



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



① Número de publicación: **2 343 996**

② Número de solicitud: 200803509

⑤ Int. Cl.:

C12Q 1/68 (2006.01)

G01N 33/50 (2006.01)

⑫

PATENTE DE INVENCION

B1

② Fecha de presentación: **11.12.2008**

④ Fecha de publicación de la solicitud: **13.08.2010**

Fecha de la concesión: **08.06.2011**

④ Fecha de anuncio de la concesión: **20.06.2011**

④ Fecha de publicación del folleto de la patente:
20.06.2011

⑦ Titular/es: **Fundación para la Investigación
Biomédica del Hospital Universitario La Paz
Paseo de la Castellana, 261
28046 Madrid, ES**

⑦ Inventor/es: **Espinosa, Enrique;
Fresno Vara, Juan Ángel;
Sánchez-Navarro, Iker y
Gámez-Pozo, Angelo**

⑦ Agente: **Pons Ariño, Ángel**

⑤ Título: **Método para la subclasificación de tumores.**

⑤ Resumen:

Método para la subclasificación de tumores.

La presente invención se encuadra dentro del campo de la biología molecular y la medicina. Específicamente, la presente invención se refiere a un método de subclasificación de tumores. Concretamente, se refiere a un método de pronóstico de la evolución de los tumores, y por tanto, un método útil para tomar decisiones con respecto al tratamiento a administrar al paciente. Además, la presente invención se refiere a un kit que permite la subclasificación de tumores mediante este método.

ES 2 343 996 B1

Aviso: Se puede realizar consulta prevista por el art. 37.3.8 LP.

DESCRIPCIÓN

Método para la subclasificación de tumores.

5 La presente invención se encuadra dentro del campo de la biología molecular y la medicina. Específicamente, la presente invención se refiere a un método de subclasificación de tumores. Concretamente, se refiere a un método de pronóstico de la evolución de los tumores, y por tanto, un método útil para tomar decisiones con respecto al tratamiento a administrar al paciente. Además, la presente invención se refiere a un kit que permite la subclasificación de tumores mediante este método.

10 **Estado de la técnica anterior**

15 El cáncer de mama es el tumor maligno más frecuente en las mujeres y la primera causa de muerte por cáncer también en mujeres. Es una enfermedad heterogénea en cuanto a las manifestaciones clínicas, el pronóstico y la sensibilidad o resistencia a los diferentes tratamientos médicos. El pronóstico del cáncer de mama es particularmente relevante porque sirve para seleccionar los tratamientos adyuvantes que la paciente recibirá después de la cirugía para reducir el riesgo de recaída.

20 En los últimos años numerosas publicaciones han puesto de manifiesto que las nuevas técnicas de alto rendimiento en genómica pueden ser de gran utilidad en la consecución de estos objetivos (Marchionni *et al.* Evid Rep Technol Assess. 2007 Dec;(160):1-105). Estas técnicas pueden proporcionar información sobre el riesgo de padecer la enfermedad, la presencia de un tumor oculto en programas de detección precoz, la detección precoz de una recaída, la posibilidad de que aparezca dicha recaída del cáncer (pronóstico) y la posibilidad de efectos secundarios y de respuesta a los tratamientos. Así, se han encontrado varios perfiles de genes relacionados con el pronóstico del cáncer de mama. Tres de estos perfiles, los conocidos como “70-gene Signature” (van’t Veer *et al.* Nature. 2002 Jan 31; 415(6871):530-6), “Recurrence Score” (Paik *et al.* N Engl J Med. 2004 Dec 30;351(27):2817-26), y “H/I Index” (Ma *et al.* J Clin Oncol. 2006 Oct 1;24(28):4611-9) han pasado una validación independiente, lo que se podría llamar ensayos en fase II.

30 El perfil de genes con valor pronóstico “70-gene Signature”, comercializado con el nombre de MammaPrint® por la empresa Agendia (van’t Veer *et al.* Nature. 2002 Jan 31; 415(6871):530-6; patente con número de publicación EP1782315) se ofrece para el diagnóstico de pacientes que presenten los siguientes criterios: tamaño del tumor menor de 5 cm, carcinoma de mama en estadio I o II, sin afectación ganglionar o hasta 3 ganglios linfáticos afectos, independientemente de la expresión del receptor de estrógenos (ER). El procedimiento requiere la medida de la expresión de 70 genes que determinan la firma génica asociada a pronóstico. El método empleado para el análisis consiste en la obtención de RNA a partir de tejido congelado en fresco (FF, siglas del inglés “Fresh Frozen”) y el empleo de microarrays de cDNA. El uso de tejido congelado para estudios moleculares a gran escala presenta varios problemas: las muestras congeladas son difíciles de recoger, complicadas de procesar y costosas de almacenar. Por el contrario, las muestras de tejido fijadas en formol y embebidas en parafina (FFPE, siglas del inglés “Formalin-Fixed Paraffin-Embedded”) son estables a temperatura ambiente, fáciles de almacenar y, lo que es más importante, existe un amplio archivo de muestras clínicas disponibles junto con su información clínica y el seguimiento de la enfermedad. Frente a estas ventajas ofrecidas por las muestras fijadas y embebidas en formol, existe la desventaja de que el RNA obtenido a partir de estas muestras está muy degradado. Mientras que los estudios mediante microarrays son muy sensibles a la degradación del RNA, el uso de la RT-PCR cuantitativa (qRT-PCR) ha demostrado ser una técnica que ofrece mejores resultados ante la degradación del RNA. Además, el análisis de expresión mediante microarrays es una técnica compleja que requiere de equipos sofisticados que no están disponibles en muchos laboratorios. Por tanto, a la vista de los motivos expuestos, el uso de la qRT-PCR a partir de RNA obtenido de muestras de tejido FFPE utilizados en la presente invención ofrece numerosas ventajas frente al análisis de expresión mediante microarrays a partir de RNA obtenido de muestras de tejido FF.

50 En los otros dos perfiles de genes con valor pronóstico mencionados, conocidos como “Recurrence Score” (Paik *et al.* N Engl J Med. 2004 Dec 30;351 (27):2817-26), y “H/I Index” (Ma *et al.* J Clin Oncol. 2006 Oct 1;24(28):4611-9), el método empleado consiste en la obtención de RNA a partir de muestras de tumores FFPE y el empleo de la técnica RT-PCR a tiempo real, con las ventajas que ello supone.

55 El “H/I Index” comercializado con el nombre de Theros H/I™ por la empresa BioTheragnostics (Ma *et al.* J Clin Oncol. 2006 Oct 1;24(28):4611-9; patente con número de publicación WO2007084220) requiere la medida de la expresión de los genes HOXB13 e IL17BR que determinan la firma génica asociada a pronóstico y 4 genes que permiten la normalización. Este método permite pronosticar la respuesta al tratamiento hormonal en mujeres con carcinoma de mama positivo para la expresión del receptor de estrógenos y sin ganglios afectos.

60 El perfil “Recurrence Score”, comercializado con el nombre de Oncotype DX™ por la empresa Genomic Health, Inc. (Paik *et al.* N Engl J Med. 2004 Dec 30;351 (27):2817-26; patente con número de publicación EP1815014) se ofrece para el pronóstico de pacientes a las que se les ha diagnosticado un carcinoma de mama invasivo en estadio I o II positivo para la expresión del receptor de estrógenos y que no presentan afectación de los ganglios linfáticos. Este perfil es uno de los pocos para los que se han realizado estudios que permiten afirmar su utilidad clínica para predecir el beneficio de un tratamiento quimioterapéutico (Gianni *et al.* J Clin Oncol. 2005;23(29):7265-7277; Paik *et al.* J Clin Oncol. 2006;24(23):3726-3724; Chang *et al.* Breast Cáncer Res Treat. 2008 Mar; 108(2):233-40).

ES 2 343 996 B1

El cáncer de ovario, aunque menos frecuente que el de mama, es el tumor más letal del tracto genital femenino, debido no sólo a su agresividad intrínseca, sino también a la dificultad del diagnóstico precoz, lo que hace que casi las dos terceras partes de las mujeres diagnosticadas estén ya en fases avanzadas de la enfermedad. En el caso concreto del cáncer de ovario también podemos encontrar trabajos que analizan perfiles de expresión mediante técnicas de alto rendimiento (Crijns *et al.*, 2006. *Int J Gynecol Cáncer*. 16:152-65).

En conclusión, el cáncer es una enfermedad heterogénea con distintos subtipos tumorales en cuanto a su pronóstico y respuesta a las diferentes opciones terapéuticas. Durante los últimos años se ha puesto de manifiesto que las técnicas de alto rendimiento en genómica son de gran utilidad para pronosticar el riesgo de recaída, la supervivencia y la respuesta a los diferentes tratamientos médicos adyuvantes. Sin embargo, los perfiles génicos descritos hasta la fecha para el pronóstico del cáncer requieren aún de ensayos de validación. Por otra parte, la capacidad predictiva de cada tipo de cada perfil es bastante limitada y existe la necesidad de perfiles génicos que permitan pronosticar la respuesta a las nuevas terapias adyuvantes (Marchionni *et al.* *Evid Rep Technol Assess*. 2007 Dec;(160):1-105).

15 Explicación de la invención

La presente invención se refiere a un método de subclasificación de tumores. Concretamente, se refiere a un método de pronóstico de la evolución de los tumores, y por tanto, un método útil para tomar decisiones con respecto al tratamiento a administrar al paciente. Además, la presente invención se refiere a un kit que permite la subclasificación de tumores mediante este método.

Por tanto, un primer aspecto de la invención se refiere a un método para la subclasificación de tumores que comprende:

- a. obtención de una muestra biológica aislada que comprende células tumorales del mamífero;
- b. detección de la cantidad del producto de la expresión de entre dos y ocho genes seleccionados de entre los siguientes: DTL, ECT2, MTDH, PRC1, RFC4, SCUBE2, STK32B o ZNF533, en la muestra obtenida en (a) y
- c. comparación de la cantidad detectada en el paso (b) con una cantidad de referencia.

Otro aspecto de la presente invención, se refiere un método de pronóstico de la evolución del tumor que comprende, además de los pasos (a)-(c) anteriormente descritos, un paso (d) donde una cantidad detectada en el paso (b) de los genes DTL, ECT2, MTDH, PRC1 o RFC4 mayor que la cantidad de referencia con la que se compara en el paso (c) o una cantidad detectada en el paso (b) de los genes SCUBE2, STK32B o ZNF533 menor que la cantidad de referencia con la que se compara en el paso (c) es indicativa de una menor supervivencia libre de recaída a distancia o una menor supervivencia global.

El término “pronóstico”, tal y como se utiliza en la presente descripción, se refiere, pero no se limita, a la probabilidad de la muerte debida al cáncer o la progresión, incluyendo recaída, capacidad de diseminación metastásica o respuesta a un determinado tratamiento de una enfermedad neoplásica.

El término “predicción”, tal y como se utiliza en la presente descripción, se refiere, pero no se limita, a la probabilidad de que un paciente responda favorable o desfavorablemente a un determinado tratamiento, y a la extensión de dichas respuestas, o de que el paciente sobreviva, tras la eliminación quirúrgica de un tumor primario y/o la quimioterapia por un periodo de tiempo sin que se produzca recaída del cáncer.

El término “supervivencia libre de recaída a distancia” (SLRD), tal y como se utiliza en la presente descripción, se refiere al tiempo transcurrido desde la fecha de la cirugía hasta la recaída a distancia o hasta la última visita. El término “supervivencia global” (SG), tal y como se utiliza en la presente descripción, se refiere al tiempo transcurrido desde la fecha de la cirugía hasta la última visita o hasta el fallecimiento de la enferma.

El término “muestra biológica aislada que comprende células tumorales”, tal y como se utiliza en la descripción se refiere, pero no se limita, a tejidos y/o fluidos biológicos de un sujeto, obtenidos mediante cualquier método conocido por un experto en la materia que sirva para tal fin. La muestra biológica puede ser un tejido, por ejemplo, pero sin limitarse, una biopsia tumoral o un aspirado por aguja fina, o puede ser un fluido biológico, por ejemplo, pero sin limitarse, una muestra de fluido, como sangre, plasma, suero, linfa, fluido ascítico, orina o exudado mamario. La muestra puede ser tomada de un mamífero humano, pero también de mamíferos no humanos, como por ejemplo, pero sin limitarse como roedores, rumiantes, felinos o caninos. La muestra biológica puede ser fresca, congelada, fijada, embebida en parafina. Preferiblemente, la muestra es fijada y embebida en parafina.

El término “producto de la expresión”, tal y como se utiliza en la descripción, hace referencia a sus productos de transcripción o expresión (ARN o proteína). O a cualquier forma resultante del procesamiento de dichos productos de transcripción o expresión.

El término “perfil de expresión génica” o “perfil de expresión de genes”, tal y como se utiliza en la presente descripción, se refiere a cualquier método que permita la cuantificación del RNA mensajero (mRNA) y/o de proteína en una muestra biológica.

ES 2 343 996 B1

El término “cantidad de referencia” o “punto de corte”, tal y como se utiliza en la presente descripción, se refiere a cualquier valor derivado de la cuantificación del producto de la expresión de los genes en una muestra biológica, que permita definir dos poblaciones con diferente riesgo de recaída a distancia.

5 La detección de la cantidad de producto de la expresión de los genes en la muestra obtenida, tal y como se utiliza en la descripción, hace referencia a la medida de la cantidad o la concentración, preferiblemente de manera semi-cuantitativa o cuantitativa. La medida puede ser llevada a cabo de manera directa o indirecta. La medida directa se refiere a la medida de la cantidad o la concentración del producto de la expresión del gen basada en una señal que se obtiene directamente del producto de la expresión del gen y que está correlacionada directamente con el número
10 de moléculas del producto de la expresión del gen presente en la muestra. Dicha señal - a la que también podemos referirnos como señal de intensidad - puede obtenerse, por ejemplo, midiendo un valor de intensidad de una propiedad química o física del producto de expresión. La medida indirecta incluye la medida obtenida de un componente secundario (por ejemplo, un componente distinto del producto de la expresión génica) o un sistema de medida biológica (por ejemplo la medida de respuestas celulares, ligandos, “etiquetas” o productos de reacción enzimática).
15

De acuerdo con la presente invención, la detección de la cantidad de producto de la expresión de los genes puede ser llevada a cabo por cualquier método de determinación de la cantidad del producto de la expresión de los genes conocido por el experto en la materia. En una realización preferida, la detección del producto de la expresión de los genes se realiza determinando el nivel de mRNA derivado de su transcripción donde el análisis del nivel de mRNA se puede realizar, a título ilustrativo y sin que limite el alcance de la invención, mediante amplificación por reacción en cadena de la polimerasa (PCR), retrotranscripción en combinación con la reacción en cadena de la ligasa (RT-LCR), retrotranscripción en combinación con la reacción en cadena de la polimerasa (RT-PCR), retrotranscripción en combinación con la reacción en cadena de la polimerasa cuantitativa (qRT-PCR), o cualquier otro método de amplificación de ácidos nucleicos; microarrays de DNA elaborados con oligonucleótidos depositados por cualquier mecanismo; microarrays de DNA elaborados con oligonucleótidos sintetizados *in situ* mediante fotolitografía o por cualquier otro mecanismo; hibridación *in situ* utilizando sondas específicas marcadas con cualquier método de mareaje; mediante geles de electroforesis; mediante transferencia a membrana e hibridación con una sonda específica; mediante RMN o cualquier otra técnica de diagnóstico por imagen utilizando nanopartículas paramagnéticas o cualquier otro tipo de nanopartículas detectables funcionalizadas con anticuerpos o por cualquier otro medio.
20
25
30

El mRNA puede ser extraído, por ejemplo, pero sin limitarse a partir de muestras de tejido fresco, tejido FF o de muestras de tejido FFPE. Los métodos de obtención de RNA total o mRNA son bien conocidos en el estado de la técnica. El uso de muestras de tejido FFPE presenta importantes ventajas con respecto a las muestras de tejido FF: son estables a temperatura ambiente, fáciles de almacenar y existe un amplio archivo de muestras clínicas disponibles junto con su información clínica y el seguimiento de la enfermedad. Por tanto, en una realización preferida el mRNA se extrae de muestras de tejido FFPE.
35

El RNA obtenido de muestras de tejido FFPE suele encontrarse muy degradado. Mientras que los estudios mediante microarrays son muy sensibles a la degradación del RNA, el uso de la RT-PCR cuantitativa (qRT-PCR) ha demostrado ser una técnica que ofrece mejores resultados ante la degradación del RNA. Además, el análisis de expresión mediante microarrays es una técnica compleja que requiere de equipos sofisticados que no están disponibles en muchos laboratorios. Por tanto, en una realización preferida la detección del mRNA se realiza mediante la técnica de qRT-PCR.
40
45

El sesgo que puede introducir la pérdida de RNA amplificable debida a la degradación debe además ser corregido mediante estrategias de normalización que puedan compensar este efecto de la degradación del RNA. Por tanto, un aspecto de la invención consiste en un método de normalización, al que de ahora en adelante nos referiremos como NorMean, que sirve para corregir los errores debidos a las diferentes eficiencias enzimáticas de las reacciones de retrotranscripción y amplificación, así como las causadas por las variaciones debidas a la diferente calidad del RNA debida a la degradación consecuencia de la fijación en formol.
50

Una vez seleccionados el gen o grupo de genes apropiados que funcionarían como genes de referencia se calcula el Factor de Normalización (NF, siglas del inglés “*Normalization Factor*”) utilizando la media geométrica de los genes seleccionados.
55

El método de normalización NorMean permite seleccionar un gen o grupo de genes que se usan como genes de referencia basándose en el valor del parámetro a_i :

$$60 \quad a_i = \Sigma CV_{ij} / \Sigma r_{ij}$$

Donde CV_{ij} es el coeficiente de variación del gen de referencia i en el material j , y r_{ij} es el coeficiente de correlación de Pearson entre la expresión del gen i y la media de expresión de todos los genes en cada muestra en el material j . El valor a_i permite ordenar los genes de referencia: aquellos genes de referencia con menor valor a_i son aquellos con la expresión más estable (bajo CV) y con mayor correlación con la expresión génica media por muestra (alto r). El NF se determina calculando la media geométrica de los valores de expresión de los genes de referencia seleccionados. El número óptimo de genes de referencia se determina comparando el porcentaje de genes con correlación significativa
65

ES 2 343 996 B1

entre las muestras de tejido FF y tejido FFPE para cada factor de normalización. Los datos se calculan mediante los valores de expresión normalizados para cada gen en cada tipo de muestra dada por el valor $2^{-\Delta Ct}$, donde: $\Delta Ct = Ct_{gen} - NF$.

5 Para determinar la concordancia entre los datos de muestras de tejido FF y tejido FFPE se utiliza el coeficiente de correlación de Pearson entre los valores de expresión normalizados.

Utilizando el método NorMean se corrigen las variaciones en la expresión del mRNA debidas a las diferentes eficiencias enzimáticas del método de detección así como las debidas a la degradación del mRNA en tejidos FFPE.

10

La expresión de 83 genes, incluidos en tres perfiles génicos relacionados con el pronóstico del cáncer de mama, en muestras de pacientes con carcinoma ductal infiltrante de mama, fue normalizada utilizando NorMean. A continuación, mediante el cálculo de una correlación de Pearson entre los datos normalizados de una serie de muestras de tejido FF con sus muestras de tejido FFPE pareadas, se eligieron aquellos genes cuya correlación era comparable entre 15 ambos tipos de tejido (53 genes). Posteriormente, un análisis univariante permitió conocer cuáles de ellos se encuentra asociado con la supervivencia libre de recaída a distancia (SLRD) en función de su p valor (17 genes). Con 17 genes se construyó un perfil génico que identificaba un grupo de pacientes de alto riesgo de recaída. Perfiles génicos con 10, 9, 8, 7 o incluso hasta 2 genes mostraron también una buena separación entre grupos de alto y bajo riesgo de recaída a distancia. Finalmente, un análisis de componentes principales y un estudio de correlación de los genes entre sí, ha permitido distinguir un perfil de 8 genes con capacidad pronostica, denominado “8-gene Score”. Por tanto, en una realización preferida, en el paso (b) del método de la presente invención, se detecta la cantidad del producto de la expresión de entre 2 y 8 genes seleccionados de entre los siguientes: DTL, ECT2, MTDH, PRC1, RFC4, SCUBE2, STK32B o ZNF533. En una realización muy preferida, en el paso (b) del método de la presente invención se detecta la cantidad del producto de la expresión de los genes DTL, ECT2, MTDH, PRC1, RFC4, SCUBE2, STK32B y ZNF533. 25 La detección de la cantidad de producto en el paso (b) de una cantidad mayor de genes, que incluyan al menos 2 de estos 8 genes podría llegar a resultados similares.

El perfil de 8 genes de la presente invención se calcula en cada muestra utilizando las medidas de expresión génica normalizadas basándose en la siguiente ecuación: “8-gene Score” = (de 0.0968 a 0.2904)*DTL + (de 0.1088 a 0.3264)*ECT2 + (de 0.0227 a 0.0681)*MTDH + (de 0.0664 a 0.1994)*PRC1 + (de 0.0278 a 0.0834)*RFC4 - (de 0.0956 a 0.2870) *SCUBE2 - (de 0.0221 a 0.0665)*STK32B - (de 0.0591 a 0.1773)*ZNF533. Con la intención de asignar a cada paciente a un grupo de riesgo se define un punto de corte, de manera que pacientes con una puntuación en el perfil menor que el punto de corte son asignadas al grupo de bajo riesgo de recaída a distancia y pacientes con una puntuación mayor que dicho punto de corte constituyen el grupo de alto riesgo de recaída a distancia. 35

Una cantidad detectada en el paso (b) de los genes DTL, ECT2, MTDH, PRC1 o RFC4 mayor que la cantidad de referencia con la que se compara en el paso (c) o una cantidad detectada en el paso (b) de los genes SCUBE2, STK32B o ZNF533 menor que la cantidad de referencia con la que se compara en el paso (c) es indicativa de una menor supervivencia libre de recaída a distancia o una menor supervivencia global. 40

Debido a la naturaleza multigénica del predictor, la sustitución de estos genes por genes que coexpresan con ellos con un coeficiente de correlación de Pearson $r \geq 0,4$ no altera la capacidad discriminativa del predictor. Ejemplos de estos genes son, pero sin limitarse: AYTL2, BIRC5, CCNB1, CCNE2, GMPS, MCM6, MELK, MYBL2, ORC6L, PGR o TGFB3. 45

Por tanto, la presente invención proporciona un método para la subclasificación de tumores y/o para el pronóstico de la evolución del tumor de una manera sencilla y eficaz.

Los términos “tumor” o “tumoral”, tal y como se utiliza en la presente descripción, se refiere a células transformadas que presentan un crecimiento incontrolado. Dependiendo de su posible evolución puede tratarse de un tumor benigno, que permanece en su lugar de inicio y no produce metástasis; o tumor maligno o cáncer, invasivo o que produce metástasis. Por tanto, el término “cáncer” o “canceroso” tal y como se utiliza en la presente descripción, se refiere a una alteración de las células tumorales que tienen capacidad de invadir tejidos o de producir metástasis en lugares distantes del tumor primario. Ejemplos de cáncer incluyen, pero sin limitarse, cáncer de mama, cánceres ginecológicos, cáncer de colon, cáncer de próstata, cáncer de piel, cáncer hepatocelular, cáncer de pulmón, cáncer de esófago, cáncer gástrico, cáncer de páncreas, cáncer de vejiga, cáncer de hígado, cáncer del tracto urinario, cáncer tiroideo, cáncer renal, melanoma, cáncer de cerebro, sarcoma, linfoma o leucemia. Dentro de los cánceres ginecológicos se engloban los tumores de ovario, de útero, de cérvix, de vagina y de vulva. Existe un grupo de tumores en los que es importante la participación de la regulación hormonal, tales como los tumores de mama, ginecológicos o de próstata. Por tanto, preferiblemente, el método de la presente invención se refiere, pero no se limita a un método para la subclasificación de tumores de mama, ginecológicos o de próstata. 50

En una realización más preferida, el método de la presente invención se refiere a un método para la subclasificación de tumores de mama. El tumor de mama es la proliferación acelerada, desordenada y no controlada de células pertenecientes a distintos tejidos de una glándula mamaria. La palabra carcinoma se aplica a los tumores malignos que se originan en estirpes celulares de origen epitelial o glandular. La mayoría de los carcinomas de mama tiene su origen en la proliferación acelerada e incontrolada de las células que tapizan el interior de los conductos que durante la lactancia, llevan la leche desde los acinos glandulares, donde se produce, hasta los conductos galactóforos, situados 65

ES 2 343 996 B1

detrás de la areola y el pezón, donde se acumula en espera de salir al exterior; este cáncer de mama se conoce como carcinoma ductal. En los casos restantes el cáncer tiene su origen en los propios acinos glandulares y se le llama carcinoma lobulillar.

5 Muchos carcinomas de mama se encuentran confinados en la luz de los ductos o de los acinos, sin invadir los tejidos vecinos; reciben el nombre de carcinomas *in situ*. Cuando proliferan en demasía pueden romper la llamada membrana basal y extenderse infiltrando los tejidos que rodean a ductos y acinos y entonces reciben nombres como carcinoma ductal infiltrante o carcinoma lobulillar infiltrante.

10 El sistema de estadificación TNM (siglas del inglés “*Tumor size, Node, Metástasis*”) es el método más común para subclasificar el carcinoma de mama. Permite incluir a los carcinomas de mama en 5 estadios posibles (0, I, II, III y IV) en función el tamaño del tumor (T), la diseminación a los ganglios linfáticos (N), y la metástasis o diseminación a otras partes del cuerpo (M).

15 Las expresiones “sin afectación de los ganglios linfáticos”, “sin ganglios linfáticos afectos”, tal y como se utiliza en la presente descripción, se refiere a un cáncer que no se ha diseminado a los ganglios linfáticos. Por el contrario, el “término con afectación de los ganglios linfáticos” o “con ganglios linfáticos afectos”, tal y como se utiliza en la presente descripción, se refiere a un cáncer que se ha diseminado al menos a un ganglio linfático.

20 Más preferiblemente el método de la presente invención, se refiere, pero no se limita a un método para la subclasificación de carcinomas de mama infiltrantes. Aún más preferiblemente, para la subclasificación de carcinomas de mama infiltrantes en estadio I o II.

25 Otro factor que permite subclasificar el cáncer de mama es la expresión de los receptores de estrógenos o progesterona. Las expresiones “positivo para receptores hormonales” tal y como se utiliza en la presente descripción, se refiere a un tumor que presenta expresión para el receptor de estrógenos y/o progesterona. La expresión “positivo para receptor de estrógenos” tal y como se utiliza en la presente descripción, se refiere a un tumor que presenta expresión para el receptor de estrógenos. La expresión “positivo para receptor de progesterona” tal y como se utiliza en la presente descripción, se refiere a un tumor que presenta expresión para el receptor de progesterona. Preferiblemente, el
30 método de la presente invención se refiere, pero no se limita, a un método para la subclasificación de tumores de mama positivos para la expresión del receptor de estrógenos.

35 El tratamiento del cáncer principal habitualmente suele ser la cirugía. En el caso del cáncer de mama, la cirugía suele consistir bien en una mastectomía o bien en una cirugía conservadora. La expresión “mastectomía”, tal y como se utiliza en la presente descripción, se refiere a la extirpación de la mama o de tanto tejido de la mama como sea posible. En la mayoría de los casos, el cirujano extirpa también los ganglios linfáticos de la axila. La “expresión cirugía conservadora”, tal y como se utiliza en la presente descripción, se refiere a la extirpación del tumor pero no de la mama; también se conoce como tumorectomía, mastectomía segmentaria o mastectomía parcial.

40 Después de la cirugía, muchos pacientes reciben terapia adyuvante. El término “terapia adyuvante”, tal y como se utiliza en la presente descripción, se refiere al tratamiento que se da después del tratamiento principal con el objetivo de prevenir la recaída del cáncer en el seno o en otro lugar. En el tratamiento del cáncer, la terapia adyuvante puede consistir, por ejemplo, pero sin limitarse en radioterapia, quimioterapia, terapia hormonal o terapia biológica.

45 Algunos pacientes antes de la cirugía reciben terapia “neoadyuvante”. El término “terapia neoadyuvante”, tal y como se utiliza en la presente descripción, se refiere a la terapia adyuvante que se administra antes del tratamiento principal con el objetivo de reducir el tamaño del tumor, para posibilitar o facilitar la cirugía. En el tratamiento del cáncer dicha terapia, puede incluir, por ejemplo, pero sin limitarse, radioterapia, quimioterapia, terapia hormonal o terapia biológica.

50 La radioterapia consiste en el uso de radiaciones ionizantes para destruir las células cancerosas. Después de la cirugía conservadora de mama, la mayoría de los pacientes reciben radioterapia. Algunos pacientes reciben radioterapia después de una mastectomía. Algunas pacientes reciben radioterapia antes de la cirugía.

55 Los tumores positivos para receptores hormonales suelen tratarse con terapia hormonal con el objetivo de reducir o detener los efectos del estrógeno sobre las células tumorales. La terapia hormonal, puede consistir, por ejemplo, pero sin limitarse, en la administración de moduladores selectivos de receptores de estrógeno, inhibidores de aromatasa, agonistas de hormonas liberadoras de hormonas luteinizantes u otros agentes hormonales. Algunos moduladores selectivos de receptores de estrógeno son, por ejemplo, pero sin limitarse tamoxifeno, raloxifeno o toremifeno. Algunos inhibidores de aromatasa son, por ejemplo, pero sin limitarse letrozol, anastrozol o exemestano. Otras sustancias anti-
60 estrogénicas: fulvestrant. Otros agentes hormonales son, por ejemplo, pero sin limitarse, los andrógenos.

65 La quimioterapia consiste en el uso de compuestos o combinaciones de compuestos con el objetivo de destruir células cancerosas. Los compuestos administrados en la quimioterapia para el cáncer de mama pueden ser por ejemplo, pero sin limitarse, agentes alquilantes, antimetabolitos, inhibidores de la mitosis o antibióticos antitumorales. Algunos agentes alquilantes son, por ejemplo, pero sin limitarse, cisplatino, ciclofosfamida, dacarbazina, ifosfamida, mecloretamina o melfalán. Algunos antimetabolitos, son por ejemplo, pero sin limitarse, fluorouracilo, metotrexato, gemcitabina, citarabina o fludarabina. Algunos inhibidores de la mitosis son, por ejemplo, pero sin limitarse paclita-

xel, docetaxel, etopósido, vinblastina, vincristina o vinorelbina. Algunos antibióticos antitumorales, son por ejemplo, pero sin limitarse bleomicina o antraciclinas. Algunas antraciclinas son, por ejemplo, daunorrubicina, doxorubicina, epirubicina o idarrubicina.

5 En una realización preferida, la muestra biológica se aísla de un mamífero que ha sido sometido a una terapia hormonal y bien a una mastectomía o bien a una cirugía conservadora seguida de radioterapia.

10 El método para la subclasificación de tumores de la presente invención es útil para predecir la respuesta del cáncer a un determinado tratamiento. En una realización preferida, el método para la subclasificación de tumores de la presente invención es útil para tomar decisiones con respecto al tratamiento adyuvante a un tratamiento principal administrado al paciente. En una realización alternativa, el método para la subclasificación de tumores de la presente invención es útil para tomar decisiones con respecto al tratamiento neoadyuvante a un tratamiento principal administrado al paciente.

15 En una realización preferida, el método para la subclasificación de tumores de la presente invención es útil para predecir la respuesta del cáncer al tratamiento con antraciclinas de un paciente con carcinoma ductal infiltrante en estadios I o II que ha recibido una terapia hormonal y una mastectomía o una cirugía conservadora seguida de radioterapia. En una realización alternativa, el método para la subclasificación de tumores de la presente invención es útil para predecir la respuesta del cáncer al tratamiento con ciclofosfamida-metotrexato-fluorouracilo (CMF) de un
20 paciente con carcinoma ductal infiltrante en estadios I o II que ha recibido una terapia hormonal y una mastectomía o una cirugía conservadora seguida de radioterapia.

