

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 504 242**

51 Int. Cl.:

C12Q 1/68 (2006.01)

G01N 33/50 (2006.01)

G01N 33/53 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **18.02.2005 E 10178199 (5)**

97 Fecha y número de publicación de la concesión europea: **02.07.2014 EP 2333112**

54 Título: **Pronóstico de cáncer de mama**

30 Prioridad:

20.02.2004 US 783271
08.12.2004 US 634430 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
08.10.2014

73 Titular/es:

JANSSEN DIAGNOSTICS, LLC (100.0%)
700 US Highway 202
Raritan, NJ 08869, US

72 Inventor/es:

WANG, YIXIN

74 Agente/Representante:

IZQUIERDO FACES, José

ES 2 504 242 T3

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

Pronóstico de cáncer de mama**Descripción**

5 CAMPO DE LA INVENCION

La presente invención se refiere al pronóstico de pacientes con cáncer de mama en base a los perfiles de expresión génica de muestras biológicas de los pacientes.

10 ANTECEDENTES DE LA TÉCNICA

15 Aproximadamente el 60%-70% de los pacientes con cáncer de mama sin compromiso de los ganglios linfáticos (LNN) se curan mediante tratamiento local o regional solamente. Lancet (1998a) y Lancet (1998b). Se han desarrollado directrices para ayudar a los médicos a seleccionar a los pacientes que deben recibir tratamiento complementario. Las recomendaciones más ampliamente utilizadas son los criterios de St. Gallen (Goldhirsch *et al.* (2003)) y los criterios de consenso de los National Institutes of Health (NIH). Eifel *et al.* (2001). Estas directrices señalan al 85%-90% de los pacientes LNN como candidatos al tratamiento complementario sistémico.

20 Existe la necesidad de identificar específicamente el riesgo de reaparición de la enfermedad de un paciente para garantizar que recibe el tratamiento apropiado. Actualmente, existen pocas herramientas de diagnóstico disponibles para identificar a los pacientes en riesgo. Se han utilizado patrones de expresión génica para clasificar los tumores de mama en diferentes subtipos clínicamente pertinentes. Perou *et al.* (2000), Sorlie *et al.* (2001), Sorlie *et al.* (2003), Gruvberger *et al.* (2001), van't Veer *et al.* (2002), van de Vijver *et al.* (2002), Ahr *et al.* (2002), Huang *et al.* (2003), Sotiriou *et al.* (2003), Woelfle *et al.* (2003), Ma *et al.* (2003), Ramaswamy *et al.* (2003), Chang *et al.* (2003), Sotiriou *et al.* (2003) y Hedenfalk *et al.* (2001).

30 Actualmente, en pacientes LNN, la decisión de aplicar un tratamiento complementario después de la extirpación quirúrgica del tumor primario, y qué tipo (endocrinoterapia y/o quimioterapia), depende en gran medida de la edad del paciente, del estado menopáusico, del tamaño del tumor, del grado de diferenciación y del estado de los receptores de hormonas esteroideas. Estos factores se tienen en cuenta en directrices tales como los criterios de St. Gallen y los criterios de consenso de los National Institutes of Health (NIH). En base a estos criterios, más del 85%-90% de los pacientes LNN serían candidatos a recibir tratamiento complementario sistémico. Existe una clara necesidad de identificar mejor los factores pronósticos para guiar la selección de las opciones de tratamiento.

35 RESUMEN DE LA INVENCION

40 Reconociendo la complejidad de la evolución de la enfermedad, los autores informan en el presente documento acerca de una completa evaluación de la expresión génica de todo el genoma para identificar marcadores pronósticos ampliamente aplicables. Ntzani *et al.* (2003) y Wang *et al.* (2004).

La presente invención abarca un método para evaluar el estado de un cáncer de mama en una muestra biológica obtenida de un paciente con cáncer de mama tal como se define en las reivindicaciones.

45 La presente invención abarca un método para estadificar un cáncer de mama en una muestra biológica obtenida de un paciente con cáncer de mama tal como se define en las reivindicaciones.

50 La presente invención abarca un método para determinar un protocolo de tratamiento del paciente con cáncer de mama en una muestra biológica obtenida de un paciente con cáncer de mama tal como se define en las reivindicaciones.

55 La presente descripción proporciona un método para tratar a un paciente con cáncer de mama obteniendo una muestra biológica de un paciente con cáncer de mama; y midiendo los niveles de expresión en la muestra de genes seleccionados de entre aquellos que codifican un ARNm: correspondiente a las SEQ ID NO: 1-111; o reconocido por los conjuntos de sondas seleccionadas de entre los psids correspondientes a las SEQ ID NO: 1-111 como se muestra en la Tabla 10, en el que los niveles de expresión génica por encima o por debajo de los niveles límite predeterminados son indicativos de un alto riesgo de reaparición y; tratando al paciente con el tratamiento complementario si es un paciente de alto riesgo.

60 La presente descripción proporciona un método para realizar la validación cruzada de un perfil de expresión génica pronóstico para pacientes con cáncer de mama obteniendo datos de expresión génica a partir de un número estadísticamente significativo de muestras biológicas de pacientes; aleatorizando el orden de las muestras; dejando fuera los datos de aproximadamente el 10%-50% de las muestras; calculando, para las muestras restantes, el factor de interés en todas las variables y seleccionando las variables que satisfagan un límite del valor p (p); seleccionando las variables que se ajustan a un modelo de predicción utilizando una búsqueda hacia adelante y evaluando el error de aprendizaje hasta que dé con una tasa de error predeterminada; ensayando el modelo de predicción en el 10%-50% de las muestras dejadas fuera; repitiendo las etapas c., -g. con un nuevo conjunto de muestras retiradas; y

siguiendo con las etapas c) -g) hasta haber ensayado el 100% de las muestras, y registrar el rendimiento de la clasificación.

5 La presente descripción proporciona un método para validar independientemente un perfil de expresión génica pronóstico para pacientes con cáncer de mama obteniendo datos de expresión génica a partir de un número estadísticamente significativo de muestras biológicas de pacientes; normalizando las variabilidades de partida en los datos de expresión génica; calculando el factor de interés en todas las variables que se han seleccionado con anterioridad; y ensayando el modelo de predicción en la muestra, y registrar el rendimiento de la clasificación.

10 La presente descripción proporciona un método para generar una Puntuación de Riesgo de Recidiva para permitir el pronóstico de los pacientes con cáncer de mama obteniendo datos de expresión génica a partir de un número estadísticamente significativo de muestras biológicas de pacientes; aplicando el análisis de regresión de Cox univariante a los datos para obtener los genes seleccionados; aplicando los niveles de expresión ponderados a los genes seleccionados con coeficientes estándar de Cox para obtener un modelo de predicción que puede aplicarse como Puntuación de Riesgo de Recidiva.

15 La presente descripción proporciona un método para generar un informe del paciente con pronóstico de cáncer de mama obteniendo una muestra biológica del paciente; midiendo la expresión génica de la muestra; aplicando una Puntuación de Riesgo de Recidiva a los resultados de la etapa b.; y utilizando los resultados obtenidos en la etapa c. para generar el informe y los informes de pacientes generados de este modo.

20 La presente invención abarca el uso de una composición que comprende un conjunto de sondas que consiste en: las SEQ ID NO: 96-111 en la Tabla 10 en un método de diagnóstico según se define en las reivindicaciones.

25 La presente invención abarca el uso de un kit para llevar a cabo un ensayo para determinar el pronóstico de cáncer de mama en una muestra biológica que contiene: materiales para detectar secuencias aisladas de ácido nucleico, sus complementos, o porciones de los mismos de una combinación de genes seleccionados de entre aquellos que codifican un ARNm identificado por las SEQ ID NO: 1-111 en la Tabla 10.

30 La presente descripción proporciona artículos para evaluar el estado de un cáncer de mama que contienen: materiales para detectar secuencias aisladas de ácido nucleico, sus complementos, o porciones de los mismos de una combinación de genes seleccionados de entre aquellos que codifican un ARNm: correspondiente a las SEQ ID NO: 96-111; o reconocido por los conjuntos de sondas correspondientes a las SEQ ID NO: 96-111 como se muestra en la Tabla 10.

35 La presente descripción proporciona una cartera de diagnóstico/pronóstico que contiene secuencias aisladas de ácido nucleico, sus complementos, o porciones de los mismos de una combinación de genes seleccionados de entre aquellos que codifican un ARNm: correspondiente a las SEQ ID NO: 1-111; o reconocido por los conjuntos de sondas seleccionadas de entre los psids correspondientes a las SEQ ID NO: 1-111 como se muestra en la Tabla 10, en la que la combinación es suficiente para caracterizar el estado de un cáncer de mama o el riesgo de recidiva en una muestra biológica.

40 BREVE DESCRIPCIÓN DE LOS DIBUJOS

45 La Figura 1 representa el perfil para la selección de muestras para el análisis y el análisis de algoritmos de agrupamiento no supervisado de los datos de expresión génica para 286 pacientes con cáncer de mama sin compromiso de los ganglios linfáticos (LNN). (A). Diagrama de flujo para seleccionar muestras de los pacientes para su análisis. Se utilizó el estado ER para identificar los subgrupos de pacientes. A continuación, se analizó cada subgrupo por separado a fin de seleccionar los marcadores. Se asignaron los pacientes en un subgrupo a un conjunto de aprendizaje o a un conjunto de prueba. Se combinaron los marcadores seleccionados de cada subgrupo para formar una sola firma para predecir la reaparición de tumores para todos los pacientes en el conjunto de prueba en su totalidad. (B). Análisis de algoritmos de agrupamiento no supervisado de los datos de expresión génica de 286 pacientes con cáncer de mama LNN. El panel de la izquierda es una vista de los 17.819 genes informativos. El color rojo indica alta expresión relativa; el verde indica baja expresión relativa. Cada columna es una muestra y cada fila es un gen. El panel de la derecha muestra las vistas ampliadas de dos clústers de genes dominantes identificados a partir de la vista de la izquierda que tenían una notoria expresión diferencial entre los dos subgrupos de pacientes. El clúster génico superior tiene un grupo de 282 genes con expresión disminuida en el subgrupo ER positivo. El clúster génico inferior está representado por un grupo de 339 genes con expresión aumentada en el subgrupo ER positivo. La barra de etiquetas en la parte inferior de cada dendrograma indica el estado ER del paciente medido mediante ensayos rutinarios.

50 Figura 2. Establecimiento del perfil de 76 genes y el análisis de Kaplan-Meier para la supervivencia libre de metástasis a distancia (DMFS) y la supervivencia global (OS). A. Selección de genes para la firma pronóstica. Los marcadores génicos para los grupos ER positivo y ER negativo se seleccionaron a partir de dos conjuntos de aprendizaje utilizando el análisis ROC. Sesenta genes fueron del grupo ER positivo y 16 genes fueron del grupo

ER negativo (*izquierda*), y la curva ROC de la firma genética de 76 genes se obtuvo a partir del análisis de los 171 pacientes en el conjunto de prueba (*derecha*). B. Análisis DMFS (*izquierda*) y análisis OS (*derecha*) en el conjunto de validación de 171 pacientes LNN. El riesgo de fracaso para cada paciente se evaluó en base a la firma genética de 76 genes y el umbral se determinó mediante el conjunto de aprendizaje. Se utilizó el ensayo de rango logarítmico para ensayar las diferencias.

Figura 3. Análisis DMFS y OS en los subgrupos de pacientes LNN. DMFS (*izquierda*) y OS (*derecha*) en 84 pacientes premenopáusicas (A), 87 pacientes posmenopáusicas (B) y 79 pacientes con tumores de un tamaño que oscila entre 10 mm-20 mm (C). Los resultados son de los pacientes independientes en el conjunto de validación. El riesgo de reaparición para cada paciente se evaluó en base a la firma genética de 76 genes y el umbral se determinó mediante el conjunto de aprendizaje. Se utilizó el ensayo de rango logarítmico para ensayar las diferencias.

La Figura 4 representa el agrupamiento jerárquico en base a 5.121 genes.

La Figura 5 es un gráfico de barras que representa los niveles de expresión de 21 genes de control (Tabla 7).

La Figura 6 representa un flujo de trabajo del análisis de datos.

La Figura 7 representa el análisis PCA con conjuntos de genes filtrados.

La Figura 8 es un gráfico circular que representa el estado ER utilizado para asignar los subgrupos de pacientes. Los genes expresados de manera diferencial entre los sub-clústers ER positivo y ER negativo en las muestras de tejido en bruto y LCM se definieron mediante la t de Student.

La Figura 9 es una serie de gráficos de barras que representan los resultados del análisis de vías mediante ontología génica para los genes asociados exclusivamente con ER en las muestras LCM, exclusivamente en los tejidos en bruto, y para aquellos que son comunes al tejido en bruto y al LCM.

La Figura 10 muestra los resultados de la validación independiente. El diagrama de flujo en la Figura 10 muestra que los 132 pacientes se recogieron en cuatro fuentes diferentes. Se calculó una Puntuación de Riesgo de Recidiva para cada paciente en base a los niveles de expresión de la firma genética de 76 genes. Los pacientes se clasificaron en grupos de resultado bueno y malo.

La Figura 11 muestra A. características clínicas y patológicas de los pacientes y sus tumores para el estudio de validación y B. una curva ROC de la firma genética de 76 genes en el estudio de validación.

La Figura 12 representa el resultado de la clasificación de los 132 pacientes en el estudio de validación.

En el gráfico se muestran la curva de supervivencia de Kaplan-Meier y el ensayo de rango logarítmico en los grupos con predicción de resultado bueno y malo.

La Figura 13 representa los resultados del análisis de vías. El diagrama de flujo en la Figura 13 muestra el método utilizado para la selección de muestras para los análisis estadísticos. A. Se dividió al azar a los 286 pacientes en un conjunto de aprendizaje de 115 pacientes y un conjunto de prueba. B. Se dividió al azar a los 286 pacientes en un conjunto de aprendizaje de un 80% de los pacientes y un conjunto de prueba de un 20% de los pacientes. El conjunto de aprendizaje se utilizó para seleccionar los marcadores génicos y para construir una firma pronóstica. El conjunto de prueba se utilizó para la validación. Ambos procedimientos se repitieron 10 veces.

La Figura 14 representa los resúmenes del análisis por ontología GO. Cada barra representa una vía significativamente sobrerrepresentada en la firma genética de 76 genes ($p < 0,05$). La desviación estándar se calcula a partir de los resultados de las firmas alternativas. A. Resultados de 10 firmas utilizando conjuntos de aprendizaje de 115 pacientes. B. Resultados de 10 firmas utilizando conjuntos de aprendizaje del 80% de los pacientes.

DESCRIPCIÓN DETALLADA

La presente invención abarca un método para evaluar el estado de un cáncer de mama en una muestra biológica obtenida de un paciente con cáncer de mama tal como se define en las reivindicaciones.

La presente invención abarca un método de estadificación de un cáncer de mama en una muestra biológica obtenida de un paciente con cáncer de mama tal como se define en las reivindicaciones. El método utiliza cualquier clasificación conocida en la técnica, incluido el sistema TNM (American Joint Committee on Cancer www.cancerstaging.org) y la comparación con los estadios correspondientes a pacientes con perfiles de expresión génica similares.

La presente invención abarca un método para determinar el protocolo de tratamiento del paciente con cáncer de mama en una muestra biológica obtenida de un paciente con cáncer de mama tal como se define en las reivindicaciones.

5 La presente descripción proporciona un método para tratar a un paciente con cáncer de mama obteniendo una muestra biológica de un paciente con cáncer de mama; y midiendo los niveles de expresión en la muestra de genes seleccionados de entre aquellos que codifican un ARNm: correspondiente a las SEQ ID NO: 1-111; o reconocido por los conjuntos de sondas seleccionadas de entre los psids correspondientes a las SEQ ID NO: 1-111 como se muestra en la Tabla 10, en el que los niveles de expresión génica por encima o por debajo de los niveles límite predeterminados son indicativos de un alto riesgo de reaparición y; tratando al paciente con un tratamiento complementario si es un paciente de alto riesgo.

10 Preferentemente, las SEQ ID NO: son 36-95 para los pacientes con receptor de estrógeno (ER) positivo y 96-111 para los pacientes ER negativo.

15 La muestra puede prepararse mediante cualquier método conocido en la técnica, incluida la preparación del tejido en bruto y la microdissección por captura con láser. Las preparaciones de tejido en bruto pueden obtenerse a partir de una biopsia o de una pieza quirúrgica.

20 Los métodos anteriormente indicados pueden incluir adicionalmente medir el nivel de expresión de al menos un gen que codifica un ARNm identificado por las SEQ ID NO: 112-132 en la Tabla 10. También pueden incluir medir el nivel de expresión de al menos un gen expresado constitutivamente en el muestra.

25 Los métodos anteriormente indicados pueden incluir adicionalmente determinar el estado del receptor de estrógeno (ER) de la muestra. El estado ER puede medirse mediante cualquier método conocido en la técnica, incluida la medición del nivel de expresión de al menos un gen indicativo del estado ER, la medición de la presencia de ER en la muestra y la medición mediante inmunohistoquímica.

30 Los métodos anteriormente indicados pueden utilizarse con muestras de cualquier fuente biológica. Preferentemente, la muestra se obtiene de un tumor primario.

Los métodos anteriormente indicados tienen preferentemente una especificidad de al menos un 40% y una sensibilidad de al menos un 90%.

35 Los métodos anteriormente indicados pueden utilizarse cuando el patrón de expresión de los genes se compara con un patrón de expresión indicativo de un paciente recidivante. La comparación puede ser mediante cualquier método conocido en la técnica, incluida la comparación de los patrones de expresión llevada a cabo con los métodos de reconocimiento de patrones. Los métodos de reconocimiento de patrones pueden ser cualesquiera de los conocidos en la técnica, incluido el análisis de riesgos proporcionales de Cox.

40 Los métodos anteriormente indicados pueden utilizarse cuando los niveles límite predeterminados son una sobreexpresión o subexpresión al menos 1,5 veces mayor en la muestra con respecto a las células benignas o al tejido normal. Preferentemente, los niveles límite predeterminados tienen al menos una sobreexpresión con un valor p estadísticamente significativo en la muestra que tiene células metastásicas con respecto a las células benignas o al tejido normal. Más preferentemente, el valor p es inferior a 0,05.

