaacgcctg 420
 gccctgagtg agcacctggt taccactgcc accttctcca tcggctccac tggcctcgtg 480
 45 gtgtatgact accagcagct gctgatgcc tacaagccag cccctggcac ctgctgctac 540
 atcatgaaga tagctccaga gagcatcccc agtcttgagg ctctcaatag aaaagtccac 600
 aacttccaga tggaatgctc tctgcaggcc aagcccgcag tgctacgtc taagctgggc 660
 50 caggcagagg ggcgagatgc aggctcagca ccctccggag gggaccggc ctctctgggc 720
 atggccgtga acaccctgtg tggcgagggt ccgctctact acatctagga cgcctccggt 780
 55 gagcagggtc agtgggaagcc ccaacgggaa aggaaacgcc ccgggcaaag ggtcttttgc 840
 agcttttgca gacgggcaag aagctgcttc tgcccacacc gcagggacaa accctggaga 900
 aatgggagct tggggagagg atgggagtg gacagagtg caccagggg cccgggaact 960
 60 cctgccacaa cagaataaag cagcctgatt g 991

65 <210> 90
 <211> 1580
 <212> ADN

ES 2 494 843 T3

<213> humano

<400> 90

5
 catcctgcc a ccctagcct tgctggggac gtgaaccctc tccccgcgcc tgggaagcct 60
 tcttggcacc gggaccgga gaatccccac ggaagccagt tccaaaaggg atgaaaaggg 120
 10 ggcgtttcgg gactgggag aagcctgtat tccagggccc ctcccagagc aggaatctgg 180
 gaccagggag tgccagcctc acccacgcag atcctggcca tgagagctcc gcacctccac 240
 ctctccgccg cctctggcgc ccgggctctg gcgaagctgc tgccgctgct gatggcgcaa 300
 15 ctctgggccg cagaggcggc gctgctcccc caaaacgaca cgcgcttggc ccccgaagcc 360
 tatggctccc cgtgcgcgcg cggctcgcag ccctggcagg tctcgtctt caacggcctc 420
 tcgttccaet gcgcgggtgt cctggtggac cagagttggg tgctgacggc cgcgcactgc 480
 20 ggaacaagc cactgtgggc tcgagtaggg gatgaccacc tgctgcttct tcagggagag 540
 cagctccgcc ggaccactcg ctctgttgtc catccaagt accaccaggg ctccagggccc 600
 atcctgcaa ggcgaacgga tgagcacgat ctcatgttgc tgaagctggc caggcccgta 660
 25 gtgctggggc cccgcgtccg ggccctgcag ctccctacc gctgtgctca gcccgagac 720
 cagtgccagg ttgctggctg gggcaccacg gccgccgga gagtgaagta caacaagggc 780
 30 ctgacctgct ccagcatcac tctctgagc cctaaagagt gtgaggtctt ctaccctggc 840
 gtggtcacca acaacatgat atgtgctgga ctggaccggg gccaggacc ttgccagagt 900
 gactctggag gccccctggt ctgtgacgag accctccaag gcatcctctc gtgggggtgtt 960
 35 taccctgtg gctctgcca gcatccagct gtctacacc agatctgcaa atacatgtcc 1020
 tggatcaata aagtcatacg ctccaactga tccagatgct acgctccagc tgatccagat 1080
 gttatgctcc tgctgatcca gatgccaga ggctccatcg tccatcctt tctccccag 1140
 40 tcggctgaac tctcccctg tctgactgt tcaaactct gccgccctcc acaccttaa 1200
 acatctcccc tctcacctca tccccacc tatccccatt ctctgcctgt actgaagctg 1260
 45 aaatgcagga agtgggtggca aaggtttatt ccagagaagc caggaagccg gtcacacccc 1320
 agcctctgag agcagttact ggggtcacc aacctgact cctctgccac tcctgctgt 1380
 gtgactttgg gcaagccaag tgccctctct gaacctcagt ttctcatct gcaaatggg 1440
 50 aacaatgacg tgctacctc ttagacatgt tgtgaggaga ctatgatata acatgtgtat 1500
 gtaaactctc atggtgattg tcatgtaagg cttaacacag tgggtggtga gttctgacta 1560
 aaggttacct gttgtcgtga 1580

55

<210> 91

<211> 3359

<212> ADN

60 <213> humano

<400> 91

65

cacaccttcg gcagcaggag ggcggcagct tctcgcaggc ggcagggcgg gcggccagga 60
 5 tcátgtccac caccacatgc caagtgggtg cgttcctcct gtccatcctg gggctggccg 120
 gctgcatcgc ggccaccggg atggacatgt ggagcaccca ggacctgtac gacaacccccg 180
 tcacctccgt gttccagtac gaagggctct ggaggagctg cgtgaggcag agttcaggct 240
 10 tcaccgaatg caggccctat ttcaccatcc tgggacttcc agccatgctg caggcagtgc 300
 gagccctgat gatcgtaggc atcgtcctgg gtgccattgg cctcctggta tccatcttg 360
 ccctgaaatg catccgcatt ggcagcatgg aggactctgc caaagccaac atgacactga 420
 15 cctccgggat catgttcatt gtctcaggtc tttgtgcaat tgctggagtg tctgtgtttg 480
 ccaacatgct ggtgactaac ttctggatgt ccacagctaa catgtacacc ggcattgggtg 540
 20 ggatgggtgca gactgttcag accaggtaca catttgggtgc ggctctgttc gtgggctggg 600
 tcgctggagg cctcacacta attgggggtg tgatgatgtg catcgctgc cggggcctgg 660
 caccagaaga aaccaactac aaagccgttt cttatcatgc ctcaggccac agtgttgcct 720
 25 acaagcctgg aggcttcaag gccagcactg gctttgggtc caacaccaa aacaagaaga 780
 tatacgatgg aggtgcccgc acagaggacg aggtacaatc ttatccttcc aagcacgact 840
 atgtgtaatg ctctaagacc tctcagcacg ggcggaagaa actcccggag agctcaccca 900
 30 aaaaacaagg agatcccatc tagatttctt cttgcttttg actcacagct ggaagttaga 960
 aaagcctcga tttcatcttt ggagaggcca aatggcttta gcctcagtct ctgtctctaa 1020
 35 atattccacc ataaaacagc tgagttattt atgaattaga ggctatagct cacattttca 1080
 atcctctatt tcttttttta aatataactt tctactctga tgagagaatg tggttttaat 1140
 ctctctctca cattttgatg atttagacag actccccctc ttctcctag tcaataaacc 1200
 40 cattgatgat ctatttccca gcttatcccc aagaaaactt ttgaaaggaa agagtagacc 1260
 caaagatggt attttctgct gtttgaattt tgtctcccca cccccaactt ggctagtaat 1320
 aaacacttac tgaagaagaa gcaataagag aaagatattt gtaatctctc cagcccatga 1380
 45 tctcggtttt cttacactgt gatcttaaaa gttaccaaac caaagtcatt ttcagtttga 1440
 ggcaacaaaa cctttctact gctgttgaca tcttcttatt acagcaacac cattctagga 1500
 50 gtttcctgag ctctccactg gagtcctctt tctgtcgcgg gtcagaaatt gtccctagat 1560
 gaatgagaaa attatttttt ttaatttaag tcctaaatat agttaaata aataatgttt 1620
 tagtaaaatg atacactatc tctgtgaaat agcctcacc ctcacatgtg atagaaggaa 1680
 55 atgaaaaaat aattgctttg acattgtcta tatggactt tgtaaagtca tgcttaagta 1740
 caaattccat gaaaagctca ctgatcctaa ttctttcctt ttgaggtctc tatggctctg 1800

60

65

5 attgtacatg atagtaagtg taagccatgt aaaaagtaaa taatgtctgg gcacagtggc 1860
 tcacgcctgt aatcctagca ctttgggagg ctgaggagga aggatcactt gagcccagaa 1920
 gttcgagact agcctgggca acatggagaa gccctgtctc tacaaaatac agagagaaaa 1980
 aatcagccag tcatggtggc ctacacctgt agtcccagca ttccgggagg ctgaggtggg 2040
 10 aggatcactt gagcccaggg aggttggggc tgcagtgagc catgatcaca ccaactgcact 2100
 ccagccaggt gacatagcga gatcctgtct aaaaaataa aaaataaata atggaacaca 2160
 gcaagtcta ggaagtaggt taaaactaat tctttaaaaa aaaaaaaaaag ttgagcctga 2220
 15 attaaatgta atgtttccaa gtgacaggta tccacatttg catggttaca agccactgcc 2280
 agttagcagt agcactttcc tggcactgtg gtcggttttg ttttgttttg ctttgtttag 2340
 agacggggc tcactttcca ggctggcctc aaactcctgc actcaagcaa ttcttctacc 2400
 20 ctggcctccc aagtagctgg aattacaggt gtgcgccatc acaactagct ggtggtcagt 2460
 tttgttactc tgagagctgt tcacttctct gaattcacct agagtggttg gaccatcaga 2520
 tgtttgggca aaactgaaag ctctttgcaa ccacacacct tccctgagct tacatcactg 2580
 25 cccttttgag cagaaagtct aaattccttc caagacagta gaattccatc ccagtaccaa 2640
 agccagatag gccccctagg aaactgaggt aagagcagtc tctaaaaact acccacagca 2700
 gcattggtgc aggggaactt ggccattagg ttattatttg agaggaaagt cctcacatca 2760
 30 atagtacata tgaaagtgac ctccaagggg attggtgaat actcataagg atcttcaggc 2820
 tgaacagact atgtctgggg aaagaacgga ttatgcccc a taaataaca agttgtgttc 2880
 aagagtcaga gcagtgagct cagaggccct tctcactgag acagcaacat ttaaaccaaa 2940
 35 ccagaggaag tatttgtgga actcactgcc tcagtttggg taaaggatga gcagacaagt 3000
 caactaaaga aaaaagaaaa gcaaggagga gggttgagca atctagagca tggagtttgt 3060
 taagtgtctt ctggatttga gttgaagagc atccatttga gttgaaggcc acagggcaca 3120
 40 atgagctctc ccttctacca ccagaaagtc cctggtcagg tctcaggtag tgcggtgtgg 3180
 ctgagctggg tttttaatta gcgcattctc tatccaacat ttaattgttt gaaagcctcc 3240
 atatagttag attgtgcttt gtaattttgt tgttgttgct ctatcttatt gtatatgcat 3300
 45 tgagtattaa cctgaatggt ttgttactta aatattaaaa acactgttat cctacagt 3359

50 <210> 92
 <211> 733
 <212> ADN
 <213> humano

55 <400> 92

60 gggatccgga gcccaaactt tctgacaaaa ctcacacatg cccaccgtgc ccagcacctg 60
 aattcgaggg tgcaccgtca gtcttctct tcccccaaa acccaaggac accctcatga 120

65

5 tctccccggac tcctgaggtc acatgCGTgG tggtggacgt aagccacgaa gaccctgagg 180
 tcaagttcaa ctggtacgtg gacggcgtgg aggtgcataa tgccaagaca aagccgcggg 240
 aggagcagta caacagcacg tacCGTgTgG tcagcgtcct caccgtcctg caccaggact 300
 ggctgaatgg caaggagtac aagtGcaagg tctccaacaa agccctccca acccccatcg 360
 10 agaaaacat ctccaaagcc aaagggcagc cccgagaacc acaggtgtac accctgcccc 420
 catccccggga tgagctgacc aagaaccagg tcagcctgac ctgcctggtc aaaggcttct 480
 atccaagcga catcgccgtg gagtgggaga gcaatgggca gccggagaac aactacaaga 540
 15 ccacgcctcc cgtgctggac tccgacggct ccttcttct ctacagcaag ctaccgtgg 600
 acaagagcag gtggcagcag gggaaCGTct tctcatgctc cgtgatgat gaggctctgc 660
 20 acaaccacta cacgcagaag agcctctccc tgtctccggg taaatgagtg cgacggccgc 720
 gactctagag gat 733

25 <210> 93
 <211> 1076
 <212> ADN
 <213> humano

30 <400> 93

atggtggttg aggttgattc catgccggct gcctcttctg tgaagaagcc atttggctc 60
 35 aggagcaaga tgggcaagtG gtgctgccgt tgcttcccc gctgcagggg gagcggcaag 120
 agcaacgtgg gcacttctgg agaccacgac gactctgcta tgaagacact caggagcaag 180
 atgggcaagt ggtgccgcca ctgcttcccc tgctgcaggg ggagtggcaa gagcaacgtg 240
 40 ggcgcttctg gagaccacga cgactctgct atgaagacac tcaggaacaa gatgggcaag 300
 tggtgctgcc actgcttccc ctgctgcagg gggagcggca agagcaaggt gggcgcttgg 360
 ggagactacg atgacagtgc ctcatggag cccaggtacc acgtccgtgg agaagatctg 420
 45 gacaagctcc acagagctgc ctggtggggg aaagtcccc gaaaggatct catcgtcatg 480
 ctcagggaca ctgacgtgaa caagaaggac aagcaaaaga ggactgctct acatctggcc 540
 tctgccaatg ggaattcaga agtagtaaaa ctctgctgg acagacgatg tcaacttaat 600
 50 gtccttgaca acaaaaagag gacagctctg ataaaggccg tacaatgcca ggaagatgaa 660
 tgtgcgttaa tgttgctgga acatggcact gatccaaata ttccagatga gtatggaaat 720
 accactctgc actacgctat ctataatgaa gataaattaa tggccaaagc actgctctta 780
 55 tatggtgctg atacgaatc aaaaaacaag catggcctca caccactggt acttgggtgta 840
 catgagcaaa aacagcaagt cgtgaaattt ttaatcaaga aaaaagcgaa tttaaatgca 900
 60 ctggatagat atggaaggac tgctctcata cttgctgtat gttgtggatc agcaagtata 960
 gtcagccttc tacttgagca aaatattgat gtatcttctc aagatctatc tggacagacg 1020