25 En otra realización preferida, el método de la presente invención se refiere, pero no se limita, a un método para la subclasificación de tumores de ovario. El tumor de ovario es la proliferación acelerada, desordenada y no controlada de células pertenecientes a distintos tejidos del ovario. Dependiendo del origen de las células, los tumores de ovario pueden ser, por ejemplo, pero sin limitarse, tumores del epitelio, tumores de células germinales, tumores de los cordones sexuales del estroma, tumores de células lipídicas o gonadoblastomas. El tipo de tumor más frecuente es el que se origina en las células epiteliales o tumor epitelial del ovario. Dependiendo del tipo de epitelio en el que se origina los tumores epiteliales del ovario pueden ser, por ejemplo, pero sin limitarse, tumores serosos, mucinosos, endometrioides, de células claras o de Brenner. Preferiblemente, el método de la presente invención se refiere, pero no se limita, a un método para la subclasificación de tumores epiteliales de ovario. Más preferiblemente, el método de la presente invención se refiere, pero no se limita, a un método para la subclasificación de tumores epiteliales serosos de ovario.

35 Otro aspecto de la invención se refiere a un kit adaptado para llevar a cabo el método para la subclasificación de tumores, que comprende los medios para determinar los productos de expresión de entre dos y ocho genes seleccionados de entre los siguientes: DTL, ECT2, MTDH, PRC1, RFC4, SCUBE2, STK32B o ZNF533, en una muestra biológica aislada que comprenda células del paciente. Dicho kit puede comprender cebadores, sondas, anticuerpos monoclonales, y todos aquellos reactivos necesarios para analizar la variación en el nivel de expresión de los genes
40 por medio de cualquiera de los métodos descritos anteriormente en este documento y/o conocidos en el estado de la técnica. El kit además puede incluir, sin ningún tipo de limitación, el uso de tampones, enzimas, enzimas polimerasas, cofactores para obtener una actividad óptima de éstas, agentes para prevenir la contaminación, etc. Por otro lado el kit puede incluir todos los soportes y recipientes necesarios para su puesta en marcha y optimización. Preferiblemente, el kit comprende además las instrucciones para llevar a cabo el primer método de la invención.

45 Salvo que se defina de otra manera, los términos científicos y técnicos usados a lo largo de la presente descripción y las reivindicaciones tienen el mismo significado que el entendido por un experto en la materia perteneciente al estado de la técnica del que forma parte la presente invención.

50 A lo largo de la descripción y las reivindicaciones la palabra “comprende” y sus variantes no pretenden excluir otras características técnicas, aditivos, componentes o pasos. Para los expertos en la materia, otros objetos, ventajas y características de la invención se desprenderán en parte de la descripción y en parte de la práctica de la invención. Las siguientes figuras y ejemplos se proporcionan a modo de ilustración, y no se pretende que sean limitativos de la presente invención.

55 Descripción de las figuras

Figura 1. Muestra el porcentaje de genes correlacionados significativamente entre muestras de tejido FF y tejido FFPE usando distintos métodos de normalización. Permite comparar la eficacia de los distintos métodos de normalización a la hora de corregir los errores debidos a las diferentes eficiencias enzimáticas de las reacciones de retrotranscripción y amplificación, así como las causadas por las variaciones debidas a la diferente calidad del RNA debida a su degradación consecuencia de la fijación en formol.

65 Figura 2. Muestra las curvas de Kaplan-Meier que muestran SLRD para los grupos de bajo riesgo y alto riesgo de recaída a distancia en los grupos definidos mediante el “8-gene Score”.

Figura 3. Muestra el subanálisis del “8-gene Score” atendiendo a la afectación ganglionar. (A) Ganglios positivos. (B) Ganglios negativos.

Figura 4. Muestra las curvas de Kaplan-Meier que muestran la SLRD en los grupos de pacientes de bajo y alto riesgo definidos por el “8-gene Score” en las bases de datos utilizadas para la validación.

Figura 5. Muestra el test de Mann-Whitney que demuestra que el “8-gene Score” es capaz de diferenciar tumores serosos de ovario con bajo potencial maligno (LMP) de tumores serosos de ovario del tipo invasivo.

Ejemplos

Los siguientes ejemplos específicos que se proporcionan en este documento de patente sirven para ilustrar la naturaleza de la presente invención. Estos ejemplos se incluyen solamente con fines ilustrativos y no han de ser interpretados como limitaciones a la invención que aquí se reivindica. Por tanto, los ejemplos descritos más adelante ilustran la invención sin limitar el campo de aplicación de la misma.

Ejemplo 1

Desarrollo de un método de normalización para compensar el efecto de la degradación del RNA

El uso de muestras de tejido fijadas en formol y embebidas en parafina (FFPE, siglas del inglés “*Formalin-Fixed Paraffin-Embedded*”) presenta importantes ventajas con respecto a las muestras de tejido congelado (FF, siglas del inglés “*Fresh Frozen*”): son estables a temperatura ambiente, fáciles de almacenar y existe un amplio archivo de muestras clínicas disponibles junto con su información clínica y el seguimiento de la enfermedad. Frente a estas ventajas ofrecidas por las muestras de tejido FFPE, existe la desventaja de que el RNA obtenido a partir de estas muestras está muy degradado. El sesgo que puede introducir esta pérdida de RNA amplificable puede ser corregido mediante estrategias de normalización que puedan compensar este efecto de la degradación del RNA.

Materiales y métodos

1. Características clinicopatológicas de las pacientes incluidas en el estudio

Para el estudio de comparación de datos de tejido FF y tejido FFPE se seleccionaron muestras de 30 pacientes diagnosticadas de carcinoma de mama infiltrante para las cuales había tanto muestra de tejido FF como de tejido FFPE, ambas obtenidas a partir de la misma cirugía.

2. Selección de genes

Para la realización del estudio de comparación de datos de expresión génica de 30 muestras de tejido FF y tejido FFPE se seleccionaron una serie de genes descritos previamente obtenidos de los siguientes perfiles génicos descritos: “70-gene Signature” (van’t Veer *et al.* Nature. 2002 Jan 31; 415(6871):530-6), “Recurrence Score” (Paik *et al.* N Engl J Med. 2004 Dec 30;351(27):2817-26), “H/I Index” (Ma *et al.* Cáncer Cell. 2004 Jun;5(6):607-16), y “76-Gene Prognostic Signature” (Wang *et al.* Lancet 2005;365:671-9), todos ellos descritos con valor pronóstico en cáncer de mama, así como varios genes de referencia (18S, ACTB, B2M, GAPDH, GUSB, HMBS, HPRT1, IPO8, PGK1, POLR2A, PPIA, RPLP0, SDHA, TBP, TFRC, UBC, YWHAZ).

En este análisis de comparación de datos de tejido FF y tejido FFPE se utilizaron dos configuraciones de tarjetas microfluídicas (“*Taqman Low Density Arrays*”, TLDAs). Una de ellas consistió en una configuración de 95 genes en la que se amplificaron 78 genes de los estudios de van’t Veer y colaboradores, Paik y colaboradores, y Ma y colaboradores, así como 17 genes de referencia. En una segunda configuración de 63 genes, se midieron 50 genes del estudio de Wang y colaboradores junto con 13 genes de referencia. En ambos casos cada tarjeta incluía cDNA de dos muestras distintas, la correspondiente a la muestra de tejido FF y la correspondiente a la muestra de tejido FFPE, ambas de la misma biopsia.

3. Aislamiento del RNA y síntesis de cDNA

Con el objeto de seleccionar las muestras de tejido FFPE adecuadas para el estudio, se obtuvieron secciones teñidas con hematoxilina/eosina que fueron analizadas por un patólogo con experiencia en mama. Se eligieron muestras que presentaran al menos un 70% de células tumorales. Secciones de 5 μm de cada una de las muestras de tejido FFPE se desparafinaron mediante extracción con xileno, y lavados con etanol en concentraciones decrecientes (100%, 90% y 70%). El RNA se extrajo mediante el kit Master Pure Purification (Epicentre), según las especificaciones del fabricante. El RNA de las muestras FF se extrajo con el reactivo TRIzol (Invitrogen) según las especificaciones del fabricante.

El RNA aislado se cuantificó mediante espectrofotometría, midiendo la absorbencia a 260 y 280 nm. Así mismo, la calidad de los distintos RNA fue evaluada a través de electroforesis nativa en geles de agarosa al 1%, y electroforesis capilar mediante el equipo Agilent 2100 Bioanalyzer (Agilent Technologies). El RNA total aislado y analizado se retrotranscribió mediante el kit High Capacity cDNA Reverse Transcription (Applied Biosystems) utilizando 1 μg de RNA en 100 μL de reacción en un termociclador modelo GeneAmp PCR System 2700 (Applied Biosystems), empleando hexámeros al azar como cebadores.

4. RT-PCR Cuantitativa

Las reacciones de qRT-PCR se llevaron a cabo en un termociclador ABI PRISM 7900 HT Sequence Detection System (Applied Biosystems). Como soporte se utilizaron las tarjetas microfluídicas (Taqman Low Density Arrays, 5 TLDAs), también diseñadas por Applied Biosystems. Cada uno de los puertos de la tarjeta microfluídica contenía un volumen final de 100 μ L, 50 μ L de Taqman Universal PCR Master Mix y 50 μ L del cDNA equivalente a 100 ng del RNA total. La expresión de cada gen se midió por duplicado en la configuración de 95 genes y por triplicado en la de 63 genes.

5. Cálculos para los perfiles de expresión génica

Se obtuvieron los valores medios de ciclo umbral crudo (Ct, del inglés “*Threshold Cycle*”), definido como el punto a partir del cual la fluorescencia es claramente distinguible de la fluorescencia basal, mediante el software SDS 2.2 (Applied Biosystems). El valor máximo de Ct se fijó en 40. Estos valores de Ct fueron anotados en Excel 2003 para cálculos posteriores. Los datos de expresión génica también fueron analizados mediante el software GenEx de MultiD Analysis (www.multid.se) y Prism4 de Graphpad (www.graphpad.com).

Los valores de expresión génica se calcularon mediante el método Δ Ct descrito previamente (Livak KJ, Schmittgen TD. *Methods*. 2001 Dec;25(4):402-8). Los valores de expresión relativa fueron calculados como diferencias entre los valores de Ct de cada gen y un factor de normalización.

6. Factores de Normalización

En este estudio, de comparación de datos de tejido FF y tejido FFPE, se analizaron y compararon varios métodos de normalización: geNorm (Vandesompele *et al.* *Genome Biol* 2002;3: RESEARCH0034), Normfinder (Andersen *et al.* *Cáncer Res* 2004;64:5245-50), la media de expresión de todos los genes, así como un método desarrollado por los inventores denominado NorMean. Una vez seleccionados el gen o grupo de genes apropiados que funcionarían como genes de referencia se calculó el Factor de Normalización (NF, de sus siglas en inglés “*Normalization Factor*”) utilizando la media geométrica de los genes seleccionados en cada método.

El método de selección de genes de referencia desarrollado por los inventores, NorMean, permite seleccionar genes de referencia basándose en el valor del parámetro a_i :

$$a_i = \Sigma CV_{ij} / \Sigma r_{ij}$$

Donde CV_{ij} es el coeficiente de variación del gen de referencia i en el material j , y r_{ij} es el coeficiente de correlación de Pearson entre la expresión del gen i y la media de expresión de todos los genes en cada muestra en el material j . El valor a_i permite ordenar los genes de referencia: aquellos genes de referencia con menor valor a_i son aquellos con la expresión más estable (bajo CV) y con mayor correlación con la expresión génica media por muestra (alto r). El factor de normalización se determina calculando la media geométrica de los valores de expresión de los genes de referencia seleccionados.

7. Presentación de los Datos

Los datos se calcularon mediante los valores de expresión normalizados para cada gen en cada tipo de muestra dada por el valor $2^{-\Delta Ct}$, donde: $\Delta Ct = Ct_{gen} - NF$.

8. Análisis Estadístico

Para determinar la concordancia entre los datos de muestras de tejido FF y tejido FFPE se utilizó el coeficiente de correlación de Pearson entre los valores de expresión normalizados con cada uno de los factores de normalización para cada gen.

Resultados

Cada método de normalización genera un NF propio, utilizando para ello genes de referencia distintos (tabla 1). El porcentaje de genes con un valor de correlación de Pearson significativo entre los dos tipos de material se utilizó para valorar la capacidad para corregir el efecto de la conservación en las muestras FFPE de cada uno de los factores de normalización (Figura 1). En la mencionada figura, se observa que el método de normalización desarrollado, NorMean, es el más estable entre los distintos métodos analizados, puesto que en las dos configuraciones de tarjetas, más del 80% de los genes mostró una buena correlación entre tejido FF y tejido FFPE.

ES 2 343 996 B1

TABLA 1

Genes de referencia seleccionados por cada método en cada material

	FF	FFPE
geNorm	18S, IPO8, POLR2A, HMBS, UBC, PPIA	GUSB, IPO8, B2M, POLR2A, UBC, SDHA
NormFinder	18S	PPIA
NormFinder2	HPRT1, B2M	
NorMean	IPO8, POLR2A, UBC, SDHA	
NorMean MI	IPO8, HMBS, PPIA, SDHA	POLR2A, IPO8, UBC, SDHA

20 Ejemplo 2

Desarrollo y validación del perfil génico con capacidad pronóstica “8-gene-Score”

Materiales y métodos

25 1. Pacientes y Datos Clínicos

Para el desarrollo del perfil se recogieron de manera retrospectiva las características clínicas y anatomopatológicas de 206 pacientes diagnosticadas entre 1996 y 2002 de un carcinoma de mama ductal infiltrante en el Hospital Universitario La Paz (HULP). Se establecieron los siguientes criterios de inclusión:

- Diagnóstico anatomopatológico de carcinoma de mama ductal infiltrante.
- Estadio I o II anatomopatológico según la clasificación TNM de 2002.
- Muestra conservada en parafina disponible.
- Receptores hormonales positivos.
- Tratamiento óptimo, definido como:
 - Mastectomía o cirugía conservadora, con márgenes quirúrgicos libres de infiltración.
 - Radioterapia mamaria tras una cirugía conservadora.
 - Hormonoterapia.
- Seguimiento clínico superior a 60 meses (excepto en casos de recidiva precoz).
- Ausencia de cáncer de mama contralateral.

Todos los tumores incluidos en el estudio correspondían a los estadios I y II, definidos según el sistema de estadificación anatomopatológico TNM (siglas del inglés “*Tumor size, Node, Metastasis*”) para tumores malignos de 2002 utilizado por la UICC (“*Union Internationale Contre le Cancer*”) y el AJCC (“*American Joint Committee on Cancer*”). Fueron excluidos los tumores con receptores hormonales negativos o desconocidos. Así mismo, también fueron excluidos los casos que no hubieran recibido hormonoterapia como parte de su tratamiento, puesto que estos casos habrían recibido un tratamiento subóptimo. Por último quedaron fuera del análisis aquellos casos para los que no había historia clínica disponible y aquellas muestras cuyo producto de extracción no presentaba suficiente calidad. Finalmente el estudio incluyó 153 casos.

Las características clínicas y anatomopatológicas que se recogieron de manera retrospectiva fueron:

- Edad al diagnóstico (años).
- Fecha del diagnóstico.
- Tipo histológico: sólo se admitieron los casos de carcinoma ductal infiltrante.

ES 2 343 996 B1

- Tamaño del tumor primario (mm).
- Estado de los ganglios axilares: afectos (positivos) o no (negativos).
- 5 ■ Número de ganglios afectos.
- Grado de diferenciación: 1, 2 o 3.
- 10 ■ Receptores hormonales: para su entrada en el análisis, el caso debía presentar positividad al menos para el receptor de estrógenos o para el de progesterona.
- Tratamiento adyuvante: quimioterapia (señalando el esquema administrado en cada caso), hormonoterapia y/o radioterapia.
- 15 ■ Fecha de la recaída.
- Localización inicial de la recaída: loco-regional o a distancia (con especificación de los órganos afectados).
- 20 ■ Supervivencia libre de recaída a distancia (SLRD): tiempo transcurrido desde la fecha de la cirugía hasta la recaída a distancia.
- Supervivencia global (SG): tiempo transcurrido desde la fecha de la cirugía hasta la última visita o hasta el fallecimiento de la enferma.
- 25 Este estudio fue aprobado por el Comité Ético de Hospital Universitario La Paz (HULP).

2. Aislamiento del RNA y síntesis de cDNA

30 Se llevó a cabo mediante el método descrito en el ejemplo 1.

3. RT-PCR Cuantitativa

35 Se llevó a cabo mediante el método descrito en el ejemplo 1.

4. Selección de Genes

40 Se seleccionaron genes que se incluyeron en una tarjeta microfluídica (TLDA) con configuración de 96 y que contenía los ensayos disponibles para los genes del “70-gene Signature” (van’t Veer *et al.* Nature. 2002 Jan 31; 415(6871):530-6.), del “Recurrence Score” (Paik *et al.* N Engl J Med. 2004 Dec 30;351 (27):2817-26), así como para los genes del “H/I Index” (Ma *et al.* J Clin Oncol. 2006 Oct 1;24(28):4611-9). Además se incluyeron varios genes de referencia (POLR2A, IPO8, SDHA, HMBS, UBC, RPLP0, ACTB, PPIA, GAPDH, B2M, GUSB, HPRT1, TFRC).

5. Cálculos para los perfiles de expresión génica

50 Se obtuvieron los valores medios de ciclo umbral crudo (Ct, del inglés “Threshold Cycle”), definido como el punto a partir del cual la fluorescencia es claramente distinguible de la fluorescencia basal mediante el software SDS 2.2 (Applied Biosystems). El valor máximo de Ct se fijó en 40. Estos valores de Ct fueron anotados en Excel 2003 para cálculos posteriores. Los valores de Ct se normalizaron utilizando los cuatro genes de referencia seleccionados (IPO8, HMBS, POLR2A y SDHA).

55 Los valores de expresión génica se calcularon mediante el método ΔCt descrito previamente (Livak and Schmittgen. Methods. 2001 Dec;25(4):402-8). Los valores de expresión relativa fueron calculados como diferencias entre los valores de Ct de cada gen y un factor de normalización (NF, de sus siglas en inglés), calculado con la media geométrica de la expresión de los cuatro genes de referencia seleccionados por NorMean en el ejemplo 1 (IPO8, POLR2A, UBC y SDHA):

$$\Delta Ct = NF - Ct_{gen}$$

65 La expresión normalizada fue ajustada tomando como 0 el menor valor de expresión, de manera que un aumento de una unidad en la expresión reflejara aproximadamente una duplicación en la cantidad de RNA.

6. Metodología empleada en la búsqueda de un perfil de expresión génica reducido

En primer lugar se descartaron aquellos genes con baja correlación entre muestras de tejido FF y tejido FFPE, en base a los resultados obtenidos en el ejemplo 1. Se calculó un nivel de significación estadística para cada gen basado en el modelo de riesgos proporcionales de Cox o modelo de Cox (Cox. J Roy Stat Soc. 1972; 34:187-220), con el objetivo de identificar aquellos genes cuya expresión estaba relacionada con supervivencia libre de recaída a distancia. Los genes relacionados con SLRD fueron posteriormente filtrados en función de su valor p y de la correlación entre ellos mismos, seleccionando aquellos genes con menor valor p en cada grupo de correlación. Estos genes seleccionados fueron utilizados para desarrollar un predictor de riesgo de recaída basado en la expresión génica utilizando el método supervisado de componentes principales de Tibshirani y Bair (Bair and Tibshirani. PLoS Biol. 2004 Apr;2(4):E108).

Para evaluar el valor predictivo del método se utilizó el LOOCV (“*Leave-One-Out-Cross-Validation*”). Para probar la significación estadística del punto de corte se evaluó el valor p del estadístico “*log-rank*” test entre los grupos de riesgo mediante 2000 permutaciones aleatorias. Los análisis se llevaron a cabo mediante BRB-ArrayTools v3.6.1 (desarrollado por Richard Simón y Amy Peng).

7. Análisis Estadístico

Para evaluar el valor pronóstico del “*8-gene Score*” en nuestra población se llevó a cabo un análisis de Kaplan-Meier y se compararon los grupos mediante el “*log-rank*” test. También se aplicó un análisis univariante y multivariante de Cox, en un modelo que incluía el Grado tumoral (1 vs 2 y 1 vs 3), tamaño (≤ 2 cm vs. > 2 cm) y estado ganglionar (ausencia de afectación ganglionar vs. uno a tres ganglios positivos). El criterio de valoración fue SLRD, como en otros estudios de perfiles génicos en cáncer de mama.

Para medir la fuerza de asociación entre el “*8-gene Score*” y el “*70-gene Signature*” y el “*Recurrence Score*” se llevaron a cabo análisis de tablas de contingencia de doble entrada y se calculó el estadístico V de Cramer (Fan *et al.* N Engl J Med. 2006 Aug 10;355(6):560-9).

Para medir la capacidad discriminativa del modelo a cinco años se calculó el índice de concordancia de Harrell corregido para el sesgo. Los modelos fueron reajustados 500 veces con la técnica de “*bootstrap resampling*”. El índice de concordancia es el porcentaje de pares de pacientes en los que lo predicho y lo observado concuerdan. De esta manera $c = 0.50$ representa la concordancia debida a la casualidad; $c = 1.0$ representa la discriminación perfecta (Harrell *et al.* Stat Med. 1996 Feb 28;15(4):361-87).

8. Bases de datos para la Validación del “*8-gene Score*”

Se utilizaron cuatro grupos de datos independientes disponibles online para validar el “*8-gene Score*”. Estos grupos de datos fueron: NKL, descargado de la página web de Rosetta Inpharmatics (<http://www.rii.com/publications/2002/nejm.html>), SWE (GSE1456 - Pawitan *et al.* Breast Cancer Res. 2005;7(6):R953-64), UPP (GSE4922 - Ivshina *et al.* Cancer Res. 2006 Nov 1;66(21): 10292-301) y LOI (GSE6532 - Loi *et al.* BMC Genomics. 2008;9:239). Los tres últimos fueron descargados del depósito de datos NCBI GEO (<http://www.ncbi.nlm.nih.gov/projects/geo/index.cgi>). Para aplicar el “*8-gene Score*” obtenido mediante qRT-PCR sobre estos datos de microarrays, los valores de expresión fueron z-transformados. A continuación fueron ajustados tomando el menor valor de expresión como 0 y escalados. Se llevó a cabo una normalización por gen dentro de las cohortes de validación utilizando los valores medios obtenidos en la cohorte de descubrimiento (Schmidt. Cancer Res. 2008 Jul 1;68(13):5405-13).

Se llevó a cabo un análisis de Kaplan-Meier y se compararon los grupos mediante el “*log-rank*” test. También se aplicó un análisis de los riesgos proporcionales de Cox. El criterio de valoración fue de nuevo SLRD.

Todos los análisis estadísticos fueron llevados a cabo mediante la versión 9.1 del paquete de software SPSS, la versión 5.00 de GraphPad Prism y la versión 2.2 de R con el paquete estadístico Design versión 2.0-12. Los valores p se consideraron bilaterales y estadísticamente significativos cuando $p < 0.05$.

Resultados

Se incluyeron 153 pacientes diagnosticadas entre febrero de 1995 y marzo de 2003; sus características clínicas se muestran en la tabla 2. La mediana de la edad fue de 58 años y la mediana de seguimiento 91 meses. Se llevó a cabo una mastectomía en 66 pacientes (43%), mientras que el resto fueron operadas con cirugía conservadora seguida de radiación adyuvante. Todas las pacientes recibieron tamoxifeno adyuvante durante cinco años y 97 de ellas (63%) fueron tratadas con quimioterapia adyuvante, consistente en la administración de ciclofosfamida-metotrexato-fluorouracilo (CMF) o un esquema similar pero incluyendo una antraciclina (FAC o FEC). Treinta y cuatro pacientes (22%) tuvieron una recaída a distancia, de las cuales 17 fallecieron, y no se tuvo seguimiento tras la recaída de 7 de ellas. De las 119 pacientes sin recaída a distancia cuatro tuvieron recaída local tratada con cirugía exitosamente.

ES 2 343 996 B1

TABLA 2

Características clínicas de las pacientes

5	Edad	Mediana 58, rango: 29-82
		Numero de pacientes (porcentaje)
10	Tamaño	
	1	77 (50,3%)
	2	76 (49,7%)
15	Estado Ganglionar	
	0	96 (62,7%)
	1	57 (37,3%)
20	Estadio	
	I	61 (39,9%)
	Ila	51 (33,3%)
25	IIb	41 (26,8%)
	Receptores Hormonales	
	RE+/RPg-	30 (19,6%)
30	RE+/RPg+	110 (71,9%)
	RE+/RPg desconocido	12 (7,9%)
	RE-/RPg+	1 (0,7%)
35	Grado	
	1	29 (18,9%)
	2	64 (41,8%)
40	3	59 (38,6%)
	x	1 (0,7%)
45	Quimioterapia	
	Sin Quimioterapia	56 (36,6%)
	CMF	42 (27,4%)
	Antraciclinas	55 (35,9%)

50 Se construyó un perfil genético reducido. En primer lugar se identificaron 53 genes con alta correlación entre los datos de expresión de las muestras de tejido FF y tejido FFPE. Posteriormente se seleccionaron 17 genes en función de su valor p relacionado con SLRD. Con estos 17 genes se construyó un modelo que identificaba adecuadamente un grupo de pacientes de alto riesgo de recaída a distancia. Modelos con 10, 9, 8, 7 o incluso menos genes mostraron también una buena separación entre grupos de alto y bajo riesgo de recaída a distancia. Se seleccionó el perfil de 8 genes porque fue el que obtuvo el mejor rendimiento. El perfil de 8 genes se calculó en cada muestra utilizando las medidas de expresión génica normalizadas basándose en la siguiente ecuación: "8-gene Score" = 0.1936*DTL + 0.2176*ECT2 + 0.0454*MTDH + 0.1329*PRC1 + 0.0556*RFC4 - 0.1913*SCUBE2 - 0.0443*STK32B - 0.1182*ZNF53. La información relativa a estos ocho genes, así como la de los genes de referencia, se encuentra codificada en la tabla 3.

60 Un aumento en la expresión de los genes DTL, ECT2, MTDH, PRC1 y RFC4 esta asociado con menor SLRD (es decir, son factores que aumentan el riesgo de recaída), mientras que un aumento en la expresión de SCUBE2, STK32B y ZNF533 esta asociado con mayor SLRD (es decir, son factores protectores frente a la recaída).

65 Del total de genes analizados hay ciertos genes que coexpresan con los componentes del predictor. Debido a la naturaleza multigénica del predictor, la sustitución de estos genes por genes que coexpresen con ellos con un coeficiente de correlación de Pearson $r \geq 0,4$ no alteraría la capacidad discriminativa del predictor. Estos genes son: AYTL2, BIRC5, CCNB1, CCNE2, GMPS, MCM6, MELK, MYBL2, ORC6L, PGR y TGFB3 (Tabla 4).

ES 2 343 996 B1

TABLA 3

Identificación de los genes incluidos en el “8-gene Score” y genes de referencia (el asterisco indica los genes de referencia).

Símbolo Gen	P valor	Coficiente en la ecuación	NM	Código Ensayo Applied Biosystems
DTL	1,2e-06	0,1936	NM_016448.2 (SEQ ID NO: 1)	Hs00212788_m1
ECT2	2,8e-06	0,2176	NM_018098.4 (SEQ ID NO: 2)	Hs00216455_m1
MTDH	2,9e-06	0,0454	NM_178812.2 (SEQ ID NO: 3)	Hs00757841_m1
PRC1	< 1e-07	0,1329	NM_003981.2 (SEQ ID NO: 4)	Hs00187740_m1
RFC4	0,0002592	0,0556	NM_181573.2 (SEQ ID NO: 5)	Hs00427469_m1
SCUBE2	0,0005634	-0,1913	NM_020974.1 (SEQ ID NO: 6)	Hs00221277_m1
STK32B	0,0004406	-0,0443	NM_018401.1 (SEQ ID NO: 7)	Hs00179683_m1
ZNF533	2,13e-05	-0,1182	NM_152520.4 (SEQ ID NO: 8)	Hs00332216_m1
IPO8*	-	-	NM_006390.2 (SEQ ID NO: 9)	Hs00183533_m1
HMBS*	-	-	NM_000190.3 (SEQ ID NO: 10)	Hs00609297_m1
POLR2A*	-	-	NM_000937.3 (SEQ ID NO: 11)	Hs00172187_m1
SDHA*	-	-	NM_004168.2 (SEQ ID NO: 12)	Hs00188166_m1

ES 2 343 996 B1

TABLA 4

Identificación de los genes que pueden sustituir a los que constituyen el “8-gene Score”

Símbolo Gen	NM	Puede sustituir al gen:	Código Ensayo Applied Biosystems
AYTL2	NM_024830.3 (SEQ ID NO: 13)	RFC4	Hs00227357_m1
BIRC5	NM_001168.2 (SEQ ID NO: 14)	DTL, PRC1, RFC4	Hs00153353_m1
CCNB1	NM_031966.2 (SEQ ID NO: 15)	DTL, ECT2, MTDH, PRC1, RFC4	Hs00259126_m1
CCNE2	NM_057749.1 (SEQ ID NO: 16)	DTL, PRC1	Hs00372959_m1
GMPS	NM_003875.2 (SEQ ID NO: 17)	PRC1, RFC4	Hs00269500_m1
MCM6	NM_005915.4 (SEQ ID NO: 18)	DTL, ECT2, PRC1, RFC4	Hs00195504_m1
MELK	NM_014791.2 (SEQ ID NO: 19)	DTL, ECT2, PRC1, RFC4	Hs00207681_m1
MYBL2	NM_002466.2 (SEQ ID NO: 20)	DTL, PRC1, RFC4	Hs00231158_m1
ORC6L	NM_014321.2 (SEQ ID NO: 21)	DTL, PRC1	Hs00204876_m1
PGR	NM_000926.4 (SEQ ID NO: 22)	SCUBE2	Hs00172183_m1
TGFB3	NM_003239.2 (SEQ NO: 23)	STK32B, ZNF533	Hs00234245_m1

Con la Intención de asignar a cada paciente a un grupo de riesgo se definió un punto de corte en 2.86. Así, pacientes con una puntuación en el perfil <2.86 fueron asignados al grupo de bajo riesgo de recaída a distancia y pacientes con una puntuación ≥ 2.86 constituyeron el grupo de alto riesgo de recaída a distancia. El valor p del estadístico “log-rank” test entre los grupos de riesgo basado en 2000 permutaciones fue 0.044. Este valor proporcionó significación estadística al punto de corte. El “8-gene Score” asigna pacientes a los grupos de bajo (60%) y alto riesgo (40%), como refleja la Figura 2.

La supervivencia libre de recaída a distancia a cinco años fue del 97.7% para el grupo de bajo riesgo y de 60.6% para el grupo de alto riesgo (HR: 20.4, IC 95%: 6.2 - 67.5; $p < 0.001$). También se calculó la supervivencia global a cinco años: 98.9% para el grupo de bajo riesgo de recaída a distancia y 86.6% para el grupo de alto riesgo (HR: 7.496, IC 95%: 2.4 - 23.4; $p < 0.001$).

Se llevó a cabo un subanálisis considerando el estado de los ganglios linfáticos, como muestra la Figura 3. En pacientes con afectación ganglionar positiva el “8-gene Score” incluía la mitad de pacientes en cada grupo (bajo riesgo vs alto riesgo): La SLRD a cinco años era de 93.3% vs. 39.5%. En mujeres con afectación ganglionar negativa los valores fueron 100% vs. 75.7%, respectivamente.

El análisis multivariante de Cox incluyó el “8-gene Score”, el tamaño tumoral, el estado de afectación ganglionar y el grado de diferenciación del tumor. El “8-gene Score” es predictor de SLRD (Tabla 5), indicando que este perfil de expresión génica añade importante información pronóstica a la aportada por los factores clínicos tradicionales. El estado ganglionar es el único factor clínico tradicional que mantiene significación estadística en este análisis multivariante.