45 Los métodos anteriormente indicados pueden utilizarse cuando la expresión génica se mide en una micromatriz o matriz génica. Las micromatrices y matrices génicas adecuadas para su uso en el presente documento también se incluyen en la invención. La micromatriz puede ser una matriz de ADNc o una matriz de oligonucleótidos y puede contener adicionalmente uno o más reactivos de control interno.

50 Los métodos anteriormente indicados pueden utilizarse cuando la expresión génica se determina mediante amplificación de ácido nucleico llevada a cabo mediante la reacción en cadena de la polimerasa (PCR) del ARN extraído de la muestra. La PCR puede ser una reacción en cadena de la polimerasa con transcriptasa inversa (RT-PCR). La RT-PCR puede contener adicionalmente uno o más reactivos de control interno.

55 Los métodos anteriormente indicados pueden utilizarse cuando la expresión génica se detecta midiendo o detectando una proteína codificada por el gen. Puede utilizarse cualquier método conocido en la técnica, incluida la detección mediante un anticuerpo específico contra la proteína y la medición de una característica del gen. Las características adecuadas incluyen, sin limitación, la amplificación, la metilación, la mutación y la variación alélica de ADN.

60 La presente descripción proporciona un método para realizar la validación cruzada de un perfil de expresión génica pronóstico para pacientes con cáncer de mama obteniendo datos de expresión génica a partir de un número estadísticamente significativo de muestras biológicas de pacientes; aleatorizando el orden de las muestras; dejando fuera los datos de aproximadamente el 10%-50% de las muestras; calculando, para las muestras restantes, el factor

de interés en todas las variables y seleccionando las variables que satisfagan un límite del valor p (p); seleccionando las variables que se ajustan a un modelo de predicción utilizando una búsqueda hacia adelante y evaluando el error de aprendizaje hasta que dé con una tasa de error predeterminada; ensayando el modelo de predicción en el 10%-50% de las muestras dejadas fuera; repitiendo las etapas c., -g. con un nuevo conjunto de muestras retiradas; y siguiendo con las etapas c) -g) hasta haber ensayado el 100% de las muestras, y registrar el rendimiento de la clasificación.

5 El método de validación cruzada puede utilizarse cuando los datos de expresión génica obtenidos en la etapa h. están representados por los genes seleccionados de entre aquellos que codifican un ARNm: correspondiente a las SEQ ID NO: 3, 4, 10, 18, 37, 40, 42, 43, 45, 55, 58, 64, 67, 72-74, 76, 81, 85-86, 89, 97, 100-101, 110-111, 125 y 132-442, o 2, 3, 5, 12, 20, 25, 36, 37, 39, 40, 41, 42, 43, 45, 46, 51, 52, 53, 54, 55, 57, 58, 59, 60, 61, 62, 63, 64, 66, 67, 69, 70, 72, 73, 74, 76, 80, 81, 83, 84, 85, 86, 87, 88, 89, 90, 94, 97, 98, 101, 102, 104, 107, 110, 111, 132, 139, 142, 150, 151, 153, 154, 158, 161, 163, 167, 170, 171, 173, 175, 181, 182, 183, 186, 188, 190, 192, 204, 206, 207, 212, 215, 218, 221, 223, 225, 228, 231, 232, 236, 238, 239, 240, 241, 242, 243, 246, 248, 15 249, 255, 257, 267, 269, 270, 271, 273, 280, 281, 282, 288, 290, 291, 299, 301, 304, 306, 311, 314, 315, 318, 327, 328, 338, 342, 346, 348, 354, 366, 368, 371, 375, 385, 388, 391, 395, 397, 402, 405, 409, 422, 424, 428, 429, 435, 436, 440, 651 y 443-650; o reconocido por los conjuntos de sondas seleccionadas del grupo que consiste en los psids correspondientes a las SEQ ID NO: 3, 4, 10, 18, 37, 40, 42, 43, 45, 55, 58, 64, 67, 72-74, 76, 81, 85-86, 89, 97, 100-101, 110-111, 125 y 132-442, o 2, 3, 5, 12, 20, 25, 36, 37, 39, 40, 41, 42, 43, 45, 46, 51, 52, 53, 54, 55, 57, 58, 20 59, 60, 61, 62, 63, 64, 66, 67, 69, 70, 72, 73, 74, 76, 80, 81, 83, 84, 85, 86, 87, 88, 89, 90, 94, 97, 98, 101, 102, 104, 107, 110, 111, 132, 139, 142, 150, 151, 153, 154, 158, 161, 163, 167, 170, 171, 173, 175, 181, 182, 183, 186, 188, 190, 192, 204, 206, 207, 212, 215, 218, 221, 223, 225, 228, 231, 232, 236, 238, 239, 240, 241, 242, 243, 246, 248, 249, 255, 257, 267, 269, 270, 271, 273, 280, 281, 282, 288, 290, 291, 299, 301, 304, 306, 311, 314, 315, 318, 327, 328, 338, 342, 346, 348, 354, 366, 368, 371, 375, 385, 388, 391, 395, 397, 402, 405, 409, 422, 424, 428, 429, 435, 25 436, 440, 651 y 443-650 como se muestra en la Tabla 10.

La presente descripción proporciona un método para validar de manera independiente un perfil de expresión génica pronóstico para pacientes con cáncer de mama obteniendo datos de expresión génica a partir de un número estadísticamente significativo de muestras biológicas de pacientes; normalizando las variabilidades de partida en los datos de expresión génica; calculando el factor de interés en todas las variables que se han seleccionado con anterioridad; y ensayando el modelo de predicción en la muestra, y registrar el rendimiento de la clasificación.

35 El método de validación independiente puede utilizarse cuando los datos de expresión génica obtenidos en la etapa d. están representados por los genes seleccionados de entre aquellos que codifican un ARNm correspondiente a las SEQ ID NO: 3, 4, 10, 18, 37, 40, 42, 43, 45, 55, 58, 64, 67, 72-74, 76, 81, 85-86, 89, 97, 100-101, 110-111, 125 y 132-442, o 2, 3, 5, 12, 20, 25, 36, 37, 39, 40, 41, 42, 43, 45, 46, 51, 52, 53, 54, 55, 57, 58, 59, 60, 61, 62, 63, 64, 66, 67, 69, 70, 72, 73, 74, 76, 80, 81, 83, 84, 85, 86, 87, 88, 89, 90, 94, 97, 98, 101, 102, 104, 107, 110, 111, 132, 139, 142, 150, 151, 153, 154, 158, 161, 163, 167, 170, 171, 173, 175, 181, 182, 183, 186, 188, 40 190, 192, 204, 206, 207, 212, 215, 218, 221, 223, 225, 228, 231, 232, 236, 238, 239, 240, 241, 242, 243, 246, 248, 249, 255, 257, 267, 269, 270, 271, 273, 280, 281, 282, 288, 290, 291, 299, 301, 304, 306, 311, 314, 315, 318, 327, 328, 338, 342, 346, 348, 354, 366, 368, 371, 375, 385, 388, 391, 395, 397, 402, 405, 409, 422, 424, 428, 429, 435, 436, 440, 651 y 443-650; o reconocido por los conjuntos de sondas seleccionadas del grupo que consiste en los psids correspondientes a las SEQ ID NO: 3, 4, 10, 18, 37, 40, 42, 43, 45, 55, 58, 64, 67, 72-74, 76, 81, 85-86, 89, 97, 100-101, 110-111, 125 y 132-442, o 2, 3, 5, 12, 20, 25, 36, 37, 39, 40, 41, 42, 43, 45, 46, 51, 52, 53, 54, 55, 57, 58, 45 59, 60, 61, 62, 63, 64, 66, 67, 69, 70, 72, 73, 74, 76, 80, 81, 83, 84, 85, 86, 87, 88, 89, 90, 94, 97, 98, 101, 102, 104, 107, 110, 111, 132, 139, 142, 150, 151, 153, 154, 158, 161, 163, 167, 170, 171, 173, 175, 181, 182, 183, 186, 188, 190, 192, 204, 206, 207, 212, 215, 218, 221, 223, 225, 228, 231, 232, 236, 238, 239, 240, 241, 242, 243, 246, 248, 249, 255, 257, 267, 269, 270, 271, 273, 280, 281, 282, 288, 290, 291, 299, 301, 304, 306, 311, 314, 315, 318, 327, 328, 338, 342, 346, 348, 354, 366, 368, 371, 375, 385, 388, 391, 395, 397, 402, 405, 409, 422, 424, 428, 429, 435, 50 436, 440, 651 y 443-650 como se muestra en la Tabla 10.

La presente descripción proporciona un método para generar una Puntuación de Riesgo de Recidiva para permitir el pronóstico de los pacientes con cáncer de mama obteniendo datos de expresión génica a partir de un número estadísticamente significativo de muestras biológicas de pacientes; aplicando el análisis de regresión de Cox univariante a los datos para obtener los genes seleccionados; aplicando los niveles de expresión ponderados a los genes seleccionados con coeficientes estándar de Cox para obtener un modelo de predicción que pueda aplicarse como Puntuación de Riesgo de Recidiva.

60 La Puntuación de Riesgo de Recidiva puede obtenerse mediante la fórmula:

65
$$\text{Puntuación de Recidiva} = A \cdot I + \sum_{i=1}^{60} I \cdot w_i \cdot x_i + B \cdot (1 - I) + \sum_{j=1}^{16} (1 - I) \cdot w_j \cdot x_j$$

en la que

5

$$I = \begin{cases} 1 & \text{si el nivel del gen X} > 10 \\ 0 & \text{si el nivel del gen X} \leq 10 \end{cases}$$

10

A y B son constantes

w_i es el coeficiente de regresión de Cox estandarizado para gen X + marcador

x_i es el valor de la expresión de ER + marcador en escala log2

w_j es el coeficiente de regresión de Cox estandarizado para gen X - marcador

x_j es el valor de la expresión de ER - marcador en escala log2

15

el gen X se selecciona del grupo que consiste en aquellos que codifican un ARNm:

i. correspondiente a las SEQ ID NO: 1-111; o

ii. reconocido por los conjuntos de sondas seleccionadas del grupo que consiste en los psids correspondientes a las SEQ ID NO: 1-111 como se muestra en la Tabla 10.

20

La presente descripción proporciona un método para generar un informe del paciente con pronóstico de cáncer de mama obteniendo una muestra biológica del paciente; midiendo la expresión génica de la muestra; aplicando una Puntuación de Riesgo de Recidiva a los resultados de la etapa b.; y utilizando los resultados obtenidos en la etapa c. para generar el informe y los informes de pacientes generados de este modo. El informe puede contener una evaluación de los resultados para el paciente y/o la probabilidad de riesgo con respecto a la población de pacientes.

25

La presente invención abarca el uso de una composición que comprende un conjunto de sondas que consiste en: las SEQ ID NO: 96-111 en la Tabla 10 en un método de diagnóstico como se define en las reivindicaciones. La composición puede contener al menos un conjunto de sondas seleccionadas de entre: las SEQ ID NO: 36-95; 96-111; ó 36-111. La composición puede contener adicionalmente reactivos para llevar a cabo un análisis de micromatrices, y un medio a través del cual se ensayan dichas secuencias de ácido nucleico, sus complementos, o porciones de los mismos.

30

La presente invención abarca el uso de un kit para llevar a cabo un ensayo para determinar el pronóstico de cáncer de mama en una muestra biológica que contiene: materiales para detectar secuencias aisladas de ácido nucleico, sus complementos, o porciones de los mismos de una combinación de genes que codifican un ARNm: identificado por las SEQ ID NO: 1-111 en la Tabla 10. El kit puede contener adicionalmente reactivos para llevar a cabo un análisis de micromatrices, y un medio a través del cual se ensayan dichas secuencias de ácido nucleico, sus complementos, o porciones de los mismos.

35

La presente descripción proporciona artículos para evaluar el estado de un cáncer de mama que contiene: materiales para detectar secuencias aisladas de ácido nucleico, sus complementos, o porciones de los mismos de una combinación de genes seleccionados de entre aquellos que codifican un ARNm: correspondiente a las

45

SEQ ID NO: 1-111; o reconocido por los conjuntos de sondas seleccionadas de entre los psids correspondientes a las SEQ ID NO: 1-111 como se muestra en la Tabla 10. Los artículos pueden contener adicionalmente reactivos para llevar a cabo un análisis de micromatrices, y un medio a través del cual se ensayan dichas secuencias de ácido nucleico, sus complementos, o porciones de los mismos.

50

Las micromatrices proporcionadas en el presente documento pueden contener secuencias aisladas de ácido nucleico, sus complementos, o porciones de los mismos de una combinación de genes seleccionados de entre aquellos que codifican un ARNm: correspondiente a las SEQ ID NO: 1-111; o reconocido por los conjuntos de sondas seleccionadas de entre los psids correspondientes a las SEQ ID NO: 1-111 como se muestra en la Tabla 10, en la que la combinación es suficiente para caracterizar el estado de un cáncer de mama o el riesgo de recidiva en una muestra biológica.

55

La medición o caracterización de las micromatrices resulta preferentemente una sobreexpresión o subexpresión al menos 1,5 veces mayor. La micromatriz puede detectar unos niveles límite predeterminado que son una sobreexpresión o subexpresión al menos 1,5 veces mayor en la muestra con respecto a las células benignas o al tejido normal. Preferentemente, los niveles límite predeterminados tienen al menos una sobreexpresión con un valor p estadísticamente significativo en la muestra que tiene células metastásicas con respecto a las células benignas o al tejido normal. Más preferentemente, el valor p es inferior a 0,05.

60

La presente descripción proporciona una cartera de diagnóstico/pronóstico que contiene secuencias aisladas de ácido nucleico, sus complementos, o porciones de los mismos de una combinación de genes seleccionados de entre aquellos que codifican un ARNm: correspondiente a las SEQ ID NO: 1-111; o reconocido por

65

- 5 los conjuntos de sondas seleccionadas de entre los psids correspondientes a las SEQ ID NO: 1-111 como se muestra en la Tabla 10, en la que la combinación es suficiente para caracterizar el estado de un cáncer de mama o el riesgo de recidiva en una muestra biológica. La cartera puede detectar unos niveles límite predeterminados que son una sobreexpresión o subexpresión al menos 1,5 veces mayor en la muestra con respecto a las células benignas o al tejido normal. Preferentemente, los niveles límite predeterminados tienen al menos una sobreexpresión con un valor p estadísticamente significativo en la muestra que tiene células metastásicas con respecto a las células benignas o al tejido normal. Más preferentemente, el valor p es inferior a 0,05.
- 10 Las SEC ID NO: 1-650 se resumen en la Tabla 10. En cada una de las SEQ ID NO: 1-650, el marcador se identifica mediante un psid o denominación GeneChip de Affymetrix y representa el gen que codifica cualquier variante, alelo, etc. correspondiente a la SEQ ID NO dada. El marcador también se define como el gen que codifica un ARNm reconocido por la sonda correspondiente al psid dado. Más adelante se describen con más detalle determinados marcadores.
- 15 M83, SEQ ID NO: 45, se menciona en las solicitudes de patente de EE.UU. con número de publicación 20020110547, 20040009491 y 20030236392, y las PCT con número de publicación WO0149716, WO0170979, WO03025138, WO03042661, WO03016475, WO2004063355, WO2004047728 y WO2004030615.
- 20 ABLIM-s, SEQ ID NO: 43, se menciona en las solicitudes de patente de EE.UU. con número de publicación 20020131971, 20030087818, 20030166064, 20030236392 y 20040005560, y las PCT con número de publicación WO0036107, WO0159063, WO0160860, WO0174405, WO0177288, WO02059271, WO0212328, WO0224956, WO0229086, WO03025138, WO2004024097, WO2004037996, WO2004039956, WO2004043361 y WO2004047728.
- 25 ATAD2, SEQ ID NO: 54 se menciona en las solicitudes de patente de EE.UU. con número de publicación 20020064872, 20020085998, 20020150581, 20020156011, 20020192678, 20030069180, 20030073623, 20030104366, 20040005560, 20040013663 y 20040058340 y las PCT con número de publicación WO0060076, WO0142467, WO0155322, WO0160860, WO0170979, WO0179286, WO196388, WO02057414, WO02059377, WO02060317, WO02086443, WO0231198, WO0250279, WO03004989, WO03025138, WO03042661, WO03062379, WO2004047728, WO2004048938, WO2004048938, WO2004063355 y WO9938972.
- 30 C11ORF9, SEQ ID NO: 48, se menciona en las solicitudes de patente de EE.UU. con número de publicación 20030065157 y 20030073623 y las PCT con número de publicación WO0056880, WO0058473, WO0159063, WO0174405, WO02059260, WO0229086, WO03101283, WO2004031413 y WO2004060270.
- 35 C3, SEQ ID NO: 46, se menciona en las solicitudes de patente de EE.UU. con número de publicación 20040005560 y 20040058340 y las PCT con número de publicación WO9101999, WO0055350, WO0160860, WO0175067, WO0188088, WO0224956, WO0294629, WO03078572, WO2004028479, WO2004030615, WO2004041170 y WO2004060270.
- 40 CGI-41, SEQ ID NO: 65, se menciona en las solicitudes de patente de EE.UU. con número de publicación 20030073623 y 20030096982 y las PCT con número de publicación WO9514772, WO9940100, WO0154472, WO0157188, WO03046152, WO2004060270.
- 45 CLN8, SEQ ID NO: 36, se menciona en el documento EP 1 104 808 B1, las solicitudes de patente de EE.UU. número 20030073623 y 20040009491 y las PCT con número de publicación WO9845437, WO9935158, WO0142451, WO0190304, WO02090526, WO02095010, WO02102993, WO02102994, WO03004622, WO04024892 y WO04041170.
- 50 CKN1, SEQ ID NO: 60, se menciona en la solicitud de patente de EE.UU. con número de publicación 20040058340 y las PCT con número de publicación WO9938972 y WO2004037996.
- 55 DUSP4, SEQ ID NO: 57, se menciona en las solicitudes de patente de EE.UU. con número de publicación 20030073623, 20030194704, 20040058340 y 20040110194 y las PCT con número de publicación WO0100828, WO0204514, WO02059377, WO02103320, WO0224956, WO03004989, WO2004018641, WO2004023973, WO2004037996, WO2004039956, WO2004043361, WO2004047728, WO2004048938 y WO9423039.
- 60 FEN1, SEQ ID NO: 62, se menciona en la patente de EE.UU. número 5.874.283, las solicitudes de patente de EE.UU. con número de publicación 20020156011, 20030069180, 20030073623 y 20030194704 y las PCT con número de publicación WO0196388, WO03016475, WO03016500, WO03072035, WO03101283, WO2004022059, WO2004030615, WO2004039956, WO2004043361, WO2004047728, WO2004048938, WO2004060270 y WO2004079014.
- 65 FKBP2, SEQ ID NO: 63, se menciona en la solicitud de patente de EE.UU. con número de publicación 20030073623 y las PCT con número de publicación WO02059377, WO02070737, WO02081731, WO02083898, WO0224956, WO03016475, WO03054152, WO2004039956, WO2004043361 y WO9219745.