65

gccagagagt atgctgtttc tagtcatcat catgtaattt gccagttact ttctga 1076

5

<210> 94
 <211> 3675
 <212> ADN
 <213> humano

10

<400> 94

15 tccgagctga ttacagacac caaggaagat gctgtaaaga gtcagcagcc acagccctgg 60

ctagctggcc ctgtgggcat ttattagtaa agttttaatg acaaaagctt tgagtcaaca 120

cacccgtggg taattaacct ggatcatccc accctggaga gccatcctgc ccatgggtga 180

20 tcaaagaagg aacatctgca ggaacacctg atgaggctgc acccttggcg gaaagaacac 240

ctgacacagc tgaaagcttg gtggaaaaaa cacctgatga ggctgcaccc ttggtggaaa 300

gaacacctga cacggctgaa agcttgggtg aaaaaacacc tgatgaggct gcacacctgg 360

25 tggaggggaa atctgacaaa attcaatggt tggagaaagc gacatctgga aagttcgaac 420

agtcagcaga agaaacacct agggaaatta cgagtcctgc-aaaagaaaca tctgagaaat 480

ttacgtggcc agcaaaagga agacctagga agatcgcctg ggagaaaaaa gaagacacac 540

30 ctaggggaaat tatgagtccc gcaaaagaaa catctgagaa atttacgtgg gcagcaaaag 600

gaagacctag gaagatcgca tgggagaaaa aagaaacacc tgtaaagact ggatgctgtg 660

35 caagagtaac atctaataaa actaaagttt tggaaaaaagg aagatctaag atgattgcat 720

gtcctacaaa agaatacctt acaaaagcaa gtgccaatga tcagaggttc ccatcagaat 780

ccaaacaaga ggaagatgaa gaatattctt gtgattctctg gagtctcttt gagagtctctg 840

40 caaagattca agtgtgtata cctgagtcta tatatcaaaa agtaatggag ataaatagag 900

aagtagaaga gcctcctaag aagccatctg ccttcaagcc tgccattgaa atgcaaaaact 960

ctgttccaaa taaagccttt gaattgaaga atgaacaaac attgagagca gatccgatgt 1020

45 tcccaccaga atccaaacaa aaggactatg aagaaaattc ttgggattct gagagtctct 1080

gtgagactgt ttacagaag gatgtgtgtt tacccaaggc tacacatcaa aaagaaatag 1140

ataaaataaa tggaaaatta gaagagtctc ctaataaaga tggcttctg aaggctacct 1200

50 gcggaatgaa agtttctatt ccaactaaag ccttagaatt gaaggacatg caaactttca 1260

aagcagagcc tccggggaag ccatctgcct tcgagcctgc cactgaaatg caaaagtctg 1320

tcccaataa agccttggaa ttgaaaaatg acaaacatt gagagcagat gagatactcc 1380

55 catcagaatc caaacaaaag gactatgaag aaagtctctg ggattctgag agtctctgtg 1440

agactgtttc acagaaggat gtgtgtttac ccaaggctcc atcaaaaaga aatagataaa 1500

60 ataaatggaa aattagaagg gtctcctggt aaagatggtc ttctgaaggc taactgcgga 1560

atgaaagttt ctattccaac taaagcctta gaattgatgg acatgcaaac tttcaaagca 1620

65

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

60

65

5 tagtctgtgt caacagaata cttatttttag aagaaaaatt catgatttct tcctgaagcc 3540
 5 tacagacata aaataacagt gtgaagaatt acttgttcac gaattgcata aagctgcaca 3600
 ggattcccat ctaccctgat gatgcagcag acatcattca atccaaccag aatctcgctc 3660
 10 tgtcactcag gctgg 3675

15 <210> 95
 <211> 2658
 <212> ADN
 <213> humano
 <400> 95

20 acccagaaga ccgtgccttg cctggaagtc ctgcctgtag gcctgaagga cttgccctaa 60
 cagagcctca acaactacct ggtgattcct acttcagccc cttggtgtga gcagcttctc 120
 25 aacatgaact acagcctcca cttggccttc gtgtgtctga gtctcttcac tgagaggatg 180
 tgcattccagg ggagtcagtt caacgtcagag gtcgagcagaa gtgacaagct ttccctgcct 240
 ggctttgaga acctcacagc aggatataac aaatttctca ggcccaattt tgggtggagaa 300
 30 cccgtacaga tagcgtgac tctggacatt gcaagtatct ctagcatttc agagagtaac 360
 atggactaca cagccacat atacctccga cagcgtgga tggaccagcg gctgggtgtt 420
 gaaggcaaca agagcttcac tctggatgcc cgcctcgtgg agttcctctg ggtgccagat 480
 35 acttacattg tggagtcaa gaagtccttc ctccatgaag tcaactgtggg aaacaggctc 540
 atccgcctct tctccaatgg cacggtcctg tatgccctca gaatcacgac aactgttgca 600
 tgtaacatgg atctgtctaa ataccatg gacacacaga catgcaagtt gcagctggaa 660
 40 agctggggct atgatggaaa tgatgtggag ttcacctggc tgagagggaa cgactctgtg 720
 cgtggactgg aacacctgcg gcttgctcag tacaccatag agcgggtattt caccttagtc 780
 accagatcgc agcaggagac aggaaattac actagattgg tcttacagtt tgagcttcgg 840
 45 aggaatgttc tgtatttcat ttggaaacc tacgttcctt ccactttcct ggtggtgttg 900
 tcctgggttt cattttggat ctctctcgat tcagtccctg caagaacctg cattggagtg 960
 acgaccgtgt tatcaatgac cacactgatg atcgggtccc gcacttctct tcccaacacc 1020
 50 aactgcttca tcaaggccat cgatgtgtac ctggggatct gcttagctt tgtgtttggg 1080
 gccttgctag aatatgcagt tgctcactac agttccttac agcagatggc agccaaagat 1140
 agggggacaa caaaggaagt agaagaagtc agtattacta atatcatcaa cagctccatc 1200
 55 tccagcttta aacggaagat cagcttgcc agcattgaaa ttccagcga caacgttgac 1260
 tacagtgact tgacaatgaa aaccagcagc aagttaaagt ttgtcttccg agaaaagatg 1320
 ggcaggattg ttgattattt cacaattcaa aaccccagta atgttgatca ctattccaaa 1380
 60 ctactgtttc ctttgatttt tatgctagcc aatgtatttt actgggcata ctacatgtat 1440

65

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

cggcggcggg gctctctact cggttctcgg ctggaagaga ctctctacag tgaccaggag 480
 ctggcctatc tccagcaggg ggaggaggcc atgcagaagg ccttgggcat ccttagcaac 540
 caagagggct ggaagaagga gagtcagcag gacaatgggg acaaagtgat gagtaaagtg 600
 gtcccagatg tgggcaaggt gttccggctg gaggtcgtgg tggaccagcc catggagagg 660
 ctctatgaag agctcgtgga gcgcatggaa gcaatggggg agtggaaccc caatgtcaag 720
 gagatcaagg tcctgcagaa gatcgggggg ccccgtgact ttgtgagcgt gcgctgtgcc 780
 aagcgccgag gctccacctg tgtgctggct ggcattggcca cagacttcgg gaacatgcct 840
 gagcagaagg gtgtcatcag ggcggagcac ggtcccactt gcatggtgct tcaccggtg 900
 gctggaagtc cctctaagac caaacttacg tggctactca gcatcgacct caaggggtgg 960
 ctgccaaga gcatcatcaa ccaggctctg tcccagacc aggtggattt tgccaaccac 1020
 ctgcgcaagc gcctggagtc ccaccctgcc tctgaagcca ggtgttgaag accagcctgc 1080
 tgttccaac tgtgcccagc tgcactggta cacacgctca tcaggagaat ccctactgga 1140
 agcctgcaag tctaagatct ccatctggtg acagtgggat ggggtggggtt cgtgtttaga 1200
 gtatgacact aggattcaga ttggtgaaag tttttagtac caagaaaaca gggatgaggc 1260
 tcttgatta aaagtaact tcattcactg attagctatg acatgagggt tcaggcccct 1320
 aaaataattg taaaactttt tttctgggcc cttatgtacc cacctaaaac catctttaa 1380
 atgctagtgg ctgatatggg tgtgggggat gctaaccaca gggcctgaga agtcttgctt 1440
 tatgggctca agaatgccat gcgctggcag tacatgtgca caaagcagaa tctcagaggg 1500
 tctctgcag ccctctgctc ctcccggccg ctgcacagca acaccacaga acaagcagca 1560
 cccacagtg ggtgccttcc agaaatatag tccaagcttt ctctgtggaa aaagacaaaa 1620
 ctcattagta gacatgtttc cctattgctt tcataggcac cagtcagaat aaagaatcat 1680
 aattcacaca aacatcagtc tttgttttaa tattgtactg ttaaaaaaat ctatgcagct 1740
 ggggtcagtg gctcacgcct gtaatcccag cattttggga ggctgaggta ggcgatcac 1800
 aaggtcagga gatcgagacc atcctggcca acatggtgaa accccgtctc tactaaaata 1860
 caaaaaatta gctgggtgtg gtggcgcaaa cctgtagtcg tagctacttg ggaggctgag 1920
 gcaggggaat cactgaacc ccggaggcgg aggtttagt gggccgagat tgtgccactg 1980
 cgctccagcg tgggcgacag agtgagactc catctcaaaa aaaaaaaaaa aaaaatctat 2040
 gctagtagat tacaacttca cactagagga gttctggaca aagcttttaa ttagtcaaac 2100
 taaattaagg ctattaataa ggaaaggaac tactgggaaa ttatgcaatt caataattta 2160
 gactctgtta ccaggatctt tcataaaaat ttaatttcca taatcataac ctaaatgagt 2220
 tcttaaagaa ttctataagc aatagctgat taatgggcc tggaagatga agattataac 2280
 tgtttattta cctaattaaa aggaaaggca gtgccaaata tgagaggata aacaatatta 2340

	gttaacattt ctgttattta tgatgccaat tagtagtaag ataattccac agctgtcaac	2400
5	tttgtttggg gctggcaact tctctgctta aacaggctaa aagtttagta ttctgggaga	2460
	agtggctgga agaaggggta atatggtgaa agcaaattcc ctttcccagg agtcaagaga	2520
10	atztatgtga g	2531
	<210> 97	
	<211> 2849	
15	<212> ADN	
	<213> humano	
	<400> 97	
20	cgggcgccgc aggagcgagt gagctgggag cgaggggaga aggcgcggag aagcccggcc	60
	gcccgggtggg cggcagaagg ctcagccgag gcggcggcgc cgactccgtt ccactctcgg	120
25	cccggatcca ggcctccggg ttcccaggcg ctcacctccc tctgacgcac tttaaagagt	180
	ctccccctt ccacctcagg gcgagtaata gcgaccaatc atcaagccat ttaccaggct	240
	tcggaggaag ctgtttatgt gatccccgca ctaattaggc tcatgaacta acaaatcgtt	300
30	tgcaacaact gtgaagaagc gaacacttcc atggattgtc cttggactta gggcgccctg	360
	cccgcctttt gcagaggaga aaaaactttt tttttttttt gcctcccccg agaactttcc	420
35	ccccttctcc tccctgcctc taactccgat cccccacgc catctcgcca aaaaaaaaaa	480
	aaaaaaaaaa aaagaaaaaa aaagaaaaaa aaagaaaaaa aattacccca atccacgcct	540
	gcaaattctt ctggaaggat tttccccct ctcttcaggc tgggcgcgtt tggtgcaaga	600
40	ttctcgggat cctcggcttt gcctctcct ctccctccc cctccttcc ttttctttt	660
	cctttctttt ctttcttctt ttccttccc ccacccccac cccacccca aacaaacgag	720
	tcccgaattc tcgtccgtcc tcgccgctgg cagcgggcgg cggaggcagc gtgcggcggc	780
45	cgccaggagc tgggagccca gggcgcccgc tcctcggcgc agcatgttcc agccggcggc	840
	caagcgtgc ttcaccatcg agtcgctggt ggccaaggac agtcccctgc ccgcctcgcg	900
50	ctccgaggac cccatccgtc ccgcggcact cagctacgct aactccagcc ccataaatcc	960
	gttctctaac ggcttccact cggccgcccgc cgccgcccgc ggtaggggag tctactcaa	1020
	cccggacttg gtgttcgccg aggcgggtctc gcacccgccc aaccccgccc tgccagtgca	1080
55	cccggtgccg ccgccgcacg ccctggcccgc ccacccccta cctcctcgc actcggcaca	1140
	ccccctattc gcctcgcagc agcgggatcc gtccatcttc taccctggc tcatccaccg	1200
60	ctaccgatat ctgggtcatc gcttccaagg gaacgacact agccccgaga gtttctttt	1260
	gcacaacgcg ctggcccga aagcccaagcg gatccgaacc gccttctccc cgtcccagct	1320
	tctaaggctg gaacacgcct ttgagaagaa tctactcgtg gtgggcggcg aaaggaagca	1380
65	gctggcacac agcctcagcc tcacggaaac tcaggtaaaa gtatggtttc agaaccgaag	1440

ES 2 494 843 T3

5 aacaaagttc aaaaggcaga agctggagga agaaggctca gattcgcaac aaaagaaaaa 1500
 agggacgcac catattaacc ggtggagaat cgccaccaag caggcgagtc cggaggaaat 1560
 agacgtgacc tcagatgatt aaaaacataa acctaacccc acagaaacgg acaacatgga 1620
 10 gcaaaagaga cagggagagg tggagaagga aaaaacccta caaaacaaaa acaaaccgca 1680
 tacacgttca ccgagaaagg gagaggggaat cggaggggagc agcggaatgc ggcgaagact 1740
 ctggacagcg agggcacagg gtcccaaacc gaggccgcgc caagatggca gaggatggag 1800
 15 gctccttcat caacaagcga ccctcgtcta aagaggcagc tgagtgagag acacagagag 1860
 aaggagaaag agggagggag agagagaaag agagagaaag agagagagag agagagagag 1920
 agaaagctga acgtgcactc tgacaagggg agctgtcaat caaacaccaa accggggaga 1980
 20 caagatgatt ggcaggatatt ccgtttatca cagtccactt aaaaaatgat gatgatgata 2040
 aaaaccacga cccaaccagg cacaggactt ttttgTTTTT tgcacttcgc tgtgtttccc 2100
 ccccatcttt aaaaataatt agtaataaaa aacaaaaaatt ccatactag ccccatccca 2160
 25 cacctgtttc aaatccttga aatgcatgta gcagttgttg ggcgaatggt gtttaaagac 2220
 cgaaaatgaa ttgtaatttt ctttccctt taaagacagg ttctgtgtgc tttttatttt 2280
 gatttttttt cccaagaaat gtgcagtctg taaacacttt ttgatacctt ctgatgtcaa 2340
 30 agtgattgtg caagctaaat gaagtaggct cagcgatagt ggtcctctta cagagaaacg 2400
 gggagcagga cgacgggggg gctgggggtg gcgggggagg gtgccacaa aaagaatcag 2460
 35 gacttgtact gggaaaaaaa ccctaaatt aattatattt ctggacatt ccctttccta 2520
 acatcctgag gcttaaaacc ctgatgcaaa cttctccttt cagtggttgg agaaattggc 2580
 cgagttcaac cattcactgc aatgcctatt ccaaacttta aatctatcta ttgcaaaacc 2640
 40 tgaaggactg tagttagcgg ggatgatggt aagtgtggcc aagcgcacgg cggcaagttt 2700
 tcaagcactg agtttctatt ccaagatcat agacttacta aagagagtga caaatgcttc 2760
 45 cttaatgtct tctataccag aatgtaata ttttTgtgtt ttgtgttaat ttgttagaat 2820
 tctaacacac tatatacttc caagaagta 2849