TABLA 5

Análisis multivariante para el “8-gene Score”, el “70-gene Signature” y el “Recurrence Score” (Tamaño Tumoral: >2 cm vs. ≤2 cm. Estado Ganglionar: Positivo (1-3 ganglios afectados) vs Negativo

		“8-Gene Score”		
		HR	IC 95%	p valor
Tamaño Tumoral		1,31	0,52 – 3,3	0,55
Estado Ganglionar		2,87	1,38 – 6,01	0,005
Grado				0,227
(2 vs. 1)		0,60	0,11 – 3,16	0,555
(3 vs. 1)		1,42	0,28 – 7,15	0,665
“8-gene Score”		15,61	4,41 – 55,17	<0,001
		“70-Gene Signature”		
		HR	IC 95%	p valor
Tamaño Tumoral		1,4	0,57 – 3,42	0,46
Estado Ganglionar		2,70	1,30 – 5,62	0,008
Grado				0,113
(2 vs. 1)		0,82	0,160 - 4,181	0,810
(3 vs. 1)		2,349	0,473 - 11,672	0,297
“70-Gene Signature”		3,505	1,33 – 9,23	0,011
		“Recurrence Score”		
		HR	IC 95%	p valor
Tamaño Tumoral		1,17	0,48 – 2,82	0,72
Estado Ganglionar		2,49	1,20 – 5,16	0,014
Grado				0,043
(2 vs. 1)		0,68	0,13 – 3,45	0,644
(3 vs. 1)		2,26	0,49 – 10,40	0,292
“Recurrence Score”				0,021
(“Intermediate” vs		4,46	0,78 – 25,33	0,091
“Low”)				
(“High” vs “Low”)		8,18	1,76 – 37,86	0,007

ES 2 343 996 B1

Se utilizó el estadístico V de Cramer para determinar la concordancia entre el “8-gene Score” y los otros perfiles estudiados. La correlación para el “8-gene Score” fue de 0.65 con el “70-gene Signature” y de 0.58 con el “Recurrence Score”. Estos resultados indican una fuerte correlación entre el “8-gene Score”, el “70-gene Signature” y el “Recurrence Score”.

Para determinar la capacidad discriminativa de cada perfil a 5 años, se calculó el índice de concordancia de Harrell corregido para el sesgo. Los valores fueron: 0.81 para el “8-gene Score”, 0.73 para el “Recurrence Score” y 0.70 para el “70-gene Signature”. Estos valores indican que el “8-gene Score” es comparable a los otros dos perfiles en cuanto a capacidad discriminativa.

Se aplicó el “8-gene Score” a una base de datos del Instituto Holandés del Cáncer (NKI) que había sido previamente utilizada para comparar varios perfiles (Fan *et al.* N Engl J Med. 2006 Aug 10;355(6):560-9). El “8-gene Score” presentó diferencias significativas en SLRD en la población completa de 295 pacientes, y también en los grupos de ganglios positivos, ganglios negativos y receptores de estrógenos positivos (Tabla 6). En una comparación directa con el “70-gene Signature”, el “8-gene Score” asignó más pacientes al grupo de bajo riesgo, mientras que la SLRD fue ligeramente menor para todos los grupos. Si el punto de corte se modificaba para incluir tantos pacientes en el grupo de bajo riesgo como hace el “70-gene Signature” los resultados eran virtualmente idénticos.

A continuación se analizó el rendimiento del “8-gene Score” en otras tres bases de datos adicionales: SWE, UPP y LOI. Se encontraron diferencias significativas en SLRD en todas ellas (Tabla 7 y Figura 4). En todos los casos el “8-gene Score” asignaba más del 55% de pacientes al grupo de bajo riesgo, con una SLRD por encima del 90% en las bases de datos LOI y SWE y por encima del 80% en la de UPP.

Diferencias en SLRD a diez años también fueron significativas (datos no mostrados).

TABLA 6

*Comparación entre el “8-gene Score” y otros perfiles génicos utilizando las base de datos del NKI. * Proporción de pacientes incluidos en los grupos de riesgo favorable y desfavorable con cada perfil génico. ** Bajo riesgo + riesgo intermedio vs alto riesgo*

Perfil	Grupo	% Bajo vs Alto Riesgo *	SLRD a los 5 años
“8-gene Score”	Todos Pacientes	52 vs 48	85,7 ± 2,7 vs 54,8 ± 4,3
	Ganglios-	53 vs 47	86,1 ± 3,9 vs 50,5 ± 6,2
	Ganglios+	52 vs 48	89,0 ± 3,7 vs 59,1 ± 6,1
	RE+	65 vs 35	88,8 ± 2,6 vs 54,2 ± 5,7
“70-Gene Signature”	Todos Pacientes	39 vs 61	94,7 ± 2,1 vs 60,5 ± 3,8
	Ganglios-	40 vs 60	93,4 ± 3,2 vs 56,2 ± 5,5
	Ganglios+	38 vs 62	95,2 ± 2,6 vs 66,3 ± 5,2
	RE+	50 vs 50	92,9 ± 2,4 vs 58,2 ± 4,7
“Recurrence Score”	Todos Pacientes	35 vs 65 **	92,2 vs 58,5
“Wound response”	Todos Pacientes	23 vs 77	92,5 vs 63,6

ES 2 343 996 B1

TABLA 7

*Funcionamiento del “8-gene Score” en otras bases de datos. *Los valores de SLRD son a 5 años*

Set	Bajo Riesgo		Alto Riesgo		Análisis Estadístico		
	Pacientes	SLRD*	Pacientes	SLRD*	Sig.	HR	95% IC
HULP153	60%	97,7	40%	60,6	<0,001	20,465	6,198 – 67,579
NKI295	53%	85,7	47%	54,8	<0,001	3,403	2,226 - 5,202
LOI336	56%	90,4	44%	71,2	<0,001	2,430	1,596 - 3,698
UPP249	58%	80,5	42%	60,5	0,001	1,975	1,302 - 2,998
SWE159	55%	90,8	45%	62,5	<0,001	4,462	2,179 - 9,139

Ejemplo 3

Ejemplo de utilización del “8-gene Score” en tumores serosos de ovario

Los tumores serosos de ovario suponen más del 50% de todos los tumores de ovario. Aproximadamente el 10-15% de estos tumores se clasifican como tumores proliferativos o de bajo potencial maligno (LMP, siglas de inglés “*Low Malignant Potential*”), considerándose el resto tumores invasivos. Se ha verificado la capacidad del “8-gene Score” de separar estas dos poblaciones de tumores serosos de ovarios utilizando la base de datos GSE12172 del National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). Los datos de los microarrays fueron procesados siguiendo el protocolo descrito para las series de validación utilizadas en el desarrollo del “8-gene Score” en el Ejemplo 2. La base de datos contiene 90 muestras, de las cuales 30 son muestras LMP y 60 invasivas. Los datos clínicos no incluyen el tiempo hasta la recaída, por lo tanto no se pueden hacer estudios de supervivencia como se realizaron para las bases de datos utilizadas en la validación. Sin embargo, como se muestra en el artículo de Anglesio y cois. (Anglesio *et al.* Mol Cáncer Res. 2008 Nov;6(11):1678-90), es posible hacer una discriminación entre los dos tipos de tumores. Aplicando el “8-gene Score” y realizando un test de Mann-Whitney, se concluye que el “8-gene Score” es capaz de diferenciar tumores serosos de ovario con bajo potencial maligno de tumores serosos de ovario del tipo invasivo en esta población (figura 5). Los resultados obtenidos muestran el potencial diagnóstico del “8-gene Score” en un tipo tumoral diferente a aquel en el que ha sido desarrollado.

ES 2 343 996 B1

REIVINDICACIONES

1. Método para la subclasificación de tumores, que comprende:

- a. obtención de una muestra biológica aislada que comprende células tumorales del mamífero;
- b. detección de la cantidad del producto de la expresión de entre dos y ocho genes siendo un gen PRC1 en combinación con los siguientes genes: DTL, ECT2, MTDH, RFC4, SCUBE2, STK32B o ZNF533, en la muestra obtenida en (a) y
- c. comparación de la cantidad detectada en el paso (b) con una cantidad de referencia.

2. Método según la reivindicación 1 donde en el paso (b) se detecta la cantidad del producto de la expresión de los genes DTL, ECT2, MTDH, PRC1, RFC4, SCUBE2, STK32B y ZNF533.

3. Método de pronóstico de la evolución del tumor que comprende los pasos (a)-(c) según cualquiera de las reivindicaciones 1 ó 2, que además comprende un paso (d) donde que la cantidad detectada en el paso (b) del producto de expresión de los genes DTL, ECT2, MTDH, PRC1 o RFC4 mayor que la cantidad de referencia con la que se compara en el paso (c) o una cantidad detectada en el paso (b) de los genes SCUBE2, STK32B o ZNF533 menor que la cantidad de referencia con la que se compara en el paso (c), es indicativa de una menor supervivencia libre de recaída a distancia o una menor supervivencia global.

4. Método según cualquiera de las reivindicaciones 1 a 3, donde el producto de expresión de los genes es el mRNA.

5. Método según cualquiera de las reivindicaciones 1 a 4, donde la detección del mRNA se realiza mediante la técnica de qRT-PCR.

6. Método según cualquiera de las reivindicaciones 1 a 5 donde la muestra biológica aislada que comprende células tumorales está fijada y embebida en parafina.

7. Método según cualquiera de las reivindicaciones 1 a 6 donde el tumor pertenece al siguiente grupo: mama, ginecológico o próstata.

8. Método según cualquiera de las reivindicaciones 1 a 7 donde el tumor es de mama.

9. Método según la reivindicación 8 donde el tumor de mama es un carcinoma de mama infiltrante.

10. Método según la reivindicación 9 donde el carcinoma de mama infiltrante está en un estadio I o II.

11. Método según cualquiera de las reivindicaciones 8 a 10 donde el tumor es positivo para la expresión de los receptores hormonales de estrógenos.

12. Método según la reivindicación 11 donde la muestra se aísla de un mamífero que ha sido sometido a una terapia hormonal y bien una mastectomía o bien una tumorectomía seguida de radioterapia.

13. Método según cualquiera de las reivindicaciones 1 a 7 donde el tumor es de ovario.

14. Método según cualquiera de las reivindicaciones 1 a 7 donde el tumor de ovario es un tumor epitelial del ovario.

15. Método según la reivindicación 14 donde el tumor epitelial del ovario es un tumor seroso.

16. Kit adaptado para llevar a cabo el método para la subclasificación de tumores, según cualquiera de las reivindicaciones 1 a 15 que comprende un cebador directo y un cebador inverso que hibridan con el fragmento definido entre los nucleótidos 106 y 238 de SEQ ID NO: 4 y los demás reactivos necesarios para:

- a. detección de la cantidad del producto de la expresión de entre dos y ocho genes siendo un gen PRC1 en combinación con los siguientes genes: DTL, ECT2, MTDH, RFC4, SCUBE2, STK32B o ZNF533, en la muestra obtenida en (a) y
- b. comparación de la cantidad detectada en el paso (b) con una cantidad de referencia.

17. Kit según la reivindicación 16 donde la detección de la cantidad del producto es de la expresión de los genes DTL, ECT2, MTDH, PRC1, RFC4, SCUBE2, STK32B y ZNF533.

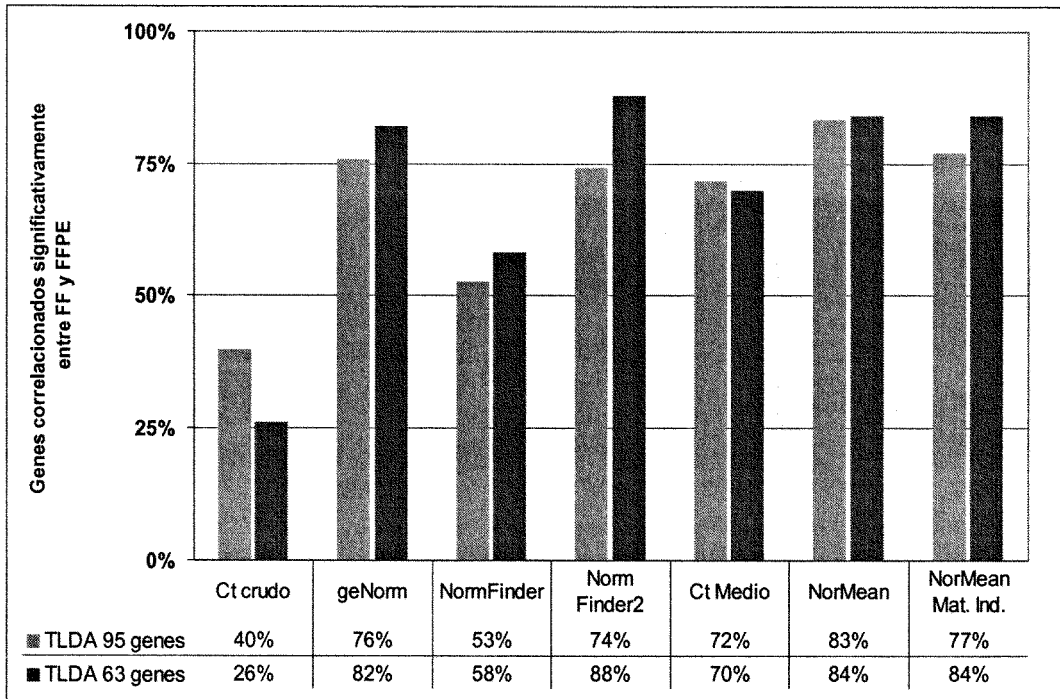


FIG. 1

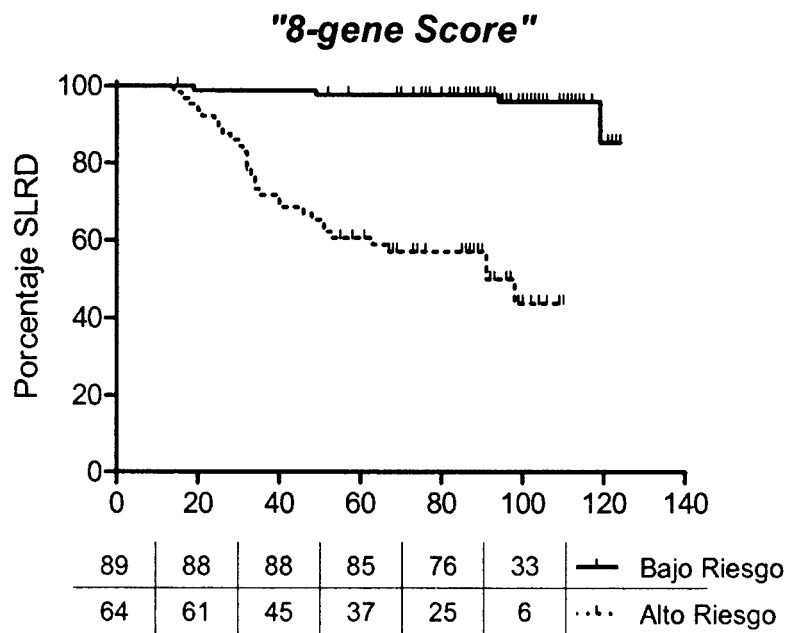


FIG. 2

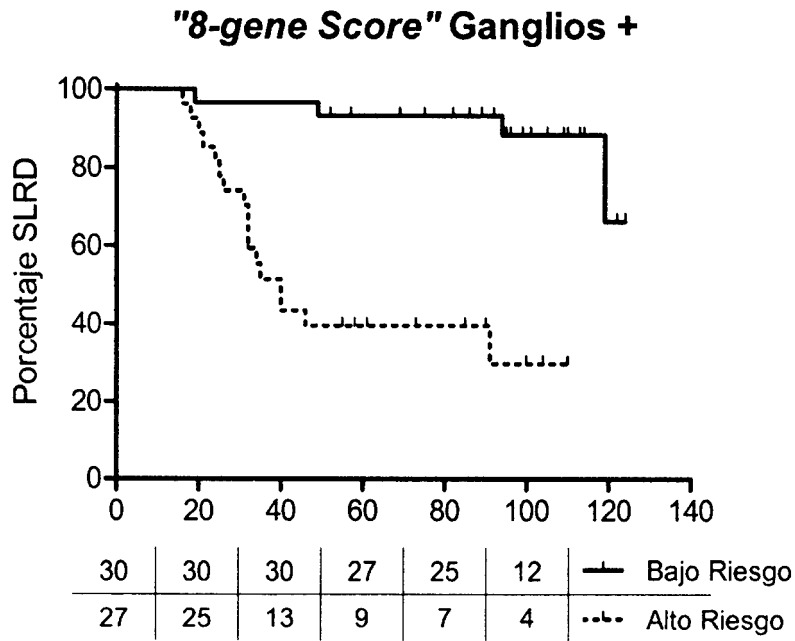


FIG. 3

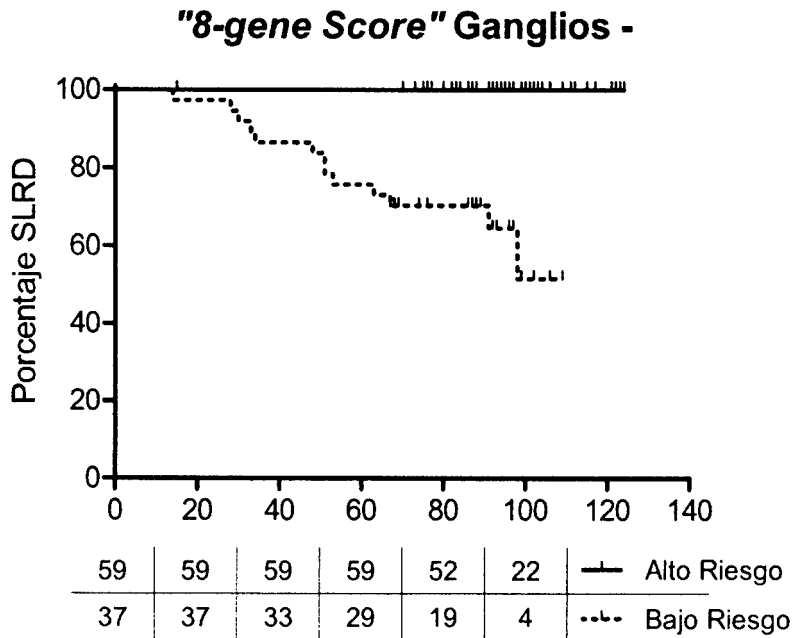


FIG. 3

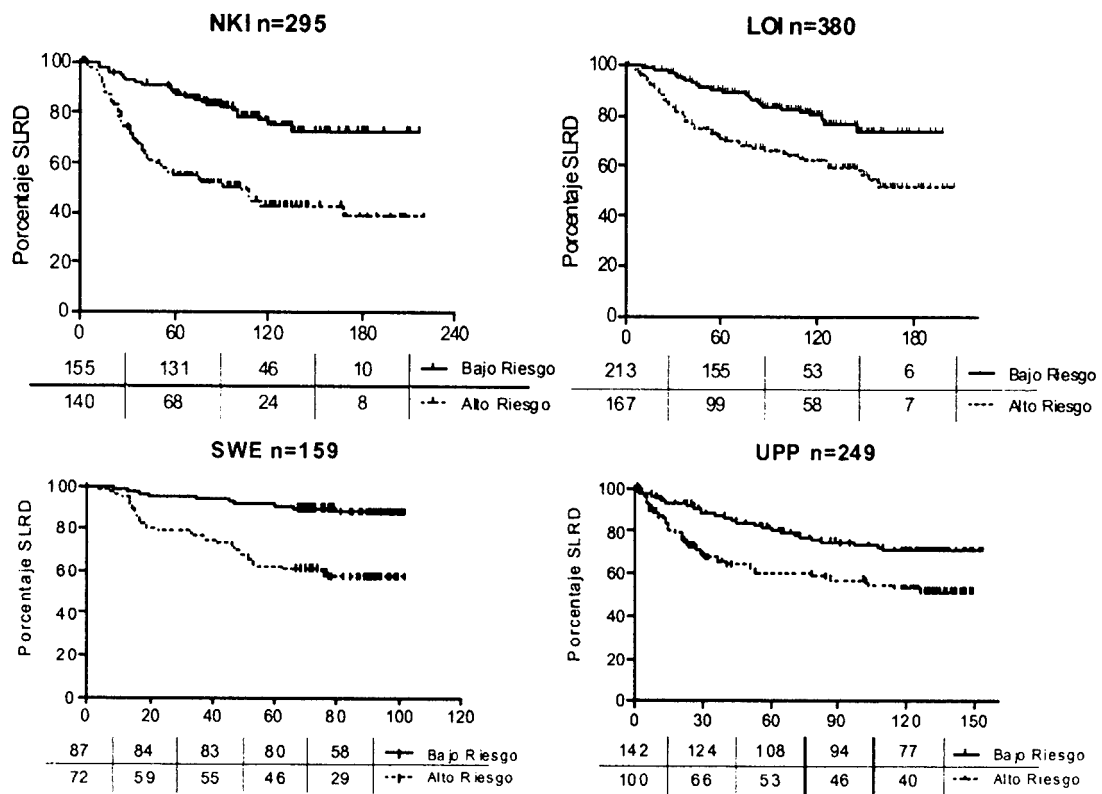


FIG. 4

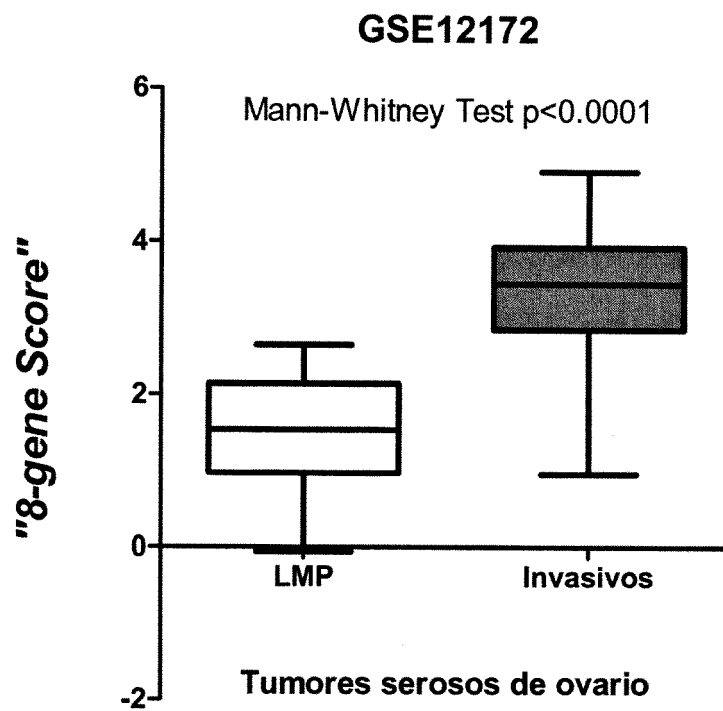


FIG. 5

ES 2 343 996 B1

LISTA DE SECUENCIAS

<110> Fundación para la investigación biomédica del Hospital Universitario La Paz

5 <120> Método para la subclasificación de tumores

<130> 1660.4

10 <160> 23

<170> PatentIn version 3.5

15 <210> 1

<211> 4428

<212> DNA

<213> *Homo sapiens*

20 <400> 1

25 cgcggtttgt ccgcgctct gctaccccct gcgcaggcgc tcaaggagct cttggactcc
 60

 aggttcccgc ggctgggaga aaaggaggcg gggatccgaa gggggaaatg actctgaggc
 120

 gcccggacgt cgctcgaag ccaatcagag agcgtgacgt cagtttgccg cggagtttgg
 180

30 cggccggggc ttacagtggc gggagttgga ggcgataacg atttgtgttg tgagaggcgc
 240

 aagctgcatg ttctgctgaa cttggaggca tttctacgac ttttctctca gctgaggcct
 300

35 ttcctccgac cctgatgctc ttcaattcgg tgctccgcca gccccagcct ggcgtcctga
 360

40 gaaatggatg gtcttcacaa taccctcttc aatcccttct gactggttat cagtgcagtg
 420

 gtaatgatga acacacttct tatggagaaa caggagtccc agttcctcct tttggatgta
 480

45 ccttctcttc tgctcccaat atggaacatg tactagcagt tgccaatgaa gaaggctttg
 540

 ttcgattgta taacacagaa tcacaaagtt tcagaaagaa gtgcttcaaa gaatggatgg
 600

50 ctcactggaa tgccgtcttt gacctggcct gggttcctgg tgaacttaaa cttgttacag
 660

55 cagcaggatg tcaaacagcc aaatthttggg acgtaaaagc tggatgagctg attggaacat
 720

 gcaaaggatc tcaatgcagc ctcaagtcag ttgccttttc taagtttgag aaagctgtat
 780

60 tctgtacggg tggaagagat ggcaacatta tggctctggga taccaggatg aacaaaaaag
 840

 atgggtttta taggcaagtg aatcaaatca gtggagctca caatacctca gacaagcaaa
 900

65 ccccttcaaa acccaagaag aaacagaatt caaaaggact tgctccttct gtggatttcc
 960

ES 2 343 996 B1

agcaaagtgt tactgtggtc ctctttcaag acgagaatac cttagtctca gcaggagctg
1020

5 tggatgggat aatcaaagta tgggatttac gtaagaatta tactgcttat cgacaagaac
1080

ccatagcatc caagtctttc ctgtaccag gtagcagcac tcgaaaactt ggatattcaa
1140

10 gtctgatttt ggattccact ggctctactt tatttgctaa ttgcacagac gataacatct
1200

acatgtttaa tatgactggg ttgaagactt ctccagtggc tattttcaat ggacaccaga
1260

15 actctacctt ttatgtaaaa tccagcctta gtccagatga ccagttttta gtcagtggct
1320

caagtgatga agctgcctac atatggaagg tctccacacc ctggcaacct cctactgtgc
20 1380

tcctgggtca ttctcaagag gtcacgtctg tgtgctggtg tccatctgac ttcacaaaga
1440

25 ttgctacctg ttctgatgac aatacactaa aaatctggcg cttgaataga ggcttagagg
1500

agaaaccagg aggtgataaa ctttccacgg tgggttgggc ctctcagaag aaaaaagagt
30 1560

caagacctgg cctagtaaca gtaacgagta gccagagtac tcctgccaaa gccccaggg
1620

caaagtgcaa tccatccaat tcttccccgt catccgcagc ttgtgccccca agctgtgctg
35 1680

gagacctccc tcttcttca aatactccta cgttctctat taaaacctct cctgccaagg
1740

40 cccgggtctcc catcaacaga agaggctctg tctcctccgt ctctcccaag ccaccttcat
1800

ctttcaagat gtcgattaga aactgggtga cccgaacacc ttctcatca ccacccatca
1860

45 ctccacctgc ttcggagacc aagatcatgt ctccgagaaa agcccttatt cctgtgagcc
1920

agaagtcac ccaagcagag gcttgctctg agtctagaaa tagagtaaag aggaggctag
50 1980

actcaagctg tctggagagt gtgaaacaaa agtgtgtgaa gagttgtaac tgtgtgactg
2040

55 agcttgatgg ccaagttgaa aatcttcatt tggatctgtg ctgccttgct ggtaaccagg
2100

aagaccttag taaggactct ctaggctccta ccaaatacaag caaaattgaa ggagctggta
2160

60 ccagtatctc agagcctccg tctcctatca gtccgtatgc ttcagaaagc tgtggaacgc
2220

tacctcttcc tttgagacct tgtggagaag ggtctgaaat ggtaggcaaa gagaatagtt
2280

65

ES 2 343 996 B1

ccccagagaa taaaaactgg ttgttggcca tggcagccaa acggaaggct gagaatccat
2340

5 ctccacgaag tccgtcatcc cagacacca attccaggag acagagcgga aagaaattgc
2400

caagcccggg caccatcacg cccagctcca tgaggaaaat ctgcacatac ttccatagaa
2460

10 agtcccagga ggactttctgt ggtcctgaac actcaacaga attatagatt ctaatctgag
2520

tgagttactg agctttgggc cactaaaaca agctgagctt tgggtccacta aaacaagatg
2580

15 aaaaatacaa gagtgactct ataactctgg tctttaagaa agctgccttt tcatttttag
2640

acaaaatctt ttcaacgctg aaatgtacct aatctgggtc tactaccata atgtatatgc
2700

20 agcttcccga ggatgaatgc tgtgtttaaa tttcataaag taaatttgtc actctagcat
2760

25 tttgaatgaa tagtcttcac tttttaaatt attcatcttc tctataataa tgacatccca
2820

gttcatggag gcaaaaaaca agtttcttgt taccctgaaa ctttctatgc tcagtggaaa
2880

30 gtatctgcca gccacagcat gaggcctgtg aaggctgact gagaaatcct ctgctgaaga
2940

cccctgggtc tgttctgcct ccaacatgta taattttatt tgaaatacat aatcttttca
3000

35 ctatgctttt gtggggtttt ttttaagtat gtgtaaaaat gtgatgctca gataagtaca
3060

40 tttatatcag ttcagtgtta aaatgcagtc tcttgagtta aagtcacctt tattttaaat
3120

gcagtgataa atgtcaactc ttcggagaaa ctaggagAAC aacaacagaa agctgtgttt
3180

45 gtcttttttc tctcaaatat atctcccgta tgagatttca ggtcccatg tttcaccaa
3240

gcaatctgct atgtcagcca acccaacatc actttctaca ggaggttatg attttgcca
3300

50 tttactagag gaagatgttt tatgaaatca atttgggggtt tgaattcagg tgcaatcatc
3360

55 agttcttttag gggctgcaat gttttaaaaa aaataagtca tcagatttta agaaaaagt
3420

gatgatttct tattgatatt tttgtaacag aatatagctc ttaactgaaa atccagaacc
3480

60 agaaacataa atcttgagtt tcttttcatg tacataaaaa gcaatagcct tttagtatag
3540

atagccctga gccaaaaagt aatagaatth tctctagata ttttaatacag agagtgtata
3600

65

ES 2 343 996 B1

gactgactct aagttaataa tgtgcaaaat atcttaaaca tccctcccct tattcaacaa
3660
5 ttatgtatca gtgatcttga accattgttt tatatTTTTc acctttgtaa cctcatggaa
3720
agaggcttta catactttct atgtactatt tacttagaag ggagccccct tccagtcatg
3780
10 aaacttcatt tgttttatcc atatccctga ggactgtgta gactttatgt cagtttttgt
3840
cattatTTTga aaatctattc tgacaacttt ttaattcctt tgatcttata agttaagct
3900
15 gtaacaactg aaattgcatg gatcaagtaa gcatagTTTT atccagggag aaaaataaaa
3960
ggaagccata gaattgctct ggtcaaaacc aagcacacca tagccttaac tgaatattta
4020
20 ggaaatctgc ctaatctgct tatatttggT gtttgTTTTt tgactgttgg gctttgggaa
4080
gatgttattt atgaccaata tctgccagta acgctgttta tctcacttgc tttgaaagcc
4140
25 aatggggggaa aaaaatccat gaaaaaaaaa agattgataa agtagatgat tttgTTTgta
4200
30 tccctaccca tctcctggca gccctactga gtgaaattgg gatacatttg gctgtcagaa
4260
attataccga gtctactggg tataacatgt ctacttTga aagctagtac ttttaaattg
4320
35 gtgccaaagg tcaactgtaa tgagataatt atccctgcct gtgtccatgt cagactTTTga
4380
gctgatcctg aataataaag ctttttacct taaaaaaaaa aaaaaaaa
4428
40

<210> 2

<211> 3916

45 <212> DNA

<213> *Homo sapiens*

<400> 2

50 agagtgctga tttagaagaa tacaatcat ggctgaaaat agtgtattaa catccactac
60
tgggaggact agcttggcag actcttccat ttttgattct aaagttactg agatttccaa
55 120
ggaaaactta cttattggat ctacttcata tgtagaagaa gagatgcctc agattgaaac
180
60 aagagtgata ttggttcaag aagctggaaa acaagaagaa cttataaaag ccttaaagga
240
cattaaagtg ggctttgtaa agatggagtc agtTgaagaa tttgaaggtt tggattctcc
300
65 ggaattTgaa aatgtatttg tagtcacgga cttcaggat tctgtcttta atgacctcta
360