GP73, SEQ ID NO: 37, se menciona en las solicitudes de patente de EE.UU. con número de publicación 20030022239, 20030073623, 20030236392, 20040009478, 20040034196 y 20040058340 y las PCT con número de publicación WO0012708, WO0061610, WO0078961, WO0140466, WO0160860, WO0177168, WO0194629, WO02059260, WO02083876, WO0224956, WO03004989, WO2004024892, WO2004031414 y WO2004053081.

5 H4FH, SEQ ID NO: 61, se menciona en la solicitud de patente de EE.UU. con número de publicación 20020151681 y las PCT con número de publicación WO0055174, WO0175067, WO0224956, WO03016476, WO2004039943.

10 IL-18, SEQ ID NO: 39, se menciona en la patente de EE.UU. número 6060283, la solicitud de patente de EE.UU. con número de publicación 20040115636 y las PCT con número de publicación WO0194629, WO02057414, WO0224956, WO0229086, WO03025138, WO03042661 y WO9724441.

15 KIAA0748, SEQ ID NO: 56, se menciona en las PCT número de publicación WO0021991, WO02094988 y WO2004003162.

20 KPNA2, SEQ ID NO: 64, se menciona en las solicitudes de patente de EE.UU. con número de publicación 20020151681, 20020156011, 20030069180 y 20040058340 y las PCT con número de publicación WO9712967, WO9964576, WO0055174, WO0146697, WO0170979, WO0196388, WO02057414, WO02059377, WO02086443, WO02102235, WO0224956, WO03010336, WO03025138, WO03027285, WO03042661, WO2004024097, WO2004024892, WO2004028479, WO2004037996, WO2004039956, WO2004043361, WO2004047728 y WO2004063355.

25 MGC11335, SEQ ID NO: 66, se menciona en las solicitudes de patente de EE.UU. con número de publicación 20030073623, 20030219744 y 20030236392 y las PCT con número de publicación WO0144448, WO0153312, WO0157182, WO0188088 y WO2004030615.

30 OR12D2, SEQ ID NO: 52, se menciona en las PCT con número de publicación WO0127158, WO02068652, WO02081498, WO02081517, WO030000735 y WO03091388.

ORC3, SEQ ID NO: 53, se menciona en la solicitud de patente de EE.UU. con número de publicación 20040013663 y las PCT con número de publicación WO9810067, WO9958642, WO0060078, WO0175067, WO02059377, WO02070737, WO02102235 y WO03042661.

35 PLK1, SEQ ID NO: 59, se menciona en el documento DE4329177, las solicitudes de patente de EE.UU. con número de publicación 20020081659 y 20020151681 y las PCT con número de publicación WO0055174, WO0055320, WO02070737, WO02086443, WO0224956, WO03003906, WO03016476, WO03018807, WO03025138, WO03042661, WO2004030615, WO2004037996, WO2004042022, WO2004047728, WO2004048938, WO2004063355 y WO2004070062.

40 PPP1CC, SEQ ID NO: 42, se menciona en las solicitudes de patente de EE.UU. con número de publicación 20030073623 y 20040013663 y las PCT con número de publicación WO03016476, WO0063438, WO0170979, WO0202623, WO02036766, WO02074237, WO0228999, WO03016475 y WO2004030615.

45 SMC4, SEQ ID NO: 55, se menciona en las solicitudes de patente de EE.UU. con número de publicación 20020123619, 20020168637 y 20040058340 y las PCT con número de publicación WO9938972, WO9964576, WO0175067, WO0190154, WO0204514, WO02059377, WO02074237, WO02086443, WO02102235, WO03003906, WO03016475, WO03025138, WO03042661, WO2003094848, WO0004024892, WO2004047728, WO2004048938, WO2004063355 y WO2004074301.

50 TRDL-1, SEQ ID NO: 44, se menciona en las patentes de EE.UU. número 6171787 y 6440694, las solicitudes de patente de EE.UU. con número de publicación 20020055474, 20020072089, 20020081659, 20030065157, 20030073623 y 20030219767 y las PCT con número de publicación WO9733902, WO9912965, WO9926976, WO9928462, WO9933980, WO9935170, WO9950416, WO9954460, WO0026244, WO0032776, WO0055320, WO0125256, WO0177291, WO2004020593, WO2004024892, WO2004033637 y 04037996.

55 Yif1p, SEQ ID NO: 38, se menciona en las solicitudes de patente de EE.UU. con número de publicación 20020131971, 20030166064 y 20040037842 y las PCT con número de publicación WO9933981, WO9947540, WO0173027, WO02060317, WO02070737, WO02076488, WO02095010, WO0224956, WO03004622, WO03038063, WO03042661, WO03054152, WO03064589, WO2004024892, WO2004037996, WO2004047728 y WO2004048938.

60 Mediante las matrices génicas Human U133a GeneChip de Affymetrix, se analizó la expresión de 22.000 transcritos utilizando ARN total de muestras de tumores congelados de 286 pacientes con cáncer de mama sin compromiso de los ganglios linfáticos (LNN) que no recibieron tratamiento complementario sistémico. Los inventores descubrieron que las medidas de la expresión génica en todo el genoma pueden identificar patrones de actividad

65

génica que subclasifican los tumores y proporcionan un medio mejorado para la evaluación del riesgo individual en pacientes con cáncer de mama sin compromiso de los ganglios linfáticos.

5 En un conjunto de aprendizaje de 115 tumores, los inventores identificaron una firma genética de 76 genes compuesta por 60 genes para pacientes ER positivo y 16 genes para pacientes ER negativo. Esta firma presentó un 93% de sensibilidad y un 48% de especificidad en un conjunto de prueba independiente posterior de 171 pacientes LNN. El perfil genético fue muy informativo a la hora de identificar pacientes que desarrollan metástasis a distancia en un plazo de 5 años (cociente de riesgos instantáneos, HR: 5,67, intervalo de confianza del 95%, IC: 2,59-12,4), incluso cuando se corrige para los factores pronósticos tradicionales en el análisis multivariable (HR: 5,55, IC: 10 2,46-12,5). El perfil de 76 genes también representó un fuerte factor pronóstico para el desarrollo de metástasis en los subgrupos de 84 pacientes premenopáusicas (HR: 9,60, $p = 1,6 \times 10^{-4}$), 87 pacientes posmenopáusicas (HR: 4,04, $p = 1,7 \times 10^{-3}$) y 79 pacientes con un tamaño de tumor de entre 10 mm y 20 mm (HR: 14,1, $p = 2,3 \times 10^{-6}$), un grupo de pacientes para los que el pronóstico de la enfermedad resulta particularmente difícil.

15 En este estudio se proporcionan los resultados de un análisis de tumores primarios de 286 pacientes con cáncer de mama LNN de todos los grupos de edad y tamaño de tumor. Los pacientes no recibieron tratamiento complementario sistémico, haciendo de ésta la primera evaluación de pronóstico multigénica sin la contribución potencialmente generadora de confusión por factores predictivos relacionados con el tratamiento sistémico, debidamente validado por un conjunto de prueba independiente y sin restricciones para el tamaño de tumor y la edad. El algoritmo basado en la expresión génica descrito puede predecir, con alto grado de confianza en pacientes LNN, la probabilidad de desarrollar una reaparición a distancia.

25 Sólo en raras ocasiones la mera presencia o ausencia de secuencias concretas de ácido nucleico en una muestra de tejido ha resultado tener un valor diagnóstico o pronóstico. Por otro lado, se considera cada vez más importante la información acerca de la expresión de diversas proteínas, péptidos o ARNm. La mera presencia de secuencias de ácido nucleico con el potencial de expresar proteínas, péptidos o ARNm (tales secuencias denominadas "genes") dentro del genoma por sí misma no resulta determinante en cuanto a si una proteína, péptido o ARNm se expresa en una determinada célula. Que un determinado gen capaz de expresar proteínas, péptidos o ARNm lo haga o no y en qué medida se produzca dicha expresión, si acaso, viene determinado por diversos factores complejos. Independientemente de las dificultades en la comprensión y la evaluación de estos factores, el ensayo de la expresión génica puede proporcionar información útil sobre la aparición de eventos importantes, tales como la tumorigénesis, la metástasis, la apoptosis y otros fenómenos clínicamente pertinentes. Las indicaciones relativas del grado de actividad o inactividad de los genes pueden encontrarse en los perfiles de expresión génica. Los perfiles de expresión génica de la presente invención se utilizan para proporcionar un pronóstico y para tratar a los 35 pacientes con cáncer de mama.

La preparación de muestras requiere la recolección de muestras de los pacientes. Las muestras de los pacientes utilizadas en el método de la invención son aquellas de las que se sospecha contienen células enfermas, tales como las células epiteliales extraídas del tumor primario en una muestra de mama. También resultan 40 preferentes las muestras tomadas de los márgenes quirúrgicos. Sin embargo, lo más preferentemente, la muestra se toma de un ganglio linfático obtenido de una cirugía de cáncer de mama. La tecnología de microdissección por captura con láser (LCM) es una manera de seleccionar las células a estudiar, minimizando la variabilidad causada por la heterogeneidad del tipo de célula. Por consiguiente, pueden detectarse fácilmente cambios moderados o pequeños en la expresión génica entre las células normales y las cancerosas. Las muestras también pueden 45 comprender células epiteliales circulantes extraídas de sangre periférica. Éstas pueden obtenerse según varios métodos, pero el método más preferente es la técnica de separación magnética descrita en la patente de EE.UU. 6.136.182. Una vez se ha obtenido la muestra que contiene las células de interés, se extrae y amplifica el ARN y se obtiene un perfil de expresión génica, preferentemente mediante micromatrices, para los genes en las carteras apropiadas.

50 Los métodos preferentes para el establecimiento de perfiles de expresión génica incluyen determinar la cantidad de ARN que es producida por un gen que puede codificar una proteína o péptido. Esto se logra mediante PCR con transcriptasa inversa (RT-PCR), RT-PCR competitiva, RT-PCR en tiempo real, RT-PCR de presentación diferencial, transferencia Northern y otros ensayos relacionados. Aunque es posible llevar a cabo estas técnicas utilizando reacciones de PCR individuales, es mejor amplificar ADN complementario (ADNc) o ARN complementario (ARNc) producido a partir de ARNm y analizarlo mediante micromatrices. Los expertos en la materia conocen varios métodos y configuraciones de matriz diferentes para su producción y se describen en patentes de EE.UU. tales como: 5.445.934, 5.532.128, 5.556.752, 5.242.974, 5.384.261, 5.405.783, 5.412.087, 5.424.186, 5.429.807, 5.436.327, 5.472.672, 5.527.681, 5.529.756, 5.545.531, 5.554.501, 5.561.071, 5.571.639, 5.593.839, 5.599.695, 60 5.624.711, 5.658.734 y 5.700.637.

La tecnología de micromatrices permite medir el nivel de ARNm en estado estacionario de miles de genes simultáneamente, presentando de este modo una potente herramienta para identificar efectos tales como la aparición, la detención o la modulación de la proliferación celular incontrolada. Actualmente son de uso generalizado 65 dos tecnologías de micromatrices. La primera son las matrices de ADNc y la segunda son las matrices de oligonucleótidos. Aunque existen diferencias en la construcción de estas matrices génicas, esencialmente los

análisis de datos de bajada y la salida son iguales. El producto de estos análisis son por lo general mediciones de la intensidad de la señal recibida de una sonda marcada utilizada para detectar una secuencia de ADNc de la muestra que hibrida con una secuencia de ácido nucleico en una ubicación conocida en la micromatriz. Por lo general, la intensidad de la señal es proporcional a la cantidad de ADNc y, por tanto de ARNm, expresado en las células de la muestra. Se dispone de, y resultan útiles, un gran número de tales técnicas. Los métodos preferentes para determinar la expresión génica pueden encontrarse en las patentes de EE.UU. 6.271.002, 6.218.122, 6.218.114 y 6.004.755.

El análisis de los niveles de expresión se lleva a cabo comparando las intensidades de señal. La mejor manera de hacerlo es generando una matriz de relación de las intensidades de expresión de los genes en una muestra de ensayo frente a las de una muestra de control. Por ejemplo, las intensidades de expresión génica de un tejido enfermo pueden compararse con las intensidades de expresión generadas a partir de tejido normal del mismo tipo (por ejemplo, muestra de tejido enfermo de mama vs. muestra de tejido normal de mama). Una relación de estas intensidades de expresión indica el número de veces que cambia la expresión génica entre las muestras de ensayo y de control.

Los perfiles de expresión génica también pueden presentarse de varias maneras. El método más común es disponer la matriz de relación o las intensidades de fluorescencia sin procesar en un dendrograma gráfico en el que las columnas indican las muestras de ensayo y las filas indican los genes. Los datos se disponen de manera que los genes con perfiles de expresión similares estén próximos entre sí. La relación de expresión para cada gen se visualiza como un color. Por ejemplo, una relación inferior a uno (que indica disminución de la expresión) puede aparecer en la porción azul del espectro mientras que una relación superior a uno (que indica aumento de la expresión) puede aparecer como un color en la porción roja del espectro. Se dispone de programas de software disponibles en el mercado para presentar tales datos, incluido GeneSpring de Agilent Technologies y el software Partek Discover™ y Partek Infer™ de Partek®.

Los genes modulados utilizados en los métodos de la invención se describen en los Ejemplos. Los genes expresados de manera diferencial presentan expresión aumentada o disminuida en pacientes con recidiva de cáncer de mama con respecto a los que no presentan recidiva. Aumento de la regulación y disminución de la regulación son expresiones relativas que significan que se encuentra una diferencia detectable (más allá de la contribución del ruido en el sistema utilizado para medirla) en la cantidad de expresión de los genes con respecto a una medida inicial. En este caso, la medida inicial es la expresión génica medida de un paciente no recidivante. Los genes de interés en las células enfermas (procedentes de pacientes recidivantes) presentan entonces expresión aumentada o disminuida con respecto al nivel inicial utilizando el mismo método de medición. Enfermo, en este contexto, se refiere a una modificación del estado de un cuerpo que interrumpe o altera, o tiene el potencial de alterar, el rendimiento adecuado de las funciones corporales tal como ocurre con la proliferación incontrolada de las células. Existe diagnóstico de enfermedad cuando algún aspecto del genotipo o fenotipo de esa persona es coherente con la presencia de la enfermedad. Sin embargo, el acto de llevar a cabo un diagnóstico o pronóstico incluye determinar cuestiones sobre la enfermedad/estado tales como determinar la probabilidad de recidiva y el seguimiento del tratamiento. En el seguimiento del tratamiento, se elaboran juicios clínicos referentes al efecto de un ciclo terapéutico determinado comparando la expresión de los genes a lo largo del tiempo para determinar si los perfiles de expresión génica han cambiado o están cambiando a patrones más coherentes con el tejido normal.

Preferentemente, los niveles de aumento y disminución de la expresión se distinguen en base a cambios en el número de veces de las mediciones de la intensidad de las sondas de micromatrices hibridadas. Resulta preferente una diferencia de 2,0 veces para hacer tales distinciones (o un valor p inferior a 0,05). Es decir, antes de decir que un gen se expresa de manera diferencial en células enfermas/recidivantes frente a las normales/no recidivantes, la célula enferma resulta producir al menos 2 veces más, o 2 veces menos intensidad que las células normales. Cuanto mayor sea la diferencia en el número de veces, más preferente es el uso del gen como herramienta de diagnóstico o pronóstico. Los genes seleccionados para los perfiles de expresión génica de la presente invención tienen niveles de expresión que dan como resultado la generación de una señal que puede distinguirse de la de los genes normales o no modulados por una cantidad que excede del nivel de fondo utilizando instrumentos de laboratorio clínico.

Pueden utilizarse valores estadísticos para distinguir con seguridad los genes modulados de los no modulados y del ruido. Las pruebas estadísticas encuentran los genes más significativamente diferentes entre los diversos grupos de muestras. La t de Student es un ejemplo de una prueba estadística robusta que puede utilizarse para encontrar diferencias significativas entre dos grupos. Cuanto menor sea el valor p, más convincente será la prueba de que el gen está mostrando una diferencia entre los diferentes grupos. Sin embargo, puesto que las micromatrices miden más de un gen a la vez, pueden realizarse al mismo tiempo decenas de miles de pruebas estadísticas. Debido a esto, es poco probable que se observen valores p pequeños sólo por casualidad y pueden hacerse ajustes para esto utilizando una corrección de Sidak, así como un experimento de aleatorización/permutación. Un valor p inferior a 0,05 mediante la t de Student es una prueba de que el gen es significativamente diferente. Una prueba más convincente es un valor p inferior a 0,05 después de tener en cuenta la corrección de Sidak. Para un gran número de muestras en cada grupo, un valor p inferior a 0,05 después del ensayo de aleatorización/permutación es la prueba más convincente de una diferencia significativa.

Otro parámetro que puede utilizarse para seleccionar los genes que generan una señal superior a la del gen no modulado o al ruido es el uso de una medición de la diferencia de señal absoluta. Preferentemente, la señal generada por la expresión del gen modulado presenta una diferencia de al menos un 20% con respecto a la del gen normal o no modulado (sobre una base absoluta). Es aún más preferente que tales genes produzcan patrones de expresión que presenten una diferencia de al menos un 30% con respecto a los de los genes normales o no modulados.