50 <210> 98
 <211> 3308
 <212> ADN
 <213> humano

55 <400> 98

60 aaaagataga tectgctcca ggagccggga agcctcggcc tggccagctg tgctgggcac 60
 ctcccctgcc tgcttctggt cccacttgca ggcaagggtga gggcatgcga atggctgcca 120
 ctgcctgggc ggggctccaa gggccacccc tccccaccct ctgtcccgca gtgaggacgg 180
 65 gactctactg ccgagaccag gctcacgctg agaggtgggc catgacctcc gagacctctt 240

ccggaagcca ctgtgccagg agcaggatgc tgcggcgacg ggcccaggaa gaggacagca 300
 5 ccgtcctgat cgatgtgagc ccccctgagg cagagaagag gggctcttac gggagcacag 360
 cccacgcctc ggagccagggt ggacagcaag cggccgcctg cagagctggg agtcctgcca 420
 agccccggat cgcagacttc gtcctcgttt gggaggagga cctgaagcta gacaggcagc 480
 10 aggacagtgc cccccgggac agaacagaca tgcacaggac ctggcgggag acttttctgg 540
 ataatcttcg tgcggctggg ctgtgtgtag accagcagga cgtccaggac gggaaacca 600
 cagtgcacta cgcctcctc agcgcctcct gggctgtgct ctgctactac gccgaagacc 660
 15 tgcgcctgaa gctgccctg caggagtac ccaaccaggc ctccaactgg tcggccggcc 720
 tgctggcatg gctgggcatc cccaacgtcc tgctggagggt tgtgccagac gtaccccccg 780
 agtactactc ctgccggtc agagtgaaca agctgccacg cttcctcggg agtgacaacc 840
 20 aggacacctt cttcacaagc accaagaggc accaaattct gtttgagatc ctggccaaga 900
 ccccgatagg ccacgagaag aaaaacctgc ttgggatcca ccagctgctg gcagaggggtg 960
 25 tcctcagtgc cgccttcccc ctgcatgacg gcccttcaa gacgccccca gagggcccgc 1020
 aggctccacg cctcaaccag cgccaagtcc tttccagca ctgggcgcgc tggggcaagt 1080
 ggaacaagta ccagcccctg gaccacgtgc gcaggtactt cggggagaag gtggccctct 1140
 30 acttcgctg gctcgggtt tacacaggct ggctcctgcc agcggcagt gtgggcacac 1200
 tgggttctt ggtgggctgc ttctggtgt tctcagacat acccacgcag gaactgtgtg 1260
 gcagcaagga cagcttcgag atgtgccac tttgcctcga ctgcccttc tggtctctct 1320
 35 ccagcgcctg tgccctggcc caggccggcc ggctgttcga ccacggcggc accgtgttct 1380
 tcagcttgtt catggcactg tgggccgtgc tgctgctgga gtactggaag cggaaagagc 1440
 ccacgctggc ctaccgctgg gactgctctg actacgagga cactgaggag aggcctcggc 1500
 40 cccagtttgc cgcctcagcc cccatgacag ccccgaacc catcacgggt gaggacgagc 1560
 cctacttccc tgagaggagc cgcgcgcgcc gcatgctggc cggctctgtg gtgatcgtgg 1620
 45 tgatggtggc cgtggtggtc atgtgcctcg tgtctatcat cctgtaccgt gccatcatgg 1680
 ccacgtgggt gtccaggctg ggcaacaccc ttctcgcagc ctgggcctct cgcacgcca 1740
 gcctcacggg gtctgtagtg aacctcgtct tcactctcat cctctccaag atctatgtat 1800
 50 ccttgccca cgtcctgaca cgatgggaaa tgaccgcac ccagaccaag ttcgaggagc 1860
 ccttcacct caaggtgttc atcttccagt tcgtcaactt ctactcctca cccgtctaca 1920
 ttgccttctt caagggcagg tttgtgggat acccaggcaa ctaccacacc ttgtttggag 1980
 55 tccgcaatga ggagtgcgcg gctggaggct gcctgatcga gctggcacag gagctcctgg 2040
 tcatcatggt gggcaagcag gtcatcaaca acatgcagga ggtcctcatc ccgaagctaa 2100

60

65

5 agggctgggtg gcagaagttc cggcttcgct ccaagaagag gaaggcggga gcttctgcag 2160
 gggctagcca gggggccctgg gaggacgact atgagcttgt gccctgtgag ggtctgtttg 2220
 acgagtacct ggaaatgggtg ctgcagttcg gcttcgtcac catcttcgtg gccgcctgtc 2280
 cgctcgcgcc gctcttcgcc ctgctcaaca actgggtgga gatccgcttg gacgcgcgca 2340
 10 agttcgtctg cgagtaccgg cggcctgtgg ccgagcgcgc ccaggacatc ggcattctggt 2400
 tccacatcct ggcgggcctc acgcacctgg cggatcatcag caacgccttc ctctggcct 2460
 tctcgtccga cttcctgccg cgcgcctact accggtggac ccgcgccac gacctgcgcg 2520
 15 gcttcctcaa cttcacgctg gcgcgagccc cgtcctcctt cggcgccgcg cacaaccgca 2580
 cgtgcaggta tcgggctttc cgggatgacg atggacatta ttcccagacc tactggaatc 2640
 20 ttcttgccat ccgcctggcc ttcgtcattg tgtttgagca tgtggttttc tccggtggcc 2700
 gcctcctgga cctcctgggtg cctgacatec cagagtctgt ggagatcaaa gtgaagcggg 2760
 agtactacct ggctaagcag gcaactggctg agaatgaggt tctttttgga acgaacggaa 2820
 25 caaaggatga gcagcccaag ggctcagagc tcagctccca ctggacaccc ttcacggttc 2880
 ccaaggccag ccagctgcag cagtgcgcc tggaggaca tctggtggtc cttaggggag 2940
 tggcccctcc tgagccctgc gagcagcgtc ctttctctt tccctcaggc agcggctgtg 3000
 30 tgaaccgctg gctgctgttg tgccctatct ctgggcacat tgccctgctt cccccagcgc 3060
 cggcttctct cctcagagcg cctgtcactc catccccggc agggagggac cgtcagctca 3120
 caaggccctc tttgtttcct gctcccagac ataagcccaa ggggcccctg cacccaaggg 3180
 35 acctgtccc tcgggtggcct ccccaggccc ctggacacga cagttctcct caggcaggtg 3240
 ggctttgtgg tcctcgcgc cctggccac atcgcctctt cctcttacac ctggtgacct 3300
 40 tcgaatgt 3308

45 <210> 99
 <211> 551
 <212> ADN
 <213> humano

50 <400> 99
 accccatccg ctggtctca cccctcggag acgctcgcc gacagcatag tacttgccgc 60
 ccagccacgc ccgcgcgcca gccaccatgc taggtaacaa gcgactgggg ctgtccggac 120
 55 tgaccctcgc cctgtccctg ctctgtgtgcc tgggtgcgct ggccgaggcg taccctcca 180
 agccggacaa cccgggagag gacgcaccag cggaggacat ggccagatac tactcggcgc 240
 tgcgacacta catcaacctc atcaccaggc agagatatgg aaaacgatcc agcccagaga 300
 60 cactgatttc agacctctg atgagagaaa gcacagaaaa tgttcccaga actcggcttg 360
 aagaccctgc aatgtgggtga tgggaaatga gacttgctct ctggcctttt cctattttca 420

65

gcccatatTT catcgtgtaa aacgagaatc cacccatcct accaatgcat gcagccactg 480
 5 tgctgaattc tgcaatgTTT tcctttgtca tcattgtata tatgtgtgTT taaataaagt 540
 atcatgcatt c 551

10 <210> 100
 <211> 1607
 <212> ADN
 <213> humano

15 <400> 100

20 aatgactcct ttcggtaagt gcagtggaag ctgtacactg cccaggcaaa gcgtccgggc 60
 agcgtaggcg ggcgactcag atcccagcca gtggacttag ccctgtttg ctctccgat 120
 aactggggTg accttggtta atattcacca gcagcctccc ccgTTgcccc tctggatcca 180
 25 ctgcttaaAT acggacgagg acagggccct gtctcctcag cttcaggcac caccactgac 240
 ctgggacagt gaatcgacaa tgccgtcttc tgtctcgtgg ggcacccctc tgctggcagg 300
 cctgtgctgc ctggTccctg tctccctggc tgaggatccc cagggagatg ctgcccagaa 360
 30 gacagataca tcccaccatg atcaggatca cccaaccttc aacaagatca cccccaacct 420
 ggctgagttc gccttcagcc tataccgcca gctggcacac cagtccaaca gcaccaatat 480
 ctctctctcc ccagtgagca tcgctacagc ctttgcaatg ctctccctgg ggaccaaggc 540
 35 tgacactcac gatgaaatcc tggagggcct gaatttcaac ctcacggaga ttccggaggc 600
 tcagatccat gaaggcttc aggaactcct ccgtaccctc aaccagccag acagccagct 660
 ccagctgacc accggcaatg gcctgttctc cagcgagggc ctgaagctag tggataagtt 720
 40 tttggaggat gttaaaaagt tgtaccactc agaagccttc actgtcaact tcggggacac 780
 cgaagaggcc aagaaacaga tcaacgatta cgtggagaag ggtactcaag ggaaaattgt 840
 45 ggattttgTc aaggagcttg acagagacac agtttttgct ctggtgaatt acatcttctt 900
 taaaggcaaa tgggagagac ctttgaagt caaggacacc gaggaagagg acttccacgt 960
 ggaccaggTg accaccgtga aggtgcctat gatgaagcgt ttaggcatgt ttaacatcca 1020
 50 gcactgtaag aagctgtcca gctgggtgct gctgatgaaa tacctgggca atgccaccgc 1080
 catcttcttc ctgcctgatg aggggaaact acagcacctg gaaaatgaac tcaccacga 1140
 tatcatcacc aagttcctgg aaaatgaaga cagaaggtct gccagcttac atttacccaa 1200
 55 actgtccatt actggaacct atgatctgaa gagcgtcctg ggtcaactgg gcatcactaa 1260
 ggtcttcagc aatggggctg acctctccgg ggtcacagag gaggcacccc tgaagctctc 1320
 caaggccgtg cataaggctg tgctgaccat cgacgagaaa gggactgaag ctgctggggc 1380
 60 catgttttta gaggccatac ccatgtctat ccccccgag gtcaagttca acaaaccctt 1440
 tgtcttctta atgattgaac aaaataccaa gtctccctc tcatgggaa aagtggTgaa 1500

65

	tccccccaa aaataactgc ctctcgctcc tcaaccctc cctccatcc ctggccccct	1560
5	ccctggatga cattaagaa gggttgagct ggtccctgcc tgcaaaa	1607
10	<210> 101 <211> 1753 <212> ADN <213> humano	
15	<400> 101	
15	cagccccgcc cctacctgtg gaagcccagc cgcccgtcc cgcggataaa aggcgcggag	60
	tgtccccgag gtcagcagat gcgcgctcct cctcgcccgc cgctaggtcc atccccggccc	120
20	agccaccatg tccatccact tcagctcccc ggtattcacc tcgcgctcag ccgccttctc	180
	gggcccgggc gccaggtgc gcctgagctc cgctcgcgcc ggcggccttg gcagcagcag	240
	cctctacggc ctcggcgcct cacggccgcg cgtggccgtg cgctctgcct atgggggccc	300
25	ggtggggccc ggcattccgc aggtcaccat taaccagagc ctgctggccc cgctgcggct	360
	ggacgcccgc cctccctcc agcgggtgcg ccaggaggag agcagcaga tcaagacct	420
	caacaacaag ttgctcct tcattcgaca ggtgcggttt ctggagcagc agaacaagct	480
30	gctggagacc aagtggacgc tgctgcagga gcagaagtcg gccaaagca gccgcctccc	540
	agacatcttt gaggccaga ttgctggcct tcggggtcag cttgaggcac tgcaggtgga	600
	tgggggcccgc ctggaggcgg agctgcggag catgcaggat gtggtggagg acttcaagaa	660
35	taagtacgaa gatgaaatta accaccgcac agctgctgag aatgagttg tgggtgctgaa	720
	gaaggatgtg gatgctgcct acatgagcaa ggtggagctg gaggccaagg tggatgcct	780
40	gaatgatgag atcaacttcc tcaggacct caatgagacg gagttgacag agctgcagtc	840
	ccagatctcc gacacatctg tgggtgctgc catggacaac agtcgctccc tggacctgga	900
	cggcatcatc gctgaggtca aggcgcagta tgaggagatg gccaaatgca gccgggctga	960
45	ggctgaagcc tgggtaccaga ccaagtttga gacctccag gccaggtcg ggaagcatgg	1020
	ggacgacctc cggaataccc ggaatgagat ttcagagatg aaccgggcca tccagaggct	1080
	gcaggctgag atcgacaaca tcaagaacca gcgtgccaag ttggaggccg ccattgccga	1140
50	ggctgaggag cgtggggagc tggcgtcaa ggatgctcgt gccaaagcagg aggagctgga	1200
	agccgcccctg cagcggggca agcaggatat ggcacggcag ctgcgtgagt accaggaact	1260
	catgagcgtg aagctggccc tggacatcga gatcgccacc taccgcaagc tgctggaggg	1320
55	cgaggagagc cggttggctg gagatggagt gggagccgtg aatatctctg tgatgaattc	1380
	cactgggtggc agtagcagtg gcggtggcat tgggctgacc ctcgggggaa ccatgggcag	1440
60	caatgccctg agcttctcca gcagtgcggg tcctgggctc ctgaaggctt attccatccg	1500
	gaccgcatcc gccagtcgca ggagtgccg cgactgagcc gcctcccacc actccactcc	1560
65		

5 tccagccacc acccacaatc acaagaagat tcccaccctt gcctcccatg cctgggtccca 1620
 agacagtgag acagtctgga aagtgatgtc agaatagctt ccaataaagc agcctcattc 1680
 tgaggcctga gtgatccacg tgaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1740
 aaaaaaaaaa aaa 1753