ES 2 343 996 B1

caaggctgat tgtagagtta ttggaccacc agttgtatta aattgttcac aaaaaggaga
420

5 gcctttgccca ttttcatgtc gcccgttgta ttgtacaagt atgatgaatc tagtactatg
480

ctttactgga tttaggaaaa aagaagaact agtcaggttg gtgacattgg tccatcacat
540

10 ggggtggagtt attcgaaaag actttaattc aaaagttaca catttgggtgg caaattgtac
600

acaaggagaa aaattcaggg ttgctgtgag tctaggtact ccaattatga agccagaatg
660

15 gatttataaa gcttgggaaa ggcggaatga acaggatttc tatgcagcag ttgatgactt
720

tagaaatgaa tttaaagttc ctccatttca agattgtatt ttaagtttcc tgggattttc
20 780

agatgaagag aaaaccaata tggaagaaat gactgaaatg caaggaggta aatatttacc
840

25 gcttggagat gaaagatgca ctcaccttgt agttgaagag aatatagtaa aagatcttcc
900

ctttgaacct tcaaagaaac tttatgttgt caagcaagag tggttctggg gaagcattca
960

30 aatggatgcc cgagctggag aaactatgta tttatatgaa aaggcaaata ctctgagct
1020

caagaaatca gtgtcaatgc tttctctaaa taccctaac agcaatcgca aacgacgtcg
35 1080

tttaaaagaa acacttgctc agctttcaag agagacagac gtgtcaccat ttccaccccg
1140

40 taagcgccca tcagctgagc attccctttc cataggggtca ctctagata tctccaacac
1200

accagagtct agcattaact atggagacac cccaaagtct tgtactaagt cttctaaaag
1260

45 ctccactcca gttccttcaa agcagtcagc aagggtggcaa gttgcaaaag agctttatca
1320

aactgaaagt aattatgtta atatattggc aacaattatt cagttatttc aagtaccatt
50 1380

ggaagaggaa ggacaacgtg gtggacctat ccttgcacca gaggagatta agactatttt
1440

55 tggtagcatc ccagatatct ttgatgtaca cactaagata aaggatgatc ttgaagacct
1500

tatagttaat tgggatgaga gcaaaagcat tggtgacatt tttctgaaat attcaaaaga
1560

60 tttggtaaaa acctaccctc cctttgtaaa cttctttgaa atgagcaagg aaacaattat
1620

taaatgtgaa aaacagaaac caagatttca tgcttttctc aagataaacc aagcaaaacc
65 1680

ES 2 343 996 B1

agaatgtgga cggcagagcc ttgttgaact tcttatccga ccagtacaga ggttacccag
1740

5 tgttgcatta cttttaaagt atcttaagaa gcatacagct gatgaaaatc cagacaaaag
1800

cactttagaa aaagctattg gatcactgaa ggaagtaatg acgcatatta atgaggataa
1860

10 gagaaaaaca gaagctcaaa agcaaatttt tgatgttggt tatgaagtag atggatgccc
1920

agctaattctt ttatcttctc accgaagctt agtacagcgg gttgaaacaa tttctctagg
1980

15 tgagcacccc tgtgacagag gagaacaagt aactctcttc ctcttcaatg attgcctaga
2040

gatagcaaga aaacggcaca aggttattgg cacttttagg agtcctcatg gccaaacccg
2100

acccccagct tctcttaagc atattcacct aatgcctctt tctcagatta agaaggtatt
2160

25 ggacataaga gagacagaag attgccataa tgcttttgcc ttgcttgtga ggccaccaac
2220

agagcaggca aatgtgctac tcagtttcca gatgacatca gatgaacttc caaaagaaaa
2280

30 ctggcetaag atgctgtgtc gacatgtagc taacaccatt tgtaaagcag atgctgagaa
2340

tcttatttat actgctgatc cagaatcctt tgaagtaa atcaaaagata tggacagtac
2400

35 attgagtaga gcatcaagag caataaaaaa gacttcaaaa aaggttacaa gagcattctc
2460

40 tttctccaaa actccaaaaa gagctcttcg aagggtcttt atgacatccc acggctcagt
2520

ggagggaaga agtccttcca gcaatgataa gcatgtaatg agtcgtcttt ctagcacatc
2580

45 atcattagca ggtatccctt ctccctccct tgtcagcctt ccttccttct ttgaaaggag
2640

aagtcatacg ttaagtagat ctacaactca tttgatatga agcgttacca aaatcttaaa
2700

50 ttatagaaat gtatagacac ctcatactca aataagaaac tgacttaaat ggtacttgta
2760

attagcactt ggtgaaagct ggaaggaaga taaataacac taaactatgc tatttgattt
2820

55 ttcttcttga aagagtaagg tttacctggt acattttcaa gtttaattcat gtaaaaaatg
2880

60 atagtgattt tgatgtaatt tatctcttgt ttgaatctgt cattcaaagg ccaataattt
2940

aagttgctat cagctgatat tagtagcttt gcaaccctga tagagtaa ataaattttatg
3000

65

ES 2 343 996 B1

ggcgggtgcc aaatactgct gtgaatctat ttgtatagta tccatgaatg aatttatgga
3060
5 aatagatatt tgtgcagctc aatttatgca gagattaaat gacatcataa tactggatga
3120
aaacttgcat agaattctga ttaaatagtg ggtctgtttc acatgtgcag tttgaagtat
3180
10 ttaaataaacc actcctttca cagtttattt tcttctcaag cgttttcaag atctagcatg
3240
tggattttaa aagatttgcc ctcatataca agaataacat ttaaaggaga ttgtttcaaa
3300
15 atatTTTTGC aaattgagat aaggacagaa agattgagaa acattgtata ttttgcaaaa
3360
acaagatggt tntagctggt tcagagagag tacggtatat ttatggtaat tttatccact
20 3420
agcaaatctt gatttagttt gatagtgtgt ggaattttat tttgaaggat aagaccatgg
3480
25 gaaaattgtg gtaaagactg tttgtaccct tcatgaaata attctgaagt tgccatcagt
3540
tttactaatc ttctgtgaaa tgcatagata tgcgcatggt caacttttta ttgtggtcct
3600
30 ataattaaat gtaaaattga aaattcattt gctgtttcaa agtgtgatat ctttcacaat
3660
agccttttta tagtcagtaa ttcagaataa tcaagttcat atggataaat gcatttttat
3720
35 ttcctatttc tttagggagt gctacaaatg tttgtcactt aaatttcaag tttctgtttt
3780
aatagttaac tgactataga ttgttttcta tgccatgtat gtgccacttc tgagagtagt
40 3840
aaatgactct ttgctacatt ttaaaagcaa ttgtattagt aagaactttg taaataaata
3900
45 cctaaaaccc aagtgt
3916

<210> 3

<211> 3729

<212> DNA

<213> *Homo sapiens*

55 <400> 3

ggcccaagcc gccattgttc cgccgagggg ggacagcggg gcctggcgct ggcgccgaga
60
cgccgcttag cggccgccac tggagacact ccctcccgcc tcccgggtct cctggcggcg
120
gcgagtgag gctgacagcg gggaacctgg gagacccttc cgccctcccc gcggtggcag
180
65 cgccgatcc ccggctccgg cgcgagggac ggccgcgatg cgctcggcct gaggttacc

ES 2 343 996 B1

240

5 ggccccggccc ttcctcgctt ccctcgacta ttccactgcg tctccgcgcc ccggcgatc
300

cctgcgagtc cctctgacgg gaggaagat ggctgcacgg agctggcagg acgagctggc
360

10 ccagcaggcc gaggagggt cggccccggt gcgggaaatg ctctcggtcg gcctaggctt
420

tctgcgacc gagctgggccc tcgacctggg gctggagccg aaacggtacc ccggctgggt
480

15 gatcctgggtg ggcaactggcg cgctcgggct gctgctgctg tttctgctgg gctacggctg
540

ggccgcggct tgcgccggct cccgcaaaaa gcggaggagc ccgccccgca agcgggagga
600

20 ggccggcggcc gtgccggccg cggccccga cgacctggcc ttgctgaaga atctccggag
660

cgaggaacag aagaagaaga accggaagaa actgtccgag aagcccaaac caaatgggcg
720

gactgtttaa gtggctgagg gtgaagctgt tcgaacacct caaagtgtaa cagcaaagca
780

30 gccaccagag attgacaaga aaaatgaaaa gtcaaagaaa aataagaaga aatcaaagtc
840

agatgctaaa gcagtgcaaa acagttcacg ccatgatgga aaggaagttg atgaaggagc
900

35 ctgggaaact aaaattagtc acagagagaa acgacagcag cgtaaactg ataaggtgct
960

gactgattct ggttcattgg attcaactat ccctgggata gaaaatacca tcacagttac
1020

caccgagcaa cttacaaccg catcatttcc tgttggttcc aagaagaata aaggtgattc
1080

45 tcatctaaat gttcaagtta gcaactttaa atctggaaaa ggagattcta cacttcaggt
1140

ttcttcagga ttgaatgaaa acctcactgt caatggagga ggctggaatg aaaagtctgt
1200

50 aaaactctcc tcacagatca gtgcaggtga ggagaagtgg aactccgttt cacctgcttc
1260

tgcaggaaag aggaaagctg agccatctgc ctggagtcaa gacactggag atgctaatac
1320

55 aaatggaaaa gactggggaa ggagttggag tgaccgttca atatcttctg gcattgggtc
1380

tactgctgag ccagtttctc agtctaccac ttctgattat cagtgggatg ttagccgtaa
1440

60 tcaaccctat atcgatgatg aatggctctgg gttaaagtgt ctgtcttctg ctgatcccaa
1500

65 ctctgattgg aatgcaccag cagaagagtg gggcaattgg gtagacgaag aaagagcttc

ES 2 343 996 B1

1560

5 acttctaaag tcccaggaac caattcctga tgatcaaaag gtctcagatg atgataaaga
1620

aaagggagag ggagctcttc caactgggaa atccaaaaag aaaaaaaga aaaagaagaa
1680

10 gcaaggtgaa gataactcta ctgcacagga cacagaagaa ttagaaaaag agattagaga
1740

agaccttcca gtgaatacct ctaaaacccg tccaaaacag gaaaaagctt tttccttgaa
1800

15 gaccataagc actagtgatc cagccgaagt actcgtcaaa aatagccagc ctatcaagac
1860

tcttccacct gctacttcta ccgagccatc tgtaatctta tcaaaaagtg attctgacaa
1920

20 gagctcttcc caagtgccgc caatactaca agagacagat aaatccaagt caaataccaa
1980

gcaaaatagt gtgcctcctt cacagaccaa gtctgaaact agctgggaat ctcccaaaca
2040

25 aataaaaaag aagaaaaaag ccagacgaga aacgtgaaat tttttttcct gaattggaca
2100

30 tgtgtttgca aacacttgtc ttgaagatta tgctgtttat gcaataatth gtgaacatgt
2160

acagagtttt atataaatth aaaccaatth ttaaaacaaa actgcggaca ccaccataaa
2220

35 aatggaatca aaagaaagtth aatthtatgaa attaaagaggt cagcagaata tactcagtga
2280

tggaagacac ttgggaaagt ctttttaata gaacaagaac gatcttaatt taagaatatt
2340

40 atcctggttt aacaacagtg cctgttttac aacagattgt gccctatctc atctgcagcc
2400

45 gaggaataaa ggattctgat tagaaagagg gttgcctaca gattagtaag caattccttg
2460

gatcttatgc acagaacttg taccatttga atctgtttta tgcttaaadc aaagtgcctt
2520

50 gatcaaatgc ataacctgcc atatctttac atatthgttg gtagcaatth gtattaaaga
2580

aatcacaagt gcaataaaaa agtcatttat catttgttta actaaactgt catggtttag
2640

55 tttacaatth ttaaaaagtth cttaaatac tgaaaatgca gttgacactt gtgtatggct
2700

60 tatgaagtta tttttgatag tcttacatta cttgaattgt tcaaagtaca gtatatttha
2760

aattaagaaa agtgaactat atgtatthgt tttatacatt taaggcttag actcataaat
2820

65 aatgctattg tttatgattt gaaaactthc aggcaaaatc caatthacat tttcccttc

ES 2 343 996 B1

2880

5 cctagcaatt acttttttcc agcttcaact cttcttagtt actaatactt tgttgacttt
2940

aaaaatgaaa tcattcacia acttttgga tatgatggag aatgaaaaac tagagtcaga
3000

10 cagctttaat tgacattgtc aacacctcca gttatcagga atacattttt ttactgcctt
3060

aacctgtagt gcgtagaata tgcatcaatt tcttgaagga gattcatggt tttataagaa
3120

15 ttttcatgta attattgcaa ttgtggtaaa ataaggaacg tttcctgctt gaaattatat
3180

tgatttaaat gatgtgtgag atgtttcacc attttcaggc actgtgtaat tctattgtaa
3240

20 taaactggca ggtatctttg taactataaa tagtgcattg tcagccatgt aactgtaaa
3300

tagcctttac caaacgtggt tgacaaggac cataattaac atcacttagt gaattgtgat
3360

25 aaagaaaaaa aagccatgat ttattcgatg tgattggctt gtttttatgt ggcgccaaga
3420

30 acgaacctgt ttaacagctg taaccaatgg tactgatcta tccatccaat gttgtcatta
3480

tatttgactg tggttcaaca gtattgcggt gtcagactag gaaagctaaa cgaacaaaat
3540

35 ggtttttagtt ttgctgaaga ctggccttat taatggacag ctttcctaac aagagattat
3600

taacttttat cagggtgtaa catctgtttc aggaacatgg cagtatgttt acatgtcaga
3660

40 agttttgttt aattctatgg tatttctaaa ttgacttggt taataaatt cagcaaaaaa
3720

45 aaaaaaaaa
3729

<210> 4

<211> 3128

50 <212> DNA

<213> *Homo sapiens*

<400> 4

55 gcttcgcccc gtggcgcggt ttgaaatfff gcggggctca acggctcgcg gagcggctac
60

gctggagtgc atcgccggtg tttgctgggtg gttgttgctc tcggggccgt gtggagttagg
60 120

tctggacctg gactcacggc tgcttgagc gtccgcatg aggagaagtg aggtgctggc
180

65 ggaggagtcc atagtatgtc tgcagaaagc cctaaatcac cttcgggaaa tatgggagct
240

ES 2 343 996 B1

aattgggatt ccagaggacc agcggttaca aagaactgag gtggtaaaga agcatatcaa
300

5 ggaactcctg gatatgatga ttgctgaaga ggaaagcctg aaggaaagac tcatcaaaag
360

catatccgtc tgtcagaaag agctgaacac tctgtgcagc gagttacatg ttgagccatt
420

10 tcaggaagaa ggagagacga ccatcttgca actagaaaaa gatttgcgca cccaagtgga
480

attgatgcga aaacagaaaa aggagagaaa acaggaactg aagctacttc aagagcaaga
540

15 tcaagaactg tgcgaaattc tttgtatgcc ccactatgat attgacagtg cctcagtgcc
600

cagcttagaa gagctgaacc agttcaggca acatgtgaca actttgaggg aaacaaaggc
20 660

ttctagcgt gaggagtttg tcagtataaa gagacagatc atactgtgta tggagaatt
720

25 agaccacacc ccagacacaa gctttgaaag agatgtggtg tgtgaagacg aagatgcctt
780

ttgtttgtct ttggagaata ttgcaacact acaaaagttg ctacggcagc tggaaatgca
840

30 gaaatcacia aatgaagcag tgtgtgaggg gctgctgact caaatccgag agctctggga
900

caggttgcaa atacctgaag aagaaagaga agctgtggcc accattatgt ctgggtcaaa
960

35 ggccaaggtc cggaaagcgc tgcaattaga agtggatcgg ttggaagaac tgaaaatgca
1020

aaacatgaag aaagtgattg aggcaattcg agtggagctg gttcagtact gggaccagtg
40 1080

cttttatagc caggagcaga gacaagcttt tgcccctttc tgtgctgagg actacacaga
1140

45 aagtctgctc cagctccacg atgctgagat tgtgcggtta aaaaactact atgaagttca
1200

caaggaactc tttgaagggtg tccagaagtg ggaagaaacc tggaggcttt tcttagagtt
1260

50 tgagagaaaa gcttcagatc caaatcgatt taaaaccga ggaggaaatc ttctaaaaga
1320

agaaaaacia cgagccaagc tccagaaaat gctgccaag ctggaagaag agttgaaggc
55 1380

acgaattgaa ttgtgggaac aggaacattc aaaggcattt atggtgaatg ggcagaaatt
1440

60 catggagtat gtggcagaac aatgggagat gcatcgattg gagaaagaga gagccaagca
1500

ggaaagacia ctgaagaaca aaaaacagac agagacagag atgctgtatg gcagcgctcc
1560

65

ES 2 343 996 B1

tcgaacacct agcaagcggc gaggactggc tccaataaca ccgggcaaag cacgtaagct
 1620
 5 gaacactacc accatgtcca atgctacggc caatagtagc attcggccta tctttggagg
 1680
 gacagtctac cactcccccg tgtctcgact tcctccttct ggcagcaagc cagtcgctgc
 1740
 10 ttccacctgt tcaggggaaga aaacaccccc tactggcagg catggagcca acaaggagaa
 1800
 cctggagctc aacggcagca tcctgagtgg tgggtaccct ggctcggccc ccctccagcg
 1860
 15 caacttcagc attaattctg ttgccagcac ctattctgag tttgcgaagg atccgtcctt
 1920
 ctctgacagt tccactgttg ggcttcagcg agaactttca aaggcttcca aatctgatgc
 1980
 20 tactttctga atcctcaatt caaccaacat ccagtcctga gaagccctga tcagtcaacc
 2040
 agctgtggct tcctgtgcct agactggacc taattatatg ggggtgactt tagtttttct
 2100
 25 tcagcttagg cgtgcttgaa acctggcca ggttccatga ccatgggctt aacttaaaga
 2160
 30 tgtgaatgag tgttacagtt gaaagccat cataggttta gtggtcctag gagacttggt
 2220
 tttgacttat atacatgaaa agtttatggc aagaagtgca aattttagca tatggggcct
 2280
 35 gacttctcta ccacataatt ctacttgctg aagcatgatc aaagcttggt ttatttcacc
 2340
 actgtaggaa aatgattgac tatgcccac cctgggggta attttggcat gtatacctgt
 2400
 40 aactagtaat taacatcttt tttgtttagg catgttcaat taatgctgta gctatcatag
 2460
 45 ctttgctctt acctgaagcc ttgtccccac cacacaggac agccttcctc ctgaagagaa
 2520
 50 tgtctttgtg tgtccgaagt tgagatggcc tgccctactg ccaagaggt gacaggaagg
 2580
 ctgggagcag ctttgttaaa ttgtgttcag ttctgttaca cagtgcattg ccctttgttg
 2640
 55 ggggtatgca tgtatgaaca cacatgcttg tcggaacgct ttctcggcgt ttgtcccttg
 2700
 gctctcatct cccccattcc tgtgcctact ttgcctgagt tcttctaccc ccgcagttgc
 2760
 60 cagccacatt gggagtctgt ttgttccaat gggttgagct gtctttgtcg tggagatctg
 2820
 gaactttgca catgtcacta ctggggaggt gttcctgctc tagcttccac gatgaggcgc
 2880
 65

ES 2 343 996 B1

cctctttacc tatectctca atcactactc ttcttgaagc actattatntt attcttccgc
2940

5 tgtctgcctg cagcagtact actgtcaaca tagtgtaaant ggttctcaaa agcttaccag
3000

tgtggacttg gtgttagcca cgctgtttac tcatacagta cgtgtcctgt ttttaaaata
3060

10 tacaattatt cttaaaaata aattaaaatc tgtatactta catttcaaaa agaaaaaaa
3120

aaaaaaa
3128

15 <210> 5
<211> 1486
<212> DNA

20 <213> *Homo sapiens*

<400> 5

25 cgcgctcacg tctgaagtgg gagcaatgca ccgggacagg gacacctcct aggccatgcc
60

tgttccagtc cagttctgcc tgaaagtccg gctggctcat cacctgccta aataaaaccg
120

30 tatacgggca aactccctcc gcaagcagcg cgccccagca ccggaagtga cgcggttacgt
180

gcccgcgtat tcctaccggc gtattcccgc cctgcttttc gcccgccggtt ccgtggcggg
240

35 aactgaggcg actgtgggga catcagtgat cgggtgaagta ccatgcaagc atttcttaaa
300

40 ggtacatcca tcagtactaa acccccgctg accaaggatc gaggagtagc tgccagtgcg
360

ggaagtagcg gagagaacaa gaaagccaaa cccgttcctt ggggtgaaaa atatcgccca
420

45 aaatgtgtgg atgaagttgc tttccaggaa gaagtggttg cagtgtctgaa aaaatcttta
480

gaaggagcag atcttcttaa tctcttgttt tacggaccac ctggaactgg aaaaacatcc
540

50 actatntttg cagcagctag agaactcttt gggcctgaac ttttccgatt aagagttctt
600

gagttaaattg catctgatga acgtggaata caagtagttc gagagaaagt gaaaaatntt
660

55 gctcaattaa ctgtgtcagg aagtcgctca gatgggaagc cgtgtccgcc tnttaagatt
720

60 gtgattctgg atgaagcaga ttctatgacc tcagctgctc aggcagcttt aagacgtacc
780

atggagaagg agtcgaaaac cacccgattc tgtcttatct gtaactatgt cagtcgaata
840

65 attgaacccc tgacctctag atgttcaaaa ttccgcttca agcctctgtc agataaaatt
900

ES 2 343 996 B1

caacagcagc gattactaga cattgccaaag aaggaaaatg tcaaaattag tgatgagggg
960

5 atagcttatac ttgttaaagt gtcagaagga gacttaagaa aagccattac atttcttcaa
1020

agcgctactc gattaacagg tggaaaggag atcacagaga aagtgattac agacattgcc
1080

10 ggggtaatac cagctgagaa aattgatgga gtatttgctg cctgtcagag tggctctttt
1140

gacaaactag aagctgtggt caaggattta atagatgagg gtcatgcagc aactcagctc
1200

15 gtcaatcaac tccatgatgt ggtttagaaa aataacttat ctgataaaca gaagtctatt
1260

atcacagaaa aacttgccga agttgacaaa tgcctagcag atgggtgctga tgaacatttg
20 1320

caactcatca gcctttgtgc aactgtgatg cagcagttat ctcagaattg ttaacgtgaa
1380

25 tataatctgga tgggggggtt tgtaaataat gaagttgtaa taaaaataaa atgacccaaa
1440

gcacctttaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaa
1486

30 <210> 6
<211> 3737
<212> DNA
35 <213> *Homo sapiens*

<400> 6

40 ggcgctccgc cacacctccc cgcgcccgcg ccgccaccgc ccgcaactccg ccgcctctgc
60

ccgcaaccgc tgagccatcc atgggggctg cgggccgcaa ccgtcccggg gcggcctggg
120

45 cggtgctgct gctgctgctg ctgctgccgc cactgctgct gctggcgggg gccgtccccg
180

cgggtcgggg ccgtgccgcg gggccgcagg aggatgtaga tgagtgtgcc caagggctag
240

50 atgactgcca tgccgacgcc ctgtgtcaga acacaccac ctctacaag tgctcctgca
300

agcctggcta ccaaggggaa ggcaggcagt gtgaggacat cgatgaatgt ggaaatgagc
360

55 tcaatggagg ctgtgtccat gactgtttga atattccagg caattatcgt tgcacttgtt
420

ttgatggctt catgttggct catgacggtc ataattgtct tgatgtggac gagtgcctgg
60 480

agaacaatgg cggctgccag catacctgtg tcaacgtcat ggggagctat gagtgcctgct
540

65 gcaaggaggg gtttttcctg agtgacaatc agcacacctg cattcaccgc tcggaagagg

ES 2 343 996 B1

600

gcctgagctg catgaataag gatcacggct gtagtcacat ctgcaaggag gccccaaagg
660
5 gcagcgtcgc ctgtgagtgc aggcctgggt ttgagctggc caagaaccag agagactgca
720
10 tcttgacctg taaccatggg aacgggtgggt gccagcactc ctgtgacgat acagccgatg
780
gcccagagtg cagctgccat ccacagtaca agatgcacac agatgggagg agctgccttg
840
15 agcgagagga cactgtcctg gaggtgacag agagcaacac cacatcagtg gtggatgggg
900
ataaacgggt gaaacggcgg ctgctcatgg aaacgtgtgc tgtcaacaat ggaggctgtg
960
20 accgcacctg taaggatact tcgacaggtg tccactgcag ttgtcctggt ggattcactc
1020
tccagttgga tgggaagaca tgtaaagata ttgatgagtg ccagacccgc aatggagggt
1080
25 gtgatcattt ctgcaaaaac atcgtgggca gttttgactg cggctgcaag aaaggattta
1140
aattattaac agatgagaag tcttgccaag atgtggatga gtgctctttg gataggacct
1200
30 gtgaccacag ctgcatcaac caccctggca catttgcttg tgcttgcaac cgagggtaca
1260
35 ccctgtatgg cttcaccac tgtggagaca ccaatgagtg cagcatcaac aacggaggct
1320
gtcagcaggt ctgtgtgaac acagtgggca gctatgaatg ccagtgccac cctgggtaca
1380
40 agctccactg gaataaaaaa gactgtgtgg aagtgaaggg gctcctgccc acaagtgtgt
1440
caccccgtgt gtccctgcac tgcggaaga gtgggtggagg agacgggtgc ttcctcagat
1500
45 gtcactctgg cattcacctc tcttcagatg tcaccacat caggacaagt gtaacctta
1560
50 agctaaatga aggcaagtgt agtttgaaaa atgctgagct gtttcccag ggtctgagc
1620
cagcactacc agagaagcac agctcagtaa aagagagctt ccgctacgta aaccttacat
1680
55 gcagctctgg caagcaagtc ccaggagccc ctggccgacc aagcaccctt aaggaaatgt
1740
ttatcactgt tgagtttgag cttgaaacta accaaaagga ggtgacagct tcttgtgacc
1800
60 tgagctgcat cgtaaagcga accgagaagc ggctccgtaa agccatccgc acgctcagaa
1860
65 aggccgtcca cagggagcag tttcacctcc agctctcagg catgaacctc gacgtggcta

ES 2 343 996 B1

1920

5 aaaagcctcc cagaacatct gaacgccagg cagagtcctg tggagtgggc cagggtcagt
1980

10 cagaaaacca atgtgtcagt tgcagggctg ggacctatta tgatggagca cgagaacgct
2040

15 gcattttatg tccaaatgga accttccaaa atgaggaagg acaaatgact tgtgaacat
2100

20 gcccaagacc aggaaattct ggggccctga agaccccaga agcttggaaat atgtctgaat
2160

25 gtggaggctt gtgtcaacct ggtgaatatt ctgcagatgg ctttgcacct tgccagctct
2220

30 gtgccctggg cacgttccag cctgaagctg gtcgaacttc ctgcttcccc tgtggaggag
2280

35 gccttgccac caaacatcag ggagctactt cctttcagga ctgtgaaacc agagttcaat
2340

40 gttcacctgg acatttctac aacaccacca ctcaccgatg tattcgttgc ccagtgggaa
2400

45 cataccagcc tgaatttgga aaaaataatt gtgtttcttg cccaggaaat actacgactg
2460

50 actttgatgg ctccacaaac ataaccagcgt gtaaaaacag aagatgtgga ggggagctgg
2520

55 gagatttcac tgggtacatt gaatcccca actaccagcag caattacca gccaacaccg
2580

60 agtgtacgtg gaccatcaac ccacccccca agcgccgcat cctgatcgtg gtccctgaga
2640

65 tcttctgcc catagaggac gactgtgggg actatctggt gatgcggaat acctcttcat
2700

70 ccaattctgt gacaacatat gaaacctgcc agacctacga acgccccatc gccttcacct
2760

75 ccaggtcaaa gaagctgtgg attcagttca agtccaatga agggaacagc gctagagggt
2820

80 tccaggtccc atacgtgaca tatgatgagg actaccagga actcattgaa gacatagttc
2880

85 gagatggcag gctctatgca tctgagaacc atcaggaaat acttaaggat aagaaactta
2940

90 tcaaggctct gtttgatgtc ctggcccatc ccagaacta tttcaagtac acagcccagg
3000

95 agtcccgaga gatgtttcca agatcgttca tccgattgct acgttccaaa gtgtccaggt
3060

100 ttttgagacc ttacaaatga ctcagcccac gtgccactca atacaaatgt tctgctatag
3120

105 ggttgggtggg acagagctgt cttccttctg catgtcagca cagtcgggta ttgctgcctc
3180

110 ccgtatcagt gactcattag agttcaattt ttatagataa tacagatatt ttggtaaatt

ES 2 343 996 B1

3240

gaacttggtt tttctttccc agcatcgtgg atgtagactg agaatggctt tgagtggcat
3300

5

cagcttctca ctgctgtggg cggatgtctt ggatagatca cgggctggct gagctggact
3360

10

ttggtcagcc taggtgagac tcacctgtcc ttctggggtc ttactcctcc tcaaggagtc
3420

tgtagtggaa aggaggccac agaataagct gcttattctg aaacttcagc ttcctctagc
3480

15

ccggccctct ctaagggagc cctctgcact cgtgtgcagg ctctgaccag gcagaacagg
3540

20

caagagggga ggggaaggaga cccctgcagg ctccctccac ccaccttgag acctgggagg
3600

actcagtttc tccacagcct tctccagcct gtgtgataca agtttgatcc caggaacttg
3660

25

agttctaagc agtgctcgtg aaaaaaaaaa gcagaaagaa ttagaaataa ataaaaacta
3720

agcacttctg gagacat
3737

30

<210> 7

<211> 3224

<212> DNA

<213> *Homo sapiens*

35

<400> 7

gtccacatc ccgcatccgg catcccagcg gccgggcatg tagcagcggc agcaacggcg
60

40

gaatatgggc gggaaaccact cccacaagcc ccccggtgtt gacgagaatg aggaagtcaa
120

45

ctttgaccat tttcagattc tgcgggcat tggtaaagg agttttggaa aggtatgcat
180

cgtgcagaag cgagacacta agaaaatgta tgcaatgaag tacatgaaca agcagaagtg
240

50

catcgagagg gatgaggttc ggaatgtttt ccgggagctg cagatcatgc aagggctgga
300

55

gcaccccttc ctggtcaatc tgtggtactc ctccaggat gaggaggaca tgttcatggt
360

ggtggacctg ctcttgggag gcgacctgcg ctaccatctg cagcagaatg tgcatttcac
420

60

agaggggact gtgaaactct acatctgtga gctggcactg gccctggagt atcttcagag
480

gtaccacatc atccacagag acatcaagcc agacaatatc ctgctggatg aacacggaca
540

65

tgttcacatt acagacttca acatagcgac ggtagtgaaa ggagcagaaa gggcttcctc
600

ES 2 343 996 B1

catggctggc accaagccct acatggctcc agaagtattc caggtgtaca tggacagagg
 660
 5 ccccgatac tcgtaccctg tcgactgggtg gtccctgggc atcacagcct atgagctgct
 720
 gcggggctgg aggccgtacg aaatccactc ggtcacgccc atcgatgaaa tcctcaacat
 780
 10 gttcaagggtg gagcgtgtcc actactcctc cacgtgggtgc aaggggatgg tggccctgct
 840
 gaggaagctc ctgaccaagg atcctgagag ccgctgtgtcc agccttcatg acatacagag
 900
 15 cgtgccctac ttggccgaca tgaactggga cgcggtgttc aagaaggcac tgatgcccg
 960
 ctttgtgccc aataaaggga ggttgaactg cgatcccaca tttgagcttg aagagatgat
 1020
 20 tctagaatcc aagccacttc acaaaaagaa gaagcgattg gcaaagaaca gatccagga
 1080
 tggcacaag gacagctgcc cgctgaatgg acacctgcag cactgtttgg agactgtccg
 1140
 25 ggaggaattc atcatattca acagagagaa gctcaggagg cagcaggac agggcagcca
 1200
 gctcttggac accgacagcc gagggggagg ccaggcccaa agcaagctcc aggacgggtg
 1260
 30 caacaacaac ctctcacc acacctgcac ccgtggctgc agcagctgag cccacacttg
 1320
 35 ttgtgctca acaggactgc actcgtctct gccctgccca cccagagccc ctctttgtgc
 1380
 cctgatggtc cctgtctcac ccctgaaaac atcagatgca gaaaagccc tggacttga
 1440
 40 gctgggaagc ctgggttctg gtccatctc catgactgat tcacgtgtga cctcagacaa
 1500
 45 gtcacgccct ctctgtgcct ccgttttctg catctgcaa aggggttaa cacttctgcc
 1560
 ccacttcaaa ttacaagatt atggggagaa cccaattagg taggaaacat gaaaaacctt
 1620
 50 tgatatttat aaaatcattt ttacgtgcaa aatataacct taatatttga agtgaccccc
 1680
 attccccaaa gcaatcaaac cgtcatgact ttgcaatttg gcacatccta gcttgttaga
 1740
 55 gggcacttcc gaaaaacaca gccctgacag caaaataaag gtctgatatg ttggcccctt
 1800
 60 ctatggaaac aacgctgcca aatcctggag caaacctga agtgtcttca tgtgcattct
 1860
 ctggcaggcc acagtcctga gcttgtaaga tgggtcagca tgcagaccag acttgtcccc
 1920
 65