Los genes pueden agruparse de manera que la información obtenida sobre el conjunto de genes en el grupo proporcione una base sólida para elaborar un juicio clínicamente pertinente, tal como un diagnóstico, un pronóstico o un tratamiento de elección. Estos conjuntos de genes componen las carteras de la invención. En este caso, los juicios respaldados por las carteras implican un cáncer de mama y su posibilidad de reaparición. Al igual que con la mayoría de los marcadores diagnósticos, suele ser deseable utilizar el menor número de marcadores suficiente para emitir un juicio médico correcto. Esto impide que un retraso en el tratamiento hasta un análisis más detallado, así como un uso inapropiado de tiempo y recursos.

Preferentemente, las carteras se establecen de manera que la combinación de los genes en la cartera presente una sensibilidad y especificidad mejoradas con respecto a genes individuales o combinaciones de genes seleccionados al azar. En el contexto de la presente invención, la sensibilidad de la cartera puede reflejarse en las diferencias del número de veces presentadas por la expresión de un gen en el estado enfermo con respecto al estado normal. La especificidad puede reflejarse en las mediciones estadísticas de la correlación de la señalización de la expresión génica con la afección de interés. Por ejemplo, puede utilizarse la desviación estándar como una medición de este tipo. Al considerar un grupo de genes para su inclusión en una cartera, una pequeña desviación estándar en las mediciones de expresión se correlaciona con una mayor especificidad. También pueden utilizarse otras mediciones de la variación tales como los coeficientes de correlación.

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

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

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

Un método de la invención implica la comparación de perfiles de expresión génica para diversos genes (o carpetas) para atribuir pronósticos. Los perfiles de expresión génica de cada uno de los genes que comprenden la cartera se fijan en un medio tal como un medio legible por ordenador. Este puede adoptar varias formas. Por ejemplo, puede establecerse una tabla en la que la entrada sea el intervalo de señales (por ejemplo, mediciones de intensidad) indicativo de enfermedad. A continuación, pueden compararse los datos reales del paciente con los valores de la tabla para determinar si las muestras del paciente son normales o enfermas. En una forma de realización más sofisticada, los patrones de las señales de expresión (por ejemplo, la intensidad de fluorescencia) se registran digitalmente o gráficamente.

A continuación, se comparan los patrones de expresión génica de las carpetas de genes utilizados junto con las muestras del paciente, con los patrones de expresión. A continuación, puede utilizarse el software de comparación de patrones para determinar si las muestras del paciente tienen un patrón indicativo de reaparición de la enfermedad. Por supuesto, estas comparaciones también pueden utilizarse para determinar si no es probable que el paciente experimente reaparición de la enfermedad. A continuación, se comparan los perfiles de expresión de las muestras con la cartera de una célula de control. Si los patrones de expresión de la muestra son coherentes con el patrón de expresión para la reaparición de un cáncer de mama, entonces (en la ausencia de consideraciones médicas contrarias) se trata al paciente como podría tratarse a un paciente recidivante. Si los patrones de expresión de la muestra son coherentes con el patrón de expresión de la célula normal/de control, se obtiene un diagnóstico negativo para cáncer de mama.

Los perfiles preferentes de la presente descripción son la cartera de 35 genes compuesta por los genes de las SEQ ID NO: 1-35, la cartera de 60 genes compuesta por los genes de las SEQ ID NO: 36-95, que se utiliza preferentemente para el pronóstico de pacientes ER positivo, y la cartera de 16 genes compuesta por genes de las SEQ ID NO: 96-111 que se utiliza preferentemente para el pronóstico de pacientes ER negativo. Lo más preferentemente, la cartera está compuesta por genes de las SEQ ID NO: 36-111. Esta cartera más preferente segrega mejor los pacientes con cáncer de mama independientemente del estado ER con alto riesgo de recidiva de los que no lo son. Una vez identificados los pacientes con alto riesgo, éstos pueden tratarse con un tratamiento complementario.

En la presente invención, el método más preferente para analizar el patrón de expresión génica de un paciente para determinar el pronóstico de cáncer de mama es mediante el uso del programa de análisis de riesgos de Cox. Lo más preferentemente, el análisis se lleva a cabo utilizando el software S-Plus (disponible en el mercado de Insightful Corporation). Mediante tales métodos, se compara un perfil de expresión génica con el de un perfil que representa con seguridad recidiva (es decir, los niveles de expresión para la combinación de genes en el perfil son indicativos de recidiva). Se utiliza el modelo de riesgo de Cox con el umbral establecido para comparar la similitud de los dos perfiles (recidiva conocida frente a paciente) y a continuación determina si el perfil del paciente supera el umbral. Si lo hace, el paciente se clasifica como recidivante en el futuro y se da un tratamiento tal como un tratamiento complementario. Si el perfil del paciente no supera el umbral, entonces se clasifica como paciente no recidivante. También pueden utilizarse otras herramientas analíticas para responder a la misma pregunta, tales como el análisis discriminante lineal, la regresión logística y los enfoques de redes neuronales.

Se dispone de muchos otros métodos bien conocidos de reconocimiento de patrones. Las siguientes referencias proporcionan algunos ejemplos:

- Votación ponderada: Golub *et al.* (1999).
- Máquinas de vectores de soporte: Su *et al.* (2001); y Ramaswamy *et al.* (2001).
- K vecinos más cercanos: Ramaswamy (2001).
- Coefficientes de correlación: van't Veer *et al.* (2002).

Los perfiles de expresión génica de la presente invención también pueden utilizarse junto con otros métodos de diagnóstico no genéticos útiles en el diagnóstico, el pronóstico o el seguimiento del tratamiento del cáncer. Por ejemplo, en algunas circunstancias resulta beneficioso combinar la potencia diagnóstica de los métodos en base a la expresión génica descritos anteriormente con los datos procedentes de marcadores convencionales tales como los marcadores seroproteicos (por ejemplo, el antígeno tumoral 27.29 ("CA 27.29")). Existe una variedad de tales marcadores, incluidos analitos tales como CA 27.29. En uno de tales métodos, se extrae periódicamente sangre de un paciente tratado y a continuación se somete a un inmunoensayo enzimático para uno de los marcadores séricos descritos anteriormente. Cuando la concentración del marcador sugiere la reaparición de tumores o el fracaso del tratamiento, se toma una fuente de muestras susceptible de análisis de la expresión génica. Cuando existe una masa sospechosa, se toma una biopsia por aspiración con aguja fina (FNA) y a continuación se analizan los perfiles de expresión génica de las células extraídas de la masa, como se ha descrito anteriormente. Como alternativa, pueden tomarse muestras de tejido de las zonas adyacentes al tejido del que se eliminó con anterioridad un tumor. Este enfoque puede ser particularmente útil cuando otros ensayos producen resultados ambiguos.

Los artículos incluyen representaciones de los perfiles de expresión génica útiles para tratar, diagnosticar, pronosticar y evaluar de otra manera las enfermedades. Estas representaciones de perfiles se reducen a un medio que puede ser leído automáticamente por una máquina tal como medios legibles por ordenador (magnéticos, ópticos, y similares). Los artículos también pueden incluir instrucciones para evaluar los perfiles de expresión génica en tales medios. Por ejemplo, los artículos pueden comprender un CD ROM con instrucciones informáticas para comparar los perfiles de expresión génica de las carteras de los genes descritos anteriormente. Los artículos también pueden tener perfiles de expresión génica registrados digitalmente en los mismos de manera que puedan compararse con los datos de expresión génica de las muestras del paciente. Como alternativa, los perfiles pueden registrarse en diferentes formatos de representación. Un registro gráfico es uno de tales formatos. Los algoritmos de clustering, tales como los incorporados en el software Partek Discover™ y Partek Infer™ de Partek® mencionados anteriormente pueden ayudar a visualizar mejor tales datos.

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

Los kits fabricados según la invención incluyen ensayos formateados para determinar los perfiles de expresión génica. Estos pueden incluir todos o algunos de los materiales necesarios para llevar a cabo los ensayos, tales como reactivos e instrucciones.

La invención se ilustra adicionalmente mediante los siguientes ejemplos no limitativos.

Ejemplos: Los genes analizados según la presente invención están por lo general relacionados con las secuencias de ácido nucleico de longitud completa que codifican la producción de una proteína o péptido. Un experto en la materia reconocerá que la identificación de secuencias de longitud completa no es necesaria desde un punto de vista analítico. Es decir, pueden seleccionarse porciones de las secuencias o EST, según principios bien conocidos, para las que pueden diseñarse sondas para evaluar la expresión génica para el gen correspondiente.

Ejemplo 1

Manipulación de muestras y trabajo con micromatrices.

Se seleccionaron muestras de tumores congelados de pacientes LNN tratados durante 1980-1995, pero no tratados con el tratamiento neoadyuvante sistémico de nuestro banco de tumores en el Erasmus Medical Center (Rotterdam, Países Bajos). Todas las muestras de tumores se enviaron a nuestro laboratorio de referencia desde 25 hospitales regionales para las mediciones del receptor de hormonas esteroideas. Las directrices para el tratamiento primario fueron similares para todos los hospitales. Se seleccionaron los tumores de manera que se evitara el sesgo. En el supuesto de un 25%-30% en 5 años, y una pérdida sustancial de los tumores debido a razones de control de calidad, se procesaron 436 muestras de tumores invasivos. Se incluyeron pacientes con un resultado clínico malo, intermedio y bueno. Se rechazaron muestras en base al contenido insuficiente del tumor (53), la mala calidad del ARN (77) o la mala calidad de la matriz génica (20), lo que dejó 286 muestras aptas para su posterior análisis.

El protocolo de estudio fue aprobado por el comité de ética médica institucional (MEC. no. 02.953). La mediana de la edad de los pacientes en el momento de la cirugía (cirugía conservadora de mama: 219 pacientes; mastectomía radical modificada: 67 pacientes) fue de 52 años (intervalo, 26-83 años). Se administró radioterapia a 248 pacientes (87%) según nuestro protocolo institucional. Las proporciones de pacientes que se sometieron a tratamiento conservador de mama y radioterapia son normales para la enfermedad LNN. Los pacientes fueron incluidos independientemente del estado de la radioterapia, ya que este estudio no tuvo como objetivo investigar los efectos potenciales de un tipo específico de cirugía o de radioterapia complementaria. Además, los estudios han demostrado que la radioterapia no tiene un efecto claro sobre la reaparición a distancia de la enfermedad. Early Breast Cancer Trialists (1995). La negatividad de los ganglios linfáticos se basó en el examen patológico por parte de patólogos regionales. Foekens *et al.* (1989a).

Antes de la inclusión, se confirmó que las 286 muestras de tumor tenían tumor suficiente (> 70%) y una implicación uniforme del tumor en secciones congeladas de 5 μ m teñidas con H&E. Se midieron los niveles de ER (y PgR) mediante ensayo de unión a ligando o inmunoensayo enzimático (EIA) (Foekens *et al.* (1989b)) o mediante inmunohistoquímica (en 9 tumores). Los valores límite utilizados para clasificar a los pacientes como positivos o negativos para ER y PR fue 10 fmol/mg de proteína o un 10% de células tumorales positivas. El seguimiento postoperatorio implicaba un examen cada 3 meses durante los primeros 2 años, cada 6 meses durante 3 a 5 años y cada 12 meses a partir del quinto año. La fecha del diagnóstico de metástasis se definió como la fecha de confirmación de metástasis después de los síntomas comunicados por el paciente, la detección de signos clínicos o en un seguimiento habitual. La mediana del período de seguimiento de los pacientes supervivientes (n = 198) fue de 101 meses (intervalo, 20-171). De los 286 pacientes incluidos, 93 (33%) presentaron pruebas de metástasis a distancia en un plazo de 5 años y se contaron como fracasos en el análisis de la supervivencia sin metástasis a distancia (DMFS). Cinco pacientes (2%) fallecieron sin pruebas de enfermedad y se censuraron en el último seguimiento. Ochenta y tres pacientes (29%) fallecieron después de una recidiva anterior. Por lo tanto, un total de 88 pacientes (31%) fueron fracasos en el análisis de la supervivencia global (OS).

Ejemplo 2

Análisis de expresión génica de los datos obtenidos en el Ejemplo 1

Se aisló ARN total de 20 a 40 secciones de criostato de 30 μ m de espesor (50 mg-100 mg) con RNazol B

(Campro Scientific, Veenendaal, Países Bajos). Las dianas biotiniladas se prepararon utilizando métodos publicados (Affymetrix, CA, Lipshutz *et al.* (1999)) y se hibridaron con la micromatriz de oligonucleótidos de Affymetrix U133a GeneChip. Las matrices se escanearon utilizando los protocolos estándar de Affymetrix. Cada conjunto de sondas se trató como un gen separado. Los valores de la expresión se calcularon utilizando el software de análisis GeneChip de Affymetrix MAS 5.0. Las matrices génicas se rechazaron si la intensidad media era <40 o si la señal de fondo era > 100. Para normalizar las señales de las matrices génicas, los conjuntos de sondas se escalaron a una intensidad diana de 600 y no se seleccionaron los archivos de máscara de escala (*scale mask files*).

Ejemplo 3

Análisis estadístico de los genes identificados en el Ejemplo 2

Se filtraron los datos de expresión génica para incluir los genes invocados como "presente" en dos o más muestras. 17.819 genes pasaron este filtro y se utilizaron para el agrupamiento jerárquico. Antes del agrupamiento, se dividió el nivel de expresión de cada gen por la mediana de su nivel de expresión en los pacientes. Esta etapa de estandarización limitaba el efecto de la magnitud de la expresión de los genes y agrupaba los genes con patrones de expresión similares en el análisis de algoritmos de agrupamiento. Para identificar los subgrupos de pacientes, se llevó a cabo el agrupamiento jerárquico con enlace medio en los genes y en las muestras utilizando GeneSpring 6.0.

Para identificar los genes que discriminan a los pacientes que desarrollaron metástasis a distancia de aquellos que siguieron libres de metástasis durante 5 años, los inventores utilizaron dos enfoques de predicción de clases supervisados. En el primer enfoque, se asignaron al azar 286 pacientes a conjuntos de aprendizaje y de prueba de 80 y 206 pacientes, respectivamente. Se examinaron las curvas de supervivencia de Kaplan-Meier (Kaplan *et al.* (1958)) para los dos conjuntos para garantizar que no hubiese diferencia significativa y la selección al azar de los conjuntos de aprendizaje y ensayo no introdujo ningún sesgo. En el segundo enfoque, se asignaron los pacientes a uno de los dos subgrupos estratificados mediante el estado ER (Fig. 1).

Se analizó por separado cada subgrupo de pacientes a fin de seleccionar los marcadores. Se asignaron al azar los pacientes en el subgrupo ER positivo a conjuntos de aprendizaje y ensayo de 80 y 129 pacientes, respectivamente. Se dividieron al azar los pacientes en el subgrupo ER negativo en conjuntos de aprendizaje y ensayo de 35 y 42 pacientes, respectivamente. Los marcadores seleccionados de cada conjunto de aprendizaje del subgrupo se combinaron para formar una sola firma para predecir la metástasis tumoral para los pacientes ER positivo y ER negativo en una validación independiente posterior.

El tamaño de la muestra del conjunto de aprendizaje se determinó mediante un método de remuestreo para garantizar su nivel de confianza estadística. En resumen, el número de pacientes en el conjunto de aprendizaje comenzó en 15 pacientes y se incrementó en etapas de 5. Para un tamaño de muestra determinado, se realizaron 10 conjuntos de aprendizaje con pacientes seleccionados al azar. Se construyó una firma genética a partir de cada uno de los conjuntos de aprendizaje y a continuación se ensayó en un conjunto de prueba establecido de pacientes mediante el análisis de la curva de eficacia diagnóstica (ROC) con metástasis a distancia en un plazo de 5 años como punto definitorio. Se calcularon la media y el coeficiente de variación (CV) del área bajo la curva (AUC) para un tamaño de muestra determinado. Se eligió un número mínimo de pacientes necesarios para el conjunto de aprendizaje en el punto en que la AUC media alcanzaba una meseta y el CV de las 10 AUC era inferior al 5%.

Los genes se seleccionaron de la siguiente manera. En primer lugar, se utilizó la regresión de riesgos proporcionales de Cox univariante para identificar los genes para los que la expresión (en la escala log₂) se correlacionaba con la longitud de DMFS. Para reducir el efecto de múltiples ensayos y para ensayar la robustez de los genes seleccionados, se construyó el modelo de Cox con el método bootstrap de los pacientes en el conjunto de aprendizaje. Efron *et al.* (1981). En resumen, se construyeron 400 muestras bootstrap del conjunto de aprendizaje, cada una con 80 pacientes elegidos al azar con reemplazo. Se ejecutó un modelo de Cox en cada una de las muestras bootstrap. Se creó una puntuación bootstrap para cada gen eliminando el 5% de los valores p superior e inferior y a continuación promediando las inversas de los restantes valores p bootstrap. Se utilizó esta puntuación para clasificar los genes. Para construir una firma genética múltiple, se ensayaron combinaciones de marcadores génicos añadiendo los genes de uno en uno según el orden de clasificación. Se realizó el análisis ROC utilizando metástasis a distancia en un plazo de 5 años como punto definitorio para calcular el área bajo la AUC para cada firma con un número cada vez mayor de genes hasta alcanzar un valor de AUC máximo.

Se utilizó la puntuación de recidiva (RS) para calcular el riesgo de metástasis a distancia de cada paciente. La puntuación se definió como la combinación lineal de señales de expresión ponderadas con el coeficiente de regresión de Cox estandarizado como coeficiente de ponderación.

$$\text{Puntuación de Recidiva} = A \cdot I + \sum_{i=1}^{60} I \cdot w_i x_i + B \cdot (1 - I) + \sum_{j=1}^{16} (1 - I) \cdot w_j x_j$$

en la que

5

$$I = \begin{cases} 1 & \text{si el nivel de ER} > 10 \text{ fmol por mg de proteína} \\ 0 & \text{si el nivel de ER} \leq 10 \text{ fmol por mg de proteína} \end{cases}$$

10

A y B son constantes

w_i es el coeficiente de regresión de Cox estandarizado para ER + marcador

x_i es el valor de la expresión de ER + marcador en una escala log2

15

w_j es el coeficiente de regresión de Cox estandarizado para ER - marcador

x_j es el valor de la expresión de ER - marcador en una escala log2

20

El umbral se determinó a partir de la curva ROC del conjunto de aprendizaje para garantizar el 100% de sensibilidad y la especificidad más alta. Se eligieron unos valores de las constantes A de 313,5 y B de 280 para centrar el umbral de la RS a cero para los pacientes ER positivo y los ER negativo. Los pacientes con puntuaciones RS positivas se clasificaron en el grupo de mal pronóstico y los pacientes con puntuaciones RS negativas se clasificaron en el grupo de buen pronóstico. La firma genética y el límite se validaron en el conjunto de prueba. Se utilizaron gráficos de supervivencia de Kaplan-Meier y ensayos de rango logarítmico para evaluar las diferencias en el tiempo hasta la metástasis a distancia de los grupos con predicción de alto y bajo riesgo. Las oportunidades relativas (OR) se calcularon como la relación de las posibilidades de metástasis a distancia entre los pacientes con predicción de recidiva y aquellos con predicción de seguir libres de recidiva.