10 <210> 102
 <211> 2276
 <212> ADN
 <213> humano
 15 <400> 102

20 aagcccagca gccccggggc ggatggctcc ggccgcctgg ctccgcagcg cggccgcgcg 60
 cgccctcctg cccccgatgc tgctgtgtct gctccagccg ccgccgctgc tggcccgggc 120
 tctgccgccc gacgcccacc acctccatgc cgagaggagg gggccacagc cctggcatgc 180
 25 agccctgccc agtagcccgg cacctgcccc tgccaagcag gaagcccccc ggctgtccag 240
 cagcctcagg cctccccgct gtggcggtgc cgaccatct gatgggctga gtgcccgcaa 300
 ccgacagaag aggttcgtgc tttctggcgg gcgctgggag aagacggacc tcacctacag 360
 30 gatccttcgg ttcccatggc agttggtgca ggagcaggtg cggcagacga tggcagaggc 420
 cctaaaggta tggagcgatg tgacgccact cacctttact gaggtgcacg agggccgtgc 480
 tgacatcatg atcgacttcg ccaggtactg gcatggggac gacctgccgt ttgatgggcc 540
 35 tgggggcatc ctggcccatg cttcttccc caagactcac cgagaagggg atgtccactt 600
 cgactatgat gagacctgga ctatcgggga tgaccagggc acagacctgc tgcaggtggc 660
 40 agcccatgaa tttggccacg tgctggggct gcagcacaca acagcagcca aggcctgat 720
 gtccgccttc tacaccttc gctaccact gagtctcagc ccagatgact gcaggggctg 780
 tcaacaccta tatggccagc cctggcccac tgtcacctcc aggaccccag ccctgggccc 840
 45 ccaggctggg atagacacca atgagattgc accgctggag ccagacgccc cgccagatgc 900
 ctgtgaggcc tcctttgacg cggctccac catccgaggc gagctctttt tcttcaaagc 960
 gggctttgtg tggcgcctcc gtgggggcca gctgcagccc ggctaccag cattggcctc 1020
 50 tcgccactgg cagggactgc ccagccctgt ggacgctgcc ttcgaggatg cccagggcca 1080
 catttggttc ttccaaggtg ctcagtactg ggtgtacgac ggtgaaaagc cagtcctggg 1140
 55 ccccgacccc ctaccgagc tgggcctggt gaggttcccg gtccatgctg ccttggctctg 1200
 gggtcccagag aagaacaaga tetacttctt ccgaggcagg gactactggc gtttccaccc 1260
 cagcaccggc cgtgtagaca gtcccgtgcc ccgagggcc actgactgga gaggggtgcc 1320
 60 ctctgagatc gacgctgcct tccaggatgc tgatggctat gcctacttcc tgcgcggccc 1380
 cctctactgg aagtttgacc ctgtgaaggt gaaggctctg gaaggcttcc cccgtctctg 1440

65

5 gggcctgac ttctttggct gtgccgagcc tgccaacact ttcctctgac catggcttgg 1500
 atgccctcag ggggtgctgac ccctgccagg ccacgaatat caggctagag acccatggcc 1560
 atctttgtgg ctgtgggcac caggcatggg actgagccca tgtctcctca gggggatggg 1620
 10 gtggggtaca accaccatga caactgccgg gagggccacg caggctcgtgg tcacctgcca 1680
 gcgactgtct cagactgggc agggaggctt tggcatgact taagaggaag ggcagtcttg 1740
 ggcccgctat gcaggctctg gcaaacctgg ctgccctgtc tccatccctg tccctcaggg 1800
 15 tagcaccatg gcaggactgg gggaaactgga gtgtccttgc tgtatccctg ttgtgaggtt 1860
 ccttccaggg gctggcactg aagcaagggg gctggggccc catggccttc agccctggct 1920
 gagcaactgg gctgtagggc agggccactt cctgaggcca ggtcttgga ggtgcctgca 1980
 20 tctgtctgcc ttctggctga caatcctgga aatctgttct ccagaatcca ggccaaaaag 2040
 ttcacagtca aatggggagg ggtattcttc atgcaggaga ccccaggccc tggaggctgc 2100
 aacatactc aatcctgtcc caggccggat cctcctgaag cccttttcgc agcactgcta 2160
 25 tcctccaaag ccattgtaaa tgtgtgtaca gtgtgtataa accttcttct tctttttttt 2220
 tttttaact gaggattgtc attaaacaca gttgttttct aaaaaaaaaa aaaaaa 2276

30 <210> 103
 <211> 7381
 <212> ADN
 <213> humano

35 <400> 103

40 tacagcccca aggtcgtccc ctctggggcc ctttcttccc cattcttccc agcagcccaa 60
 agctctgggtg ggacaggggc agccccctggg gagggaggag aggaccagg aaccgggcta 120
 ggaggggtggc ccaccattt ccagtgtgac ctgttccat tccccatgt ctctcccat 180
 45 cctccccgc actcagctca ggctgatgag aagcagagca acgggtgtat cgggtgtttc 240
 tttcctgggtg gggtagtggg gtggggctga ggagagaaaa ggggtgattag cgtggggccc 300
 cgccctcttt tgtctcttc ccaggttccc tggccccttc ggagaaacgc acttggttcg 360
 50 ggccagccgc ctgaggggac gggctcacgt ctgctcctca cactgcagct gctgggcccgt 420
 ggagcttccc cagggagcca gggggacttt tgccgcagcc atgaaggggg cacgctggag 480
 gaggggtccc tgggtgtccc tgagctgcct gtgtctctgc ctcttccgc atgtggtccc 540
 55 aggaaccaca gaggacacat taataactgg aagtaaaact cctgccccag tcacctcaac 600
 aggctcaaca acagcgacac tagagggaca atcaactgca gcttcttcaa ggacctctaa 660
 tcaggacata tcagcttcat ctcagaacca ccagactaag agcacggaga ccaccagcaa 720
 60 agctcaaacc gacaccctca cgcagatgat gacatcaact ctttttctt cccaagtgt 780
 acacaatgtg atggagactg ttacgcagga gacagctcct ccagatgaaa tgaccacatc 840

65

	atttcctcc	agtgtcacca	acacactcat	gatgacatca	aagactataa	caatgacaac	900
5	ctccacagac	tccactcttg	gaaacacaga	agagacatca	acagcaggaa	ctgaaagtcc	960
	taccccagtg	acctcagcag	tctcaataac	agctggacag	gaaggacaat	cacgaacaac	1020
	ttcctggagg	acctctatcc	aagacacatc	agcttcttct	cagaaccact	ggactcggag	1080
10	cacgcagacc	accaggggaat	ctcaaaccag	caccctaaca	cacagaacca	cttcaactcc	1140
	ttctttctct	ccaagtgtac	acaatgtgac	agggactgtt	tctcagaaga	catctccttc	1200
15	aggtgaaaca	gctacctcat	ccctctgtag	tgtcacaaac	acatccatga	tgacatcaga	1260
	gaagataaca	gtgacaacct	ccacaggctc	cactcttgga	aaccaggggg	agacatcatc	1320
	agtacctgtt	actggaagtc	ttatgccagt	cacctcagca	gccttagtaa	cagttgatcc	1380
20	agaaggacaa	tcaccagcaa	ctttctcaag	gacttctact	caggacacaa	cagctttttc	1440
	taagaaccac	cagactcaga	gcgtggagac	caccagagta	tctcaaatca	acaccctcaa	1500
	caccctcaca	ccggttacia	catcaactgt	ttatcctca	ccaagtggat	tcaacccaag	1560
25	tggaacagtt	tctcaggaga	cattcccttc	tggtgaaaca	accatctcat	ccccttcag	1620
	tgtcagcaat	acattcctgg	taacatcaaa	gggtttcaga	atgccaatct	ccagagactc	1680
30	tactcttgga	aacacagagg	agacatcact	atctgtaagt	ggaaccattt	ctgcaatcac	1740
	ttccaaagtt	tcaaccatat	ggtggtcaga	cactctgtca	acagcactct	ccccagttc	1800
	tctacctcca	aaaatatcca	cagctttcca	caccagcag	agtgaagggtg	cagagaccac	1860
35	aggacggcct	catgagagga	gctcattctc	tccaggtgtg	tctcaagaaa	tatttactct	1920
	acatgaaaca	acaacatggc	cttctctatt	ctccagcaaa	ggccacacaa	cttggtcaca	1980
	aacagaactg	ccctcaacat	caacaggtgc	tgccactagg	cttgtcacag	gaaatccatc	2040
40	tacaagggca	gctggcacta	ttccaagggg	cccctctaag	gtctcagcaa	taggggaacc	2100
	aggagagccc	accacatact	cctcccacag	cacaactctc	ccaaaaacaa	caggggcagg	2160
45	cgcccagaca	caatggacac	aagaaacggg	gaccactgga	gaggctcttc	tcagcagccc	2220
	aagctatagt	gtgattcaga	tgataaaaac	ggccacatcc	ccatcttctt	cacctatgct	2280
	ggatagacac	acatcacaac	aaattacaac	ggcaccatca	acaaatcatt	caacaataca	2340
50	ttccacaagc	acctctctc	aggaatcacc	agctgtttcc	caaaggggtc	acactcgagc	2400
	cccgcagacc	acacaagaat	cacaaaccac	gaggtccgtc	tccccatga	ctgacaccaa	2460
	gacagtcacc	accccaggtt	cttctttcac	agccagtggg	cactcgccct	cagaaattgt	2520
55	tcctcaggac	gcaccacca	taagtgcagc	aacaaccttt	gccccagctc	ccaccgggaa	2580
	tggtcacaca	accagggccc	cgaccacagc	actgcaggca	gcaccagca	gccatgatgc	2640
60	cacctggggg	ccctcaggag	gcacgtcact	ttccaaaaca	ggtgccctta	ctctggccaa	2700
	ctctgtagtg	tcaacaccag	ggggcccaga	aggacaatgg	acatcagcct	ctgccagcac	2760

65

5 ctcacctgac acagcagcag ccatgaccca taccaccag gctgagagca cagaggcctc 2820
 tggacaaaca cagaccagcg aaccggcctc ctcaggggtca cgaaccacct cagcgggac 2880
 agctacccct tcctcatccg gggcgagtgg cacaacacct tcaggaagcg aaggaatattc 2940
 cacctcagga gagacgacaa ggttttcatc aaaccctcc agggacagtc acacaacca 3000
 10 gtcaacaacc gaattgctgt ccgcctcagc cagtcatggt gccatcccag taagcacagg 3060
 aatggcgtct tcgatcgtcc ccggcacctt tcatcccacc ctctctgagg cctccactgc 3120
 agggagaccg acaggacagt caagcccaac ttctcccagt gcctctctc aggagacagc 3180
 15 cgccatttcc cggatggccc agactcagag gacaggaacc agcagaggggt ctgacactat 3240
 cagcctggcg tcccaggcaa ccgacacctt ctcaacagtc ccaccacac ctccatcgat 3300
 cacatccagt gggcttacat ctccacaaac ccagaccac actctgtcac cttcagggtc 3360
 20 tggtaaaacc ttcaccacgg ccctcatcag caacgccacc cctcttctg tcaccagcac 3420
 ctctcagcc tccacaggtc acgccacccc tcttgtgtc agcagtgcta cctcagctc 3480
 cacagtatcc tcggactccc ctctgaagat ggaaacatca ggaatgacaa caccgtcact 3540
 25 gaagacagac ggtgggagac gcacagccac atcaccaccc cccacaacct cccagaccat 3600
 catttcacc attcccagca ctgccatgca caccgctcc acagctgccc ccatcccat 3660
 30 cctgcctgag agaggagttt ccctcttccc ctatggggca ggcgccgggg acctggagtt 3720
 cgtcaggagg accgtggact tcacctccc actcttcaag ccggcgactg gcttccccct 3780
 tggctcctct ctccgtgatt ccctctactt cacagacaat ggccagatca tcttcccaga 3840
 35 gtcagactac cagattttct cctaccccaa cccactcca acaggcttca caggccggga 3900
 ccctgtggcc ctggtggctc cgttctggga cgatgctgac ttctccactg gtcgggggac 3960
 cacatthtat caggaatacg agacgttcta tggatgaacac agcctgctag tccagcaggc 4020
 40 cgagtcttgg attagaaaga tgacaaaca cgggggctac aaggccaggt gggccctaaa 4080
 ggtcacgtgg gtcaatgcc acgcctatcc tgcccagtg accctcggga gcaacaccta 4140
 45 ccaagccatc ctctccacgg acgggagcag gtcctatgcc ctgtttctct accagagcgg 4200
 tgggatgcag tgggacgtgg cccagcgtc aggcaaccg gtgctcatgg gcttctctag 4260
 tggagatggc tatttcgaaa acagcccact gatgtcccag ccagtgtggg agaggtatcg 4320
 50 ccctgataga ttctgaatt ccaactcagg cctccaaggg ctgcagttct acaggctaca 4380
 ccgggaagaa aggcccaact accgtctcga gtgcctgcag tggctgaaga gccagcctcg 4440
 gtggcccagc tggggctgga accaggtctc ctgccctgt tcctggcagc agggacgacg 4500
 55 ggacttacga ttccaacccg tcagcatagg tcgctggggc ctcggcagta ggcagctgtg 4560
 cagcttcacc tcttggcgag gaggcgtgtg ctgcagctac gggccctggg gagagttctg 4620