ES 2 343 996 B1

aaggctctcag cgctgcggtc tcaactcctcc cctcatttaa gaagactatc cttacctttt
1980

5 agtttcagca gtcctcacca ccaccatata cccagtgctg ggatggcaca caggtgtcca
2040

ttcagatgag agttgggtcg ctgagcattg gttactcctg cagagtgtaa tcagcacccc
2100

10 atccaactgg cccgaaagcc cagacctgca gcagaactct ccaactctct atcagctttc
2160

agggttttct ctcttgggaa ggggtgtaaaa tcagcttgct agattcttct tacagagagt
2220

15 atccaatcgg tattggtgga gcggctcctt attatacaa taggaagcat ggggtccttag
2280

aaagtttatt tcaggaggaa aatgggttca cacaaaaagc aaactacatt ctgatctgct
20 2340

cagggagaag cttgcctttg aactggaaga tgttgggatg agcagggaaa gcttagactt
2400

25 tggagtcagg tttgtgttca gaatccagcc ctgctggcta ctaactaact gggagacctt
2460

aggcaaagca tgcaatcgct ctgaatggca gtttctcat ttttaaacag ggataataaa
2520

30 actaatattg caggggagtt acaggggttaa ataagatcct gtgtgtaacc ccaagcattg
2580

gatgactcat agaatggcct tttttgtcag cataatcgtc atcattattt agatactttc
35 2640

ttccttcaact caccagcag gtcagttttc tgtgcaaaca aacctgttta ggattcttcc
2700

40 aaatgttctt cctggggctt ttgatatttg tttgttacat cctgctgaag ttcgactgtg
2760

tttttatttt ttcacccaac ttccattttt cactttttac atgattactc aatccttggg
2820

45 gctgtccatg tcatctctta gatttcttaa aagacatttt aatgtatggt taggttttat
2880

atttttattt tttaaaaaag aaatagtcag tgttttcctc ctttcaaccg agactatttc
2940

50 tggattgtgt gtcctcgtc agttgacttg ttttgcacac ttttctttac ttcattgtccc
3000

catcaacaac cgtcctgctc cccacctccc ccaggaaata aggggcctgc tcctctcctt
55 3060

actgtgaccc tggaggctct taagatgatg atggtttttt ttattgggct gagttcacga
3120

60 attaggggca ggagctggaa gtcgccctag gaacaccaga tttcctgggt ctgttcaagt
3180

tggcatttct tgtttggaat aaactatttc ttggacattc cttc
3224

65 <210> 8
<211> 3320

ES 2 343 996 B1

<212> DNA

<213> *Homo sapiens*

5 <400> 8
gcacgtccaa accccgggggt cctcggccag gccagggggg cgggcgtgcg gcggggccgg
60
10 gccgggcccgg gcgcgcccag cgggctccgg gcacacgccc ctcacctggg cgaggcccgg
120
tgcgcaggcg ctcagcgcgg gggcagcggc cgggcgctta agcccggcgc gggcaagagt
180
15 tggacagttc aggggcccgg ccgggcccag ctgtcgcgccg cgccacctcc accgcccggc
240
ctgctaccgc caccgccacc gcctgggctg cagcagatcg cgggagaacg tctgcgcgaa
300
20 gccccggtgc agccctggcg ggaccgcagc aaccaccgga gcgtgagcaa cagcttcccg
360
gctggaggag ccaactgcccc acgcgcccgc ggcccgcgcg ccgagatggt ctacatttca
25 420
actccatcaa atattaaatt gatgtttatg ggctaacctg gaaaccgaaa aacagtttgt
480
30 ttactataca gaataagaaa atatgtccaa attcctgtca gctgacttat aaagggactt
540
ttagaacaca acccaatgca tgaggtactc cttaagccct gacaaccatc ttgaagatgg
35 600
aatcatgaat atggcaaatt ttctacgggg ctttgaagaa aaggggataa agaacgacag
660
gcctgaggac cagttgagca aagagaaaaa gaaaattctt ttctccttct gtgaggtgtg
40 720
caacatccag ctgaactctg cagcccaggc tcaggtgcat tccaacggca aatcccaccg
780
45 caaacgagtg aagcagctga gtgatgggca gccccacca cccgcccagg ccagccccag
840
cagcaacagc agcacaggca gtacatgcca cactactacc cttcctgctc ttgtgcgcac
900
50 acctaccctg atgatgcagc cttcactgga tatcaaacca tttatgtctt ttccagtgga
960
cagtatgtct gctgttgggc tctttccaaa ttttaacaca atggatcctg tgcaaaaagc
55 1020
ggttattaac catacatttg gagtatccat tccccaaag aagaaacaag ttatttcttg
1080
60 taatgtctgt cagcttcgct ttaactcaga tagccaggcc gaggcccact acaaaggaag
1140
taaacatgcc aagaaggcca aagcactaga cgcaacgaaa aataaaccca aaatggttcc
1200

65

ES 2 343 996 B1

ttccaaggac agcgcaaagg ctaatcccag ctgctccatc actccaatca caggcaacaa
 1260
 5 ctctgacaaa tcagaagata aaggggaagtt aaaagccagc agttccagtc agccatcaag
 1320
 ctctgaaagt ggctcatttc tcctcaaadc tggcacaaca cccctgccac ctggagcagc
 1380
 10 cacttctccc tccaagagca caaatggagc tcccgggtact gttgttgaat cagaagaaga
 1440
 aaaagccaaa aaattacttt attgttcact atgcaaagtg gctgtgaaact ccctgtcaca
 1500
 15 gctagaggca cacaacacag gatctaaaca caagaccatg gttgaagctc gtaatggggc
 1560
 tgggtccaatt aaatcctatc ctagacctgg atcaagatta aagatgcaga atggcagtaa
 20 1620
 ggggtcagga ctacagaaca agacatttca ttgtgaaatc tgtgatgttc atgttaattc
 1680
 25 agaaattcaa ctcaaacagc acatttctag ccgaaggcat aaagatcgag ttgcagggaa
 1740
 accactgaag ccaaaatata gcccttataa caaactccag cggagcccga gtattctagc
 1800
 30 agcaaaactt gcattccaga aagatatgat gaagcctttg gccccagcct tctgtcctc
 1860
 acctctcgca gcggcggcag ccgtgtcctc agcgctgtca ctcccacccc ggcctctgc
 35 1920
 ctgctcttc caggctccag ccattcctcc agctcttctg aggctgggc atgggcccac
 1980
 40 ccgcgccact cctgcctcca tcctctttgc tccgtactaa cgtctgcaaa cccaagacg
 2040
 tttgaggcca gtaggaagtt gagaaagaaa gccacaagga ttcagcaatt taagcagttt
 2100
 45 gtgttgagcgt acccccccatt cacccacaaa atgcagccca ccaccagtc cccaattcc
 2160
 agagagaaac acatcactag attgacgagg gcttttagag actgctccta taattgagga
 2220
 50 atcagcagca taggctttat ctccctactt cctccccac cctaccgcgt tcaatttgtg
 2280
 tatctcagat atttgatttc ttgagcatgt attggtcaga aagaaaaaga aagaaaacca
 2340
 55 ttctctatth ttccatgggt atgatctggc ttaaaaacaat atggaatatt ttctgacag
 2400
 60 acttgaaaac taggatcttc ctcatagctc tgtaaataac tctgggtatcc aactgggtgt
 2460
 attctagtgt ggaacaaaaa acagtgaaca caagtgtcta cttgtggata ctatcttacc
 2520
 65

ES 2 343 996 B1

aacggatcac aaagttctct ttttgggtga ctggtgttga cagggattgt gtactgtttc
2580

5 gaccaagta ctgctgtatc ctgtgtagta ctagatctgt atctcttgtc tgaattaaac
2640

atTTTTtagtt gctgtgtaaa tgtaggaag caaagtagct ttgaattttc tctgtaagag
2700

10 aataaaagat aaaataatgt atTTTcttta tttcacattt tatttggaga tatttaaag
2760

aaatagaaag ccatataaca tagctataat tactacctgt ttgctatTTT tgtttaactt
2820

15 atTTTTgtagc ttcctcactg gtttgtgttg ctaagcaaac gtatacaaac agaaaaagag
2880

20 cttcctaata tgcacatttt tgcacctctt gtgcattctg caagaagaca atgaatccat
2940

gtatttctat ataagtatct tcttcatttt taacctgttt tcatttghtaa tgatgctaca
3000

25 taattttgtg gctccctaata gcaataatgt acagaagata aaaaaattta taattgaaca
3060

atTTgtcttt ctgttcgctg aatatattgc aggtcatgct cccatgcccc ctttctaag
3120

30 ccgatatcaa caagactgga atgtttgtga tatcaggggc aagagttatc atagaacaac
3180

35 aaaatagagt ggactctttt agagcatcta tatctgcaac ctgtaaatgg taaaatgtct
3240

gtgtattttt ttaaattgtac cttttcaata aaagtacaaa aaataaaaaa aaaaaaaaaa
3300

40 aaaaaaaaaa aaaaaaaaaa
3320

<210> 9

<211> 5178

45 <212> DNA

<213> *Homo sapiens*

<400> 9

50 ggcgcggagt cagagtggcg cagcaagtgg ccgcaggtgg cgacggtggc ggggggtggg
60

gtgtgaggta atccaggggt cgcggaagag gaggctgaga gggTcaaaag aaaactaaag
55 120

ctgcagtccg gcctactgtt ccgggggccg cggagcccc acccggggag atggacctca
180

60 accggatcat ccaggcgctg aagggcacca tcgaccCGaa gttgCGgatt gcagccgaga
240

acgagctcaa ccagtcctac aagattatca atTTTgcccc cagtttactt cggattatag
300

65 tctctgacca tgtggaattc ccagTAcgac aggcagctgc cattTAcctg aagaacatgg

ES 2 343 996 B1

360

5 tgacacaata ctggccagat cgagaacctc caccaggaga agcaatattt ccattcaaca
420

10 ttcacgaaaa cgatcgccag caaatcgtg ataacattgt ggaaggaata attcgggtctc
480

15 cagatttagt gagagtccaa ttaacaatgt gtctccgtgc catcataaaa catgattttc
540

20 ctggtcactg gccaggagtg gtcgacaaga tagactatta cttgcaatca cagagcagtg
600

25 caagctggct tggcagttta ttatgcctgt atcaactggg gaagacatat gaatataaga
660

30 aagcagaaga gagagaacct cttataatag caatgcagat attcctgcct cgtattcagc
720

35 aacaaattgt tcagctcctt cctgattcct cctattattc tgtattactg cagaaacaaa
780

40 ttctgaaaat cttttatgca cttgttcagt atgcattgcc tcttcagcta gtgaataacc
840

45 aaaccatgac aacatggatg gagatcttcc gaactattat cgacaggacc gttcctcctg
900

50 agactctgca cattgatgag gatgatagac cagaactggg atgggtggaag tgtaagaagt
960

55 gggcactgca tattgtagct cggctctttg aacgatatgg aagcccagga aatgtcacia
1020

60 aagaatactt tgaattttct gaattctttt tgaaaaccta tgcagtgggc attcagcagg
1080

65 tgctactaaa aatttttagat caatatagac agaaagaata tgtagctccc cgtgttcttc
1140

70 agcaagcatt caactatctc aaccaagggg tggttcattc tataacctgg aagcagatga
1200

75 agccacacat acagaatatc tctgaagatg tgattttttc tgtgatgtgt tataaagatg
1260

80 aggatgaaga gctgtggcaa gaagatccat atgagtatat aaggatgaaa tttgatattt
1320

85 ttgaagatta tgcttctccc accacagcag cccagactct cttatatact gctgcaaaga
1380

90 aaagaaaaga ggtgttgcca aaaatgatgg cattctgtta tcaaactctg acagacccga
1440

95 actttgacct taggaagaaa gatggagccc tgcattgtgat tggttcccta gctgagattt
1500

100 tactgaagaa gagtttattc aaggaccaa tggagctggt tctacaaaat catgtatttc
1560

105 cattattatt gtctaacctg ggatatcttc gagctagatc ttgctgggta cttcatgcat
1620

110 ttagttcttt gaagttccat aatgagctca atctaagaaa tgccgttgaa ttagcgaaga

ES 2 343 996 B1

1680

5 agagcctgat tgaagataaa gagatgcctg tcaaagttga agctgccctt gctcttcagt
1740

ctttaatttc taaccagata caagctaagg aatatatgaa gccacatgtg aggcctatta
1800

10 tgcaggaact gttgcacatt gttagagaga cagaaaatga tgatgttact aatgtcatcc
1860

agaagatgat atgtgaatac agtcaagagg tagcctcaat tgctgttgat atgaccaac
1920

15 acttggtcga gatatttggc aaagttcttc aaagtgatga atatgaagaa gttgaagaca
1980

aaacagtaat ggctatggga attttacata ccattgatac tatcttaaca gttgtagaag
2040

20 atcataaaga gattaccag cagttagaga atatctgtct acggatcatt gatcttgttc
2100

tgcagaaaca tgtaattgaa ttctatgaag aaattctttc cctggcatac agtttaacct
2160

25 gccacagtat ttcccctcaa atgtggcagc ttctaggtat actatatgaa gtgtttcagc
2220

aggattgctt tgaatacttt acagacatga tgcctctcct gcataattat gtgacaatag
2280

30 atacagatac cttactatca aatgcaaaac atttagaaat tctttttaca atgtgtagga
2340

35 aggtactatg tggagatgca ggagaagatg cagagtgca tgcagctaaa cttctggaag
2400

40 tcatcattct tcagtgcaa ggaaggggaa ttgatcagtg cattccactc ttcgttcaac
2460

ttgttttggga gagattaact cgaggggtca aaactagtga gcttcgtact atgtgtcttc
2520

45 aggttgcaat tgctgccttg tactacaacc ctgatttgct gctacatact ttagaacgaa
2580

ttcagttgcc tcacaaccct ggacctatca ctgtacagtt tataaatcaa tggatgaatg
2640

50 atacagattg ttttcttggg catcatgacc ggaagatgtg tataatagga ctgagtatcc
2700

ttttggaatt gcaaaatcga cctcctgcag tagatgctgt ggtgggacag attgttcct
2760

55 caattctttt ccttttcctt ggccctaaagc aggtctgtgc tactagacaa ctggtaaacc
2820

60 gggagatcg ttcaaaagca gagaaagctg atatggaaga aaatgaggag atttcaagtg
2880

atgaagagga gacaaatgta actgctcaag caatgcagtc aaataatgga agaggtgaag
2940

65 atgaggagga ggaagatgat gactgggatg aagaagtatt ggaagaaacc gcgcttgagg

ES 2 343 996 B1

3000

5 ggttcagtac tccacttgac cttgacaata gtgtggatga atatcagttt tttacacaag
3060

ctctgataac tgtgcagagt cgagatgcag cctggtacca gctgctgatg gcaccactca
3120

10 gccgaggatca gaggacagca ctgcaggagg tgtacacact ggcagagcac cgacggacgg
3180

tggcagaggc aaagaagaag attgaacaac agggaggctt cacctttgaa aacaaaggag
3240

15 tcctctccgc atttaatttt gggactgtgc ccagcaacaa ctgaaggaaa gaacatcagc
3300

tgaccaaagt tcatcgctgc atttatttc acaagaggag tgtgagggtc aaggggatga
3360

20 aatgaggggc tgcttttagg gccctcctgc tgtgccagtt accatctggc attaggcagc
3420

acttttatct actctttccc ctttgacctt tgtcacctg aaatatatat tttaaacagc
25 3480

tactgtaagt atgaaatgaa agaaaaacaa tcattggacg gaaaaggaca acccatatgt
3540

30 tccaaaggct gaatgcccaa ggttgtttta gaggattgga tagacttgca cgtctcaggt
3600

ttttgccatg cagaatcaat ggatttatgc ggataacagt gccttctggt gtacatgaat
3660

35 tatcagaaaa aaatttttgg agtgcattgc aatttttttt aaagcataaa acatatttct
3720

agatacaaca taaaccttgg ttatatgtca actattctgc attttactct gtgaatttat
40 3780

tgttaggcag ttactgcaag tcaactcgtt ttcaaatct tcacggctcc ctctgctcac
3840

45 cctgctgctg gggggctttt ttcaggggtt gtattataaa atatgactg gctgtggttt
3900

ttcataagat gttttgtggc tttttaaaga agtgtcttac tccctcttct ctcatTTTTT
3960

50 tctgccttga aaaggggtgg tatttctttt ggggtcaaca aatacacata tcagtttcac
4020

tcctaacctt gtaagttcag gactcatttt cttggcagca agtggcaggg gctctttagt
4080

55 cactggagtt aacagtaagt ctgagtatat tctgaataat gtaattatgc aattaattga
4140

taatcatacc taaagcacat tgaacttcta agagacaggt gctacgtaag cacactgtct
60 4200

ttctggatgg ggacttgtat tttaaataac ttatcattcc agctatgttg gactagggtc
4260

65 caaagcattt tataatattt ttttttaatc taggaaaaag acccaaacaa atctaacttc

ES 2 343 996 B1

4320

5 ttgctttctc acctatttga attttctct acctaatttg tttgtgtctt tattatagct
4380

cagtttgact tctcaaactt ttcaagggtt ggggcagctg cacttttagg ttgcctatag
4440

10 gaaatattgc atatggagat tacagtgttt ttctgaccta gttcaagacc agaacatggg
4500

tcataggggtt tttattcagc aaaatgaaaa cgtatcttca gaacttaaca tattactgga
4560

15 tgtgatacag attttgcttt ctgtggaatt gaaaattcac aaaaattcac agcttaaatt
4620

20 tcccatcagt gactggagga atttttttca ggtgcttcct atattactat ccctcatgca
4680

tattaactct ttagaatttt aggttaagtg atgtctatag aagtagcctg gaaaaccatg
4740

25 agttttggag ttcagtgacc ccctgctttc ctctgctcct cccttcccaa ggcattgaag
4800

ctgaatgtgc caactggcag ttagaagcga aagatggcat tgggagaaat tttagagagc
4860

30 tttcaaactc tttattctca tgttccacat ggtctaattt taacataaat aatatgcctt
4920

cacactggat tgtaaaaggc atgtgatttt tgagatttta ttttgtgatg tattgtcttc
4980

35 tgcagtatta aagggaaaga gatattaatg tgcattacc tatcttgttt ttgaagccag
5040

40 ggtagttgta tgattttggt accagcagtg ctaacctgaa tgtgacctgg ttaccttgga
5100

aatgcaggaa cttatatgaa tgtactataa aataaaatgc ggactgattc ccaggaaaaa
5160

45 aaaaaaaaaa aaaaaaaaa
5178

<210> 10

<211> 1526

50 <212> DNA

<213> *Homo sapiens*

<400> 10

55 ccggaagtga cgcgaggctc tgcggagacc aggagtcaga ctgtaggacg acctcgggtc
60

60 ccacgtgtcc ccggtactcg ccggccggag cccccggctt cccggggccg ggggacctta
120

gcggcaccca cacacagcct actttccaag cggagccatg tctggtaacg gcaatgcggc
180

65 tgcaacggcg gaagaaaaca gcccaaagat gagagtgatt cgcgtgggta cccgcaagag
240

ES 2 343 996 B1

ccagcttgct cgcatacaga cggacagtgt ggtggcaaca ttgaaagcct cgtaccctgg
 300
 5 cctgcagttt gaaatcattg ctatgtccac cacaggggac aagattcttg atactgcaact
 360
 ctctaagatt ggagagaaaa gcctgtttac caaggagctt gaacatgccc tggagaagaa
 420
 10 tgaagtggac ctggttgttc actccttgaa ggacctgccc actgtgcttc ctctggctt
 480
 caccatcgga gccatctgca agcgggaaaa ccctcatgat gctgttgctt ttcacccaaa
 540
 15 atttgttggg aagaccctag aaaccctgcc agagaagagt gtggtgggaa ccagctccct
 600
 20 gcgaagagca gcccagctgc agagaaagtt cccgcatctg gagttcagga gtattcgggg
 660
 aaacctcaac acccggttc ggaagctgga cgagcagcag gagttcagtg ccatcatcct
 720
 25 ggcaacagct ggcctgcagc gcatgggctg gcacaaccgg gtggggcaga tcctgcaccc
 780
 tgaggaatgc atgtatgctg tgggccaggg ggccttgggc gtggaagtgc gagccaagga
 840
 30 ccaggacatc ttggatctgg tgggtgtgct gcacgatccc gagactctgc ttcgctgcat
 900
 cgctgaaagg gccttcctga ggcacctgga aggaggctgc agtgtgccag tagccgtgca
 960
 35 tacagctatg aaggatgggc aactgtacct gactggagga gtctggagtc tagacggctc
 1020
 40 agatagcata caagagacca tgcaggctac catccatgtc cctgcccagc atgaagatgg
 1080
 ccctgaggat gaccacagt tggtaggcat cactgctcgt aacattccac gagggcccca
 1140
 45 gttggctgcc cagaacttgg gcatcagcct ggccaacttg ttgctgagca aaggagccaa
 1200
 aaacatcctg gatgttgac ggcagcttaa cgatgcccac taactggttt gtggggcaca
 1260
 50 gatgcctggg ttgctgctgt ccagtgccta catcccgggc ctcaagtccc cattctcact
 1320
 55 gctatctggg gagtgattac cccgggagac tgaactgcag ggttcaagcc ttccagggat
 1380
 ttgcctcacc ttggggcctt gatgactgcc ttgcctcctc agtatgtggg ggcttcatct
 1440
 60 ctttagagaa gtccaagcaa cagccttga atgtaaccaa tcctactaat aaaccagttc
 1500
 tgaaggtgta aaaaaaaaaa aaaaaa
 1526

65 <210> 11
 <211> 6584

ES 2 343 996 B1

<212> DNA

<213> *Homo sapiens*

5 <400> 11

tttttttttc ctttttggga gctgaaaaat ttccggtaag ggaaagaagg gctcctttcg
60

ctccttattt ccccgctccc ttccctcccc caccttcccc tcctccggct ttttctcccc
10 120

aactcgggga ggtccttccc ggtggccgcc ctgacgaggt ctgagcacct aggcggaggc
180

ggcgcaggct tttttagtg aggtttgcgc ctgctgagcg cgctgcctc cgccatgcac
15 240

gggggtggcc cccctcggg ggacagcgca tgcccgtgc gcaccatcaa gagagtccag
300

ttcggagtcc tgagtccgga tgaactgaag cgaatgtctg tgacggaggg tggcatcaaa
20 360

taccagaga cgactgaggg aggccgcccc aagcttgggg ggctgatgga cccgaggcag
25 420

ggggtgattg agcggactgg ccgctgcaa acatgtgcag gaaacatgac agagtgtcct
480

ggccactttg gccacattga actggccaag cctgtgtttc acgtgggctt cctggtgaag
30 540

acaatgaaag ttttgcgctg tgtctgcttc ttctgctcca aactgcttgt ggactctaac
600

aacccaaaga tcaaggatat cctggctaag tccaagggac agcccaagaa gcggtcaca
35 660

catgtctacg acctttgcaa gggcaaaaac atatgagagg gtggggagga gatggacaac
40 720

aagttcggtg tggaacaacc tgagggtgac gaggatctga ccaagaaaa gggccatggt
780

ggctgtgggc ggtaccagcc caggatccgg cgttctggcc tagagttgta tgcggaatgg
45 840

aagcacgtta atgaggactc tcaggagaag aagatcctgc tgagtccaga gcgagtgcac
900

gagatcttca aacgcatctc agatgaggag tgttttgtgc tgggcatgga gccccgctat
50 960

gcacggccag agtggatgat tgtcacagtg ctgcctgtgc ccccgctctc cgtgcggcct
1020

gctgtttgta tgcagggctc tgcccgtaac caggatgacc tgactcacia actggctgac
55 1080

atcgtgaaga tcaacaatca gctgctggcg aatgagcaga acggcgcagc ggcccatgtc
60 1140

attgcagagg atgtgaagct cctccagttc catgtggcca ccatggtgga caatgagctg
1200

65

ES 2 343 996 B1

cctggcttgc cccgtgccat gcagaagtct gggcgcccc tcaagtcctt gaagcagcgg
1260

5 ttgaagggca aggaaggccg ggtgcgaggg aacctgatgg gcaaaagagt ggacttctcg
1320

gcccgtactg tcatcacccc cgaccccaac ctctccattg accaggttgg cgtgccccgc
1380

10 tccattgctg ccaacatgac ctttgcggag attgtcacc ccttcaacat tgacagactt
1440

caagaactag tgcgcagggg gaacagccag taccagggcg ccaagtacat catccgagac
1500

15 aatgggtgatc gcattgactt gcgtttccac cccaagccca gtgaccttca cctgcagacc
1560

ggctataagg tggaacggca catgtgtgat ggggacattg ttatcttcaa ccggcagcca
20 1620

actctgcaca aaatgtccat gatggggcat cgggtccgca ttctcccatg gtctaccttt
1680

25 cgcttgaatc ttagtgtgac aactccgtac aatgcagact ttgacgggga tgagatgaac
1740

ttgcacctgc cacagtctct ggagacgcga gcagagatcc aggagctggc catggttctt
1800

30 cgcatgattg tcacccccca gagcaatcgg cctgtcatgg gtattgtgca ggacacactc
1860

acagcagtgc gcaaattcac caagagagac gtcttcctgg agcgggggta agtgateaac
35 1920

ctcctgatgt tcctgtcgac gtgggatggg aagggtccac agccggccat cctaaagccc
1980

40 cggccccctgt ggacaggcaa gcaaattctt tcctcatca tacctggtca catcaattgt
2040

atccgtaccc acagcaccca tcccgatgat gaagacagtg gcccttacia gcacatctct
2100

45 cctgggggaca ccaaggtggt ggtggagaat ggggagctga tcatgggcat cctgtgtaag
2160

aagtctctgg gcacgtcagc tggctccctg gtccacatct cctacctaga gatgggtcat
50 2220

gacatcactc gcctcttcta ctccaacatt cagactgtca ttaacaactg gctcctcatc
2280

gagggtcata ctattggcat tggggactcc attgctgatt ctaagactta ccaggacatt
55 2340

cagaacacta ttaagaaggc caagcaggac gtaatagagg tcatcgagaa ggacacacaac
2400

60 aatgagctgg agcccacccc agggaaact ctgcggcaga cgtttgagaa tcaggatgaac
2460

cgcattctta acgatgcccg agacaagact ggctcctctg ctcagaaatc cctgtctgaa
2520

65

ES 2 343 996 B1

tacaacaact tcaagtctat ggtcgtgtcc ggagctaaag gttccaagat taacatctcc
 2580
 5 caggtcattg ctgtcgttgg acagcagaac gtcgagggca agcggattcc atttggcttc
 2640
 aagcaccgga ctctgcctca cttcatcaag gatgactacg ggcctgagag ccgtggcttt
 2700
 10 gtggagaact cctacctagc cggcctcaca cccactgagt tctttttcca cgccatgggg
 2760
 ggtcgtgagg ggctcattga cacggctgtc aagactgctg agactggata catccagcgg
 2820
 15 cggctgatca agtccatgga gtcagtgatg gtgaagtacg acgcgactgt gcggaactcc
 2880
 atcaaccagg tgggtgcagct gcgctacggc gaagacggcc tggcaggcga gagcgttgag
 2940
 20 ttccagaacc tggctacgct taagccttcc aacaaggctt ttgagaagaa gttccgcttt
 3000
 gattatacca atgagagggc cctgcggcgc actctgcagg aggacctggt gaaggacgtg
 3060
 25 ctgagcaacg cacacatcca gaacgagttg gagcgggaat ttgagcggat gcgggaggat
 3120
 30 cgggaggtgc tcagggcat cttcccaact ggagacagca aggtcgtcct cccctgtaac
 3180
 ctgctgcgga tgatctggaa tgctcagaaa atcttccaca tcaaccacg ccttccctcc
 3240
 35 gacctgcacc ccatcaaagt ggtggagggg gtcaaggaat tgagcaagaa gctggtgatt
 3300
 gtgaatgggg atgaccact aagtcgacag gcccaggaaa atgccacgct gctcttcaac
 3360
 40 atccacctgc ggtccacgtt gtgttcccgc cgcattggcag aggagtctcg gctcagtggg
 3420
 45 gaggccttcg actggctgct tggggagatt gagtccaagt tcaaccaagc cattgcgcat
 3480
 cccggggaaa tgggtggggc tctggctgcg cagtcccttg gagaacctgc caccagatg
 3540
 50 accttgaata cttccacta tgctggtgtg tctgccaaga atgtgacgct ggggtgtgcc
 3600
 cgacttaagg agctcatcaa catttccaag aagccaaaga ctcttcgct tactgtcttc
 3660
 55 ctggtgggcc agtccgctcg agatgctgag agagccaagg atattctgtg ccgtctggag
 3720
 60 catacaacgt tgaggaaggt gactgccaac acagccatct actatgacct caacccccag
 3780
 agcacggtgg tggcagagga tcaggaatgg gtgaatgtct actatgaaat gcctgacttt
 3840
 65