25

30

Los análisis univariable y multivariable con la regresión de riesgos proporcionales de Cox se realizaron en las variables clínicas individuales con y sin la firma genética. El HR y su intervalo de confianza del 95% (CI) se obtuvieron a partir de estos resultados. Todos los análisis estadísticos se realizaron utilizando el software S-Plus 6.1 (Insightful, VA).

Ejemplo 4

35

Análisis de vías de los genes identificados en el Ejemplo 3

40

Se asignó una clase funcional a cada uno de los genes en el gen de firma pronóstica. Se realizó el análisis de vías con el software Ingenuity 1.0 (Ingenuity Systems, CA). Se utilizaron sondas de Affymetrix como entrada para buscar las redes biológicas construidas por el software. Las redes biológicas identificadas por el programa se evaluaron en el contexto de las clases funcionales generales mediante clasificación por ontología GO. Se seleccionaron y evaluaron las vías con dos o más genes en la firma pronóstica.

Ejemplo 5

45

Resultados para los Ejemplos 1-4

Características de los pacientes y de los tumores

50

Las características clínicas y patológicas de los 286 pacientes se resumen en la Tabla 1.

55

60

65

Tabla 1. Características clínicas y patológicas de los pacientes y de sus tumores

Características	Todos los pacientes (%)	Conjunto de aprendizaje ER positivo (%)	Conjunto de aprendizaje ER negativo (%)	Conjunto de validación (%)
Número	286	80	35	171
Edad (media ± DE)	54±12	54±13	54±13	54±12
≤ 40 años	36(13)	12(15)	3(9)	21(12)
41-55 años	129(45)	30(38)	17(49)	82(48)
56-70 años	89(31)	28(35)	11(31)	50(29)
>70 años	32(11)	1(13)	4(11)	18(11)
Estado menopáusico				
Premenopáusica	139(49)	39(49)	16(46)	84(49)
Postmenopáusica	147(51)	41(51)	19(54)	87(51)
Estadio T				
T1	146(51)	38(48)	14(40)	94(55)
T2	132(46)	41(51)	19(54)	72(42)
T3/4	8(3)	1(1)	2(6)	5(3)
Grado de diferenciación				
Escaso	148(52)	37(46)	24(69)	87(51)
Moderado	42(15)	12(15)	3(9)	27(16)
Bueno	7(2)	2(3)	2(6)	3(2)
Desconocido	89(31)	2(36)	6(17)	54(32)
ER*				
Positivo	209(73)	80(100)	0(0)	129(75)
Negativo	77(27)	0(0)	35(100)	42(25)
PgR*				
Positivo	165(58)	59(74)	5(14)	101(59)
Negativo	111(39)	19(24)	29(83)	63(37)
Desconocido	10(3)	2(2)	1(3)	7(4)
Metástasis <5 años				
Sí	93(33)	24(30)	13(37)	56(33)
No	183(64)	51(64)	17(49)	115(67)
Censurado si <5 años	10(3)	5(6)	5(14)	0(0)

*ER positivo y PgR positivo: > 10 fmol/mg de proteína o > 10% de células tumorales positivas

No hubo diferencias en la edad o el estado menopáusico. El grupo de aprendizaje ER negativo tuvo una proporción ligeramente mayor de tumores más grandes y, como se esperaba, más tumores con escaso grado de diferenciación que el grupo de aprendizaje ER positivo. El grupo de validación de 171 pacientes (129 ER positivo, 42 ER negativo) no se diferenció del grupo total de 286 pacientes con respecto a ninguno de las características de los pacientes o de los tumores.

Se utilizaron dos enfoques para identificar los marcadores predictivos de recidiva de la enfermedad. En primer lugar, se dividieron al azar los 286 pacientes (ER positivo y ER negativo combinados) en un conjunto de aprendizaje y un conjunto de prueba. Se seleccionaron treinta y cinco genes de 80 pacientes en el conjunto de aprendizaje y se construyó un modelo de Cox para predecir la aparición de metástasis a distancia. Se observó un valor pronóstico moderado. Tabla 2. El análisis de algoritmos de agrupamiento no supervisado mostró dos subgrupos distintos altamente correlacionados con estado ER del tumor (ensayo de chi cuadrado: $p < 0,0001$). Figura 1B, que respaldaba el segundo enfoque en el que los pacientes se colocaban primero en subgrupos en base al estado ER.

Tabla 2

	SEQ ID NO:	Coefficiente de Cox	Valor p
5	1	4,008	0,00006
	2	-3,649	0,00026
	3	4,005	0,00006
	4	-3,885	0,00010
	5	-3,508	0,00045
10	6	-3,176	0,00150
	7	3,781	0,00016
	8	3,727	0,00019
	9	-3,570	0,00036
	10	-3,477	0,00051
15	11	3,555	0,00038
	12	-3,238	0,00120
	13	-3,238	0,00120
	14	3,405	0,00066
	15	3,590	0,00033
20	16	-3,157	0,00160
	17	-3,622	0,00029
	18	-3,698	0,00022
	19	3,323	0,00089
	20	-3,556	0,00038
25	21	-3,317	0,00091
	22	-2,903	0,00370
	23	-3,338	0,00085
	24	-3,339	0,00084
	25	-3,355	0,00079
30	26	3,713	0,00021
	27	-3,325	0,00088
	28	-2,984	0,00284
	29	3,527	0,00042
	30	-3,249	0,00116
35	31	-2,912	0,00360
	32	3,118	0,00182
	33	3,435	0,00059
	34	-2,971	0,00297
	35	3,282	0,00103

Se analizó cada subgrupo con el fin de seleccionar los marcadores. Se seleccionaron setenta y seis genes de pacientes en los conjuntos de aprendizaje (60 para el grupo ER positivo, 16 para el grupo ER negativo) (Fig. 2A, izquierda). Con los genes seleccionados y el estado ER en conjunto, se construyó un modelo de Cox para predecir la reaparición del cáncer para todos los pacientes LNN. La validación del predictor de 76 genes en el conjunto de prueba de 171 pacientes produjo una ROC con un valor de AUC de 0,694, una sensibilidad del 93% (52/56) y una especificidad del 48% (55/115) (Fig. 2A, derecha). Los pacientes con una puntuación de recidiva por encima del umbral de la firma pronóstica tienen una OR 11•9 veces superior (IC del 95%: 4,04-35,1, p <0,0001) de desarrollar metástasis a distancia en un plazo de 5 años. Como control, se generaron conjuntos de 76 genes seleccionados al azar. Estos produjeron una ROC con un valor de AUC medio de 0,515, una sensibilidad del 91% y una especificidad del 12% en el grupo de prueba. Los pacientes estratificados mediante un conjunto de genes de este tipo tendrían una oportunidad relativa de 1,3 (0,50-3,90, p = 0,8) de desarrollar metástasis, lo que indica una clasificación aleatoria. Además, el análisis de Kaplan-Meier para la supervivencia libre de metástasis a distancia (DMFS) y la supervivencia global (OS) en función de la firma genética de 76 genes mostró diferencias muy significativas en el tiempo hasta la metástasis entre los grupos con predicción de buen y mal pronóstico (Fig. 2). A los 60 y 80 meses, las respectivas diferencias absolutas en DMFS entre los grupos con predicción de buen y mal pronóstico eran del 40% (93% vs. 53%) y del 39% (88% vs. 49%) y aquellos en OS eran del 27% (97% vs. 70%) y del 32% (95% vs. 63%), respectivamente.

El perfil de 76 genes también representó un factor pronóstico sólido para el desarrollo de metástasis a distancia en los subgrupos de 84 pacientes premenopáusicas (HR: 9,60), 87 pacientes posmenopáusicas (HR: 4,04) y 79 pacientes con tamaños de tumor de 10 mm a 20 mm (HR: 14,1) (Fig. 3.).

La regresión de Cox univariante y multivariable se resumen en la Tabla 3.

65

Tabla 3: Análisis univariante y multivariable para DMFS en el conjunto de prueba de 171 pacientes LNN

	Análisis univariante			Análisis multivariable*		
	HR†	(95% CI)†	Valor p	HR†	(95%CI)†	Valor p
5	Edad ‡					
	Edad2 vs. Edad1	1,16 (0,51 - 2,65)	0,7180	1,14 (0,45 - 2,91)	0,7809	
	Edad3 vs. Edad1	1,32 (0,56 - 3,10)	0,5280	0,87 (0,26 - 2,93)	0,8232	
	Edad4 vs. Edad1	0,95 (0,32 - 2,82)	0,9225	0,61 (0,15 - 2,60)	0,5072	
10	Menopáusica					
	Estado§	1,24 (0,76 - 2,03)	0,3909	1,53 (0,68 - 3,44)	0,3056	
	Estadio □	1,08 (0,66 - 1,77)	0,7619	2,57 (0,23 - 29,4)	0,4468	
	Diferenciación ¶¶	0,38 (0,16 - 0,90)	0,0281	0,60 (0,24 - 1,46)	0,2590	
	Tamaño de tumor**	1,06 (0,65 - 1,74)	0,8158	0,34 (0,03 - 3,90)	0,3849	
15	ER††					
	ER††	1,09 (0,61 - 1,98)	0,7649	1,05 (0,54 - 2,04)	0,8935	
	PR††	0,83 (0,51 - 1,38)	0,4777	0,85 (0,47 - 1,53)	0,5882	
	Firma genética de 76 genes	5,67 (2,59 - 12,4)	1,5 x 10 ⁻⁵	5,55 (2,46 - 12,5)	3,6 x 10 ⁻⁵	

* El modelo multivariable incluía 162 pacientes, debido a que faltaban valores en 9 pacientes
 † Cociente de riesgos instantáneos e intervalo de confianza del 95%
 ‡ Edad1 es ≤ 40 años, Edad2 es de 41 a 55 años, Edad3 es de 56 a 70 años, Edad4 es > 70 años
 § Posmenopáusica vs. premenopáusica
 □ Estadio: II y III vs. I
 ¶¶ Grado de diferenciación: moderado/bueno vs. escaso, grado de diferenciación desconocido se incluyó como grupo distinto
 ** Tamaño de tumor > 20 mm vs. ≤ 20 mm
 †† Positivo vs. negativo

Aparte de la firma genética de 76 genes, sólo de grado de diferenciación fue significativo en el análisis univariante y se asoció una diferenciación moderada/buena con una DMFS favorable. La estimación de regresión multivariable de HR para la aparición de metástasis tumoral en un plazo de 5 años fue de 5,55 (p < 0,0001), lo que indica que el conjunto de 76 genes representa una firma pronóstica independiente fuertemente asociada con un mayor riesgo de metástasis tumoral. Los análisis univariante y multivariable también se realizaron por separado para los pacientes ER positivo y ER negativo. La firma genética de 76 genes también fue una variable pronóstica independiente en los subgrupos estratificados por el estado ER.

Se analizó la función de los 76 genes (Tabla 4) en la firma pronóstica para relacionar los genes con las vías biológicas.

Tabla 4

	Estado ER	SEQ ID NO.	Coefficiente de Cox est.	Valor p de Cox
40	+	36	-3,83	0,00005
	+	37	-3,865	0,00001
	+	38	3,63	0,00002
45	+	39	-3,471	0,00016
	+	40	3,506	0,00008
	+	41	-3,476	0,00001
	+	42	3,392	0,00006
	+	43	-3,353	0,00080
50	+	44	-3,301	0,00038
	+	45	3,101	0,00033
	+	46	-3,174	0,00128
	+	47	3,083	0,00020
	+	48	3,336	0,00005
55	+	49	-3,054	0,00063
	+	50	-3,025	0,00332
	+	51	3,095	0,00044
	+	52	-3,175	0,00031
	+	53	-3,082	0,00086
60	+	54	3,058	0,00016
	+	55	3,085	0,00009
	+	56	-2,992	0,00040

65

ES 2 504 242 T3

	Estado ER	SEQ ID NO.	Coefficiente de Cox est.	Valor p de Cox
	+	57	-2,791	0,00020
5	+	58	-2,948	0,00039
	+	59	2,931	0,00020
	+	60	-2,896	0,00052
	+	61	2,924	0,00050
	+	62	2,915	0,00055
10	+	63	-2,968	0,00099
	+	64	2,824	0,00086
	+	65	-2,777	0,00398
	+	66	-2,635	0,00160
	+	67	-2,854	0,00053
15	+	68	2,842	0,00051
	+	69	-2,835	0,00033
	+	70	2,777	0,00164
	+	71	-2,759	0,00222
	+	72	-2,745	0,00086
20	+	73	2,79	0,00049
	+	74	2,883	0,00031
	+	75	-2,794	0,00139
	+	76	-2,743	0,00088
	+	77	-2,761	0,00164
25	+	78	-2,831	0,00535
	+	79	2,659	0,00073
	+	80	-2,715	0,00376
	+	81	2,836	0,00029
	+	82	-2,687	0,00438
30	+	83	-2,631	0,00226
	+	84	-2,716	0,00089
	+	85	2,703	0,00232
	+	86	-2,641	0,00537
	+	87	-2,686	0,00479
35	+	88	-2,654	0,00363
	+	89	2,695	0,00095
	+	90	-2,758	0,00222
	+	91	2,702	0,00084
	+	92	-2,694	0,00518
40	+	93	2,711	0,00049
	+	94	-2,771	0,00156
	+	95	2,604	0,00285
	-	96	-3,495	0,00011
	-	97	3,224	0,00036
45	-	98	-3,225	0,00041
	-	99	-3,145	0,00057
	-	100	-3,055	0,00075
	-	101	-3,037	0,00091
	-	102	-3,066	0,00072
50	-	103	3,06	0,00077
	-	104	-2,985	0,00081
	-	105	-2,983	0,00104
	-	106	-3,022	0,00095
	-	107	-3,054	0,00082
55	-	108	-3,006	0,00098
	-	109	-2,917	0,00134
	-	110	-2,924	0,00149
	-	111	-2,882	0,0017

60 Aunque 18 de los 76 genes tienen una función desconocida, se identificaron varias vías o actividades bioquímicas que fueron bien representadas, tales como la muerte celular, el ciclo celular y la proliferación, la replicación y reparación del ADN y la respuesta inmunitaria (Tabla 5).

65

Tabla análisis 5. Análisis de vías de los 76 genes de la firma pronóstica

	Clase funcional	Firma genética de 76 genes
5	Muerte celular Ciclo celular	TNFSF10, TNFSF13, MAP4, CD44, IL18, GAS2, NEFL, EEF1A2, BCLG, C3 CCNE2, CD44, MAP4, SMC4L1, TNFSF10, AP2A2, FEN1, KPNA2, ORC3L, PLK1
	Proliferación	CD44, IL18, TNFSF10, TNFSF13, PPP1CC, CAPN2, PLK1, SAT
10	Recombinación/replicación, replicación de ADN Respuesta inmunitaria Crecimiento Ensamblaje y organización celular	TNFSF10, SMC4L1, FEN1, ORC3L, KPNA2, SUPT16H, POLQ, ADPRTL1 TNFSF10, CD44, IL18, TNFSF13, ARHGDI1B, C3 PPP1CC, CD44, IL18, TNFSF10, SAT, HDGFRP3 MAP4, NEFL, TNFSF10, PLK1, AP2A2, SMC4L1
15	Transcripción Señalización e interacción intercelular Supervivencia	KPNA2, DUSP4, SUPT16H, DKFZP434E2220, PHF11, ETV2 CD44, IL18, TNFSF10, TNFSF13, C3 TNFSF10, TNFSF13, CD44, NEFL
20	Desarrollo Morfología celular Síntesis de proteínas Unión a ATP Unión a ADN	IL18, TNFSF10, COL2A1 CAPN2, CD44, TACC2 IL18, TNFSF10, EEF1A2 PRO2000, URKL1, ACACB HIST1H4H, DKFZP434E2220, PHF11
25	Formación de colonias Adhesión Neurogénesis Aparato de Golgi Actividad quinasa Actividad transferasa	CD44, TNFSF10 CD44, TMEM8 CLN8, NEURL GOLPH2, BICD1 CNK1, URKL1 FUT3, ADPRTL1

30 Se descubrieron genes implicados en la evolución de la enfermedad, incluidos calpaina2, proteína de reconocimiento del origen, fosfatasa de doble especificidad, inhibidor de la disociación Rho-GDP, proteína de la superfamilia TNF, componente 3 del complemento, proteína asociada a microtúbulos, proteína fosfatasa 1 y regulador de la apoptosis BCL-G. Además, en la firma genética había genes pronósticos caracterizados con anterioridad tales como ciclina E2 (Keyomarsi *et al.* (2002)) y CD44 (Herrera-Gayol *et al.* (1999)).

35 **Ejemplo 6**

Análisis de los Ejemplos 1-5

40 Los inventores proporcionan resultados de un análisis de tumores primarios de 286 pacientes con cáncer de mama sin compromiso de los ganglios linfáticos de todos los grupos de edad y tamaños de tumor. Los pacientes no habían recibido tratamiento complementario sistémico, por lo que la evaluación de pronóstico multigénica no estaba sujeta a las contribuciones potencialmente generadoras de confusión por factores predictivos relacionados con el tratamiento sistémico.