60

65

tgaaggctgg cacgtgcagc gtccttgga gttggcccag gaactggagc cacagagctg 4680
 5 gtgctgcccg tggaatgaca agccctacct ctgtgccctg taccagcaga ggcggcccca 4740
 cgtgggctgt gctacataca ggccccaca gccgcctgg atgttcgggg acccccacat 4800
 caccaccttg gatggtgtca gttacacctt caatgggctg ggggacttcc tgctggtcgg 4860
 10 ggccaagac gggaaactcct ctttctgct tcagggccgc accgcccaga ctggctcagc 4920
 ccaggccacc aacttcatcg cttttgccc tcagtaccgc tccagcagcc tgggccccgt 4980
 cacggtcaa tggctcctg agcctcacga cgcaatccgt gtcctgctgg ataaccagac 5040
 15 tgtgacattt cagcctgacc atgaagacgg cggaggccag gagacgttca acgccaccgg 5100
 agtctctctg agccgcaacg gctctgaggt ctcggccagc ttcgacggct gggccaccgt 5160
 20 ctggtgatc gcgctctcca acatctcca cgcctccgc agcctcccgc ccgagtacca 5220
 gaaccgcacg gaggggctcc tgggggtctg gaataacaat ccagaggacg acttcaggat 5280
 gcccaatggc tccaccattc ccccaggag ccctgaggag atgcttttcc actttggaat 5340
 25 gacctggcag atcaacggga caggcctcct tggcaagagg aatgaccagc tgccttcaa 5400
 cttcacccct gttttctact cacaactgca aaaaaacagc tcctgggctg aacatttgat 5460
 ctccaactgt gacggagata gctcatgcat ctatgacacc ctggccctgc gcaacgcaag 5520
 30 catcggactt cacacgaggg aagtcagtaa aaactacgag caggcgaacg ccaccctcaa 5580
 tcagtacccg cctccatca atggtggtcg tgtgattgaa gcctacaagg ggcagaccac 5640
 35 gctgattcag tacaccagca atgctgagga tgccaactc acgctcagag acagctgcac 5700
 cgacttggag ctctttgaga atgggacgtt gctgtggaca cccaagtcgc tggagccatt 5760
 cactctggag attctagcaa gaagtgcaa gattggcttg gcatctgcac tccagcccag 5820
 40 gactgtggtc tgccattgca atgcagagag ccagtgttg tacaatcaga ccagcagggg 5880
 gggcaactcc tccctggagg tggctggctg caagtgtgac gggggcacct tcggccgcta 5940
 ctgagggggc tccgaggatg cctgtgagga gccgtgctc ccgagtgtcc actgcttcc 6000
 45 tgggaagggc tgcgaggcct gccctcaaaa cctgactggg gatgggcggc actgtgcggc 6060
 tctggggagc tcttctctgt gtcagaacca gtcctgccct gtgaattact gctacaatca 6120
 50 aggccactgc tacatctccc agactctggg ctgtcagccc atgtgcacct gccccccagc 6180
 cttcactgac agccgctgct tcctggctgg gaacaacttc agtccaactg tcaacctaga 6240
 acttccctta agagtcatcc agctcttget cagtgaagag gaaaatgcct ccatggcaga 6300
 55 ggtcaacgcc tcggtggcat acagactggg gaccctggac atgcgggcct ttctccgcaa 6360
 cagccaagtg gaacgaatcg attctgcagc accggcctcg ggaagcccca tccaactg 6420
 gatggtcatc tcggagtcc agtaccgcc tcggggcccg gtcattgact tcctgaacaa 6480
 60 ccagctgctg gccgcggtgg tggaggcgtt cttataccac gttccacgga ggagtgagga 6540

65

5 gcccaggaac gacgtggtct tccagcccat ctcccggggaa gacgtgcgcg atgtgacagc 6600
 cctgaacgtg agcacgctga aggcttactt cagatgcat ggctacaagg gctacgacct 6660
 ggtctacagc ccccagagcg gcttcacctg cgtgtccccg tgcagtaggg gctactgtga 6720
 10 ccatggaggc cagtgccagc acctgcccag tgggccccgc tgcagctgtg tgtccttctc 6780
 catctacagc gcctggggcg agcactgtga gcacctgagc atgaaactcg acgcgttctt 6840
 cggcatcttc tttggggccc tgggcggcct cttgctgctg ggggtcggga cgttcgtggt 6900
 15 cctgcgcttc tggggttgct ccggggccag gttctcctat ttctgaact cagctgaggc 6960
 cttgccttga aggggcagct gtggcctagg ctacctcaag actcacctca tccttaccgc 7020
 acatttaagg cgccattgct tttgggagac tggaaaaggg aagggtgactg aaggctgtca 7080
 20 ggattcttca aggagaaatga atactgggaa tcaagacaag actatacctt atccataggc 7140
 gcaggtgcac agggggaggc cataaagatc aaacatgcat ggatgggtcc tcacgcagac 7200
 acaccacag aaggacacta gcctgtgcac gcgcgcgtgc acacacacac acacacacac 7260
 25 gagttcataa tgtggtgatg gccctaagt aagcaaatg cttctgcaca caaaactctc 7320
 tggtttactt caaattaact ctatttaaat aaagtctctc tgactttttg tgtctccaaa 7380
 30 a 7381

35 <210> 104
 <211> 2323
 <212> ADN
 <213> humano

40 <400> 104

40 agctatgatc gcaacacctt ggtggccatc gtggtgggtg tggggcgcct catcactggc 60
 atggaccgag gcctcatggg catgtgtgtc aacgagcggc gacgcctcat tgtgcctccc 120
 45 cacctgggct atgggagcat cggcctggcg gggctcattc caccggatgc caccctctac 180
 ttcgatgtgg ttctgctgga tgtgtggaac aaggaagaca ccgtgcaggt gagcacattg 240
 ctgcgccccg cccactgccc ccgcatggtc caggacggcg actttgtccg ctaccactac 300
 50 aatggcacc cttgctggacgg cacctccttc gacaccagct acagtaaggg cggcacttat 360
 gacacctagc tgggctctgg ttggctgatc aagggcatgg accaggggct gctgggcatg 420
 tgtcctggag agagaaggaa gattatcatt cctccattcc tggcctatgg cgagaaaggc 480
 55 tatgggacag tgatcccccc acaggcctcg ctggctcttc acgtcctcct gattgacgtg 540
 cacaacccga aggacgctgt ccagctagag acgctggagc tcccccccg ctgtgtccgc 600
 60 agagccgggg ccggggactt catgcgctac cactacaatg gtccttgat ggacggcacc 660
 ctcttcgatt ccagctactc ccgcaaccac acctacaata cctatatcgg gcagggttac 720
 atcatccccg ggatggacca ggggctgcag ggtgcctgca tgggggaacg ccggagaatt 780

65

5 accatcccc cgcacctcgc ctatggggag aatggaactg gagacaagat ccctggctct 840
 gccgtgctaa tcttcaacgt ccatgtcatt gacttccaca accctgcgga tgtggtggaa 900
 atcaggacac tgtcccggcc atccgagacc tgcaatgaga ccaccaagct tggggacttt 960
 gttcgatacc attacaactg ttctttgctg gacggcacc agctgttcac ctcgcatgac 1020
 10 tacggggccc cccaggaggc gactctcggg gccacaagg tgatcgaagg cctggacacg 1080
 ggctgcagg gcatgtgtgt gggagagagg cggcagctca tcgtgcccc gcacctggcc 1140
 cacggggaga gtggagcccg gggagtccca ggtagtctg tgctgctgtt tgaggtggag 1200
 15 ctggtgtccc gggaggatgg gctgccaca ggctacctgt ttgtgtggca caaggacct 1260
 cctgccaaacc tgtttgaaga catggacctc aacaaggatg gcgaggtccc tccggaggag 1320
 ttctccacct tcatcaaggc tcaagtgagt gagggcaaag gacgcctcat gcctgggag 1380
 gacctgaga aaaccatagg agacatgttc cagaaccagg accgcaacca ggacggcaag 1440
 atcacagtcg acgagctcaa gctgaagtca gatgaggacg aggagcgggt ccacgaggag 1500
 25 ctctgagggg cagggagcct ggccaggcct gagacacaga ggccccactgc gagggggaca 1560
 gtggcggtg gactgacctg ctgacagtca ccctccctct gctgggatga ggtccaggag 1620
 ccaactaaaa caatggcaga ggagacatct ctggtgttcc caccacccta gatgaaaatc 1680
 30 cacagcacag acctctaccg tgtttctctt ccatccctaa accacttct taaaatgttt 1740
 ggatttgcaa agccaatttg gggcctgtgg agcctggggg tggatagggc catggctggt 1800
 cccccaccat acctcccc cacaactc acacagctga gcttgttatc catctccca 1860
 35 aactttctct ttctttgtac ttcttgcatt cccactccc agcccctatt cctctatgtg 1920
 acagctggct aggaccctc tgccttctt cccaatcctg actggctcct aggggaagggg 1980
 40 aaggctcctg gagggcagcc ctacctctcc catgcccctt gccctcctcc ctcgctcca 2040
 gtggaggctg agctgacctt gggctgctgg aggccagact gggctgtagt tagcttttca 2100
 tccctaaaga aggctttccc taaggaacca tagaagagag gaagaaaaca aagggcatgt 2160
 45 gtgaggggag ctgcttgggt ggggtttagg gctatgaaat cttggatttg gggctgaggg 2220
 gtgggagggg gggcagagct ctgcacactc aaaggctaaa ctggtgtcag tcttttttc 2280
 50 ctttgttcca aataaaagat taaacaaaa aaaaaaaaaaaa aaa 2323

<210> 105
 <211> 741
 <212> ADN
 55 <213> humano
 <400> 105

60 tcacgtgacc cgggcgcgct gcggccgcc gcgaggacc ggcgagaggc ggcggcggga 60
 gcggcgggta tggacgggtc cggggagcag cccagaggcg gggggccac cagctctgag 120

65

5 cagatcatga agacaggggc ccttttgctt caggggatga ttgccgccgt ggacacagac 180
 tccccccgag aggtcttttt ccgagtggca gctgacatgt tttctgacgg caacttcaac 240
 tggggccggg ttgtcgcctt tttctacttt gccagcaaac tgggtgctcaa ggccctgtgc 300
 10 accaaggtgc cggaaactgat cagaaccatc atgggctgga cattggactt cctccgggag 360
 cggctgttgg gctggatcca agaccagggg ggttgggacg gcctcctctc ctactttggg 420
 acgcccacgt ggcagaccgt gaccatcttt gtggcgggag tgctcaccgc ctcgctcacc 480
 15 atctggaaga agatgggctg aggccccag ctgccttggga ctgtgttttt cctccataaa 540
 ttatggcatt tttctgggag ggggtggggat tgggggacat gggcattttt cttacttttg 600
 taattattgg ggggtgtggg gaagagtggg cttgaggggg taataaacct ccttcgggac 660
 20 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 720
 aaaaaaaaaa aaaaaaaaaa a 741

25 <210> 106
 <211> 2373
 <212> ADN
 <213> humano

30 <400> 106

35 cccaggccca cccaccag caccctggc gcagggactg ctggaacctg gctgtgcgcg 60
 ctgtcgttt aagacagact ctgccggcgc cgtccggagc cttagaaacc ggccccggat 120
 cgcgagccgg agccggagcc ggagccgggg ccggccgggc tgctgaggcc cgagcggcag 180
 40 gagcgcagcg cggagcgtg agccaggcgc ccagtcgcga gaagctgccg ccgcctctgc 240
 ccgccccggc cgcagcccc gggcggcca tggggcgggc acggcgtcgc tgcaggcgcc 300
 ggcagccctg gagggcagcc gcttaggcgc tgcgctcttg tccccgcagg tgcagccag 360
 45 ggcggcgggg cgcgcccagc cccggcccc ggagcggccc ccgcggtccc cacctccatg 420
 gacgccttca aggggggcat gacgctggag cggctgccgg aggggctccg gccgcgccc 480
 ccgccacccc atgacatggg gccgccttc cacctggccc ggccccccga cccccgcgag 540
 50 ccgctcgaga actccgccag cgagtcgtct gacacggagc tgccagagaa ggagcgcggc 600
 ggggaaccca aggggcccga ggacagtggg gcgggaggca cgggctgagg cggcgcagac 660
 gaccagcca agaagaaga gcagcggcgg caacgtacgc acttcacaag ccagcagttg 720
 55 caagagctag aggccacggt ccagaggaac cgctacccc acatgagcat gaggaggagg 780
 atcggcgtgt ggaccaacct caccgagccg cgcgtgcggg tctggttcaa gaaccggcga 840
 60 gccagtggc gtaagcgcga gcgtaaccag cagctggacc tgtgcaaggg tggctacgtg 900
 ccgagttca gcggcctagt gcagccctac gaggacgtgt acgccgccgg ctactcctac 960
 aacaactggg ccgccaagag cctggcgcca gcgccgctct ccaccaagag cttcaccttc 1020

65

5 ttcaactcca tgagcccgct gtcgtcgcag tccatgttct cagcaccag ctccatctcc 1080
 tccatgacca tgccgtccag catgggcca ggcgccgtgc ctggcatgcc caactcgggc 1140
 ctcaacaaca tcaacaacct caccggctcc tcgctcaact cggccatgtc gccgggct 1200
 10 tgcccgtacg gcactcccgc ctgcacctac agcgtctacc gggacacgtg caactcgagc 1260
 ctagccagcc tgcggctcaa gtccaaacag cactcgtcgt ttggctacgg cgccctgcag 1320
 ggcccggcct cgggcctcaa cgcgtgccag tacaacagct gaccgccccg ccgcaccacg 1380
 15 cgggcccggcg gccggagcgg ggaagggcgc gggcgcgag gacgcacgcg gggccccggc 1440
 tcgcaagccc cagctcaccg cgccgcggac ctcacacctg cgcagcccc tctctccact 1500
 tcccactccg ggttggtttt gtgtttgctt ttccggacc cactctgcc tccaaaaaga 1560
 20 caaaaaaaaa aaaaaaaaaa aaagcaaaaa gacgtcggag aaaagtgcc cgaaaaaatg 1620
 gatgagttgc aatttctctc gggatggcgc gggtggtgtg tgtgtgttcc cacgggcccc 1680
 ggaggcccac tccgcggagg gcacgcggcg cggtaggcga gcgccgaggc ccagcggccg 1740
 25 ggggaggacg acctcgtatc ccgcgtcccc gccgcgtgg atccggactg agtgccggg 1800
 cctgcggact ggatgtgcgg ggccctggact tgcctaggat tccccgacc cgtacaaacc 1860
 30 aagttgccct ctccgagcta ggcccggccg agagcgcctt agctcgagtc ggatccgtgt 1920
 tggggcgggc gttgggttg gggggacggt gccccagcc caggatcggg cactcagtgg 1980
 agccgcacac ggccccggcg gccttgtag agcctcgtg gccccgcgc ccggagccct 2040
 35 atattaaggc cacggagcga cagcgggcag tgcgggcctg gcgggaggtg ggggaggtcc 2100
 atctcagaac accccagcct tgagcttagc tgcaggcca ggccctctgc tctgctcccg 2160
 40 ggctaggagg tggccctctg tctgggcgaa cagccccctc ctaccgccc gccgtgcaag 2220
 agtcgagccg gcagagcaag gggcgcgccc ccagggcct gcgcccactt tgcacaccgg 2280
 ctctccggcc cgcgcccctg ttacagcgt ccctgtgtat gttggactga ctgtaataaa 2340
 45 tctgtctata tcgactaaaa aaaaaaaaaa aaa 2373