ES 2 343 996 B1

gatgtggccc gaatctcccc ctggctgttg cgggtggagc tggatcggaa gcacatgact
3900

5 gaccggaagc tcaccatgga gcagattgct gaaaagatca atgctggttt tggtgacgac
3960

ttgaactgca tctttaatga tgacaatgca gagaagctgg tgctccgtat tcgcatcatg
4020

10 aacagcgatg agaacaagat gcaagaggag gaagagggtg tggacaagat ggatgatgat
4080

gtcttcctgc gctgcatcga gtccaacatg ctgacagata tgaccctgca gggcatcgag
4140

15 cagatcagca aggtgtacat gcacttgcca cagacagaca acaagaagaa gatcatcatc
4200

acggaggatg ggggaattcaa ggccctgcag gagtggatcc tggagacgga cggcgtgagc
20 4260

ttgatgcggg tgctgagtga gaaggacgtg gaccccgtac gcaccacgtc caatgacatt
4320

25 gtggagatct tcacggtgct gggcattgaa gccgtgcgga aggccctgga gcgggagctg
4380

taccacgtca tctcctttga tggctcctat gtcaattacc gacacttggc tctcttgtgt
4440

30 gataccatga cctgtcgtgg ccacttgatg gccatcacc gacacggagt caaccgccag
4500

gacacaggac cactcatgaa gtgttccttt gaggaaacgg tggacgtgct tatggaagca
35 4560

gccgcacacg gtgagagtga ccccatgaag ggggtctctg agaatatcat gctgggccag
4620

40 ctggctccgg ccggcactgg ctgctttgac ctctgcttg atgcagagaa gtgcaagtat
4680

ggcatggaga tccccaccaa tatccccggc ctgggggctg ctggaccac cggcatgttc
4740

45 tttggttcag caccagtc ccaggggtgga atctctcctg ccatgacacc ttggaaccag
4800

ggtgcaaccc ctgcctatgg cgcctgttcc ccagtggtg ggagtggaat gaccccaggg
4860

50 gcagccggct tctctcccag tgctgctgca gatgccagcg gcttcagccc aggttactcc
4920

cctgcctggt ctccccacacc gggctccccg gggctccccag gtccctcaag ccctacatc
55 4980

ccttcaccag gtggtgccat gtctcccagc tactcgccaa cgtcacctgc ctacgagccc
5040

60 cgctctcctg ggggctacac accccagagt ccctcttatt cccccacttc accctctac
5100

tcccctacct ctccatccta ttctccaacc agtcccaact atagtcccac atcaccagc
5160

65

ES 2 343 996 B1

tattcgcaa cgtcaccag ctactaccg acctctcca gctactcacc cacctctccc
5220

5 agctactcgc ccacctctcc cagctactcg cccacctctc ccagctactc acccacttcc
5280

cctagctact cgcccacttc ccctagctac tcgccaacgt ctcccagcta ctcaccgaca
5340

10 tctcccagct actcgccaac ttcaccagc tattctcca cttctcccag ctactcacct
5400

acctctcaa gctattcacc cacctcccc agctactcac ccacttcccc aagttactca
5460

15 cccaccagcc cgaactattc tccaaccagt cccaattaca cccaacatc acccagctac
5520

agccccgat caccagcta ttcacctact agtcccaact acacacctac cagcccctaac
20 5580

tacagccaa cctctccaag ctactctcca acatcaccca gctattcccc gacctacca
5640

25 agttactccc cttccagccc acgatacaca ccacagtctc caacctatac cccaagctca
5700

cccagctaca gccccagctc gcccagctac agcccaacct cacccaagta caccccaacc
5760

30 agtccttctt acagtcccag ctccccagag tataccccaa cctctcccaa gtactcacct
5820

accagtccca aatattcacc cacctctccc aagtactcgc ctaccagtcc cacctattca
35 5880

cccaccacc caaatactc cccaacatct cctacttatt cccaacctc tccagtctac
5940

40 accccaacct ctccaagta ctcacctact agccccactt actcgccac tcccccaag
6000

tactcgcca ccagccccac ctactcgccc acctccccca aaggctcaac ctactctccc
6060

45 acttcccctg gttactcgcc caccagcccc acctacagtc tcacaagccc ggctatcagc
6120

ccggatgaca gtgacgagga gaactgaggg cacgtggggg gcggcagcgg gctagggccc
50 6180

agggcagctt gcccgtgctg ctgtgcagtt cttgcctccc tcacggggcg tcacccccag
6240

cccagctccg ttgtacataa atgccttgctg gcagagctcc cggatgaactt ctggatcccg
55 6300

tttctgatgc agactcttgt cttgttctcc acttgtgctg ttagaactca ctggcccagt
6360

60 ggtgttctca ctctacccc acccaccccc tgctgtccc caaattgaag atccttctt
6420

gcctgtggct tgatgcgggg cgggtaaagg gtattttaac ttaggggtag ttctgtctgt
6480

65

ES 2 343 996 B1

gagtggttac agctgaccc cgggaagaac aaagctaaag ctgccttttg tctgttattt
6540

5 ttttttttg aagtttaaat aaagtttact aattttgacc aaaa
6584

<210> 12

<211> 2405

10 <212> DNA

<213> *Homo sapiens*

<400> 12

15 tccggcgtgg tgcgcaggcg cggatatccc cctccccgc cagctcgacc ccggtgtggt
60

gcgcaggcgc agtctgcgca gggactggcg ggactgcgcg gcggcaacag cagacatgct
120

20 ggggggtccgg ggcctgtcgc ggctgctgag cgctcggcgc ctggcgctgg ccaaggcgtg
180

gccaacagtg ttgcaaacag gaacccgagg ttttcaactc actggtgatg ggaacaagag
240

25 ggcattctgct aaagtttcag attccatttc tgctcagtat ccagtagtgg atcatgaatt
300

30 tgatgcagtg gtggtaggcg ctggaggggc aggcttgcca gctgcatttg gcctttctga
360

ggcagggttt aatacagcat gtgttaccac gctgtttcct accaggtcac aactgtttgc
420

35 agcacagggg ggaatcaatg ctgctctggg gaacatggag gaggacaact ggaggtggca
480

tttctacgac accgtgaagg gctccgactg gctgggggac caggatgcca tccactacat
540

40 gacggagcag gccccgccc ccgtggtcga gctagaaaat tatggcatgc cgtttagcag
600

aactgaagat gggaagattt atcagcgtgc atttggtgga cagagcctca agtttgga
660

gggcgggcag gcccatcggg gctgctgtgt ggctgatcgg actggccact cgctattgca
720

50 caccttatat ggaaggcttc tgcgatatga taccagctat tttgtggagt attttgcctt
780

ggatctcctg atggagaatg gggagtgccg tgggtgcatc gcactgtgca tagaggacgg
840

55 gtccatccat cgcataagag caaagaacac tgttggtgcc acaggaggct acgggcgcac
900

ctacttcagc tgcacgtctg cccacaccag cactggcgac ggcacggcca tgatcaccag
960

60 ggcaggcctt ccttgccagg acctagagtt tgttcagttc caccctacag gcatatatgg
1020

65 tgctggttgt ctattacgg aaggatgtcg tggagagggg ggcattctca ttaacagtca

ES 2 343 996 B1

1080

5 aggcgaaagg tttatggagc gatacgcccc tgtcgcgaag gacctggcgt ctagagatgt
1140

gggtgtctcg tccatgactc tggagatccg agaaggaaga ggctgtggcc ctgagaaaga
1200

10 tcacgtctac ctgcagctgc accacctacc tccagagcag ctggccacgc gcctgcctgg
1260

catttcagag acagccatga tcttcgctgg cgtggacgtc acgaaggagc cgatccctgt
1320

15 cctccccacc gtgcattata acatgggicgg cattccccacc aactacaagg ggcaggctct
1380

gaggcacgtg aatggccagg atcagattgt gcccggcctg tacgcctgtg gggaggccgc
1440

20 ctgtgcctcg gtacatggtg ccaaccgcct cggggcaaac tcgctcttgg acctggttgt
1500

ctttggtcgg gcatgtgccc tgagcatcga agagtcatgc aggcctggag ataaagtccc
1560

25 tccaattaa ccaaacgctg gggaagaatc tgtcatgaat cttgacaaat tgagatttgc
1620

tgatggaagc ataagaacat cggaactgcg actcagcatg cagaagtcaa tgcaaatca
1680

tgctgccgtg ttccgtgtgg gaagcgtggt gcaagaaggt tgtgggaaaa tcagcaagct
1740

35 ctatggagac ctaaagcacc tgaagacggt cgaccgggga atggtctgga acacggacct
1800

ggtggagacc ctggagctgc agaacctgat gctgtgtgcg ctgcagacca tctacggagc
1860

40 agaggcacgg aaggagtcac ggggcgcgca tgccagggaa gactacaagg tgccgattga
1920

tgagtacgat tactccaagc ccatccaggg gcaacagaag aagccctttg aggagcactg
1980

gaggaagcac accctgtcct atgtggacgt tggcactggg aaggctactc tggaatatag
2040

50 acccgtgatc gacaaaactt tgaacgaggc tgactgtgcc accgtccccg cagccattcg
2100

ctcctactga tgagacaaga tgtggtgatg acagaatcag cttttgtaat tatgtataat
2160

55 agctcatgca tgtgtccatg tcataactgt cttcatacgc ttctgcactc tggggaagaa
2220

ggagtacatt gaagggagat tggcacctag tggctgggag cttgccagga acccagtggc
2280

60 cagggagcgt ggcacttacc tttgtccctt gcttcattct tgtgagatga taaaactggg
2340

65 cacagctctt aaataaaata taaatgaaca aactttcttt tatttccaaa aaaaaaaaaa

ES 2 343 996 B1

2400

aaaaa
2405

5 <210> 13
<211> 3994
<212> DNA
10 <213> *Homo sapiens*
<400> 13
15 gcccgctcca gccgcccgc atcctcggcc cgcgccccga gacccgcgcc cagctagccc
60
cggccccgct cggcgcccca ggcagctcgg ctgcgctcgc cgcgggacgg cgcggccatg
120
20 aggctgcggg gatgctggacc ccgggccgcc cctgcctcca gcgcaggggc cagcgcgct
180
cggctgctgg cccccccggg gcggaacccc ttcgtgcacg agctgcgcct cagcgcctg
240
25 cagaaggccc aggtggccct catgacactg acgctcttcc cggctccggt cctggttgcc
300
gctgccatga tgctgctggc ctggcccctc gcacttgtcg catccctggg ctctgcggag
360
aaggaacccg agcagcccc ggccctgtgg aggaaggttg tggacttctt gctgaaggcc
420
35 atcatgcgca ccatgtggtt cgcggcgcc ttccaccggg tggccgtgaa ggggcggcag
480
gcgctgcccc ccgagggcgc catcctcagc ctgcgcctc actcgtccta cttcgacgcc
540
40 atccctgtga ccatgacgat gtcctccatc gtgatgaagg cagagagcag agacatcccc
600
atctggggaa ctctgatcca gtatatacgg cctgtgttcg tgtcccggtc agaccaggat
660
45 tctcgcagga aaacagtaga agaaatcaag agacgggccc agtccaacgg aaagtggcca
720
cagataatga tttttccaga aggaacttgt acaaacagga cctgcctaata taccttcaaa
780
cctggtgcat tcatccctgg agcgcctcgc cagcctgtgg ttttacgata tccaaataaa
840
55 ctggacacca tcacatggac gtggcaagga cctggagcgc tggaaatcct gtggctcacg
900
ctgtgtcagt ttcacaacca agtggaaatc gagttccttc ctgtgtacag cccttctgag
960
60 gaggagaaga ggaacccccg cctgtatgcc agcaacgtgc ggcgagtcac ggccgaggcc
1020
ttgggtgtct ccgtgactga ctacacgttc gaggactgcc agctggccct ggcggaagga
1080
65

ES 2 343 996 B1

cagctccgtc tccccgctga cacttgccct ttagaatttg ccaggctcgt gcggggcctc
 1140
 5 gggctaaaaac cagaaaagct tgaaaaagat ctggacagat actcagaaaag agccaggatg
 1200
 aagggaggag agaagatagg tattgaggag tttgccgctt ccctggaagt ccccgtttct
 1260
 10 gacttgctgg aagacatggt ttcactgttc gacgagagcg gcagcggcga ggtggacctg
 1320
 cgagagtgtg tggttgccct gtctgtcgtc tgccggccgg cccggaccct ggacaccatc
 1380
 15 cagctggcctt tcaagatgta cggagcgcaa gaggacggca gcgtcggcga aggtgacctg
 1440
 tcctgcatcc tcaagacggc cctgggggtg gcagagctca ccgtgaccga cctattccga
 1500
 20 gccattgacc aagaggagaa ggggaagatc acattcgctg acttccacag gtttgagaa
 1560
 25 atgtaccctg ctttcgaga ggaatacctg taccggatc agacacattt cgaaagctgt
 1620
 gcagagacct cacctgcgcc aatcccaaac ggcttctgtg ccgatttcag cccggaaaac
 1680
 30 tcagacgctg ggcggaagcc tgttcgcaag aagctggatt aggaccagg gttgcggaga
 1740
 35 gacgcggccc ctcccgcgtg gacatcaccg ccatgagcct ctttgcgagt gacctctggg
 1800
 ctccgctcct cactcctgct gtacaggcac tgtcttcagc ccgagttcca ggggcctcgg
 1860
 40 gggctgtttg tatcttgttc ctttgtgaag tgtgttgagc aaccgacgct tactgtgcga
 1920
 gaatcggagg gcgcgcacgc ggatcccccg cctggcctgg acccctggg gtcaggttcc
 1980
 45 ctgcccggcg gggggcaccg gtgccgcccc gtgttctccc acggggccct ggtttcgagt
 2040
 ctctgtcaca gcctcttccg gcggcagcgt gcaccgggcg ggcctccgtg cacactcagc
 2100
 50 acacgcctgc cacacagcgt gcgcttgctg gtcactctgg cacgaaacct gtctgcctct
 2160
 55 gtggatccac agcctggcag agccgagccg tcacctgatt tttcagtgtt tctacctgtg
 2220
 tgctggagct catgagtatt ttataaactc catttaggta cttcaggaaa catgcagcat
 2280
 60 tttttaaaaa atgaaaattg tttttctact tcatttttcc ttttagagtc aaaggatatt
 2340
 tttttatagg cttttttttt ttttaatatag aatctgaggc tgtttgggct ttgacttaaa
 2400
 65

ES 2 343 996 B1

tttccatcag gcctctctcc agcaggtaat ccctctcctt ccgctgggtc ccctggggag
2460

5 gtgtgaactc aagggcctag ccccaaaaca ctttttctgc ttttcttaat ctttttccag
2520

tccccctctt tttataaac gttggcagtt tgatgtttct gtttcggcat aacgtaatcc
2580

10 atttcactgt agcctaaact ccagtccgag gttggatatt gttcaaatga gcagggccccg
2640

agctggaagc gcaaggcagc cgccgccgtg ccgctcctcc cttgccctca ggccaggtcc
2700

15 ctgctggaag cggctgcatc ttctgtcag ccctggtttc catggtgact ggcgtcacgc
2760

20 agccacccga gtatggctga cttcctgca gagagaggag ccgcagtctt ttgcttgtgg
2820

aaggagacgc tgggctgtgc ggtgcggagg gtgatgagga tgtctggtga cagccgtgcg
2880

25 gacaccactc ctctctgcag cactgcctcc cagcgccagg gtcgcgggca catcccactg
2940

agagcggggg tcctgcccc a tcttagagtc aaaggcagag gggcttccag gccctggatg
3000

30 gggatatttg gtgtcacctg aagtcctctt gacatcacct tgtttcatca tttttatga
3060

cagaattaga aacccatcct tcaagcaca taatcatcac agacttgagt ttgcttccta
3120

35 aagcaaaggc tccgggtttg tttggaaaat ttttttgatt tctgaaatga attgattttt
3180

40 atatttgggg catctctata gaaagtgacc accaaggcca gtaagtacgg gaaaaaatgt
3240

ttactaactt cctcagagat tcgtgatacg cgtttctcca ctgacagaca tttaaaaaca
3300

45 accttcagct ccgtttcaat caatcacctc gacttgtttt ttagcatgga cactgccagc
3360

aggacagaca gggatggagt aaaccgaagt caatttcagg gctcttggcg tgttgacac
3420

50 agaagaaatc ctagtgcagc ctttggtagc taacagtcac tgattttata attggagaat
3480

55 gcgtaaagat tcatttttca aggagaagag cctgcaaag gccaatgaag gaggtaaata
3540

aactaagata ttccgagggg agggacccag gccacctccc ttccgcaggt ctgcagatga
3600

60 aggggtttttt gaatgaaatg ccaactgtgca ttttcagaaa aaaaaatctc tgataaacag
3660

actttgaaatg gatgtttgtt cctcctgatt ctcttttctc ttcgtggcga cttagagtgt
3720

65

ES 2 343 996 B1

g³⁷⁸⁰cggatattc ggaactgtga atgtacatag cgttgagtta aacccttgt gtgtgagaca
5 ggacgcagcg ggcccctggt ggcctggggg ccagaccctg gggcaggtgg ggcattgggccc
3840
ctggcctgcg gggacctgct ggggtgtgag ggcagagggg gggttgccat gaaggaactt
3900
10 gggattttca atggaataag taaaacataa agtctatact tgggaaaaaa aaaaaaaaaa
3960
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaa
15 3994
<210> 14
<211> 2655
<212> DNA
20 <213> *Homo sapiens*
<400> 14
25 cccagaaggc cgcgggggggt ggaccgccta agagggcgtg cgctcccgcac atgccccgcg
60
g¹²⁰cgccatt aaccgccaga tttgaatcgc gggaccctgt ggcagaggtg gcggcggcgg
120
30 catgggtgcc ccgacgttgc cccctgcctg gcagcccttt ctcaaggacc accgcatctc
180
tacattcaag aactggcctt tcttggaggg ctgcgcctgc accccggagc ggatggccga
240
35 ggctggcttc atccactgcc ccaactgagaa cgagccagac ttggcccagt gtttcttctg
300
40 cttcaaggag ctggaaggct gggagccaga tgacgacccc atagaggaac ataaaaagca
360
ttcgtccggt tgcgctttcc tttctgtcaa gaagcagttt gaagaattaa cccttgggtga
420
45 atttttgaaa ctggacagag aaagagccaa gaacaaaatt gcaaaggaaa ccaacaataa
480
gaagaaagaa tttgaggaaa ctg⁵⁴⁰cggagaa agtgcgccgt gccatcgagc agctggctgc
540
50 catggattga ggcctctggc cggagctgcc tgggtcccaga gtggctgcac cacttccagg
600
gtttattccc tgggtccacc agccttctctg tgggcccctt agcaatgtct taggaaagga
55 660
gatcaacatt ttcaaattag atgtttcaac tgtgctcttg ttttgtcttg aaagtggcac
720
60 cagaggtgct tctgcctgtg cagcgggtgc tgctggtaac agtggctgct tctctctctc
780
tctctctttt ttgggggctc atttttgctg ttttgattcc cgggcttacc aggtgagaag
840
65 tgagggagga agaaggcagt gtcccttttg ctagagctga cagctttggt cgcgtgggca
900

ES 2 343 996 B1

gagccttcca cagtgaatgt gtctggacct catgttggtg aggctgtcac agtcctgagt
960

5 gtggacttgg caggtgcctg ttgaatctga gctgcagggt ccttatctgt cacacctgtg
1020

cctcctcaga ggacagtttt tttgttggtg tgtttttttg tttttttttt tttggtagat
1080

10 gcatgacttg tgtgtgatga gagaatggag acagagtccc tggctcctct actgtttaac
1140

aacatggctt tcttattttg tttgaattgt taattcacag aatagcacia actacaatta
1200

15 aaactaagca caaagccatt ctaagtcatt ggggaaacgg ggtgaacttc aggtggatga
1260

20 ggagacagaa tagagtgata ggaagcgtct ggcagatact ccttttgcca ctgctgtgtg
1320

attagacagg cccagtgagc cgcggggcac atgctggccg ctccctccctc agaaaaaggc
1380

25 agtggcctaa atccttttta aatgacttgg ctcgatgctg tgggggactg gctgggctgc
1440

tgcaggccgt gtgtctgtca gcccaacctt cacatctgtc acgttctcca cacgggggag
1500

30 agacgcagtc cgcccaggtc cccgctttct ttggaggcag cagctcccgc agggctgaag
1560

tctggcgtaa gatgatggat ttgattcgcc ctccctccctg tcatagagct gcagggtgga
1620

35 ttgttacagc ttcgctggaa acctctggag gtcattctcg ctgttcctga gaaataaaaa
1680

40 gcctgtcatt tcaaactctg ctgtggacct tactggggtt ttaaaatatt gtcagttttt
1740

catcgtctgc cctagcctgc caacagccat ctgcccagac agccgcagtg aggatgagcg
1800

45 tcctggcaga gacgcagttg tctctgggcg cttgccagag ccacgaacct cagacctgtt
1860

tgtatcatcc gggctcctt cgggcagaaa caactgaaaa tgcacttcag acccacttat
1920

50 ttctgccaca tctgagtcgg cctgagatag acttttccct ctaaactggg agaatatcac
1980

agtggttttt gttagcagaa aatgcactcc agcctctgta ctcatctaag ctgcttattt
2040

55 ttgatatttg tgtcagtctg taaatggata cttcacttta ataactgttg cttagtaatt
2100

60 ggctttgtag agaagctgga aaaaaatggt tttgtcttca actcctttgc atgccaggcg
2160

gtgatgtgga tctcggcttc tgtgagcctg tgctgtgggc agggctgagc tggagccgcc
2220

65

ES 2 343 996 B1

cctctcagcc cgctgccac ggcctttcct taaaggccat ccttaaaacc agaccctcat
 2280
 5 ggctaccagc acctgaaagc ttcctcgaca tctgttaata aagccgtagg cccttgtcta
 2340
 agtgcaaccg cctagacttt ctttcagata catgtccaca tgtccatttt tcaggttctc
 2400
 10 taagttggag tggagtctgg gaagggttgt gaatgaggct tctgggctat gggtgaggtt
 2460
 ccaatggcag gttagagccc ctctggccaa ctgccatcct ggaaagtaga gacagcagtg
 2520
 15 cccgctgccc agaagagacc agcaagccaa actggagccc ccattgcagg ctgtcgccat
 2580
 20 gtggaaagag taactcacia ttgccaataa agtctcatgt ggttttatct aaaaaaaaaa
 2640
 aaaaaaaaaa aaaaa
 2655
 25 <210> 15
 <211> 2101
 <212> DNA
 <213> *Homo sapiens*
 30
 <400> 15
 35 acgaacaggc caataaggag ggagcagtgc ggggtttaaa tctgaggcta ggctggctct
 60
 tctcggcgtg ctgcggcgga acggctgttg gtttctgctg ggtgtaggtc cttggctggt
 120
 40 cgggcctccg gtgttctgct tctccccgct gagctgctgc ctggtgaaga ggaagccatg
 180
 gcgctccgag tcaccaggaa ctcgaaaatt aatgctgaaa ataaggcgaa gatcaacatg
 240
 45 gcagggcga aagcgcgttc tacggcccct gctgcaacct ccaagcccgg actgaggcca
 300
 agaacagctc ttggggacat tggttaacaaa gtcagtgaac aactgcaggc caaatgcct
 360
 50 atgaagaagg aagcaaaacc ttcagctact ggaaaagtca ttgataaaaa actaccaaaa
 420
 cctcttgaaa aggtacctat gctggtgcca gtgccagtgt ctgagccagt gccagagcca
 480
 55 gaacctgagc cagaacctga gcctgttaaa gaagaaaac tttcgctga gcctatcttg
 540
 60 gttgatactg cctctccaag cccaatggaa acatctggat gtgcccctgc agaagaagac
 600
 ctgtgtcagg ctttctctga tgtaattctt gcagtaaatg atgtggatgc agaagatgga
 660
 65 gctgatccaa acctttgtag tgaatatgtg aaagatattt atgcttatct gagacaactt

ES 2 343 996 B1

720

5 gaggaagagc aagcagtcag accaaaatac ctactgggtc ggggaagtcac tggaaacatg
780

agagccatcc taattgactg gctagtagcag gttcaaatga aattcagggtt gttgcaggag
840

10 accatgtaca tgactgtctc cattattgat cggttcatgc agaataattg tgtgcccgaag
900

aagatgctgc agctggttgg tgtcactgcc atgtttattg caagcaaata tgaagaaatg
960

15 taccctccag aaattggtga ctttgctttt gtgactgaca acacttatac taagcaccaa
1020

atcagacaga tggaaatgaa gattctaaga gctttaaact ttggctctggg tcggcctcta
1080

20 cctttgcact tccttcggag agcatctaag attggagagg ttgatgtcga gcaacatact
1140

ttggccaaat acctgatgga actaactatg ttggactatg acatgggtgca ctttctctct
1200

25 tctcaaattg cagcaggagc tttttgctta gcactgaaaa ttctggataa tggatgaatgg
1260

acaccaactc tacaacatta cctgtcatat actgaagaat ctcttcttcc agttatgcag
1320

30 cacctggcta agaattgtagt catggtaaat caaggactta caaagcacat gactgtcaag
1380

35 aacaagtatg ccacatcgaa gcatgctaag atcagcactc taccacagct gaattctgca
1440

ctagttcaag atttagccaa ggctgtggca aagggtgtaac ttgtaaactt gagttggagt
1500

40 actatatatta caaataaaat tggcaccatg tgccatctgt acatattact gttgcattta
1560

cttttaataa agcttgtggc cctttttact tttttatagc ttaactaatt tgaatgtggg
1620

45 tacttcctac tgtagggtag cggaaaagtt gtcttaaaag gtatgggtggg gatattttta
1680

50 aaaactcctt ttggtttacc tggggatcca attgatgtat atgtttatat actgggttct
1740

55 tgttttatat acctggcttt tactttatta atatgagtta ctgaagggtga tggagggtatt
1800

tgaaaatfff acttccatag gacatactgc atgtaagcca agtcatggag aatctgctgc
1860

60 atagctctat tttaaagtaa aagtctacca ccgaatccct agtccccctg ttttctgttt
1920

cttcttgtga ttgctgcat aattctaagt tatttacttt taccactatt taagttatca
1980

65 acttttagcta gtatcttcaa acttttactt tgaaaaatga gaattttata ttctaagcca

ES 2 343 996 B1

2040

gttttcattt tggttttgtg ttttggtaa taaaacaata ctcaaataca aaaaaaaaaa
2100

a
2101

<210> 16

<211> 2748

<212> DNA

<213> *Homo sapiens*

<400> 16

agcgggtgcg gggcgggacc ggcccggcct atatattggg ttggcgccgg cgccagctga
60

gccgagcggg agctggctctg gcgaggtttt atacacctga aagaagagaa tgtcaagacg
120

aagtagccgt ttacaagcta agcagcagcc ccagcccagc cagacggaat ccccccaaga
180

agcccagata atccaggcca agaagaggaa aactaccag gatgtcaaaa aaagaagaga
240

ggaggtcacc aagaaacatc agtatgaaat taggaattgt tggccacctg tattatctgg
300

ggggatcagt cttgcatta tcattgaaac acctcacaaa gaaataggaa caagtgattt
360

ctccagattt acaaattaca gatttaaaaa tctttttatt aatccttcac ctttgcctga
420

tttaagctgg ggatgttcaa aagaagtctg gctaaacatg ttaaaaaagg agagcagata
480

tgttcatgac aacattttg aagttctgca ttctgacttg gaaccacaga tgaggtccat
540

acttctagac tggcttttag aggtatgtga agtatacaca cttcataggg aacattttta
600

tcttgacaaa gacttttttg atagatttat gttgacacaa aaggatataa ataaaaatat
660

gcttcaactc attggaatta cctcattatt cattgcttcc aaacttgagg aaatctatgc
720

tcctaaactc caagagtttg cttacgtcac tgatggtgct tgcagtgaag aggatatctt
780

aaggatggaa ctattatat taaaggcttt aaaatgggaa ctttgtcctg taacaatcat
840

ctcctggcta aatctctttc tccaagttga tgctcttaaa gatgctccta aagttcttct
900

acctcagtat tctcaggaaa cattcattca aatagctcag cttttagatc tgtgtattct
960

agccattgat tcattagagt tccagtacag aatactgact gctgctgcct tgtgccattt
1020

ES 2 343 996 B1

tacctccatt gaagtggtta agaaagcctc aggtttggag tgggacagta tttcagaatg
 1080
 5 tgtagattgg atggtacctt ttgtcaatgt agtaaaaagt actagtccag tgaagctgaa
 1140
 gacttttaag aagattccta tggaagacag acataatatac cagacacata caaactatctt
 1200
 10 ggctatgctg gaggaagtaa attacataaa caccttcaga aaagggggac agttgtcacc
 1260
 agtgtgcaat ggaggcatta tgacaccacc gaagagcact gaaaaaccac caggaaaaca
 1320
 15 ctaaagaaga taactaagca aacaagttgg aattcaccaa gattgggtag aactggtatc
 1380
 actgaactac taaagtttta cagaaagtag tgctgtgatt gattgcccta gccaatcac
 1440
 20 aagttacact gccattctga ttttaaaact tacaattggc actaaagaat acatttaatt
 1500
 25 atttcctatg ttagctgtta aagaaacagc aggacttggt tacaagatg tcttcattcc
 1560
 caaggttact ggatagaagc caaccacagt ctataccata gcaatgtttt tcctttaatc
 1620
 30 cagtgttact gtgtttatct tgataaacta ggaattttgt cactggagtt ttggactgga
 1680
 taagtgctac cttaaagggg atactaagtg atacagtact ttgaatctag ttgttagatt
 1740
 35 ctcaaaattc ctacactctt gactagtgca atttggttct tgaaaattaa atttaaactt
 1800
 40 gtttacaaag gtttagtttt gtaataaggt gactaattta tctatagctg ctatagcaag
 1860
 ctattataaa acttgaattt ctacaaatgg tgaaatttaa tgttttttaa actagtttat
 1920
 45 ttgccttgcc ataacacatt ttttaactaa taaggcttag atgaacatgg tgttcaacct
 1980
 gtgctctaaa cagtgggagt accaaagaaa ttataaaca gataaatgct gtggctcctt
 2040
 50 cctaactggg gctttcttga catgtaggtt gcttggaat aacctttttg tatatcacia
 2100
 55 tttgggtgaa aaacttaagt accctttcaa actatttata tgaggaagtc actttactac
 2160
 tctaagatat ccctaaggaa tttttttttt taatttagtg tgactaaggc tttatttatg
 2220
 60 tttgtgaaac tgtaagggtc ctttctaaat tcctccattg tgagataagg acagtgtcaa
 2280
 agtgataaag cttaacactt gacctaaact tctattttct taaggaagaa gagtattaaa
 2340
 65