45 Este estudio puso de manifiesto una firma genética de 76 genes que predice con exactitud la reaparición de tumores a distancia. Esta firma es aplicable a todos los pacientes de cáncer de mama LNN independientemente de la edad, el tamaño y el grado de diferenciación del tumor y del estado ER. En el análisis multivariable de Cox para DMFS la firma genética de 76 genes fue la única variable significativa, reemplazando a las variables clínicas, incluido el grado de diferenciación. Al cabo de 5 años, las diferencias absolutas en DMFS y OS entre los pacientes con las firmas genéticas de 76 genes buenas y malas fueron del 40% y del 27%, respectivamente. De los pacientes con una firma de pronóstico bueno, el 7% desarrolló metástasis a distancia y el 3% murió en un plazo de 5 años. Si se valida más a fondo, esta firma pronóstica producirá un valor predictivo positivo del 37% y un valor predictivo negativo del 95%, en el supuesto de una tasa de reaparición de la enfermedad del 25% en pacientes LNN. En concreto, esta firma puede ser valiosa para definir el riesgo de reaparición para la creciente proporción de tumores T1 (< 2 cm). La comparación con las directrices del St. Gallen y del NIH fue instructiva. Aunque garantizando que el mismo número de los pacientes de alto riesgo reciba el tratamiento necesario, la firma genética de 76 genes recomendaría la quimioterapia complementaria sistémica sólo al 52% de los pacientes de bajo riesgo, en comparación con el 90% y el 89% según las directrices del St. Gallen y del NIH, respectivamente (Tabla 6).

65

Tabla 6. Comparación de la firma genética de 76 genes y el actual consenso convencional sobre el tratamiento del cáncer de mama

Método	Pacientes orientados a recibir quimioterapia complementaria en el conjunto de prueba	
	Enfermedad metastásica a los 5 años (%)	Libre de enfermedad metastásica a los 5 años (%)
St. Gallen	52/55 (95)	104/115 (90)
NIH	52/55 (95)	101/114 (89)
Firma genética de 76 genes	52/56 (93)	60/115 (52)

Los criterios de consenso convencionales. St. Gallen: tumor \geq 2 cm, ER negativo, grado de diferenciación 2-3, paciente <35 años (cualquiera de estos criterios); NIH: tumor > 1 cm.

Por lo tanto, la firma genética de 76 genes puede dar como resultado una reducción del número de pacientes LNN de bajo riesgo a los que se recomienda un tratamiento complementario sistémico innecesario.

Los 76 genes en la firma pronóstica pertenecen a muchas clases funcionales, lo que sugiere que diferentes vías podrían conducir a la evolución de la enfermedad. La firma incluía genes bien caracterizados y 18 genes desconocidos. Este hallazgo podría explicar el rendimiento superior de esta firma en comparación con otros factores pronósticos. Aunque se encontraron genes implicados en la muerte celular, la proliferación celular y la regulación de la transcripción en ambos grupos de pacientes estratificados por el estado ER, los 60 genes seleccionados para el grupo ER positivo y los 16 genes seleccionados para el grupo ER negativo no tuvieron ninguna superposición. Este resultado respalda la idea de que la medida de la heterogeneidad y los mecanismos subyacentes para la evolución de la enfermedad podrían diferir para los dos subgrupos basados en el ER de los pacientes con cáncer de mama.

La comparación de nuestros resultados con los del estudio de van de Vijver *et al.* (2002) es difícil debido a las diferencias en los pacientes, las técnicas y los materiales utilizados. Van de Vijver *et al.* incluyeron pacientes sin compromiso de los ganglios linfáticos y con compromiso de los ganglios linfáticos, que habían recibido o no tratamiento complementario sistémico, y sólo mujeres menores de 53 años. Además, las plataformas de micromatrices utilizadas en los estudios son diferentes, Affymetrix vs. Agilent. De los 70 genes del estudio de van't Veer (2002), sólo 48 están presentes en la matriz Affymetrix U133A, mientras que de nuestros 76 genes sólo 38 están presentes en la matriz Agilent. Hay un solapamiento de 3 genes entre las dos firmas (ciclina E2, complejo de reconocimiento del origen y proteínas de la superfamilia TNF). A pesar de la aparente diferencia, ambas firmas incluían genes que identificaban varias vías comunes que podrían estar implicados en la reaparición de tumores. Este hallazgo respalda la idea de que si bien pudiera haber redundancia en los miembros génicos, podrían ser necesarias firmas eficaces para incluir la representación de las vías específicas.

Las fortalezas de nuestro estudio en comparación con el estudio de van de Vijver *et al.* (2002) son el mayor número de pacientes LNN no tratados (286 vs. 141) y la independencia de nuestra firma genética de 76 genes con respecto a la edad, el estado menopáusico y el tamaño de los tumores. El conjunto de validación de nuestros pacientes está completamente libre de solapamiento con el conjunto de aprendizaje a diferencia del 90% de los demás informes. Ransohoff (2004). En conclusión, ya que sólo aproximadamente el 30%-40% de los pacientes LNN no tratados desarrollan la reaparición de tumores, la firma pronóstica podría proporcionar una poderosa herramienta para identificar a los pacientes con bajo riesgo, evitando el tratamiento excesivo en un número considerable de pacientes. La recomendación de un tratamiento complementario sistémico en pacientes con cáncer de mama primario podría ser orientada en el futuro por esta firma pronóstica. El valor predictivo de nuestra firma genética con respecto a la eficacia de los diferentes modos de tratamiento sistémico podría ensayarse en el contexto de tratamiento complementario o en pacientes con enfermedad metastásica.

Ejemplo 7

Comparación de un perfil genético de un tumor de mama generado a partir de microdissección por captura con láser y tejido en bruto en un cáncer de mama en estadio I/II

La determinación de perfiles de expresión génica ha demostrado ser una potente herramienta de diagnóstico y pronóstico para diversos tipos de cáncer. Casi exclusivamente en todos los casos se utilizó ARN de tumor en bruto para la hibridación en la matriz génica. Los estrógenos juegan un papel importante en el desarrollo y crecimiento de los tumores dependientes de hormonas.

Aproximadamente el 75% de los cánceres de mama expresan el receptor de estrógeno (ER), que es un indicador para el tratamiento con tamoxifeno (complementario) y se asocia con los resultados del paciente.

Para comprender mejor los mecanismos desencadenados por el estrógeno en las células epiteliales de mama y su asociación con la tumorigénesis, se utilizó la microdissección por captura con láser (LCM) para conseguir una población histológicamente homogénea de células tumorales de 29 tumores de mama primario en estadio temprano, en combinación con el análisis de expresión GeneChip. De estos 29 pacientes, 11 eran ER negativo y 17

eran ER positivo en base a la unión a ligando cuantitativa o inmunoensayos enzimáticos en citosoles tumorales. Para la comparación, también se obtuvieron los perfiles de expresión génica utilizando ARN de tejido en bruto aislado a partir del mismo grupo de 29 pacientes.

5 Las muestras nuevas de tejido congelado se recogieron de 29 pacientes con cáncer de mama sin compromiso de los ganglios linfáticos (para las características de los tumores, Tabla 2) que habían sido tratados quirúrgicamente de un tumor de mama y no habían recibido tratamiento neocomplementario sistémico. Para cada muestra de tejido del paciente, se utilizó en primer lugar el portaobjetos H&E para evaluar la morfología celular. Se aisló ARN de las células tumorales obtenidas mediante LCM (PALM) realizada en secciones de criostato y a partir de secciones de criostato enteras, es decir, tejido en bruto del mismo tumor. Se analizó la calidad de la muestra de ARN mediante un Agilent Bioanalyzer. Las muestras de ARN se hibridaron con la matriz génica Affymetrix human U133A que contiene aproximadamente 22.000 conjuntos de sondas. Se cuantificó la fluorescencia y se normalizaron las intensidades. Se utilizaron el análisis de algoritmos de agrupamiento y el análisis de componentes principales para agrupar a los pacientes con perfiles de expresión génica similares. Se seleccionaron los genes que se expresan de manera diferencial entre las muestras ER positivo y ER negativo.

20 El ARN total aislado a partir de las células de cáncer de mama obtenidas mediante LCM se sometió a dos rondas de amplificación con T7 en la preparación diana, frente a una ronda de amplificación con ARN de tejido en bruto. Los niveles de expresión de 21 genes de control (Tabla 7) se compararon entre conjunto de datos LCM y el conjunto de tejido en bruto para demostrar la fidelidad de la amplificación lineal.

Tabla 7: Lista de genes de control

SEQ ID NO:	Nombre
112	proteína fosfatasa 2, subunidad reguladora B (B56), isoforma delta
113	factor de unión a CCCTC (proteína con dedos de cinc)
114	familia 4 de transportadores de solutos (intercambiador de aniones), miembro 1, proteína adaptadora
115	ribonucleasa P
116	proteína hipotética FLJ20188
117	proteína KIAA0323
118	ADNc FLJ12469
119	factor de iniciación de la traducción eIF-2b subunidad delta
120	ribonucleoproteína nuclear heterogénea K
121	hidroximetilbilano sintasa
122	ADNc DKFZp586O0222
123	fase de lectura abierta 4 del cromosoma 20
124	interaccionador 4 del receptor de la hormona tiroidea
125	hipoxantina fosforribosiltransferasa 1 (síndrome de Lesch-Nyhan)
126	homólogo de DnaJ (Hsp40), subfamilia C, miembro 8
127	fosfatasa de doble especificidad 11 (de interacción con el complejo 1 ARN/RNP)
128	autoantígeno 1 relacionado con la atopia de unión al calcio
129	factor 2 derivado de células del estroma
130	región del punto de rotura 1 del sarcoma Ewing
131	complejo de transcripción CCR4-NO, subunidad 2
132	proteína única 7 con caja F

50 Los resultados obtenidos se representan en la Tabla 8 y en las Figuras 4-9.

Tabla 8. Características clínicas de los pacientes

Característica	Nº de pacientes (%)
Edad en años	
< 40	1 (3)
40-44	5 (17)
45-49	8 (28)
≥ 50	15 (52)
Diámetro del tumor en mm	
≤ 20	11 (40)
> 20	17 (59)
Grado de diferenciación histológica	
II (intermedio)	5 (17)
III (escaso)	12 (41)

Característica	Nº de pacientes (%)
Estado del receptor de estrógeno	
Negativo	11 (40)
Positivo	17 (59)
Cirugía	
Tratamiento conservador de mama	26 (90)
Mastectomía	3 (10)
Quimioterapia	
No	29 (100)
Tratamiento hormonal	
No	29 (100)
Supervivencia libre de enfermedad en meses	
≤ 48	13 (45)
> 48	16 (55)

La Figura 4 es un agrupamiento jerárquico en base a 5.121 genes y muestra que las muestras LCM y de tejido en bruto se separan completamente en base a los perfiles globales de expresión de ARN. La Figura 5 es un gráfico de barras que representa los niveles de expresión de 21 genes de control (Tabla 6) en aislados de ARN de muestras LCM y de tejidos en bruto y muestra que una ronda de amplificación lineal adicional utilizada para el ARN obtenido mediante LCM no ocasionó la expresión diferencial de los genes de control. La Figura 6 es una vía de análisis de datos y la Figura 7 muestra el análisis PCA con conjuntos de genes filtrados. La Figura 8 es un gráfico circular en el que se utilizó el estado ER para asignar los subgrupos de pacientes. Los genes expresados de manera diferencial entre los sub-clústers ER positivo y ER negativo en las muestras LCM y de tejido en bruto se definieron mediante la t de Student. La Figura 9 es una serie de gráficos de barras que muestran los resultados del análisis de vías mediante ontología génica para los genes asociados exclusivamente con ER en muestras LCM, exclusivamente en los tejidos en bruto y para aquellos que son comunes en LCM y en tejido en bruto.

En resumen, los resultados obtenidos muestran varias conclusiones importantes. En primer lugar, genes relacionados con la proliferación celular y el metabolismo energético se vieron expresados de manera diferencial en pacientes ER-/ER+ en los conjuntos de datos de tejido en bruto y en los conjuntos de datos LCM. En segundo lugar, debido al enriquecimiento de las células de cáncer de mama mediante LCM, los genes implicados en la transducción de señales por unión a receptor de superficie celular, la transducción de señales RAS, la transducción de señales JAK-STAT y la apoptosis se encontraron asociados al estado ER. Estos genes no se identificaron en el conjunto de datos masivos. En tercer lugar, la microdissección proporciona un enfoque sensible al estudio de las células tumorales epiteliales y una mejor comprensión de la vía de señalización asociada con los receptores de estrógeno. Por lo tanto, es evidente que la aplicación del perfil de expresión génica descrito en el presente documento a las células tumorales aisladas mediante LCM es acorde con los resultados obtenidos en el tejido en bruto heterogéneo.

40 Ejemplo 8

Validación y análisis de vías de la firma pronóstica de 76 genes en el cáncer de mama

Este ejemplo presenta los resultados de un estudio de validación en el que se utilizó la firma genética de 76 genes para predecir los resultados de 132 pacientes obtenidos de 4 fuentes independientes.

Además, con el fin de evaluar la robustez de esta firma genética, este ejemplo proporciona adicionalmente la identificación de componentes sustituibles de la firma y describe cómo las sustituciones conducen a la identificación de las vías claves en una firma eficaz.

Se obtuvieron muestras nuevas de tejido congelado de 132 pacientes que habían sido tratados quirúrgicamente por un tumor de mama y no habían recibido tratamiento complementario sistémico. Las muestras de los pacientes utilizadas se recogieron entre 1980 y 1996 (Fig. 10). Para cada muestra de tejido del paciente, se utilizó un portaobjetos H&E para evaluar la morfología celular. A continuación, se prepararon muestras de ARN total y se analizó la calidad de la muestra mediante Agilent Bioanalyzer. Las muestras de ARN se analizaron mediante análisis de micromatrices. Se cuantificó la fluorescencia y se normalizaron las intensidades. Se calculó una Puntuación de Riesgo de Recidiva para cada paciente en base a los niveles de expresión de la firma genética de 76 genes. Los pacientes se clasificaron en grupos de resultado bueno y malo. Figuras 11 y 12.

Con el fin de evaluar la robustez de esta firma genética, se diseñaron y utilizaron dos análisis estadísticos. Figura 13. En primer lugar, se repitieron los procedimientos de selección de genes y construcción de la firma que se utilizaron para descubrir la firma genética de 76 genes. Como se muestra en la Tabla 8, se seleccionaron al azar diez conjuntos de aprendizaje de 115 pacientes cada uno, de un total de 286 pacientes. El resto de los pacientes se utilizó como conjunto de prueba.

En segundo lugar, se aumentó el número de pacientes en un conjunto de aprendizaje hasta el 80% de los

286 pacientes y se utilizó el 20% restante de los pacientes como conjunto de prueba. Este procedimiento de selección también se repitió 10 veces. En ambos procedimientos, se utilizaron las curvas de supervivencia de Kaplan-Meier para garantizar que no hubiera diferencia significativa en la supervivencia libre de enfermedad entre el par de aprendizaje y de prueba. Se seleccionaron los genes y se construyó una firma a partir de cada uno de los conjuntos de aprendizaje utilizando la regresión de riesgos proporcionales de Cox. Se validó cada firma en el correspondiente conjunto de prueba. Además, se asignó la firma pronóstica de 76 genes a grupos funcionales utilizando la clasificación por ontología GO. Se seleccionaron las vías que incluyen un número significativo de genes en la firma (valor $p < 0,05$ y > 2 coincidencias). Las vías seleccionadas también se evaluaron en todas las firmas pronósticas obtenidas de diferentes conjuntos de aprendizaje.

En la Tabla 9, A. contiene los resultados de 10 firmas que utilizan conjuntos de aprendizaje de 115 pacientes y B. contiene los resultados de 10 firmas que utilizan conjuntos de aprendizaje del 80% de los pacientes

	A		B	
AUC de ROC	0,62 (0,55-0,70)	AUC de ROC	0,62 (0,53-0,72)	
Sensibilidad	86% (0,84 - 0,88)	Sensibilidad	83% (0,81 - 0,85)	
Especificidad	34% (0,21 - 0,56)	Especificidad	46% (0,28 - 0,62)	
Frec. de recidiva	33%	Frec. de recidiva	33%	
PPV	40% (0,35 - 0,49)	PPV	47% (0,32 - 0,58)	
NPV	81% (0,75 - 0,89)	NPV	82% (0,78 - 0,89)	
Oportunidad relativa	3,5 (1,7-7,9)	Oportunidad relativa	5,6 (1,7-15)	

Los resultados obtenidos en este Ejemplo demuestran que:

- La firma genética de 76 genes se valida con éxito en 132 pacientes independientes, dando un valor de AUC de 0,757 en los 132 pacientes con cáncer de mama LNN de 4 fuentes independientes. La firma presenta un 88% de sensibilidad y un 41% de especificidad.
- El AUC media para las firmas con sustitución es 0,64 (IC del 95%: 0,53-0,72). Este resultado es coherente con la del predictor de 76 genes (AUC de 0,69). Veintiún vías sobrerrepresentadas en la firma genética de 76 genes también se encontraron en las demás firmas pronósticas, lo que sugiere que en la reaparición de tumores están implicadas vías biológicas comunes.
- Estos resultados sugieren que los perfiles de expresión génica proporcionan un enfoque de gran alcance para realizar la evaluación del riesgo del resultado para el paciente. Los datos destacan la viabilidad de un ensayo pronóstico molecular que proporciona a los pacientes una medición cuantitativa de recidiva del tumor.