50 <210> 107
 <211> 1314
 <212> ADN
 <213> humano

55 <400> 107

60 aattcccggc tcggggacct ccacgcaccg cggctagcgc cgacaaccag ctagcgtgca 60
 agggcggcgc gctcagcgcg taccggcggg cttcgaaacc gcagtcctcc ggcgaccccg 120
 aactccgctc cggagcctca gccccctgga aagtgatccc ggcacccgag agccaagatg 180
 ccggcccact tgctgcagga cgatatctct agctcctata ccaccaccac caccattaca 240
 65 gcgcctccct ccagggctct gcagaatgga ggagataagt tggagacgat gcccctctac 300

	ttggaagacg acattcgccc tgatataaaa gatgatatat atgacccccac ctacaaggat	360
5	aaggaaggcc caagcccaa ggttgaatat gtctggagaa acatcatcct tatgtctctg	420
	ctacacttgg gagccctgta tgggatcact ttgattccta cctgcaagtt ctacacctgg	480
	ctttgggggg tattctacta ttttgtcagt gccctgggca taacagcagg agctcatcgt	540
10	ctgtggagcc accgctctta caaagctcgg ctgcccctac ggctctttct gatcattgcc	600
	aacacaatgg cattccagaa tgatgtctat gaatgggctc gtgaccaccg tgcccaccac	660
	aagttttcag aaacacatgc tgatcctcat aattcccagc gtggcttttt cttctctcac	720
15	gtgggttggc tgcttgtgcg caaacaccca gctgtcaaag agaaggggag tacgctagac	780
	ttgtctgacc tagaagctga gaaactggtg atgttccaga ggaggtacta caaacctggc	840
20	ttgctgatga tgtgcttcat cctgcccacg cttgtgccct ggtatttctg gggtgaaact	900
	tttcaaaaaca gtgtgttctg tgccactttc ttgcgatatg ctgtggtgct taatgccacc	960
	tggctggtga acagtgctgc ccacctcttc ggatatcgtc cttatgacaa gaacattagc	1020
25	ccccgggaga atatcctggt ttcacttggg gctgtgggtg agggcttcca caactaccac	1080
	cactcctttc cctatgacta ctctgccagt gagtaccgct ggcacatcaa cttcaccaca	1140
	ttcttcattg attgcatggc cgccctcggc ctggcctatg accggaagaa agtctccaag	1200
30	gccgccatct tggccaggat taaaagaacc ggagatggaa actaaaaaaaa aaaaaaaaaa	1260
	aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaataaaa aaaaaaaaaa aaaa	1314
35		
40		
45		
50		
55		
60		
65		

REIVINDICACIONES

1. Método de identificación del origen de una metástasis de origen desconocido que comprende las etapas de
- 5 a. medir los Biomarcadores asociados con tres carcinomas diferentes en una muestra que contiene células metastásicas, en el que los Biomarcadores son los niveles de expresión de genes Marcadores y en el que los genes Marcadores están seleccionados de entre:
- 10 i. SP-B, TTF y DSG3, que son marcadores para el cáncer de pulmón;
 ii. F5 y PSCA, que son marcadores para el cáncer de páncreas; y
 iii. CDH17, que es un marcador para el cáncer de colon;
- b. combinar los datos de los Biomarcadores en un algoritmo en el que el algoritmo
- 15 i. normaliza los Biomarcadores frente a una referencia; e
 ii. impone un punto de corte que optimiza la sensibilidad y la especificidad de cada Biomarcador, pondera la prevalencia de los carcinomas y selecciona un tejido de origen;
- 20 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 concreto de carcinomas; y
 d. opcionalmente medir los Biomarcadores específicos para uno o más carcinomas diferentes adicionales, y repetir las etapas c) y d) para los Biomarcadores adicionales;
- 25 2. Método según la reivindicación 1 en el que la expresión génica se mide utilizando al menos una de las SEQ ID NO: 11-54 y 57-58.
3. Método según la reivindicación 1, que comprende adicionalmente obtener información clínica adicional, incluido el sitio de metástasis, para determinar el origen del carcinoma.
- 30 4. Método según la reivindicación 1 que comprende adicionalmente medir la expresión de al menos un gen expresado constitutivamente en la muestra.
5. Método de orientación para el tratamiento determinando el origen de una metástasis de origen desconocido según la reivindicación 1 e identificando el tratamiento apropiado para ello.
- 35 6. Método para proporcionar un pronóstico determinando el origen de una metástasis de origen desconocido, según la reivindicación 1 e identificando el correspondiente pronóstico para ello.
- 40 7. Uso de una composición que comprende al menos una secuencia aislada seleccionada de entre las SEQ ID NO: 11-54 y 57-58 en el método según cualquiera de las reivindicaciones 1-6.
8. Uso de un kit que comprende ARN o ADNc que hibrida con los genes marcadores según la reivindicación 1 en un método según cualquiera de las reivindicaciones 1-6.
- 45 9. Uso de una micromatriz o matriz génica para llevar a cabo el método según la reivindicación 1.

50

55

60

65

Figura 1

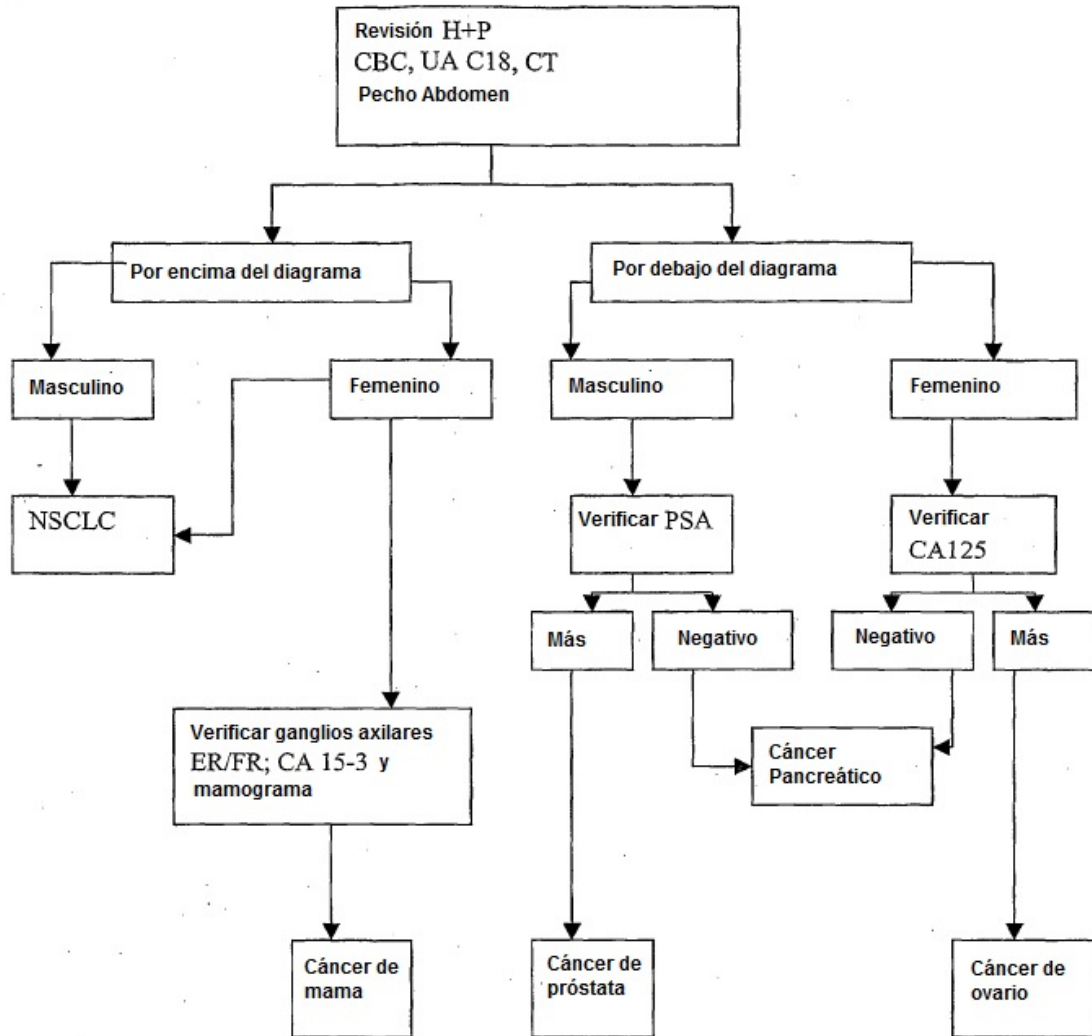


Figura 2

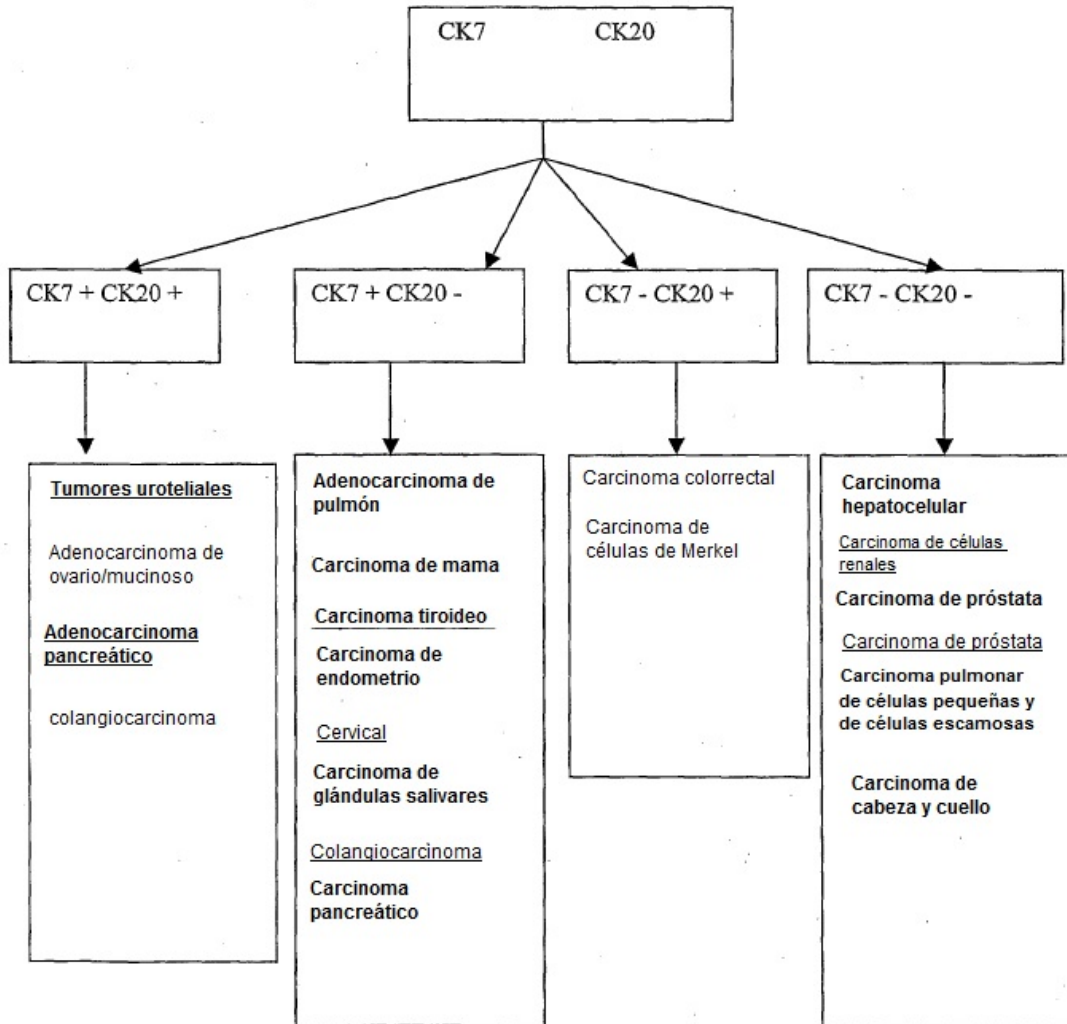


Figura 3

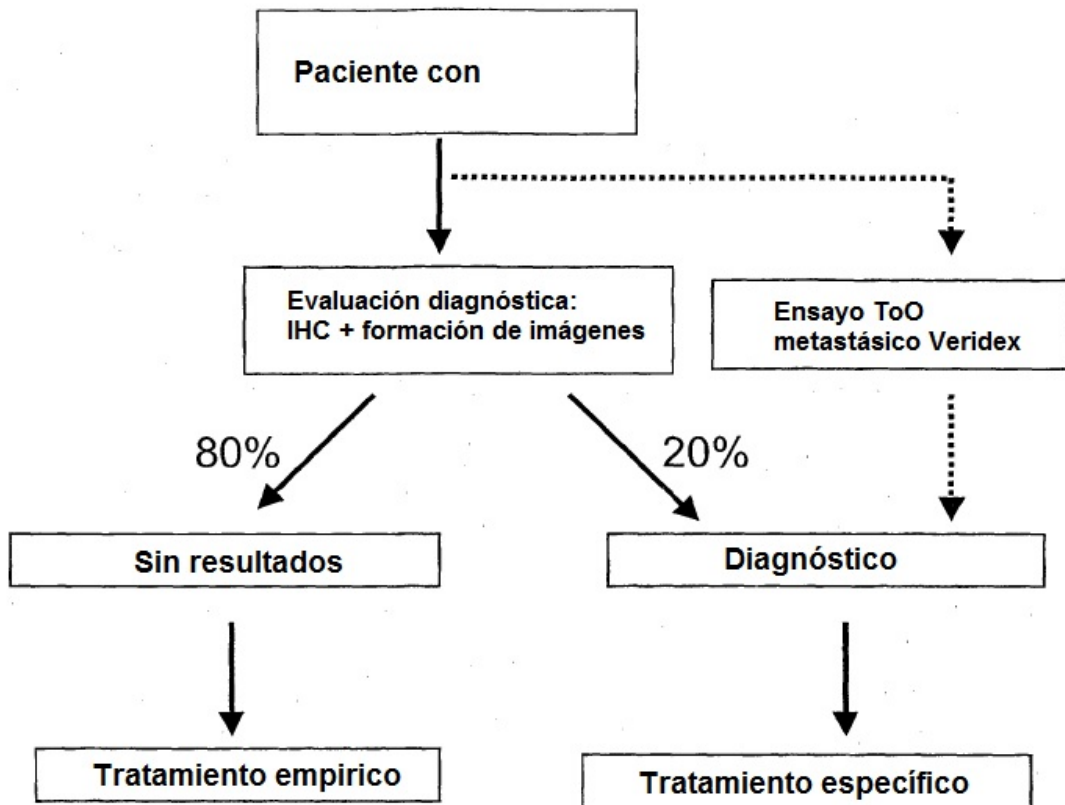
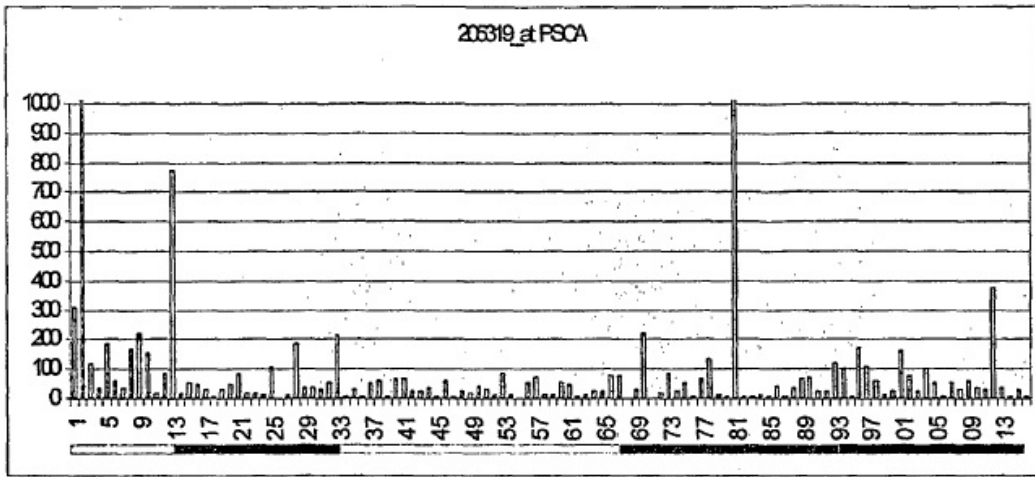