ES 2 343 996 B1

tatatactga ctcctagaaa tctatttatt aaaaaaagac atgaaaactt gctgtacata
2400

5 ggctagctat ttctaaatat tttaaattag cttttctaaa aaaaaaatcc agcctcataa
2460

agtagattag aaaactagat tgctagttta ttttgttatc agatatgtga atctcttctc
2520

10 cttttgaaga aactatacat ttattgttac ggtatgaagt cttctgtata gtttgttttt
2580

aaactaatat ttgtttcagt attttgtctg aaaagaaaaac accactaatt gtgtacatat
2640

15 gtattatata aacttaacct ttttaatactg tttattttta gccattggtt taaaaaataa
2700

20 aagttaaaaa aatttaactg cttaaaagta aaaaaaaaaa aaaaaaaaa

<210> 17

<211> 2457

25 <212> DNA

<213> *Homo sapiens*

<400> 17

30 gagggccggc cggtttgccg gcggctgcgg agggatatctg aggctcggct gccaagccga
60

35 acgggctctg gtcctcccca gtggccggcc ggggcggaag caggaggcgg gggcagcgtg
120

cgcgctgctg gtcttctctc ccgcggcgct ggggcccgcg ctccgctgct gttgctccat
180

40 tcggcgcttt tctggcggct ggctcctctc cgctgccggc tgctcctcga ccaggcctcc
240

ttctcaacct cagcccgcgg cgccgaccct tccggcacc cccgccccg tctcgtactg
300

45 tcgcccgtcac cgccgcggct ccggccctgg ccccgatggc tctgtgcaac ggagactcca
360

50 agctggagaa tgctggagga gaccttaagg atggccacca ccactatgaa ggagctggtg
420

tcattctgga tgctgggtgct cagtacggga aagtcataga ccgaagagtg agggaactgt
480

55 tcgtgcagtc tgaaattttc cccttggaac caccagcatt tgctataaag gaacaaggat
540

tccgtgctat tatcatctct ggaggaccta attctgtgta tgctgaagat gctccctggt
600

60 ttgatccagc aatattcact attggcaagc ctgttcttgg aatttgctat ggtatgcaga
660

tgatgaataa ggtatttgga ggtactgtgc acaaaaaaag tgtcagagaa gatggagttt
720

65 tcaacattag tgtggataat acatgttcat tattcagggg ccttcagaag gaagaagttg
780

ES 2 343 996 B1

ttttgcttac acatggagat agtgtagaca aagtagctga tggattcaag gttgtggcac
840

5 gttctggaaa catagtagca ggcatagcaa atgaatctaa aaagttatat ggagcacagt
900

tccaccctga agttggcctt acagaaaatg gaaaagtaat actgaagaat ttcctttatg
960

10 atatagctgg atgcagtgga accttcaccg tgcagaacag agaacttgag tgtattcgag
1020

agatcaaaga gagagtaggc acgtcaaaag ttttggtttt actcagtggg ggagtagact
1080

15 caacagtttg tacagctttg ctaaactcgtg ctttgaacca agaacaagtc attgctgtgc
1140

acattgataa tggctttatg agaaaacgag aaagccagtc tgttgaagag gccctcaaaa
1200

agcttggaaat tcaggtcaaa gtgataaatg ctgctcattc tttctacaat ggaacaacaa
1260

25 ccctaccaat atcagatgaa gatagaacct cacggaaaag aattagcaaa acgttaaata
1320

tgaccacaag tcctgaagag aaaagaaaaa tcattgggga tacttttggt aagattgcc
1380

30 atgaagtaat tggagaaatg aacttgaac cagaggagggt tttccttgcc caaggtactt
1440

tacggcctga tctaattgaa agtgcacccc ttgttgcaag tggcaaagct gaactcatca
1500

35 aaacccatca caatgacaca gagctcatca gaaagttgag agaggaggga aaagtaatag
1560

40 aacctctgaa agattttcat aaagatgaag tgagaatttt gggcagagaa cttggacttc
1620

cagaagagtt agtttccagg catccatttc caggctcctgg cctggcaatc agagtaatat
1680

45 gtgctgaaga accttatatt tgtaaggact ttcctgaaac caacaatatt ttgaaaatag
1740

tagctgattt ttctgcaagt gttaaaaagc cacataccct attacagaga gtcaaagcct
1800

50 gcacaacaga agaggatcag gagaagctga tgcaaattac cagtctgcat tcaactgaatg
1860

55 ctttcttgct gccaatataa actgtaggtg tgcaggggta ctgtcgttcc tacagttacg
1920

tgtgtggaat ctccagtaaa gatgaacctg actgggaatc acttattttt ctggctaggc
1980

60 ttatacctcg catgtgtcac aacgttaaca gagttgttta tataatttggc ccaccagtta
2040

aagaacctcc tacagatggt actcccactt tcttgacaac aggggtgctc agtactttac
2100

65

ES 2 343 996 B1

gccaagctga ttttgaggcc cataacattc tcagggagtc tgggtatgct gggaaaatca
2160

5 gccagatgcc ggtgattttg acaccattac attttgatcg ggaccacctt caaaagcagc
2220

cttcatgcca gagatctgtg gttattcgaa cctttattac tagtgacttc atgactggta
2280

10 tacctgcaac acctggcaat gagatccctg tagaggtggt attaaagatg gtcactgaga
2340

ttaagaagat tcctggtatt tctcgaatta tgtatgactt aacatcaaag cccccaggaa
2400

15 ctactgagtg ggagtaataa acttcttggt ctattaaagt accgtgtgca gtttaaa
2457

<210> 18

20 <211> 3769

<212> DNA

<213> *Homo sapiens*

25 <400> 18

aaagctgcag cgtctggaaa aaagcgactt gtggcggctg agcgtggcgc aggcgaatcc
60

30 tcggcactaa gcaaataatgg acctcgcggc ggcagcggag ccgggcgccg gcagccagca
120

cctggaggtc cgcgacgagg tggccgagaa gtgccagaaa ctgttctctg acttcttgga
180

35 ggagtttcag agcagcgatg gagaaattaa atacttgcaa ttagcagagg aactgattcg
240

40 tcctgagaga aacacattgg ttgtgagttt tgtggacctg gaacaattta accagcaact
300

ttccaccacc attcaagagg agttctatag agtttacctt tacctgtgtc gggccttgaa
360

45 aacattcgtc aaagaccgta aagagatccc tcttgccaag gatttttatg ttgcattcca
420

agacctgcct accagacaca agattcgaga gctcacctca tccagaattg gtttgctcac
480

50 tcgcatcagt gggcaggtgg tgcggactca cccagttcac ccagagcttg tgagcggaac
540

55 ttttctgtgc ttggactgtc agacagtgat cagggatgta gaacagcagt tcaaatacac
600

acagccaaac atctgccgaa atccagtttg tgccaacagg aggagattct tactggatac
660

60 aaataaatca agatttgttg attttcaaaa ggttcgtatt caagagacct aagctgagct
720

tcctcgaggg agtatcccc gcagtttaga agtaatttta agggctgaag ctgtggaatc
780

65 agctcaagct ggtgacaagt gtgactttac agggacactg attgttgtgc ctgacgtctc

ES 2 343 996 B1

840

5 caagcttagc acaccaggag cacgtgcaga aactaattcc cgtgtcagtg gtgttgatgg
900

10 atatgagaca gaaggcattc gaggactccg ggcccttggt gttagggacc tttcttatag
960

15 gctggctctt cttgcctgct gtgttgcgcc aaccaacca aggtttgggg ggaaagagct
1020

20 cagagatgag gaacagacag ctgagagcat taagaaccaa atgactgtga aagaatggga
1080

25 gaaagtgttt gagatgagtc aagataaaaa tctataccac aatctttgta ccagcctggt
1140

30 ccctactata catggcaatg atgaagtaaa acgggggtgct ctgctgatgc tctttgggtg
1200

35 cgttccaaag acaacaggag aagggacctc tcttcgaggg gacataaatg tttgcattgt
1260

40 tggtgacca agtacagcta agagccaatt tctcaagcac gtggaggagt tcagccccag
1320

45 agctgtctac accagtggta aagcgtccag tgctgctggc ttaacagcag ctgttgtgag
1380

50 agatgaagaa tctcatgagt ttgtcattga ggctggagct ttgatgttgg ctgataatgg
1440

55 tgtgtgttgt attgatgaat ttgataagat ggacgtgcgg gatcaagttg ctattcatga
1500

60 agctatggaa cagcagacca tatccatcac taaagcagga gtgaaggcta ctctgaacgc
1560

65 ccggacgtcc attttggcag cagcaaacc aatcagtgga cactatgaca gatcaaaatc
1620

70 attgaaacag aatataaatt tgtcagctcc catcatgtcc cgattcgatc tcttctttat
1680

75 ccttgtggat gaatgtaatg aggttacaga ttatgccatt gccaggcgca tagtagattt
1740

80 gcattcaaga attgaggaat caattgatcg tgtctattcc ctcgatgata tcagaagata
1800

85 tcttctcttt gcaagacagt ttaaaccxaa gatttccaaa gagtcagagg acttcattgt
1860

90 ggagcaatat aaacatctcc gccagagaga tggttctgga gtgaccaagt cttcatggag
1920

95 gattacagtg cgacagcttg agagcatgat tcgtctctct gaagctatgg ctcggatgca
1980

100 ctgctgtgat gaggccaac ctaaacaatgt gaaggaagct ttccggttac tgaataaatc
2040

105 aatcatccgt gtggaaacac ctgatgtcaa tctagatcaa gaggaagaga tccagatgga
2100

110 ggtagatgag ggtgctgggt gcatcaatgg tcatgctgac agccctgctc ctgtgaacgg

ES 2 343 996 B1

2160

5 gatcaatggc tacaatgaag acataaatca agagtctgct cccaaagcct ccttaaggct
2220

gggcttctct gagtactgcc gaatctctaa ccttattgtg cttcacctca gaaaggtgga
2280

10 agaagaagag gacgagtcag cattaagag gagcgagctt gttaactggt acttgaagga
2340

aatcgaatca gagatagact ctgaagaaga acttataaat aaaaaaagaa tcatagagaa
2400

15 agttattcat cgactcacac actatgatca tgttctaatt gagctcaccc aggctggatt
2460

gaaaggctcc acagagggaa gtgagagcta tgaagaagat ccctacttgg tagttaacct
2520

20 taactacttg ctcgaagatt gagatagtga aagtaactga ccagagctga ggaactgtgg
2580

cacagcacct cgtggcctgg agcctggctg gagctctgct agggacagaa gtgtttctgg
2640

25 aagtgatgct tccaggattt gttttcagaa acaagaattg agttgatggt cctatgtgtc
2700

30 acattcatca caggtttcat accaacacag gcttcagcac ttcctttggt gtgtttcctg
2760

tcccagtgaa gttggaacca aataatgtgt agtctctata accaataacct ttgttttcat
2820

35 gtgtaagaaa aggcccatta cttttaaggt atgtgctgtc ctattgagca aataactttt
2880

40 tttcaattgc cagctactgc ttttattcat caaataaaa taacttgttc tgaagttgtc
2940

tattggattt ctttctactg taccctgatt attacttcca tctacttctg aatgtgagac
3000

45 tttccctttt tgcttaacct ggagtgaaga ggtagaactg tggattatg gatgaggttt
3060

ctatgagaag gagtcattag agaactcata tgaaagctag aggccttaga gatgactttc
3120

50 caaggttaat tccagttggt tttttttttt ttttaagtta taaaagtta ttatactttt
3180

ttaaaattac tctttagtaa tttattttac ttctgtgtcc taagggtaat ttctcaggat
3240

55 tgttttcaaa ttgctttttt aggggaaata ggtcatttgc tatattaca gcaatcccca
3300

60 aattttatgg tcttccagga aaagttatta ccgtttatga tactaacagt tcctgagact
3360

tagctatgat cagtatgtc atgaggtgga gcagttcctg tgttgagct ttaacaaca
3420

65 gatggcattc attaaatcac aaagtatggt aaaggtcaca aaagcaaat aactgtctga

ES 2 343 996 B1

3480

5 ggctaaggcc cacgtgggac agtctaatac ccatgagtac tcaacttgcc ttgatgtctg
3540
agctttccag tgcaatgtga atttgagcag ccagaaatct attagtagaa agcaagacag
3600
10 attaatatag gttaaaacaa tgatttaa atgtttctcc caataattat ctctttccct
3660
ggaatcaact tgtatgaaac cttgtcaaaa tgtactccac aagtatgtac aattaagtat
3720
15 tttaaaaata aatggcaaac attaaaaaca aaaaaaaaaa aaaaaaaaaa
3769

<210> 19

20 <211> 2501

<212> DNA

<213> *Homo sapiens*

25 <400> 19

cgaaaagatt cttaggaacg ccgtaccagc cgcgtctctc aggacagcag gcccctgtcc
60
30 ttctgtcggg cgccgctcag ccgtgcctc cgcccctcag gttctttttc taattccaaa
120
taaacttgca agaggactat gaaagattat gatgaacttc tcaaatatta tgaattacat
180
35 gaaactattg ggacaggtgg ctttgcaaag gtcaaacttg cctgccatat ccttactgga
240
gagatggtag ctataaaaat catggataaa aacacactag ggagtgattt gccccggatc
300
40 aaaacggaga ttgaggcctt gaagaacctg agacatcagc atatatgtca actctaccat
360
gtgctagaga cagccaacaa aatattcatg gttcttgagt actgccctgg aggagagctg
45 420
tttgactata taatttccca ggatcgctg tcagaagagg agaccgggt tgccttccgt
480
50 cagatagtat ctgctgttgc ttatgtgcac agccagggct atgctcacag ggacctcaag
540
ccagaaaatt tgctgtttga tgaatatcat aaattaaagc tgattgactt tggctctctg
600
55 gcaaaacca aggtaacaa ggattaccat ctacagacat gctgtgggag tctggcttat
660
gcagcacctg agttaataca aggcaaatca tatcttgat cagaggcaga tgtttggagc
60 720
atgggcatac tgttatatgt tcttatgtgt ggatttctac catttgatga tgataatgta
780
65 atggctttat acaagaagat tatgagagga aaatatgatg ttccaagtg gctctctccc
840

ES 2 343 996 B1

agtagcattc tgcttcttca acaaatgctg caggtggacc caaagaaacg gatttctatg
 900
 5 aaaaatctat tgaaccatcc ctggatcatg caagattaca actatcctgt tgagtggcaa
 960
 agcaagaatc cttttattca cctc gatgat gattgcgtaa cagaactttc tgtacatcac
 1020
 10 agaaacaaca ggcaacaat ggaggattta atttactgt ggcagtatga tcacctcacg
 1080
 gctacctatc ttctgcttct agccaagaag gctcggggaa aaccagttcg ttttaaggctt
 1140
 15 tcttctttct cctgtggaca agccagtgt accccattca cagacatcaa gtcaaataat
 1200
 tggagtctgg aagatgtgac cgcaagtgat aaaaattatg tggcgggatt aatagactat
 1260
 20 gattggtgtg aagatgattt atcaacaggt gctgctactc cccgaacatc acagtttacc
 1320
 25 aagtactgga cagaatcaaa tggggtggaa tctaaatcat taactccagc cttatgcaga
 1380
 acacctgcaa ataaattaa gaacaaagaa aatgtatata ctctaagtc tgctgtaaag
 1440
 30 aatgaagagt actttatgtt tcctgagcca aagactccag ttaataagaa ccagcataag
 1500
 agagaaatac tcaactacgcc aaatcgttac actacaccct caaaagctag aaaccagtgc
 1560
 35 ctgaaagaaa ctccaattaa aataccagta aattcaacag gaacagacaa gttaatgaca
 1620
 40 ggtgtcatta gccctgagag gcggtgccgc tcagtggaat tggatctcaa ccaagcacat
 1680
 atggaggaga ctccaaaag aaaggagacc aaagtgtttg ggagccttga aagggggttg
 1740
 45 gataaggtta tcaactgtgct caccaggagc aaaaggaagg gttctgccag agacgggccc
 1800
 agaagactaa agcttacta taatgtgact acaactagat tagtgaatcc agatcaactg
 1860
 50 ttgaatgaaa taatgtctat tcttcaaag aagcatgttg actttgtaca aaagggttat
 1920
 aactgaagt gtcaaacaca gtcagatttt gggaaagtga caatgcaatt tgaattagaa
 1980
 55 gtgtgccagc ttcaaaaacc cgatgtggtg ggtatcagga ggcagcggct taagggcgat
 2040
 60 gcctgggttt acaaaagatt agtggagac atcctatcta gctgcaaggt ataattgatg
 2100
 gattcttcca tcctgccgga tgagtgtggg tgtgatacag cctacataaa gactgttatg
 2160
 65

ES 2 343 996 B1

atcgctttga ttttaaagtt cattggaact accaacttgt ttctaaagag ctatcttaag
2220
5 accaatatct ctttgTTTT aaacaaaaga tattatTTTg tgtatgaatc taaatcaagc
2280
ccatctgtca ttatgttact gtctTTTTta atcatgtggT tttgtatatt aataattgtt
2340
10 gactttctta gattcacttc catatgtgaa tgtaagctct taactatgtc tctttgtaat
2400
gtgtaatttc tttctgaaat aaaaccattt gtgaaataa aaaaaaaaaa aaaaaaaaaa
2460
15 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa a
2501

<210> 20
20 <211> 2731
<212> DNA
<213> *Homo sapiens*
25 <400> 20

gcgcttggcg ggagatagaa aagtgcttca acccgcgccg gcggcgactg cagttcctgc
60
30 gagcgaggag cgcgggacct gctgacacgc tgacgccttc gagcgcgcc cggggcccgg
120
agcggccgga gcagcccggg tcctgacccc ggcccggctc ccgctccggg ctctgccggc
180
35 gggcgggcga gcgcggcgcg gtccggggccg gggggatgtc tcggcggacg cgctgcgagg
240
atctggatga gctgcactac caggacacag attcagatgt gccggagcag agggatagca
300
40 agtgcaaggt caaatggacc catgaggagg acgagcagct gagggccctg gtgaggcagt
360
45 ttggacagca ggactggaag ttcttgcca gccacttccc taaccgact gaccagcaat
420
gccagtacag gtggctgaga gttttgaatc cagaccttgt caaggggcca tggaccaaaag
480
50 aggaagacca aaaagtcac gagctgggta agaagtatgg cacaagcag tggactga
540
ttgccaagca cctgaagggc cggtggggga agcagtgccg tgaacgctgg cacaaccacc
600
55 tcaaccctga ggtgaagaag tcttgctgga ccgaggagga ggaccgcatc atctgcgagg
660
cccacaaggt gctgggcaac cgctgggccg agatcgccaa gatggtgcca gggaggacag
720
60 acaatgctgt gaagaatcac tggaactcta ccatcaaaaag gaaggtggac acaggaggct
780
65 tcttgagcga gtccaaagac tgcaagcccc cagtgtactt gctgctggag ctcgaggaca
840

ES 2 343 996 B1

aggacggcct ccagagtgcc cagcccacgg aaggccaggg aagtcttctg accaactggc
900

5 cctccgtccc tcctaccata aaggaggagg aaaacagtga ggaggaactt gcagcagcca
960

ccacatcgaa ggaacaggag cccatcggta cagatctgga cgcagtgcga acaccagagc
1020

10 ccttgaggga attcccgaag cgtgaggacc aggaaggctc cccaccagaa acgagcctgc
1080

cttacaagtg ggtggtggag gcagctaacc tcctcatccc cgctgtgggt tctagcctct
1140

15 ctgaagccct ggacttgatc gagtcggacc ctgatgcttg gtgtgacctg agtaaatttg
1200

acctcccctga ggaaccatct gcagaggaca gtatcaaca cagcctagtg cagctgcaag
20 1260

cgtcacatca gcagcaagtc ctgccacccc gccagccttc cgccctgggtg cccagtgtga
1320

25 ccgagtaccg cctggatggc cacaccatct cagacctgag ccggagcagc cggggcgagc
1380

tgatccccat ctccccagc actgaagtcg ggggctctgg cattggcaca ccgccctctg
1440

30 tgctcaagcg gcagaggaag aggcgtgtgg ctctgtcccc tgctactgag aatagcacca
1500

gtctgtcctt cctggattcc tgtaacagcc tcacgcccac gagcacacct gttaagacct
35 1560

tgcccttctc gccctcccag tttctgaact tctggaacaa acaggacaca ttggagctgg
1620

40 agagcccctc gctgacatcc accccagtgt gcagccagaa ggtggtggtc accacaccac
1680

tgaccggga caagacacc ctgcaccaga aacatgctgc gtttgtaacc ccagatcaga
1740

45 agtactccat ggacaacact cccacacgc caacccggtt caagaacgcc ctggagaagt
1800

acggaccctt gaagcccctg ccacagacct cgcacctgga ggaggacttg aaggaggtgc
50 1860

tgcgttctga ggctggcatc gaactcatca tcgaggacga catcaggccc gagaagcaga
1920

55 agaggaagcc tgggctgcgg cggagcccca tcaagaaagt ccggaagtct ctggctcttg
1980

acattgtgga tgaggatgtg aagctgatga tgtccacact gcccaagtct ctatccttgc
2040

60 cgacaactgc cccttcaaac tcttcagcc tcaccctgtc aggtatcaaa gaagacaaca
2100

gcttgctcaa ccagggcttc ttgcaggcca agcccagaaa ggcagcagtg gcccagaagc
2160

65

ES 2 343 996 B1

cccgaagcca cttcacgaca cctgccccta tgtccagtgc ctggaagacg gtggcctgcg
2220

5 gggggaccag ggaccagctt ttcattgcagg agaaagcccg gcagctcctg ggccgcctga
2280

agcccagcca cacatctcgg accctcatct tgtcctgagg tgttgagggt gtcacgagcc
2340

10 cattctcatg ttacagggg ttgtgggggc agaggggggtc tgtgaatctg agagtcattc
2400

aggtgacctc ctgcagggag ctttctgcca ccagcccctc cccagactct caggtggagg
2460

15 caacagggcc atgtgctgcc ctgttgccga gccagctgt gggcggctcc tggtgctaac
2520

20 aacaaagttc cacttccagg tctgcctggt tccctcccca aggccacagg gagctccgtc
2580

agcttctccc aagcccacgt caggcctggc ctcatctcag accctgctta ggatggggga
2640

25 tgtggccagg ggtgctcctg tgctcacct ctcttggtgc atttttttgg aagaataaaa
2700

ttgcctctct cttaaaaaaa aaaaaaaaaa a
2731

30 <210> 21
<211> 1647
<212> DNA
35 <213> *Homo sapiens*

<400> 21

40 gcgcgcgggg ttcggtgacc cgcggcggtc acgggaattg ttcgctttag tgccggcgcc
60

atggggctcg agctgatcgg gcgcctagcc ccgcgctgg gcctcgccga gcccgacatg
120

45 ctgaggaaag cagaggagta cttgcgcctg tcccgggtga agtggtgctcg cctctccgca
180

cgaccacagg agaccagcag tgcagtcatg tgccctggacc ttgcagcttc ctggatgaag
240

50 tgccccttgg acagggctta ttttaattaa ctttctgggt tgaacaagga gacatatcag
300

agctgtctta aatcttttga gtgtttactg ggcctgaatt caaatattgg aataagagac
360

55 ctagctgtac agtttagctg tatagaagca gtgaacatgg cttcaaagat actaaaaagc
420

60 tatgagtcca gtcttcccca gacacagcaa gtggatcttg acttatccag gccacttttc
480

acttctgctg cactgctttc agcatgcaag attctaaagc tgaaagtgga taaaaacaaa
540

65 atggtagcca catccggtgt aaaaaaagct atatttgatc gactgtgtaa acaactagag

ES 2 343 996 B1

600

aagattggac agcaggtcga cagagaacct ggagatgtag ctactccacc acggaagaga
660

5

aagaagatag tggttgaagc cccagcaaag gaaatggaga aggtagagga gatgccacat
720

10

aaaccacaga aagatgaaga tctgacacag gattatgaag aatggaaaag aaaaattttg
780

gaaaatgctg ccagtgctca aaaggctaca gcagagtgat ttcagcttcc aaactggat
840

15

acattccaaa ctgatagtac attgccatct ccaggaagac ttgacggctt tgggattttg
900

tttaaacttt tataataagg atcctaagac tgttgccttt aaatagcaaa gcagcctacc
960

20

tggaggctaa gtctgggcag tgggctggcc cctggtgtga gcattagacc agccacagtg
1020

cctgattggt atagccttat gtgctttcct acaaaatgga attggaggcc gggcgcagtg
1080

25

gctcacgcct gtaatcccag cactttggga ggccaagggtg ggtggatcac ctgaggtcag
1140

gagctcgaga ccagcctggc caacatgggtg aaaccccatc tctactaaaa atacaaaaat
1200

30

tagccagggtg tgatggtgca tgcctgtaat cccagctcct cagtaggctg agacaggagc
1260

atcacttgaa cgtgggaggc agaggttgca gtgagccgag attgcaccac cgcactccag
1320

35

cctgggtgac agagcgagac ttatctcata aataaataga tagatactcc agcctgggtg
1380

40

acagagcgag acttatagat agatagatag atagatggat agatagatag atagatagat
1440

agatagataa acggaattgg agccattttg cttaaagtga atggcagtc cttgtcttat
1500

45

tcagaatata aaattcagtc tgaatggcat cttacagatt ttacttcaat ttttgtgtac
1560

ggatattttt atttgactaa atcaatatat tgtacagcct aagttaataa atgttattta
1620

50

tatatgcaaa aaaaaaaaa aaaaaaa
1647

55

<210> 22

<211> 13037

<212> DNA

60 <213> *Homo sapiens*

<400> 22

agtccacagc tgtcactaat cggggtaagc cttgttgtat ttgtgcgtgt ggggtggcatt
60

65

ES 2 343 996 B1

ctcaatgaga actagcttca cttgtcattt gagtgaatc tacaacccga ggcggctagt
120
5 gctcccgcac tactgggatc tgagatcttc ggagatgact gtcgccccga gtacggagcc
180
agcagaagtc cgacccttcc tgggaatggg ctgtaccgag aggtccgact agccccaggg
240
10 ttttagtgag ggggcagtgg aactcagcga gggactgaga gcttcacagc atgcacgagt
300
ttgatgccag agaaaaagtc gggagataaa ggagccgcgt gtcactaaat tgccgtcgca
360
15 gccgcagcca ctcaagtgcc ggacttgtga gtactctgcg tctccagtcc tcggacagaa
420
gttggagaac tctcttggag aactccccga gttaggagac gagatctcct aacaattact
20 480
actttttctt gcgctcccca cttgccgctc gctgggacaa acgacagcca cagttcccct
540
25 gacgacagga tggaggccaa gggcaggagc tgaccagcgc cgccctcccc cgccccgac
600
ccaggagggtg gagatccctc cgggccagcc acattcaaca cccactttct cctcccctctg
660
30 cccctatatt cccgaaacct cctcctcctt cccttttccc tcctcctgga gacgggggag
720
gagaaaaggg gagtccagtc gtcattgactg agctgaaggc aaagggctcc cgggctcccc
35 780
acgtggcggg cggccccgcc tccccgagg tcggatcccc actgctgtgt cgcccagccg
840
40 cagggtccgtt cccggggagc cagacctcgg acaccttgcc tgaagtttcg gccataccta
900
tctccctgga cgggctactc ttccctcggc cctgccaggg acaggacccc tccgacgaaa
960
45 agacgcagga ccagcagtcg ctgtcggacg tggagggcgc atattccaga gctgaagcta
1020
caaggggtgc tggaggcagc agttctagtc cccagaaaa ggacagcgga ctgctggaca
50 1080
gtgtcttggg cactctgttg gcgcccctcag gtccccggca gagccaacct agccctcccg
1140
cctgcgaggt caccagctct tggcgcctgt ttggccccga acttcccga gatccaccgg
55 1200
ctgccccgc caccagcgg gtgttgtccc cgctcatgag ccggtccggg tgcaaggttg
1260
60 gagacagctc cgggacggca gctgcccata aagtgctgcc ccggggcctg tcaccagccc
1320
ggcagctgct gctccccgcc tctgagagcc ctactggct cggggcccca gtgaagccgt
1380
65

ES 2 343 996 B1

ctccgcaggc cgctgcggtg gaggttgagg aggaggatgg ctctgagtcc gaggagtctg
1440

5 cgggtccgct tctgaagggc aaacctcggg ctctgggtgg cgcggcggct ggaggaggag
1500

ccgcggctgt cccgccgggg gcggcagcag gaggcgtcgc cctggtcccc aaggaagatt
1560

10 cccgcttctc agcgcccagg gtcgccttgg tggagcagga cgcgccgatg gcgcccgggc
1620

gctccccgct ggccaccacg gtgatggatt tcatccacgt gcctatcctg cctctcaatc
1680

15 acgccttatt ggagccccgc actcggcagc tgctggaaga cgaaagttac gacggcgggg
1740

ccggggctgc cagcgccttt gccccgccgc ggagttcacc ctgtgcctcg tccaccccgg
1800

20 tgcgtgtagg cgacttcccc gactgcgcgt acccgcccga cgccgagccc aaggacgacg
1860

25 cgtaccctct ctatagcgac ttccagccgc ccgctctaaa gataaaggag gaggaggaag
1920

gcgcggaggc ctccgcgcgc tccccgcgtt cctaccttgt ggccggtgcc aaccccgcag
1980

30 ccttcccgga tttcccgttg gggccaccgc ccccgtgcc gccgcgagcg accccatcca
2040

gacccgggga agcggcggtg acggccgcac ccgccagtgc ctcagtctcg tctgcgtcct
2100

35 cctcggggtc gaccctggag tgcatcctgt acaaagcggg gggcgcgccg ccccagcagg
2160

gcccgttcgc gccgccgcc tgcaaggcgc cgggcgcgag cggctgcctg ctcccgcggg
2220

40 acggcctgcc ctccacctcc gcctctgcc cgcgcccg ggccggcccc gcgctctacc
2280

45 ctgcactcgg cctcaacggg ctcccgcagc tcggctacca ggccgccgtg ctcaaggagg
2340

gcctgccgca ggtctaccgg ccctatctca actacctgag gccggattca gaagccagcc
2400

50 agagcccaca atacagcttc gagtcattac ctcagaagat ttgtttaatc tgtggggatg
2460

aagcatcagg ctgtcattat ggtgtcctta cctgtgggag ctgtaaggtc ttctttaaga
2520

55 gggcaatgga agggcagcac aactacttat gtgctggaag aaatgactgc atcgttgata
2580

60 aaatccgcag aaaaaactgc ccagcatgtc gccttagaaa gtgctgtcag gctggcatgg
2640

tccttggagg tcgaaaattt aaaaagttca ataaagtcag agttgtgaga gcactggatg
2700

65

ES 2 343 996 B1

ctgttgctct cccacagcca gtgggcgttc caaatgaaag ccaagcccta agccagagat
2760

5 tcactttttc accagggtcaa gacatacagt tgattccacc actgatcaac ctgttaatga
2820

gcattgaacc agatgtgatc tatgcaggac atgacaacac aaaacctgac acctccagtt
2880

10 ctttgctgac aagtcttaat caactaggcg agaggcaact tctttcagta gtcaagtgg
2940

ctaaatcatt gccagggtttt cgaaacttac atattgatga ccagataact ctcatcagtt
3000

15 attcttggat gagcttaatg gtgtttggtc taggatggag atcctacaaa cacgtcagtg
3060

ggcagatgct gtattttgca cctgatctaa tactaaatga acagcggatg aaagaatcat
20 3120

cattctattc attatgcctt accatgtggc agatcccaca ggagtttgtc aagcttcaag
3180

25 ttagccaaga agagttcctc tgtatgaaag tattgttact tcttaataca attcctttgg
3240

aagggctacg aagtcaaacc cagtttgagg agatgaggtc aagctacatt agagagctca
3300

30 tcaaggcaat tggtttgagg caaaaaggag ttgtgtcgag ctcacagcgt ttctatcaac
3360

ttacaaaact tcttgataac ttgcatgatc ttgtcaaaca acttcatctg tactgcttga
3420

35 atacatttat ccagtcccgg gcaactgagtg ttgaatttcc agaaatgatg tctgaagtta
3480

ttgctgcaca attaccaag atattggcag ggatggtgaa accccttctc tttcataaaa
40 3540

agtgaatgtc atctttttct tttaaagaat taaattttgt ggtatgtctt tttgttttgg
3600

45 tcaggattat gaggtcttga gtttttataa tgttcttctg aaagccttac atttataaca
3660

tcatagtgtg taaatttaaa agaaaaattg tgaggttcta attattttct tttataaagt
3720

50 ataattagaa tgtttaactg ttttgtttac ccatattttc ttgaagaatt tacaagattg
3780

aaaaagtact aaaattgtta aagtaaaacta tcttatccat attatttcat accatgtagg
55 3840

tgaggatttt taacttttgc atctaacaaa tcatcgactt aagagaaaaa atcttacatg
3900

60 taataacaca aagctattat atgttatttc taggtaactc cctttgtgtc aattatattt
3960

ccaaaaatga acctttaaaa tggatgcaa aattttgtct atatatattt gtgtgaggag
4020

65

ES 2 343 996 B1

gaaattcata actttcctca gattttcaaa agtattttta atgcaaaaaa tgtagaaaga
 4080
 5 gtttaaaacc actaaaatag attgatgttc ttcaaactag gcaaaacaac tcatatgtta
 4140
 agaccatttt ccagattgga aacacaaatc tcttaggaag ttaataagta gattcatatc
 4200
 10 attatgcaaa tagtattgtg ggttttgtag gtttttaaaa taaccttttt tggggagaga
 4260
 attgtcctct aatgaggat tgcgagtgga cataagaaat cagaagatta tggcctaact
 4320
 15 gtactcctta ccaactgtgg catgctgaaa gttagtcact cttactgatt ctcaattctc
 4380
 tcacctttga aagtagtaaa atatctttcc tgccaattgc tcctttgggt cagagcttat
 20 4440
 taacatcttt tcaaatcaaa ggaaagaaga aagggagagg aggaggaggg aggtatcaat
 4500
 25 tcacatacct ttctcctctt tctcctccac tatcatgaat tcatattatg tttcagccat
 4560
 gcaaatcttt ttaccatgaa atttcttcca gaattttccc ctttgacac aaattccatg
 4620
 30 catgtttcaa ccttcgagac tcagccaaat gtcatttctg taaaatcttc cctgagtctt
 4680
 ccaagcagta atttgccttc tcctagagtt tacctgccat tttgtgcaca tttgagttac
 35 4740
 agtagcatgt tattttacaa ttgtgactct cctgggagtc tgggagccat ataaagtgg
 4800
 40 caatagtgtt tgctgactga gagttgaatg acattttctc tctgtcttgg tattactgta
 4860
 gatttcgatc attccttggg tacatttctg catatttctg taccatgac tttatcactt
 4920
 45 tcttctccca tgctttatct ccatcaatta tcttcattac ttttaattt tccacctttg
 4980
 cttcctactt tgtgagatct ctccctttac tgactataac atagaagaat agaagtgtat
 50 5040
 tttatgtgtc ttaaggacaa tactttagat tccttgttct aagtttttaa actgaatgaa
 5100
 55 tggaatatta tttctctccc taagcaaaat tccacaaaac aattatttct tatgtttatg
 5160
 tagccttaaa ttgttttgta ctgtaaact cagcataaaa actttcttca tttctaattt
 5220
 60 cattcaacaa atattgattg aatacctggg attagcacia gaaaaatgtg ctaataagcc
 5280
 ttatgagaat ttggagctga agaaagacat ataactcagg aaagttacag tccagtagta
 5340
 65