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

Tabla 10

Identificación de secuencias				
SEQ ID NO:	psid	Nombre del gen	Número de registro	Descripción del gen
1	213165_at	CDABP0086	AI041204	
2	217432_s_at		AF179281	iduronato 2-sulfatasa (síndrome de Hunter)
3	221500_s_at		BE782754	sintaxina 16 /
4	208452_x_at	MYO9B	NM_004145	miosina IXB
5	220234_at	CA8	NM_004056	anhidrasa carbónica VIII
6	207865_s_at	BMP8	NM_001720	proteína morfogenética ósea 8 (proteína osteogénica 2)
7	201769_at	KIAA0171	NM_014666	producto del gen KIAA0171
8	218940_at	FLJ13920	NM_024558	proteína hipotética FLJ13920
9	209018_s_at	BRPK	BF432478	proteína quinasa BRPK
10	216647_at	DKFZp586L1824	AL117663	del clon DKFZp586L1824
11	213405_at	DKFZp564E122	N95443	del clon DKFZp564E122
12	202921_s_at	ANK2	NM_001148	anquirina 2, neuronal, variante de transcrito 1
13	208401_s_at		U01157	receptor del péptido 1 similar a glucagón con repetición de dinucleótido CA
14	218090_s_at	WDR11	NM_018117	proteína WD40 con dominio de repetición 11
15	218139_s_at	FLJ10813	NM_018229	proteína hipotética FLJ10813
16	202485_s_at	MBD2	NM_003927	proteína 2 con dominio de unión a metil-CpG, variante de transcrito 1
17	201357_s_at	SF3A1	NM_005877	factor de corte y empalme 3a, subunidad 1, 120 kD
18	214616_at	H3FD	NM_003532	familia de histonas H3, miembro D

35

40

45

50

55

60

65

ES 2 504 242 T3

19	207719_x_at	KIAA0470	NM_014812	producto del gen KIAA0470
20	202734_at	TRIP10	NM_004240	interaccionador 10 del receptor de la hormona tiroidea
21	202175_at	FLJ22678	NM_024536	proteína hipotética FLJ22678
22	213870_at		AL031228	clon 1033B10 en el cromosoma 6p21.2-21.31
23	208967_s_at	adk2	U39945	adenilato quinasa 2
24	204312_x_at		AI655737	proteína 1 de unión a elemento sensible a cAMP
25	203815_at	GSTT1	NM_000853	glutación S-transferasa ζ 1
26	207996_s_at	C18ORF1	NM_004338	fase de lectura abierta 1 del cromosoma 18
27	221435_x_at	HT036	NM_031207	proteína hipotética HT036
28	219987_at	FLJ12684	NM_024534	proteína hipotética FLJ12684
29	221559_s_at	MGC:2488	BC000229	clon MGC: 2488
30	207007_at	NR1I3	NM_005122	subfamilia de receptores nucleares 1, grupo I, miembro 3
31	219265_at	FLJ13204	NM_024761	proteína hipotética FLJ13204
32	40420_at		AB015718	ARNm de lok para la proteína quinasa
33	202266_at	AD022	NM_016614	proteína asociada a receptor de TRAF y TNF
34	219522_at	FJX1	NM_014344	supuesto ligando secretado homólogo a fix1
35	212334_at	AKAP350C	BE880245	AKAP350C, con corte y empalme alternativo
36	219340_s_at	CLN8	AF123759	Supuesta proteína transmembrana
37	217771_at	GP73	NM_016548	proteína de membrana de Golgi (LOC51280)
38	202418_at	Yif1p	NM_020470	supuesta proteína transmembrana; homóloga de la proteína de membrana de Golgi de levadura
39	206295_at	IL-18	NM_001562	interleucina 18
40	201091_s_at		BE748755	proteína similar a heterocromatina
41	204015_s_at	DUSP4	BC002671	fosfatasa de doble especificidad 4
42	200726_at	PPP1CC	NM_002710	proteína fosfatasa 1, subunidad catalítica, isoforma y
43	200965_s_at	ABLIM-s	NM_006720	proteína LIM de unión a actina 1, variante de transcrito
44	210314_x_at	TRDL-1	AF114013	Ligando de muerte 1 y relacionado con factor de necrosis tumoral
45	221882_s_at	M83	AI636233	proteína transmembrana de cinco cruces
46	217767_at	C3	NM_000064	componente 3 del complemento
47	219588_s_at	FLJ20311	NM_017760	proteína hipotética
48	204073_s_at	C11ORF9	NM_013279	fase de lectura abierta 9 del cromosoma 11
49	212567_s_at		AL523310	supuesto factor de iniciación de la traducción
50	211382_s_at	TACC2	AF220152	
51	201663_s_at	CAP-C	NM_005496	polipéptido C asociado a cromosoma
52	221344_at	OR12D2	NM_013936	receptor olfativo, familia 12, subfamilia D, miembro 2
53	210028_s_at	ORC3	AF125507	subunidad 3 del complejo de reconocimiento del origen
54	218782_s_at	PRO2000	NM_014109	proteína PRO2000
55	201664_at	SMC4	AL136877	(mantenimiento estructural del cromosoma 4, similar a levadura)
56	219724_s_at	KIAA0748	NM_014796	producto del gen KIAA0748
57	204014_at	DUSP4	NM_001394	fosfatasa de doble especificidad 4
58	212014_x_at	CD44	AI493245	CD44
59	202240_at	PLK1	NM_005030	quinasa 1 similar a Polo (<i>Drosophila</i>)
60	204740_at	CNK1	NM_006314	potenciador de conector de similar a KSR (supresor quinasa de ras de <i>Drosophila</i>)
61	208180_s_at	H4FH	NM_003543	familia de histonas H4, miembro de H
62	204768_s_at	FEN1	NM_004111	endonucleasa específica de estructura Flap
63	203391_at	FKBP2	NM_004470	proteína 2 de unión a FK506
64	211762_s_at	KPNA2	BC005978	carioferina α 2 (cohorte 1 de RAG, importina α 1)
65	218914_at	CGI-41	NM_015997	proteína CGI-41
66	221028_s_at	MGC11335	NM_030819	proteína hipotética MGC11335
67	211779_x_at	MGC13188	BC006155	Clon MGC: 13188
68	218883_s_at	FLJ23468	NM_024629	proteína hipotética FLJ23468
69	204888_s_at		AA772093	Similar a <i>Neuralized</i> (<i>Drosophila</i>)
70	217815_at	FACTP140	NM_007192	factor de elongación de la transcripción específico de cromatina, subunidad de 140 kD
71	201368_at	Tis11d	U07802	
72	201288_at	ARHGDIβ	NM_001175	Inhibidor de la disociación de Rho GDP (GDI) β
73	201068_s_at	PSMC2	NM_002803	Proteasoma (prosome, macropaina) subunidad 26S, ATPasa, 2

ES 2 504 242 T3

	74	218478_s_at	DKFZP434E2220	NM_017612	proteína hipotética DKFZP434E2220
	75	214919_s_at	KIAA1085	R39094	
5	76	209835_x_at		BC004372	Similar a CD44
	77	217471_at		AL117652	
	78	203306_s_at	SLC35A1	NM_006416	Familia de portador de soluto 35 (transportador de CMP-ácido siálico), miembro 1
	79	205034_at	CCNE2	NM_004702	ciclina E2
10	80	221816_s_at		BF055474	supuesto antígeno proteína dedo de zinc NY-REN-34
	81	219510_at	POLQ	NM_006596	polimerasa (dirigida a ADN) ζ
	82	217102_at		AF041410	proteína asociada al cáncer
	83	208683_at	CANP	M23254	subunidad grande de proteasa neutra activada por Ca ²
15	84	215510_at		AV693985	gen variante ets 2
	85	218533_s_at	FLJ20517	NM_017859	proteína hipotética FLJ20517
	86	215633_x_at	LST-1 N	AV713720	ARNm para la proteína LST-1 N
	87	221928_at		AI057637	Hs234898 ESTs, débilmente similar al factor de crecimiento de linfocitos B 2109260A
20	88	214806_at	BICD	U90030	Bicaudal-D
	89	204540_at	EEF1A2	NM_001958	factor de elongación de la traducción eucariota 1 α 2
	90	221916_at		BF055311	proteína hipotética
	91	216693_x_at	DKFZp434C1722	AL133102	
25	92	209500_x_at		AF114012	ligando de muerte 1β relacionado con el factor de necrosis tumoral
	93	209534_at	FLJ10418	AK001280	moderadamente similar al factor de crecimiento derivado de hepatoma
	94	207118_s_at	MMP23A	NM_004659	metaloproteinasas de matriz 23A
30	95	211040_x_at		BC006325	expresada en fase S y G-2, 1
	96	218430_s_at	FLJ12994	NM_022841	proteína hipotética FLJ12994
	97	217404_s_at		X16468	colágeno tipo II, α-1
	98	205848_at	GAS2	NM_005256	específica de detención del crecimiento 2
	99	214915_at	FLJ11780	AK021842	clon HEMBA1005931, débilmente similar a la proteína con dedos de zinc 83
35	100	216010_x_at		D89324	α (1,31,4) fucosiltransferasa
	101	204631_at	MYH2	NM_017534	miosina, polipéptido pesado 2, músculo esquelético adulto
	102	202687_s_at		U57059	ARNm del ligando Apo-2
40	103	221634_at		BC000596	Similar a proteína ribosómica L23a, clon MGC:2597
	104	220886_at	GABRQ	NM_018558	receptor de ácido γ-aminobutírico (GABA), ζ
	105	202237_at	ADPRTL1	NM_006437	similar a ADP-ribosiltransferasa (NAD ⁺ ; poli (ADP-ribosa) polimerasa), 1
	106	204218_at	DKFZP564M 082	NM_014042	proteína DKFZP564M082
45	107	221241_s_at	BCLG	NM_030766	regulador de la apoptosis BCL-G
	108	209862_s_at		BC001233	Similar al producto del gen KIAA0092, clon MGC:4896
	109	217019_at	RPS4X	AL137162	Contiene nuevo gen y 5 parte de gen para nueva proteína similar a proteína ribosómica 4 unida a X
50	110	210593_at		M55580	espermidinespermina N1-acetiltransferasa
	111	216103_at	KIAA0707	AB014607	KIAA0707
	112	202513_s_at	PPP2R5D	NM_006245	proteína fosfatasa 2, subunidad reguladora B (B56), isoforma delta
	113	202521_at	CTCF	NM_006565	factor de unión a CCCTC (proteína con dedos de cinc)
55	114	218682_s_at	SLC4A1AP	NM_018158	familia 4 de transportadores de solutos (intercambiador de aniones), miembro 1, proteína adaptadora
	115	203436_at	RPP30	NM_006413	ribonucleasa P
60	116	220127_s_at	FLJ20188	NM_017703	proteína hipotética FLJ20188
	117	212355_at	KIAA0323	AI075450	proteína KIAA0323
	118	215158_s_at	FLJ12469	AK022531	ADNc FLJ12469
	119	209429_x_at		AF112207	subunidad delta del factor de iniciación de la traducción eIF-2b
65	120	200097_s_at		AI701949	ribonucleoproteína nuclear heterogénea K

ES 2 504 242 T3

	121	203040_s_at	HMBS	NM_000190	hidroximetilbilano sintasa
	122	221647_s_at	DKFZp586O0222	AL136935	clon DKFZp586O0222
5	123	218089_at	C20orf4	NM_015511	fase de lectura abierta 4 del cromosoma 20 DKFZP564N1363
	124	203732_at	TRIP4	NM_016213	interaccionador 4 del receptor de la hormona tiroidea
	125	202854_at	HPRT1	NM_000194	hipoxantina fosforribosiltransferasa 1 (síndrome de Lesch-Nyhan)
10	126	205545_x_at	DNAJC8	NM_014280	homólogo de DnaJ (Hsp40), subfamilia C, miembro 8
	127	202703_at	DUSP11	NM_003584	fosfatasa de doble especificidad 11 (de interacción con el complejo 1 ARN/RNP)
	128	216903_s_at	CBARA1	AK022697	autoantígeno 1 relacionado con la atopía de unión a calcio
15	129	203090_at	SDF2	NM_006923	factor 2 derivado de células del estroma
	130	209214_s_at	EWSR1	BC004817	región del punto de rotura 1 del sarcoma de Ewing
	131	217798_at	CNOT2	AI123426	complejo de transcripción CCR4-NOT, subunidad 2
	132	201178_at	FBX07	NM_012179	proteína única 7 con caja F
20	133	212160_at		AI984005	exportina, ARNt (receptor de exportación nuclear para los ARNt)
	134	201111_at	CSE1	AF053641	proteína cerebral de susceptibilidad a apoptosis celular
	135	201112_s_at	CSE1L	NM_001316	Similar a segregación de cromosomas 1 (homólogo de levadura)
25	136	204817_at	KIAA0165	NM_012291	polos del huso extra, <i>S. cerevisiae</i> , homólogo de (KIAA0165)
	137	215623_x_at	FLJ11338	AK002200	muy similar al ARNm del polipéptido C asociado a cromosoma
30	138	38158_at	KIAA0165	D79987	
	139	201076_at	NHP2L1	NM_005008	similar a proteína cromosómica no histona 2 (<i>S.</i> <i>cerevisiae</i>), 1
	140	201947_s_at	CCT2	NM_006431	chaperonina que contiene TCP1, subunidad 2 (beta)
	141	202647_s_at	NRAS	NM_002524	homólogo de oncogén viral RAS de neuroblastoma (v- ras)
35	142	202705_at	CCNB2	NM_004701	ciclina B2
	143	204009_s_at		NM_004985	homólogo de oncogén viral del sarcoma de rata v-Ki- ras2 Kirsten 2
40	144	204566_at	PPM1D	NM_003620	proteína fosfatasa 1D dependiente de magnesio, isoforma delta
	145	214710_s_at		BE407516	
	146	202095_s_at	BIRC5	NM_001168	que contiene repetición de IAP de baculovirus 5 (survivina)
	147	204900_x_at	SAP30	NM_003864	polipéptido asociado a sin3, 30 kD
45	148	201986_at	KIAA0593	AB011165	ARNm de proteína KIAA0593, cds parcial
	149	201987_at		AI984051	proteína asociada al receptor de hormona tiroidea, subunidad de 240 kDa
	150	203605_at	SRP54	NM_003136	partícula de reconocimiento de señales 54kD
50	151	213226_at		AI346350	autoantígeno 1 de la polimiositis-esclerodermia (75kD)
	152	205757_at	ENTPD5	NM_001249	ectonucleósido trifosfato difosfohidrolasa 5
	153	212062_at	KIAA0611	AB014511	ARNm para la proteína KIAA0611
	154	213007_at		W74442	polimerasa (dirigida a ADN), Y
55	155	203362_s_at	MAD2L1	NM_002358	similar a MAD2 (deficiente para la detención de la mitosis, levadura, homólogo), 1
	156	204641_at	NEK2	NM_002497	quinasa 2 relacionada con NIMA (gen a nunca en mitosis)
	157	206983_at	CCR6	NM_004367	receptor de quimiocinas (motivo C-C) 6
	158	210375_at	EP3a2	X83858	receptor de prostaglandina E
60	159	213933_at	DKFZp586M 0723	AW242315	del clon DKFZp586M0723
	160	201756_at	RPA2	NM_002946	proteína de replicación A2 (32 kDa)
	161	208688_x_at	eIF3	U78525	factor de iniciación de la traducción eucariota
65	162	212655_at	KIAA0579	AB011151	proteína KIAA0579

ES 2 504 242 T3

	163	213124_at		BG538800	proteína DKFZP434N043
	164	213520_at	RECQL4	NM_004260	similar a proteína RecQ, 4
	165	218277_s_at	FLJ22060	NM_024612	proteína hipotética FLJ22060
5	166	201938_at	DOC1	NM_004642	delecionado en el cáncer de boca (ratón, homólogo) 1
	167	202692_s_at	UBTF	NM_014233	factor de transcripción de unión cadena arriba, ARN polimerasa I
	168	203616_at	POLB	NM_002690	polimerasa (dirigida a ADN), β
	169	204407_at		AF080255	proteína <i>lodestar</i>
10	170	206188_at	KIAA0628	NM_014789	producto del gen KIAA0628
	171	209831_x_at		AB004574	desoxirribonucleasa II, lisosómica
	172	211980_at		AI922605	colágeno, tipo IV, alfa 1
	173	214853_s_at		AI091079	proteína transformadora 1 (que contiene dominio de homología 2 con Src)
15	174	215888_at	FLJ23236	AK026889	clon COL00725
	175	216037_x_at		AA664011	similar a factor de transcripción 7, 2 (específica de linfocitos T, caja HMG)
	176	202666_s_at	BAF53A	NM_004301	BAF53
20	177	204146_at		BE966146	proteína de interacción con RAD51
	178	203920_at	NR1H3	NM_005693	subfamilia 1 de receptores nucleares, grupo H, miembro 3
	179	205322_s_at		AW182367	factor de transcripción 1 de regulación de metales
	180	206644_at	NR0B1	NM_000475	subfamilia 0 de receptores nucleares, grupo B, miembro 1
	181	201558_at	RAE1	NM_003610	(exportador de ARN 1, <i>S. pombe</i>) homólogo
25	182	209448_at		BC002439	proteína de interacción con Tat (30 kD)
	183	220960_x_at	RPL22	NM_000983	proteína ribosómica L22
	184	207320_x_at	STAU	NM_004602	variante de transcrito T4 staufen (<i>Drosophila</i> , proteína de unión a ARN)
	185	208948_s_at	MGC:4921	BC000830	MGC:4921
30	186	213037_x_at		AJ132258	proteína staufen, <i>Drosophila</i> parcial, proteína de unión a ARN)
	187	200725_x_at	RPL10	NM_006013	proteína ribosómica L10
	188	200937_s_at	RPL5	NM_000969	proteína ribosómica L5
35	189	208696_at	PNAS-102	AF275798	ARNm de chaperonina PNAS-102 que contiene TCP1, subunidad 5 (épsilon)
	190	209593_s_at	FKSG18	AF317129	FKSG18
	191	209619_at		K01144	cadena Y de antígeno MHC II
	192	218336_at	PFDN2	NM_012394	prefoldina 2
40	193	219390_at	FLJ20731	NM_017946	proteína hipotética FLJ20731
	194	206976_s_at	HSP105B	NM_006644	choque térmico 105 kDa
	195	204444_at	KNSL1	NM_004523	similar a quinesina, 1
	196	206364_at	KIAA0042	NM_014875	producto del gen KIAA0042
	197	209408_at		U63743	quinesina asociada a centrómero mitótico
45	198	202629_at	APPBP2	AV681579	proteína de unión a proteína precursora de β amiloide (cola citoplasmática), 2
	199	202630_at	APPBP2	AA046411	proteína de unión a proteína precursora de β amiloide (cola citoplasmática), 2
50	200	202631_s_at	APPBP2	NM_006380	proteína de unión a proteína precursora de β amiloide (cola citoplasmática), 2
	201	210629_x_at		AF000425	variante de corte y empalme cLST1A
	202	204670_x_at	HLA-DRB5	NM_002125	MHC, clase II, DR β 5
55	203	208306_x_at	HLA-DRB4	NM_021983	MHC, clase II, DR β 4
	204	206654_s_at	RPC32	NM_006467	polimerasa (ARN) III (dirigida a ADN) (32 kD)
	205	218360_at	RAB22A	NM_020673	RAB22A, miembro de la familia de oncogenes RAS
	206	209380_s_at	CFTRMRP	AF146074	proteína ABC, casete de unión a TP, subfamilia C
60	207	201114_x_at	PSMA7	NM_002792	subunidad de proteasoma (prosoma, macropaína), tipo α , 7
	208	202243_s_at	PSMB4	NM_002796	subunidad de proteasoma (prosoma, macropaína), tipo β , 4
	209	202244_at	PSMB4	NM_002796	subunidad de proteasoma (prosoma, macropaína), tipo β , 4
	210	203878_s_at	MMP11	NM_005940	metaloproteinasas de matriz 11 (estromelisin 3)
	211	216474_x_at		AF206667	β I triptasa de mastocitos, con corte y empalme alternativo
65	212	217009_at		AL121974	secuencia de ADN del clon RP3-417L20 en el cromosoma 6p12-21.3