Figura 4

A



B

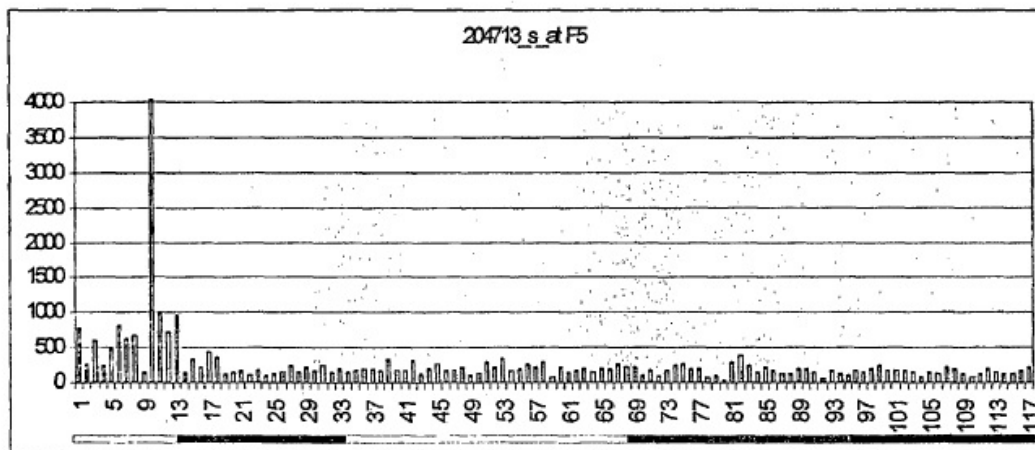
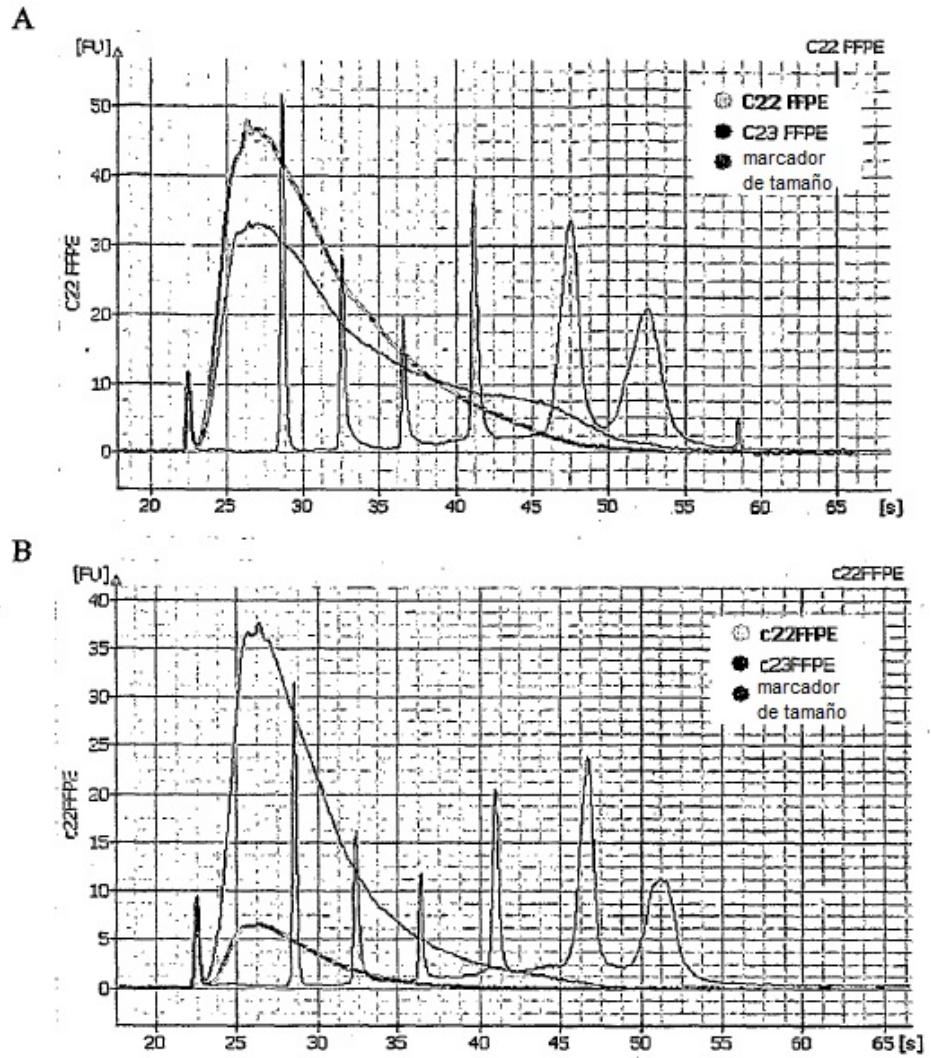


Figura 5



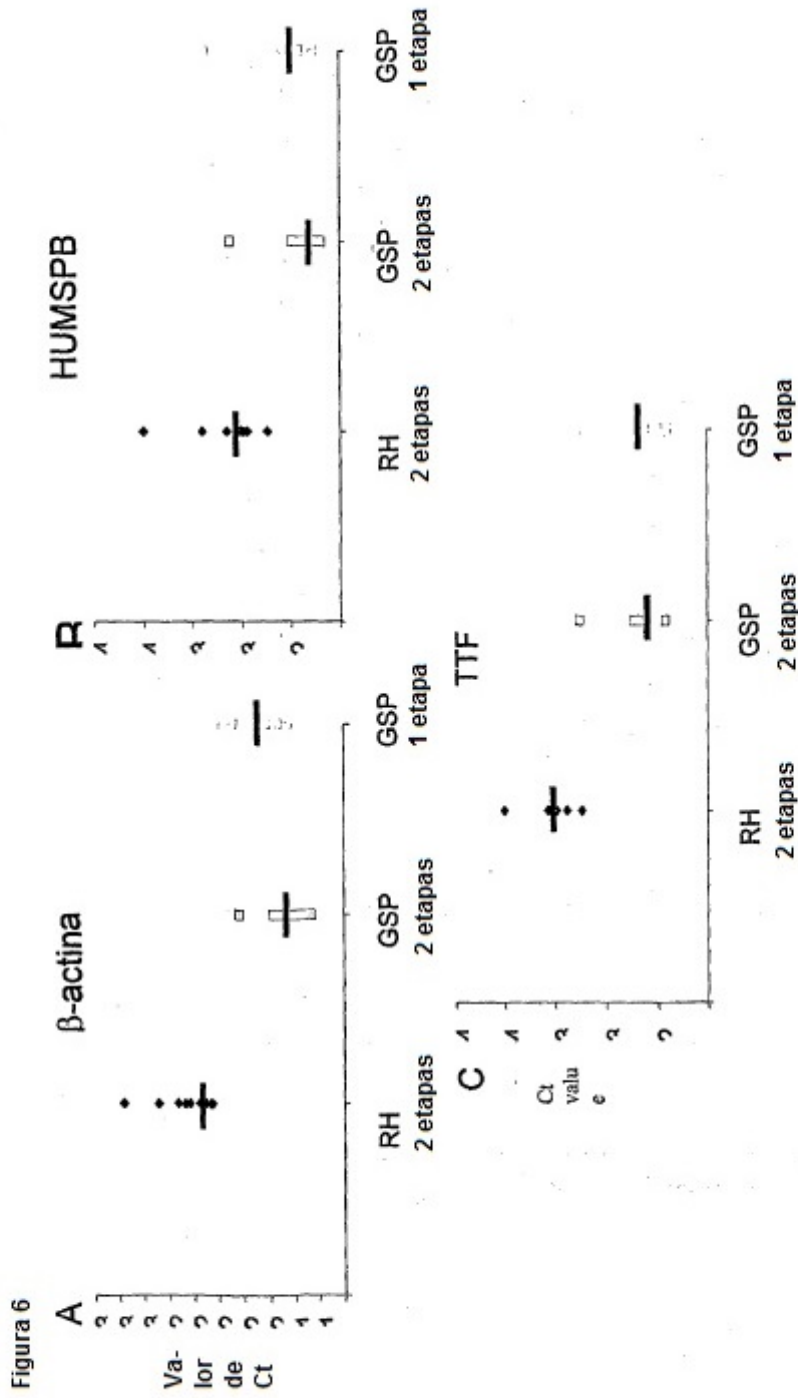
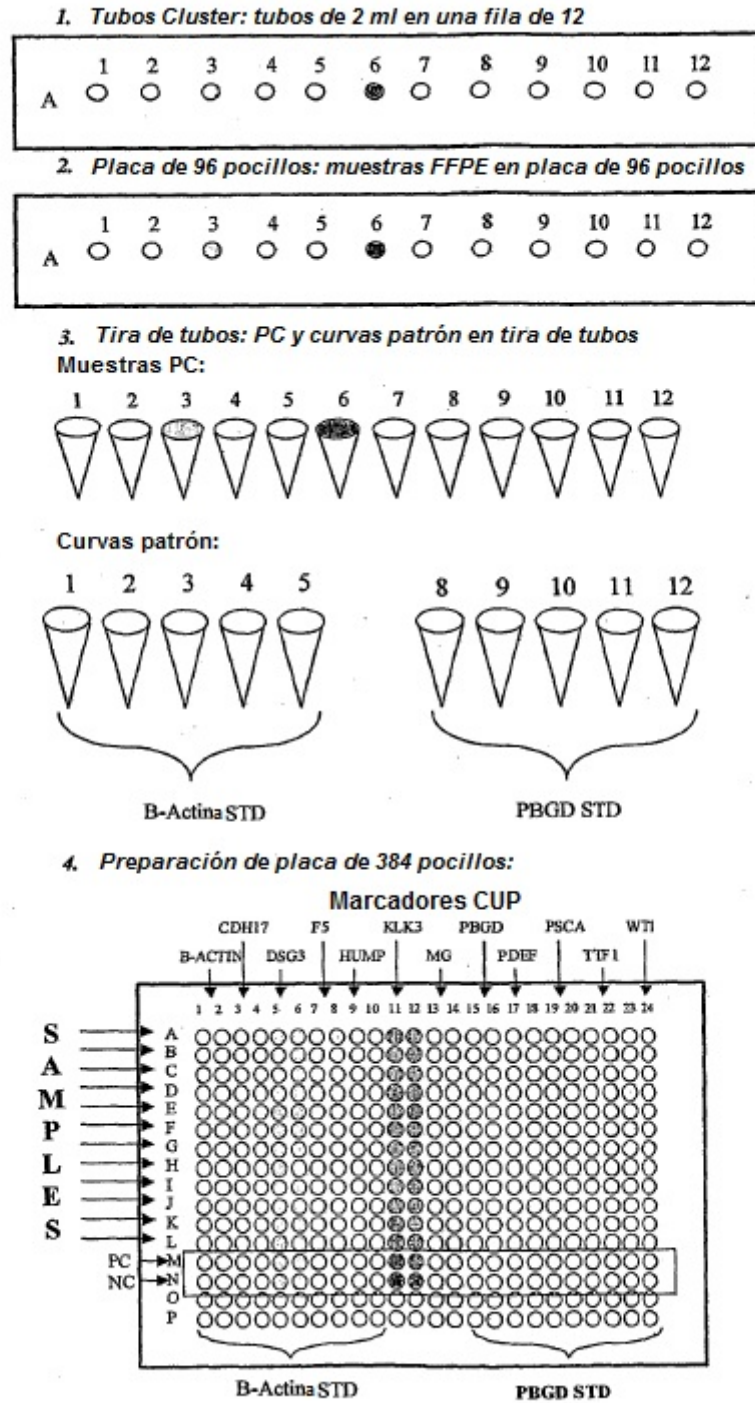