ES 2 343 996 B1

ggataaatt acagtgccctg ataaataggc attttaatat ttgtacactc aacgtatact
5400

5 aggtaggtgc aaaacattta catataatth tactgatacc catgcagcac aaaggtaacta
5460

actttaaata ttaaataaca cctttatgtg tcagtaattc atttgcatta aatccttattg
5520

10 aaaaggcttt caatatatth tccccacaaa tgtcatccca agaaaaaagt atttttaaca
5580

tctcccaaat ataatagtta caggaaatct acctctgtga gagtgacacc tctcagaatg
5640

15 aactgtgtga cacaagaaaa tgaatgtagg tctatccaaa aaaaacccca agaaacaaaa
5700

acaatattat tagcccttha tgcttaagtg atggactcag ggaacagttg atgttgtgat
20 5760

cattttatta tctgattcct gttactttga attaaaccaa tattttgatg atataaatca
5820

25 tttccaccag catatattta atttccataa taactttaaa attttctaata ttcactcaac
5880

tatgagggaa tagaatgtgg tggccacagg tttggctttt gttaaaatgt ttgatatcct
5940

30 cgatgttgat ctctgtctgc aatgtagatg tctaaacact aggatttaata atttaaggct
6000

aagctttaaa aataaagtag ctttttaaaa agaataatggc ttcaccaaata ggaataatacc
35 6060

taattttctaa atctttttct ctacaaagtc ctatctacta atgtctccat tactatttag
6120

40 tcatcataac cattatcttc attttacatg tcgtgttctt tctggtagct ctaaaatgac
6180

actaaatcat aagaagacag gttacatatc aggaaatact tgaaggttac tgaataatagat
6240

45 tcttgagtta atgaaaaatata tttctgtaaa aaggtttgaa aagccatttg agtctaaagc
6300

attatacctc cattatcagt agttatgtga caattgtgtg tgtgtttaaata gtttaaagat
50 6360

gtggcacttt ttaataaggc aatgctatgc tattttttcc catttaacata taagataatt
6420

55 tattgctata cagatgatata ggaaatatga tgaacaataata tttttttgcc aaaactatgc
6480

cttgtaagta gccatggaata gtcaacctgt aacttaaata atccacagat agtcatgtgt
6540

60 ttgatgatgg gcactgtgga gataactgac ataggactgt gcccccttc tctgccactt
6600

actagctgga tgagattaag caagtcattt aactgctctg attaaacctg ctttcccaa
6660

65

ES 2 343 996 B1

gtgctttgta atgaatagaa atggaaacca aaaaaaacgt atacaggcct tcagaaatag
6720

5 taattgctac tattttgttt tcattaagcc atagtctgg ctataatfff atcaaactca
6780

ccagctatat tctacagtga aagcaggatt ctagaaagtc tcaactgtttt atttatgtca
6840

10 ccatgtgcta tgatatattt ggttgaattc atttgaaatt agggctggaa gtattcaagt
6900

aatttcttct gctgaaaaaa tacagtgttt tgagtttagg gcctgtttta tcaaagttct
6960

15 aaagagccta tcaactcttc attgtagaca ttttaaata atgacactga ttttaacatt
7020

tttaagtgtc tttttagaac agagagcctg actagaacac agcccctcca aaaacccatg
7080

ctcaaattat ttttactatg gcagcaattc cacaaaaggg aacaatgggt ttagaaatta
7140

25 caatgaagtc atcaaccaa aaaacatccc tatccctaag aaggttatga tataaaatgc
7200

ccacaagaaa tctatgtctg ctttaactctg tcttttattg ctttggaagg atggctatta
7260

30 catttttagt ttttgctgtg aatacctgag cagtttctct catccatact tctcttcac
7320

acatcagaag tcaggataga atatgaatca ttttaaaaac ttttacaact ccagagccat
7380

35 gtgcataaga agcattcaaa acttgccaaa acatacattt tttttcaaat ttaaagatac
7440

tctatttttg tattcaatag ctcaacaact gtggccccca ctgataaagt gaagtggaca
7500

aggagacaag taatggcata agtttgtttt tcccaaagta tgcctgttca atagccattg
7560

45 gatgtgggaa atttctacat ctcttaaaat tttacagaaa atacatagcc agatagtcta
7620

gcaaaagttc accaagtctt aaattgctta tccttacttc actaagtcatt gaaatcattt
7680

50 taatgaaaag aacatcacct aggttttggtg gtttcttttt ttcttattca tggctgagtg
7740

aaaacaacaa tctctgtttc tccctagcat ctgtggacta tttaatgtac cattattcca
7800

cactctatgg tccttactaa atacaaaatt gaacaaaaag cagtaaaaca actgactctt
7860

60 cacccatatt ataaaatata atccaagcca gattagtcaa catccataag atgaatccaa
7920

gctgaaactgg gcctagatta ttgagttcag gttggatcac atccctattt attaataaac
7980

65

ES 2 343 996 B1

ttaggaaaga aggccttaca gaccatcagt tagctggagc taatagaacc tacacttcta
8040

5 aagttcggcc tagaatcaat gtggccttaa aagctgaaaa gaagcaggaa agaacagttt
8100

tcttcaataa tttgtccacc ctgtcactgg agaaaattta agaatttggg ggtgttggtta
8160

10 gtaagttaaa cacagcagct gttcatggca gaaattattc aatacatacc ttctctgaat
8220

atcctataac caaagcaaag aaaaacacca aggggtttgt tctcctcctt ggagttgacc
8280

15 tcattccaag gcagagctca ggtcacaggc acaggggctg cgcccaagct tgtccgcagc
8340

cttatgcagc tgtggagtct ggaagactgt tgcaggactg ctggcctagt cccagaatgt
8400

20 cagcctcatt ttcgatttac tggctcttgt tgctgtatgt catgctgacc ttattgttaa
8460

25 acacaggttt gtttgctttt tttccactca tggagacatg ggagaggcat tatttttaag
8520

ctggttgaaa gctttaaccg ataaagcatt tttagagaaa tgtgaatcag gcagctaaga
8580

30 aagcatactc tgtccattac ggtaaagaaa atgcacagat tattaactct gcagtgtggc
8640

attagtgtcc tggatcaatat tcggatagat atgaataaaa tatttaaatg gtattgtaaa
8700

35 tagttttcag gacatatgct atagcttatt tttattatct tttgaaattg ctcttaatac
8760

atcaaatcct gatgtattca atttatcaga tataaattat tctaaatgaa gcccagttaa
8820

atgtttttgt cttgtcagtt atatgttaag tttctgatct ctttgtctat gacgtttact
8880

45 aatctgcatt tttactgtta tgaattattt tagacagcag tggtttcaag ctttttgcca
8940

ctaaaaatac cttttatttt ctctccccc agaaaagtct ataccttgaa gtatctatcc
9000

50 accaaactgt acttctatta agaaatagtt attgtgtttt cttaatgttt tgttattcaa
9060

agacatatca atgaaagctg ctgagcagca tgaataacaa ttatatccac acagatttga
9120

tatattttgt gcagccttaa cttgatagta taaaatgtca ttgcttttta aataatagtt
9180

60 agtcaatgga cttctatcat agctttccta aactaggtta agatccagag ctttggggct
9240

ataatatatt acatacaatt aagttatctt tttctaaggg ctttaaaatt catgagaata
9300

65

ES 2 343 996 B1

accaaaaaag gtatgtggag agttaataca aacataccat attcctgttg aaacagagat
 9360
 5 gtggctctgc ttgttctcca taaggtagaa atactttcca gaatttgcct aaactagtaa
 9420
 gccctgaatt tgctatgatt agggatagga agagattttc acatggcaga ctttagaatt
 9480
 10 cttcacttta gccagtaaag tatctccttt tgatcttagt attctgtgta ttttaacttt
 9540
 tctgagttgt gcatgtttat aagaaaaatc agcacaagg gtttaagtta aagccttttt
 9600
 15 actgaaatth gaaagaaaca gaagaaaata tcaaagttct ttgtattttg agaggattaa
 9660
 20 atatgattta caaaagttac atggagggct ctctaaaaca ttaaattaat tattttttgt
 9720
 tgaaaagtct tacttttagc atcattttat tcctcagcaa ctagctgtga agcctttact
 9780
 25 gtgctgtatg ccagtcactc tgctagattg tggagattac cagtgttccc gtcttctccg
 9840
 agcttagagt tggatgggga ataaagacag gtaaacagat agctacaata ttgtactgtg
 9900
 30 aatgcttatg ctggaggaag tacaggaac tattggagca cctaagagga gcacctacct
 9960
 tgaatttagg ggtagcaga ggcacctga aaaaagtcaa agctaagcca caatctataa
 10020
 35 gcagtttagg aattagcaga acgtgcgtgg tgaggagatg ccaaaggcaa gaagagaaga
 10080
 40 gtattccaaa caggaggat tccaaagaga gaagagtatc ccaaacaaca tttgcacaaa
 10140
 cctgatgggg agagagaatg tggggtgggg atggatgatg agactgaaga agaaagccag
 10200
 45 gtctagataa tcagtggcct tgtacacat gttaaagagt gtagacttga ttctgttgta
 10260
 aacaggaaag cagcacaatt catatgaata ttttagaaga ctcccactgg aatatggaga
 10320
 50 ataaagttgg agatgactaa tcctggaagc agggagaaca tttttgagga agttgacta
 10380
 55 ttttggtgaa aatgatgatc ataaacatga agaattgtag gtgatcatga cctcctctct
 10440
 aatthttccag aagggttttg gaagatataa cataggaaca ttgacaggac tgacgaaagg
 10500
 60 agatgaaata caccatataa attgtcaaac acaaggccag atgtctaatt atthttgctta
 10560
 tgtgttgaaa ttacaaatth tcatcagga aacaaaaaac tacaaaactt agthtttcca
 10620
 65

ES 2 343 996 B1

agtcccagaa ttctatctgt ccaaacaatc tgtaccactc cacctatatc cctacctttg
10680

5 catgtctgtc caacctcaaa gtccaggtct atacacacgg gtaagactag agcagttcaa
10740

gtttcagaaa atgagaaaga ggaactgagt tgtgctgaac ccatacaaaa taaacacatt
10800

10 ctttgtatag attcttggaa cctcgagagg aattcaccta actcataggt atttgatggt
10860

atgaatccat ggctgggctc ggcttttaaa aagccttatac tgggattcct tctatggaac
10920

15 caagttccat caaagcccat ttaaaagcct acattaataaa caaaattcct gctgcattgt
10980

atacaataa tgatgtcatg atcaataat cagatgccat tatcaagtgg aattacaaaa
20 11040

tggtataccc actccaaaaa aaaaaaaaaa gctaaattct cagtagaaca ttgtgacttc
11100

25 atgagccctc cacagccttg gagctgagga gggagcactg gtgagcagta ggttgaagag
11160

aaaacttggc gcttaataat ctatccatgt tttttcatct aaaagagcct tctttttgga
11220

30 ttaccttatt caatttccat caaggaaatt gttagttcca ctaaccagac agcagctggg
11280

aaggcagaag cttactgtat gtacatggta gctgtgggaa ggaggtttct ttctccaggt
35 11340

cctcactggc catacaccag tcccttgta gttatgcctg gtcataagacc cccgttgcta
11400

40 tcatctcata ttttaagtctt tggcttgtga atttatctat tctttcagct tcagcactgc
11460

agagtgtctg gactttgcta acttccattt cttgctggct tagcacattc ctcataggcc
11520

45 cagctctttt ctcatctggc cctgctgtgg agtcaccttg ccccttcagg agagccatgg
11580

cttaccactg cctgctaagc ctccactcag ctgccaccac actaaatcca agcttctcta
11640

50 agatgttgca gactttacag gcaagcataa aaggcttgat cttcctggac ttccctttac
11700

ttgtctgaat ctcacctct tcaactttca gtctcagaat gtaggcattt gtcctctttg
55 11760

ccctacatct tccttcttct gaatcatgaa agcctctcac ttctcttgc tatgtgctgg
11820

60 aggcttctgt caggttttag aatgagttct catctagtcc tagtagcttt tgatgcttaa
11880

gtccaccttt taaggatacc tttgagattt agaccatggt tttcgcttga gaaagcccta
11940

65

ES 2 343 996 B1

atctccagac ttgcctttct gtggatttca aagaccaact gaggaagtca aaagctgaat
12000

5 gttgactttc tttgaacatt tccgctataa caattccaat tctcctcaga gcaatatgcc
12060

tgctccaac tgaccaggag aaaggtccag tgccaaagag aaaaacacaa agattaatta
12120

10 tttcagttga gcacatactt tcaaagtggg ttgggtattc atatgagggt ttctgtcaag
12180

agggtgagac tcttcatcta tccatgtgtg cctgacagtt ctctggcac tggctggtaa
12240

15 cagatgcaaa actgtaaaaa ttaagtgatc atgtatttta acgatatcat cacatactta
12300

20 ttttctatgt aatgttttaa atttccccta acatactttg actgttttgc acatggtaga
12360

tattcacatt tttttgtggt gaagttgatg caatcttcaa agttatctac cccgttgctt
12420

25 attagtaaaa ctagtgttaa tacttggcaa gagatgcagg gaatctttct catgactcac
12480

gccctattta gttattaatg ctactaccct attttgagta agtagtaggt ccctaagtac
12540

30 attgtccaga gttatacttt taaagatatt tagccccata tacttcttga atctaaagtc
12600

35 atacaccttg ctctcattt ctgagtggga aagacatttg agagtatggt gacaattggt
12660

ctgaagggtt ttgccaagaa ggtgaaactg tcctttcatc tgtgtatgcc tggggctggg
12720

40 tccctggcag tgatgggggtg acaatgcaaa gctgtaaaaa ctagggtgcta gtgggcacct
12780

aatatcatca tcatatactt attttcaagc taatatgcaa aatcccatct ctgtttttaa
12840

45 actaagtgta gatttcagag aaaatatttt gtggttcaca taagaaaaca gtctactcag
12900

cttgacaagt gttttatggt aaattggctg gtggtttgaa atgaatcatc ttcacataat
12960

50 gttttcttta aaaatattgt gaatttaact ctaattcttg ttattctgtg tgataataaa
13020

55 gaataaacta atttcta
13037

<210> 23

<211> 3183

60 <212> DNA

<213> *Homo sapiens*

<400> 23

65 gacagaagca atggccgagg cagaagacaa gccgaggtgc tggtgaccct gggcgtctga
60

ES 2 343 996 B1

gtggatgatt ggggctgctg cgctcagagg cctgcctccc tgccttccaa tgcataatac
120

5 cccacacccc agccaatgaa gacgagagggc agcgtgaaca aagtcattta gaaagccccc
180

gaggaagtgt aaacaaaaga gaaagcatga atggagtgcc tgagagacaa gtgtgtcctg
240

10 tactgcccc accttttagct gggccagcaa ctgcccggcc ctgcttctcc ccacctactc
300

actgggtgatc tttttttttt tacttttttt tcccttttct tttccattct cttttcttat
360

15 tttctttcaa ggcaaggcaa ggattttgat tttgggacct agccatggtc cttctgcttc
420

ttcttttaaaa taccactttt ctccccatcg ccaagcggcg tttggcaata tcagatatcc
20 480

actctattta tttttaccta aggaaaaact ccagctccct tcccactccc agctgccttg
540

25 ccaccctcc cagccctctg cttgccctcc acctggcctg ctgggagtca gagcccagca
600

aaacctgttt agacacatgg acaagaatcc cagcgtaca aggcacacag tccgcttctt
660

30 cgtcctcagg gttgccagcg cttcctggaa gtcctgaagc tctcgcagtg cagtgagttc
720

atgcaccttc ttgccaagcc tcagtctttg ggatctgggg aggccgctg gttttcctcc
35 780

ctccttctgc acgtctgctg gggctctctc ctctccaggc cttgccgtcc ccctggcctc
840

40 tcttcccagc tcacacatga agatgcactt gcaaagggct ctggtgggtcc tggccctgct
900

gaactttgcc acggtcagcc tctctctgtc cacttgcacc accttggact tcggccacat
960

45 caagaagaag aggggtggaag ccattagggg acagatcttg agcaagctca ggctcaccag
1020

ccccctgag ccaacggtga tgaccacgt ccctatcag gtcctggccc tttacaacag
50 1080

caccgggag ctgctggagg agatgcatgg ggagagggag gaaggctgca cccaggaaaa
1140

caccgagtcg gaatactatg ccaaagaaat ccataaattc gacatgatcc aggggctggc
55 1200

ggagcacaac gaactggctg tctgccctaa aggaattacc tccaaggttt tccgcttcaa
1260

60 tgtgtcctca gtggagaaaa atagaaccaa cctattccga gcagaattcc gggctctgctg
1320

ggtgcccaac cccagctcta agcggaatga gcagaggatc gagctcttcc agatccttcg
1380

65

ES 2 343 996 B1

gccagatgag cacattgcca aacagcgcta tatcgggtggc aagaatctgc ccacacgggg
1440

5 cactgccgag tggctgtcct ttgatgtcac tgacactgtg cgtgagtggc tgttgagaag
1500

agagtccaac ttaggtctag aaatcagcat tcactgtcca tgtcacacct ttcagcccaa
1560

10 tggagatatt ctggaaaaca ttcacgaggt gatggaaatc aaattcaaag gcgtggacaa
1620

tgaggatgac catggccgtg gagatctggg ggcctcaag aagcagaagg atcaccacaa
1680

15 ccctcatcta atcctcatga tgattcccc acaccggctc gacaacccgg gccagggggg
1740

tcagaggaag aagcgggctt tggacaccaa ttactgcttc cgcaacttgg aggagaactg
1800

20 ctgtgtgctc cccctctaca ttgacttccg acaggatctg ggctggaagt gggctccatga
1860

25 acctaagggc tactatgcca acttctgctc aggcccttgc ccatacctcc gcagtgcaga
1920

cacaaccac agcacggtgc tgggactgta caacactctg aaccctgaag catctgcctc
1980

30 gccttgctgc gtgccccagg acctggagcc cctgaccatc ctgtactatg ttgggaggac
2040

35 ccccaaagtg gagcagctct ccaacatggt ggtgaagtct tgtaaagtga gctgagaccc
2100

cacgtgctgc agagagaggg gagagagaac caccactgcc tgactgcccg ctctcggga
2160

40 aacacacaag caacaaacct cactgagagg cctggagccc acaaccttcg gctccgggca
2220

aatggctgag atggaggttt ctttttgaa cttttcttc ttgctggctc tgagaatcac
2280

45 ggtggtaaag aaagtgtggg tttggttaga ggaaggctga actcttcaga acacacagac
2340

tttctgtgac gcagacagag gggatgggga tagaggaaag ggatggtaag ttgagatggt
2400

50 gtgtggcaat gggatttggg ctaccctaaa gggagaagga agggcagaga atggctgggt
2460

cagggccaga ctggaagaca cttcagatct gaggttggat ttgctcattg ctgtaccaca
2520

55 tctgctctag ggaatctgga ttatgttata caaggcaagc atttttttt tttttttaa
2580

60 gacaggttac gaagacaaag tcccagaatt gtatctcata ctgtctggga ttaagggcaa
2640

atctattact tttgcaaact gtcctctaca tcaattaaca tcgtgggtca ctacagggag
2700

65

ES 2 343 996 B1

aaaatccagg tcatgcagtt cctggcccat caactgtatt gggccttttg gatatgctga
2760

5 acgcagaaga aaggggtggaa atcaaccctc tcctgtctgc cctctgggtc cctcctctca
2820

cctctccctc gatcatattt ccccttggac acttggttag acgccttcca ggtcaggatg
2880

10 cacatttctg gattgtggtt ccatgcagcc ttggggcatt atgggttctt cccccacttc
2940

ccctccaaga ccctgtgttc atttgggtgtt cctggaagca ggtgctacaa catgtgaggc
3000

15 attcggggaa gctgcacatg tgccacacag tgacttggcc ccagacgcat agactgaggt
3060

20 ataaagacaa gtatgaatat tactctcaaa atctttgtat aaataaatat ttttggggca
3120

tcctggatga tttcatcttc tggaatattg tttctagaac agtaaaagcc ttattctaag
3180

25 gtg
3183

30

35

40

45

50

55

60

65



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

① ES 2 343 996

② N° de solicitud: 200803509

③ Fecha de presentación de la solicitud: 11.12.2008

④ Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TÉCNICA

⑤ Int. Cl.: **C12Q 1/68** (2006.01)
G01N 33/50 (2006.01)

DOCUMENTOS RELEVANTES

Categoría	⑥ Documentos citados	Reivindicaciones afectadas
X	US 20040058340 A1 (DAI et al.) 25.03.2004, resumen; página 2, párrafo 15; página 5, párrafo 27 - página 6, párrafo 28; tablas 1,5; reivindicaciones 1,19,31.	1,4,7,8
Y		2,17
Y	WO 2006052218 A1 (AGENCY FOR SCIENCE, TECHNOLOGY AND RESEARCH [SG/SG]). 18.05.2006, resumen; tabla 1; reivindicaciones 1-3,10.	2,17
X	ESPINOSA et al. Breast cancer prognosis determined by gene expression profiling: a quantitative reverse transcriptase polymerase chain reaction study. Journal of Clinical Oncology. 2005, Vol 23(29); páginas 7278-7285, especialmente páginas 7279 (pacientes y métodos) y 7281 (discusión).	4-12
X	LI et al. Astrocyte Elevated Gene-1 is a novel prognostic marker for breast cancer progression and overall patient survival. Clinical Cancer research. 2008, Vol 14(11); páginas 3319-3326. Todo el documento.	3
X	SKUBITZ et al. Differential gene expression identifies subgroups of ovarian carcinoma. Translational research. 2005, Vol 148(5); páginas 223-248, especialmente página 228 y tabla 4.	13-15

Categoría de los documentos citados

X: de particular relevancia

Y: de particular relevancia combinado con otro/s de la misma categoría

A: refleja el estado de la técnica

O: referido a divulgación no escrita

P: publicado entre la fecha de prioridad y la de presentación de la solicitud

E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

para todas las reivindicaciones

para las reivindicaciones nº:

Fecha de realización del informe

30.07.2010

Examinador

Mª D. García Grávalos

Página

1/5

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

C12Q, G01N

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

EPODOC, WPI, USPTO PATENT DATABASE, INVENES, BIOSIS, MEDLINE, EMBASE, SCIENCEDIRECT

Fecha de Realización de la Opinión Escrita: 30.07.2010

Declaración

Novedad (Art. 6.1 LP 11/1986)	Reivindicaciones	2, 17	SÍ
	Reivindicaciones	1, 3-16	NO
Actividad inventiva (Art. 8.1 LP 11/1986)	Reivindicaciones		SÍ
	Reivindicaciones	1-17	NO

Se considera que la solicitud cumple con el requisito de **aplicación industrial**. Este requisito fue evaluado durante la fase de examen formal y técnico de la solicitud (Artículo 31.2 Ley 11/1986).

Base de la Opinión:

La presente opinión se ha realizado sobre la base de la solicitud de patente tal y como ha sido publicada.

1. Documentos considerados:

A continuación se relacionan los documentos pertenecientes al estado de la técnica tomados en consideración para la realización de esta opinión.

Documento	Número Publicación o Identificación	Fecha Publicación
D01	US 20040058340 A1	25-03-2004
D02	WO 2006052218 A1	18-05-2006
D03	ESPINOSA et al. 2005, Vol 23(29); páginas 7278-7285.	2005
D04	LI et al. 2008, Vol 14(11); páginas 3319-3326.	01-06-2008
D05	SKUBITZ et al. 2006, Vol 148(5), páginas 223-248.	2006

2. Declaración motivada según los artículos 29.6 y 29.7 del Reglamento de ejecución de la Ley 11/1986, de 20 de marzo, de patentes sobre la novedad y la actividad inventiva; citas y explicaciones en apoyo de esta declaración

La presente solicitud divulga un método de pronóstico de la evolución de tumores, en especial de tumores de mama, ginecológicos o próstata (reivindicaciones 1-15). Dicho método permite la subclasificación de los tumores lo que es útil para tomar decisiones con respecto al tratamiento a administrar al paciente. Asimismo divulga un kit para llevar a cabo dicho método (reivindicaciones 16 y 17). El método de la invención comprende la obtención de una muestra biológica con células tumorales, la detección en dicha muestra de la cantidad del producto de la expresión de entre 2 y 8 genes siendo uno de ellos PRC1 en combinación con DTL, ECT2, MTDH, RFC4, SCUBE2, STK32B o ZNF533 y por último la comparación de la cantidad detectada con una cantidad de referencia (reivindicación 1). La reivindicación 2 divulga el mismo método pero especificando que se detecte el nivel de expresión de los 8 genes. Una etapa adicional a dicho método especifica que una cantidad de producto de expresión de los genes PRC1, DTL, ECT2, MTDH o RFC4 mayor que la de referencia, o una cantidad de producto de expresión de los genes SCUBE2, STK32B o ZNF533 menor que la de referencia es indicativa de una menor supervivencia libre de recaída a distancia o una menor supervivencia global (reivindicación 3). Preferiblemente, el producto de expresión es mRNA, su detección se realiza mediante RT-PCR cuantitativa y la muestra biológica aislada que comprende las células tumorales está fijada y embebida en parafina (reivindicación 4-6). Las reivindicaciones 7-15 especifican el tipo de tumores de mama, ginecológicos o de próstata para los que está indicado el método de subclasificación de la presente solicitud. Por último, el kit para la subclasificación de tumores comprende tanto los cebadores necesarios para detectar PRC1, que hibridan entre los nucleótidos 106 y 238 de dicho gen, como los demás reactivos necesarios para la detección de los otros 7 genes (reivindicaciones 16 y 17).

El documento D01 divulga grupos de marcadores genéticos cuyo patrón de expresión puede ser utilizado para diferenciar distintas condiciones clínicas asociadas con el cáncer de mama. Asimismo, dicho patrón de expresión puede ser usado para proporcionar información acerca de la posibilidad de sufrir metástasis a distancia en los 5 años siguientes al diagnóstico inicial. Igualmente, divulga métodos para la clasificación y el tratamiento de pacientes en base al pronóstico obtenido mediante la medida de la expresión de algunos de dichos marcadores. Por último, divulga kits que contienen microarrays para la detección de los marcadores y software para el análisis de los datos obtenidos. Entre los múltiples marcadores divulgados se encuentran PRC1, DTL, ECT2, MTDH, RFC4, SCUBE2 y STK32B (ver resumen; página 2, párrafo 15; página 5, párrafo 27 - página 6, párrafo 28; tablas 1 y 5, secuencias 977, 1645, 1739, 219, 838, 1844 y 1782; reivindicaciones 1, 19, 31).

El documento D02 divulga métodos, sistemas y composiciones para predecir la susceptibilidad de un paciente a padecer cáncer, basados en la correlación del estado mutacional de p53 y el perfil de expresión génica de ciertos genes. En particular, divulga métodos para la clasificación, pronóstico y diagnóstico de cánceres, incluyendo el cáncer de mama. Específicamente, describen un método que comprende varios pasos, siendo uno la obtención del perfil de expresión de distintos genes, entre ellos PRC1, ECT2, SCUBE2 y ZNF533, a partir de muestras de tumores (ver resumen, tabla 1, reivindicaciones 1-3, 10).

El documento D03 divulga el análisis de la expresión de 60 de los 70 genes descritos en Van't Veer et al. (2002, mencionado en la solicitud de la presente invención) y conocidos como "70-gene signature". Las muestras biológicas fueron aisladas de pacientes con carcinoma de mama invasivo en estadio temprano (estadio I o II), siendo un 75% de ellos positivos para la expresión de los receptores hormonales de estrógenos. Las muestras podían provenir también de pacientes sometidos a mastectomía o lumpectomía seguida de radioterapia. Para la detección de la expresión génica emplean RT-PCR cuantitativa y muestras fijadas y embebidas en parafina (ver páginas 7279 (pacientes y métodos) y 7281 (discusión)).

Hoja adicional

El documento D04 divulga MTDH (AEG1) como un nuevo marcador de pronóstico de la progresión del cáncer de mama y de la supervivencia global del paciente. En concreto, divulga que la sobreexpresión de MTDH está asociada con una menor supervivencia global del paciente (ver todo el documento).

El documento D05 divulga una subclasificación de carcinomas de ovario, en particular carcinomas serosos de ovario, en base a la expresión diferencial de ciertos genes. Entre ellos se encuentran PRC1, RFC4 y ECT2 (ver página 228 y tabla 4).

1. NOVEDAD (Art. 6.1 LP 11/1986).

1.1. Reivindicaciones 1-17.

El documento D01 se considera el más cercano al estado de la técnica en cuanto a la medición de la expresión de diferentes genes en muestras tumorales de pacientes para la subclasificación de cánceres, en particular cáncer de mama. El documento D01 anticipa el uso de la medición de la expresión de PRC1, DTL, ECT2, MTDH, RFC4, SCUBE2 y STK32B, mediante microarrays, como marcadores para el pronóstico del cáncer de mama.

El documento D03 se considera el más cercano al estado de la técnica en relación tanto a la técnica de detección del nivel de expresión génica como al tipo de tumores clasificables por el método de la invención. La expresión génica se mide por RT-PCR cuantitativa y los cánceres analizados cumplen los requisitos especificados en las reivindicaciones 7-12.

El documento D04 anticipa la sobreexpresión de MTDH como marcador de una menor supervivencia global de pacientes con cáncer de mama.

El documento D05 anticipa la detección de la expresión de diversos genes, entre ellos PRC1, RFC4 y ECT2, para la subclasificación de cánceres de ovario.

Por el contrario, ninguno de estos documentos anticipa la detección de los ocho genes PRC1, DTL, ECT2, MTDH, RFC4, SCUBE2, STK32B y ZNF533 para la subclasificación de tumores.

En consecuencia, el objeto de la invención tal y como se reivindica en las reivindicaciones 1,3-16 se encuentra anticipado en los documentos D01, D03, D04 y D05. Por lo tanto, dichas reivindicaciones 1,3-16 carecen de novedad (Art. 6.1 LP 11/1986).

Sin embargo, las reivindicaciones 2 y 17 cumplen con el requisito de novedad (Art. 6.1 LP 11/1986).

2. ACTIVIDAD INVENTIVA (Art. 8.1 LP 11/1986).

2.1. Reivindicaciones 2, 17

La diferencia entre el documento D01 y el objeto técnico de las reivindicaciones 2 y 17 de la presente solicitud radica en la detección del nivel de expresión de los ocho genes PRC1, DTL, ECT2, MTDH, RFC4, SCUBE2, STK32B y ZNF533.

El documento D02 anticipa la medida de la expresión de, entre otros, PRC1, ECT2, SCUBE2 y ZNF533 como marcadores para la clasificación y pronóstico de cánceres.

Por lo tanto, según lo divulgado en el documento D01 en combinación con el D02, la medida del nivel de expresión de los ocho genes divulgados en la presente invención podría considerarse una alternativa tal que un experto en la materia la tendría en consideración sin esfuerzo inventivo.

De acuerdo con lo expuesto, las reivindicaciones 2 y 17 no cumplen con el requisito de actividad inventiva (Art. 8.1 LP 11/1986).