Figura 7



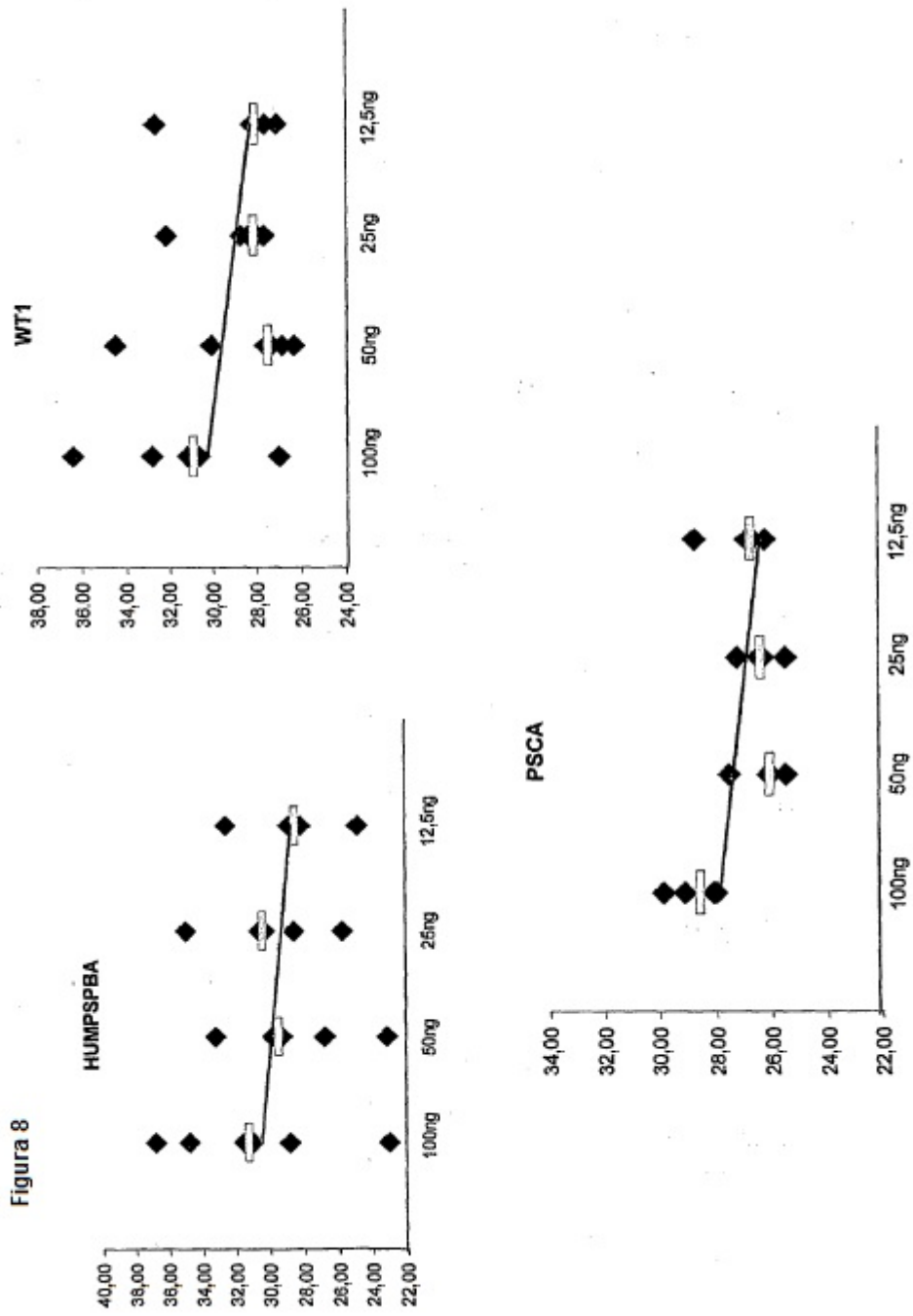


Figura 9A

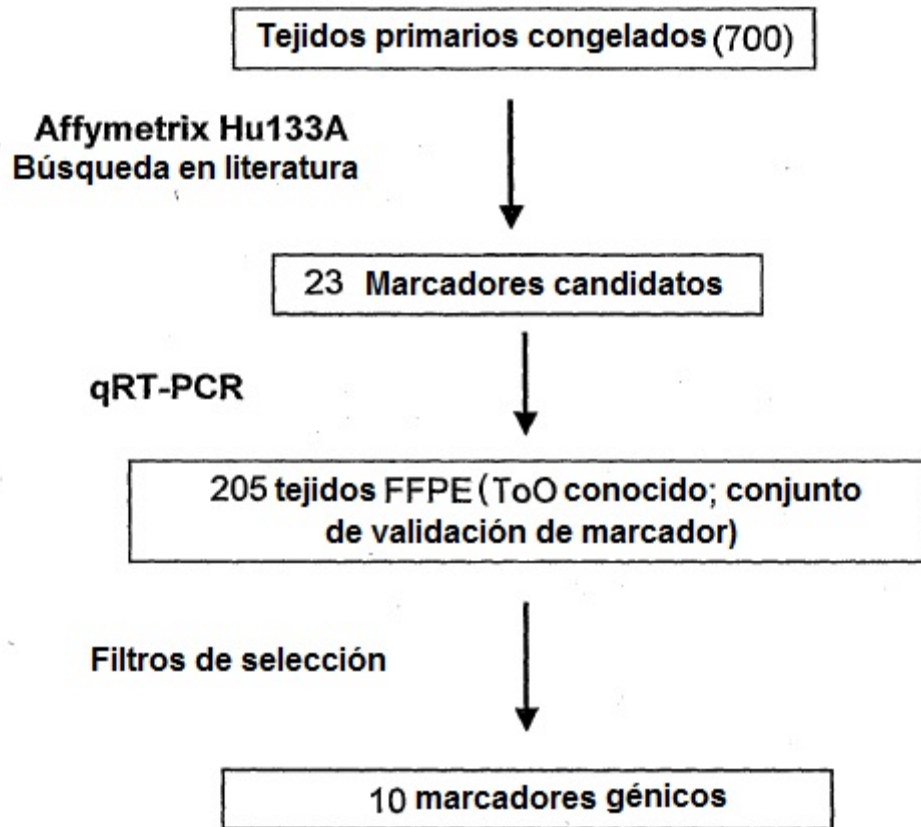
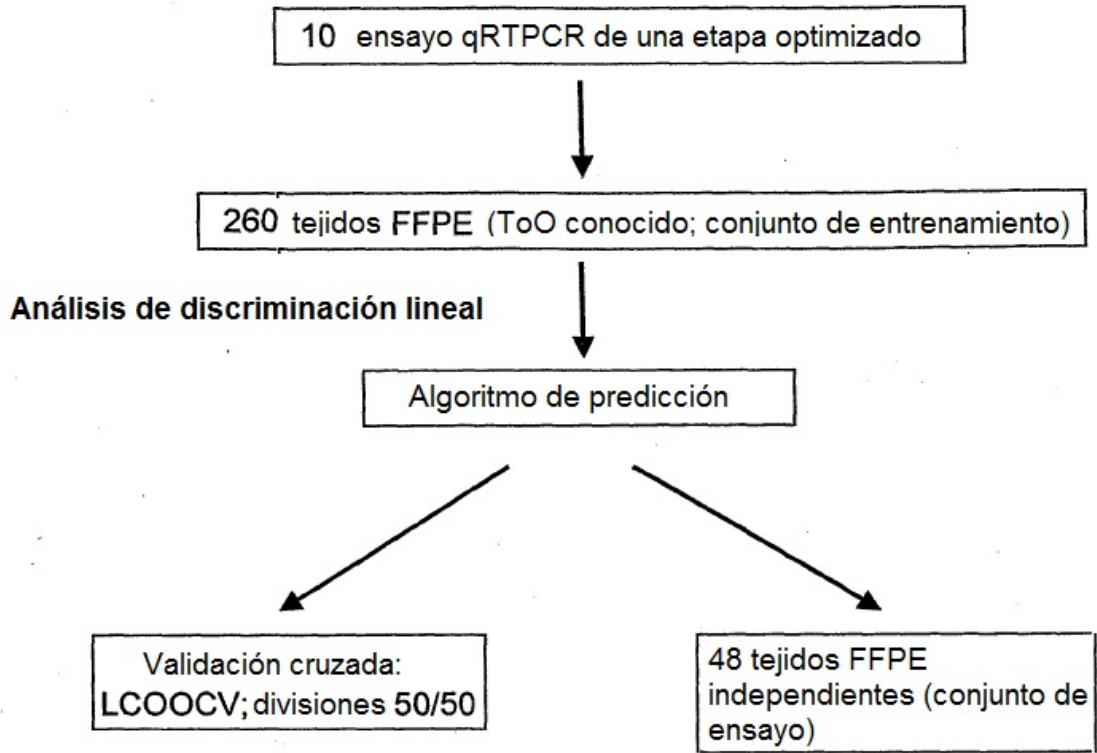


Figura 9B



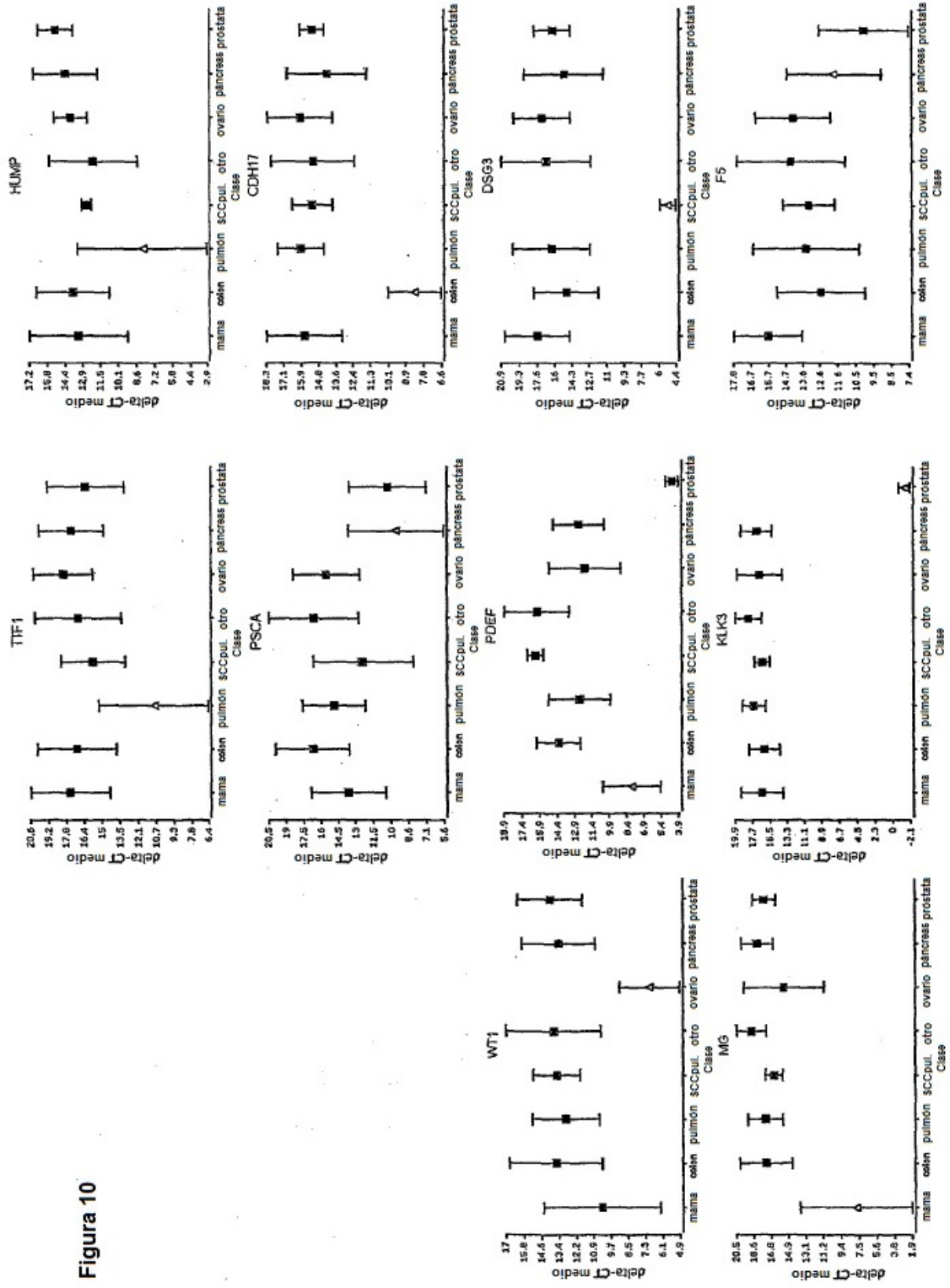


Figura 10

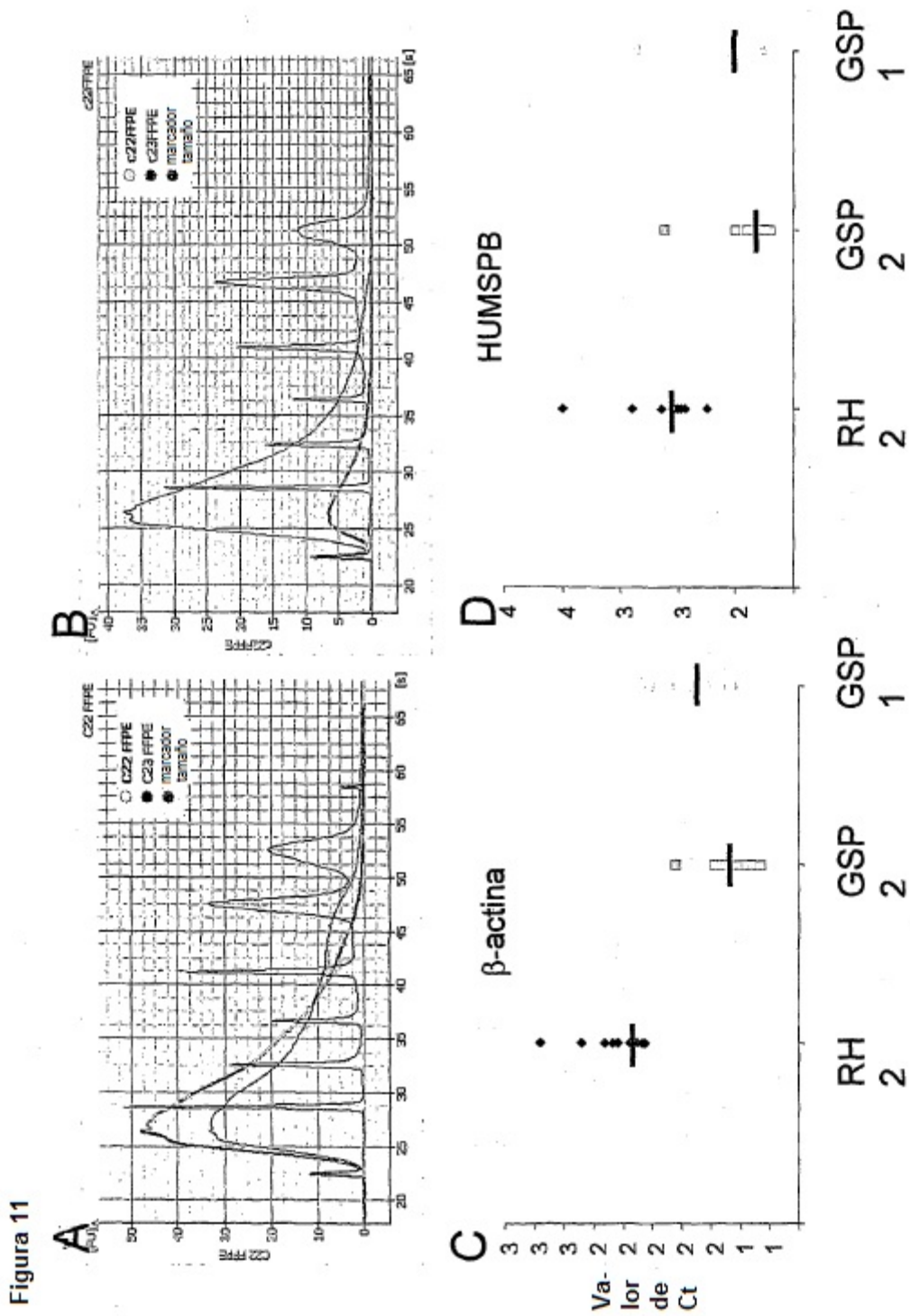


Figura 12