ES 2 504 242 T3

	213	202968_s_at	Dyrk2	Y09216	ARNm para la proteína quinasa, Dyrk2
	214	202969_at		AI216690	quinasa regulada por fosforilación en tirosina (Y) de doble especificidad, 2
5	215	204092_s_at	STK15	NM_003600	serina treonina quinasa 15
	216	204171_at	RPS6KB1	NM_003161	proteína quinasa S6 ribosómica, 70 kD, polipéptido 1
	217	204825_at	KIAA0175	NM_014791	producto del gen KIAA0175
	218	208079_s_at	STK6	NM_003158	serina treonina quinasa 6
10	219	219148_at	TOPK	NM_018492	quinasa de unión a PDZ; proteína quinasa de linfocitos T
	220	219813_at	LATS1	NM_004690	homólogo de LATS (supresor tumoral grande, <i>Drosophila</i>) 1
	221	202779_s_at	E2-EPF	NM_014501	proteína transportadora de ubiquitina
	222	217978_s_at	HSA243666	NM_017582	proteína NICE-5
	223	210413_x_at	SCCA2	U19557	antígeno de carcinoma de células escamosas 2
15	224	219478_at	WFDC1	NM_021197	dominio 1 de núcleo con cuatro puentes disulfuro WAP
	225	204319_s_at	RGS10	NM_002925	regulador de la señalización de proteína G 10
	226	204017_at	KDEL3	NM_006855	receptor de retención de proteínas en el retículo endoplasmático KDEL (Lys-Asp-Glu-Leu), 3
20	227	206150_at	TNFRSF7	NM_001242	superfamilia de receptores del factor de necrosis tumoral, miembro 7
	228	205926_at	WSX-1	NM_004843	receptor de citocinas de clase I
	229	205400_at	WAS	NM_000377	Síndrome de Wiskott-Aldrich (eczema-trombocitopenia)
	230	209539_at	KIAA0006	D25304	ARNm para el gen KIAA0006
	231	221922_at		AW195581	proteína KIAA0761
25	232	200614_at	CLTC	NM_004859	clatrina, polipéptido pesado (Hc)
	233	202550_s_at	VAPB	NM_004738	proteína B y C asociada a VAMP (proteína de membrana asociada a vesículas)
	234	212159_x_at		AI125280	complejo proteico 2 relacionado con el adaptador, subunidad α 2
30	235	202733_at	P4HA2	NM_004199	procolágeno-prolina 2-oxoglutarato 4-dioxigenasa (prolina 4-hidroxilasa) α polipéptido II
	236	208905_at		BC005299	citocromo C, clon MGC:12367
	237	32137_at	JAG2	AF029778	Jagged2
	238	201088_at	KPNA2	NM_002266	carioferina α 2 (cohorte 1 de RAG, importina α 1)
35	239	202824_s_at	TCEB1	NM_005648	factor de elongación de la transcripción B (SIII), polipéptido 1 (15 kD, elongina C)
	240	201584_s_at	DDXL	NM_005804	helicasa de ARN nuclear, variante DECD de la familia con cajas DEAD
40	241	218461_at	LOC51184	NM_016301	proteína x 0004 (LOC51184)
	242	204489_s_at	CD44	NM_000610	CD44
	243	204490_s_at	CDw44	M24915	antígeno CDw44
	244	207165_at	HMMR	NM_012485	receptor para motilidad mediada por hialuronano (RHAMM)
	245	210916_s_at	CD44	AF098641	CD44 isoforma RC
45	246	212063_at	CMPX1	BE903880	proteína con dedos de zinc 6
	247	204470_at	GRO1	NM_001511	oncogén GRO1 (actividad estimuladora del crecimiento de melanoma, alfa)
	248	207430_s_at	MSMB	NM_002443	microseminoproteína, β
	249	210297_s_at		U22178	proteína secretora prostática 57
50	250	213009_s_at	FLJ12639	AK022701	ADNc FLJ12639
	251	203536_s_at	CIAO1	NM_004804	proteína Ciao1 WD40
	252	204026_s_at	ZWINT	NM_007057	interaccionador ZW10
	253	204435_at	KIAA0410	NM_014778	producto del gen KIAA0410
55	254	209271_at		AB032251	ARNm de BPTF para el factor de transcripción con dedo PHD bromodominio
	255	212074_at	KIAA0810	BE972774	proteína KIAA0810
	256	218009_s_at	PRC1	NM_003981	regulador proteico de la citocinesis 1
	257	218768_at	NUP107	NM_020401	proteína de complejo de poros nucleares
60	258	M33197_3_at	GAPDH	M33197	gliceraldehído-3-fosfato deshidrogenasa
	259	203524_s_at	MPST	NM_021126	mercaptopiruvato sulfurtransferasa
	260	206335_at	GALNS	NM_000512	galactosamina (N-acetil)-6-sulfato sulfatasa, (síndrome de Morquio, mucopolisacaridosis tipo IVA)
	261	203503_s_at	PEX14	NM_004565	factor de biogénesis de peroxisomas 14

65

ES 2 504 242 T3

5	262	202673_at	DPM1	NM_003859	polipéptido 1 de dolicol-fosfato manosiltransferasa, subunidad catalítica 1
	263	207543_s_at	P4HA1	NM_000917	procolágeno-prolina, 2-oxoglutarato 4-dioxigenasa (prolina 4-hidroxilasa), α polipéptido I
	264	204192_at	CD37	NM_001774	antígeno CD37
	265	204960_at	PTPRCAP	NM_005608	proteína tirosina fosfatasa, de tipo receptor, proteína asociada a C
10	266	211991_s_at		M27487	MHC de clase II DPw3-cadena alfa-1
	267	200822_x_at	TPI1	NM_000365	triosafosfato isomerasa 1
	268	219502_at	FLJ10858	NM_018248	proteína hipotética FLJ10858
	269	219499_at	FLJ10578	NM_018144	proteína hipotética FLJ10578
	270	209238_at		BE966922	sintaxina 3A
15	271	212593_s_at		N92498	
	272	201598_s_at	INPPL1	NM_001567	similar a inositol polifosfato fosfatasa, 1
	273	201760_s_at	LOC55884	NM_018639	proteína WD que contiene caja CS
	274	222077_s_at		AU153848	proteína activadora de GTPasa
20	275	203764_at	KIAA0008	NM_014750	producto del gen KIAA0008
	276	59705_at		AA911739	
	277	204070_at	RARRES3	NM_004585	respondedor al receptor de ácido retinoico (inducido por tazaroteno) 3
	278	212149_at		AW470003	
25	279	214039_s_at	DKFZp586E1124	T15777	DKFZp586E1124
	280	217956_s_at	MASA	NM_021204	enzima E-1
	281	200994_at		BG291787	proteína 7 de unión a RAN
	282	200995_at		AI741392	
30	283	205106_at	MTCP1	NM_014221	proliferación de linfocitos T maduros 1
	284	209035_at	hMK-1	M69148	Midquina (factor promotor del crecimiento de neuritas 2)
	285	210311_at	FGF5	AF171928	factor de crecimiento de fibroblastos 5 variante corta
	286	211251_x_at	NFY-C	U78774	ARNm de NFY-C
	287	220406_at	TGFB2	NM_003238	factor de crecimiento transformante, β 2
35	288	206967_at	CCNT1	NM_001240	ciclina T1
	289	204252_at		M68520	ARNm de proteína quinasa relacionada con cdc2
	290	205955_at	FLJ11136	NM_018336	proteína hipotética FLJ11136
	291	203258_at	DRAP1	NM_006442	proteína 1 asociada a DR1 (cofactor negativo 2 α)
	292	204022_at		AI668780	
40	293	207660_at		NM_004019	distrofina (distrofia muscular, tipos Duchenne y Becker),
	294	215050_x_at		BG325734	proteína quinasa 2 activada por proteína quinasa activada por mitógenos
	295	204337_at		AL514445	regulador 4 de señalización de proteína G
	296	217687_at		AA224446	
45	297	221546_at		BC000794	factor de corte y empalme de pre-ARNm similar a Prp de <i>S. cerevisiae</i> , 18
	298	206546_at	SYCP2	NM_014258	proteína del complejo sinaptonemal 2
	299	206278_at		D10202	ARNm para el receptor del factor activador de plaquetas
50	300	206429_at	F2RL1	NM_005242	similar al receptor del factor de coagulación II (trombina), 1
	301	216408_at		AJ302584	gen para el receptor olfativo, línea celular BM28.7
	302	221306_at	GPR27	NM_018971	receptor acoplado a proteína G, 27
	303	221442_at	MC3R	NM_019888	receptor de melanocortina 3
	304	201446_s_at		BF692742	
55	305	205018_s_at		NM_005757	proteína con dedos de zinc tipo C3H; similar a proteína B muscleblind de <i>D. melanogaster</i> (MBLL)
	306	214379_at		AI954458	AI954458
	307	214698_at		AW190873	
	308	219336_s_at		NM_015947	proteína CGI-18 (LOC51008)
60	309	220760_x_at	FLJ14345	NM_024733	proteína hipotética FLJ14345
	310	221480_at		BG180941	
	311	221615_at	PPIE	AF104013	peptidil-prolil cis-trans isomerasa E
	312	221923_s_at		AA191576	nucleofosmina (fosfoproteína nucleolar B23, numatrina)

65

ES 2 504 242 T3

	313	201211_s_at	DDX3	AF061337	ARN helicasa DDX3 con caja DEAD
	314	205638_at	BAI3	NM_001704	inhibidor de la angiogénesis específico del cerebro 3
	315	205881_at	ZNF74	NM_003426	proteína con dedos de zinc 74 (Cos52)
5	316	206179_s_at		NM_007030	proteína específica del cerebro p25 alfa (p25)
	317	206308_at		AJ223333	ARNm para supuesta ADN metiltransferasa
	318	207361_at	HBP1	NM_012257	proteína 1 que contiene caja HMG
	319	208902_s_at		BF431363	
	320	209603_at		AI796169	
10	321	214174_s_at		BE043700	
	322	215747_s_at		X06130	ARNm para gen del ciclo celular RCC1
	323	216480_x_at	AF10CALM	AF060927	proteína de fusión AF10CALM tipo I
	324	216711_s_at		M73444	ARNm de CCG1p
	325	222115_x_at		BC003693	similar a gen 3930401 K13 de ADNc de RIKEN
15	326	221686_s_at	DKFZp434J0450	AL136869	DKFZp434J0450
	327	210533_at	MSH4	AF104243	homólogo de MutS específico de meiosis
	328	217485_x_at	hPMS3	D38435	ARNm de hPMS3
	329	202162_s_at		AI769416	complejo de transcripción CCR4-NO, subunidad 8
20	330	202401_s_at	SRF	NM_003131	factor de respuesta sérica (factor de transcripción de unión al elemento de respuesta sérica c-fos)
	331	206067_s_at		NM_024426	tumor de Wilms 1 (WT1), variante de transcrito D
	332	206127_at	ELK3	NM_005230	proteína con dominio ETS (proteína accesoria a SRF 2)
	333	207402_at	ZNF132	NM_003433	proteína con dedos de zinc 132
25	334	207768_at	EGR4	NM_001965	respuesta de crecimiento precoz 4
	335	208414_s_at	HOXB3	NM_002146	caja homeótica B3
	336	214879_x_at		AY007087	clon TCCCIA00046, factor de transcripción cadena arriba 2, de interacción con c-fos
	337	219314_s_at	ZNF219	NM_016423	proteína con dedos de zinc 219
30	338	219779_at	FLJ20980	NM_024721	proteína hipotética FLJ20980
	339	220653_at	ZIM2	NM_015363	dedo de zinc, sellado 2
	340	219778_at	FOG2	NM_012082	Friend of GATA2
	341	203947_at	CSTF3	NM_001326	factor de estimulación de la escisión, 3 pre-ARN, subunidad 3, 77kD
35	342	220096_at	FLJ20378	NM_017795	proteína hipotética FLJ20378
	343	201326_at		BE737030	TCP1 que contiene chaperonina, subunidad 6A (zeta 1)
	344	206769_at	TMSB4Y	NM_004202	Timosina, beta 4, cromosoma Y
	345	211197_s_at	KIAA0653	AL355690	EST del clon 34465, inserto completo
40	346	204994_at	MX2	NM_002463	resistencia a mixovirus (influenza) 2, homólogo de murino
	347	201662_s_at		D89053	Acil-CoA sintetasa 3
	348	206141_at	MOCS3	NM_014484	molibdopterina sintasa sulfurilasa
	349	209992_at		AB044805	6-fosfofructo-2-quinasa isoforma cardíaca
	350	210160_at		BC000398	acetilhidrolasa del factor activador de plaquetas, isoforma lb, subunidad beta (30 kD),
45	351	218016_s_at	FLJ10509	NM_018119	proteína hipotética FLJ10509
	352	220582_at	FLJ12190	NM_025071	proteína hipotética FLJ12190
	353	222294_s_at		AW971415	
	354	202239_at	ADPRTL1	NM_006437	similar a ADP-ribosiltransferasa (NAD ⁺ ; poli(ADP-ribosa)polimerasa), 1
50	355	205342_s_at		AF026303	familia sulfotransferasa, citosólica, 1C, miembro 1
	356	202294_at		AI126490	
	357	201597_at	COX7A2	NM_001865	Cyt C oxidasa subunidad VIIa polipéptido 2 (hígado)
	358	206353_at	COX6A2	NM_005205	Cyt C oxidasa subunidad VIa polipéptido 2
55	359	218739_at	LOC51099	NM_016006	proteína CGI-58
	360	217557_s_at		AV710357	
	361	202413_s_at	USP1	NM_003368	proteasa específica de ubiquitina 1
	362	213661_at		AI671186	proteína DKFZP586H2123
	363	212729_at		AI916274	proteína KIAA1232
60	364	202951_at		BE048506	proteína serina treonina quinasa
	365	207667_s_at	MAP2K3	NM_002756	proteína quinasa activada por mitógenos quinasa 3
	366	212565_at		BE302191	proteína KIAA0965
	367	212740_at		BF740111	fosfoinosítido-3-quinasa, subunidad reguladora 4, p150
65	368	213490_s_at		AI762811	proteína quinasa activada por mitógenos quinasa 2

ES 2 504 242 T3

	369	213595_s_at	KIAA0451	AA127643	producto del gen KIAA0451
	370	220640_at	CSNK1G1	NM_022048	caseína quinasa 1, gamma 1
5	371	207569_at	ROS1	NM_002944	homólogo de oncogén del virus del sarcoma UR2 aviar v-ros, 1
	372	209041_s_at		BG395660	enzima conjugadora de ubiquitina E2G 2
	373	207214_at	PEC-60	NM_014471	péptido gastrointestinal
	374	214425_at		AV645756	precursor de alfa-1-microglobulina/bicunina
10	375	203650_at	EPCR	NM_006404	receptor de proteína C, endotelial
	376	210733_at	PRO1292	AF130055	FLB4941 PRO1292, proteína de membrana de asociación a cadena de translocación
	377	220056_at	IL22R	NM_021258	receptor de interleucina 22
15	378	205049_s_at	CD79A	NM_001783	antígeno CD79A (alfa asociado a inmunoglobulina) variante de transcrito 1
	379	211245_x_at		AF002256	homólogo del receptor inhibidor de linfocitos citolíticos, cl-9, ARNm, receptor similar a Fean, dos dominios, cola citoplasmática de largo, 4
	380	212128_s_at		AW411370	distroglicano 1 (glicoproteína asociada a distrofina 1)
20	381	205019_s_at	VIPR1	NM_004624	receptor del péptido intestinal vasoactivo 1
	382	206001_at	NPY	NM_000905	neuropéptido Y
	383	216289_at	MAMMA1002427	AU148039	MAMMA1002427
	384	208250_s_at	DMBT1	NM_004406	delecionado en tumores cerebrales malignos 1
25	385	202091_at		BC003087	ligante de Arl Two, clon MGC:1121
	386	205068_s_at		BE671084	regulador de GTPasa asociado con la quinasa de adhesión focal pp125 (FAK)
	387	221136_at	GDF2	NM_016204	factor de diferenciación de crecimiento 2
30	388	202688_at	TNFSF10	NM_003810	superfamilia de factores de necrosis tumoral (ligando), miembro 10
	389	214336_s_at	TNFSF10	AI621079	complejo proteico coatómero, subunidad α , superfamilia de TNF (ligando), miembro 10
	390	213055_at		BF693956	antígeno CD47 (antígeno relacionado con Rh, transductor de señales asociado a integrina)
35	391	201719_s_at	EPB41 L2	NM_001431	similar a proteína banda 4.1 de membrana de eritrocitos, 2
	392	208353_x_at	ANK1	NM_020480	anquirina 1, variante de transcrito eritrocítica 7
	393	215717_s_at		X62009	ARNm parcial para fibrilina 5
40	394	206826_at	PMP2	NM_002677	proteína de mielina periférica 2
	395	207542_s_at	AQP1	NM_000385	acuaporina 1 (proteína integral de formación de canal, 28 kD)
	396	207596_at	PRO2176	NM_018515	proteína hipotética PRO2176
	397	208297_s_at	EVI5	NM_005665	sitio de integración de virus ecotrópicos 5
45	398	217289_s_at	G6PT	AF097831	transportador de glucosa-6-fosfato
	399	205972_at		NM_006841	proteína transportadora; transportador de glutamina acoplado a Na ⁺ y H ⁺ , sistema N1 (G17)
	400	218835_at	SFTPA2	NM_006926	proteína asociada a surfactante pulmonar A2
	401	219716_at	APOL6	NM_030641	apolipoproteína L, 6
50	402	218928_s_at	SLC37A1	NM_018964	familia de transportadores de solutos 37 (transportador de glicerol-3-fosfato), miembro 1
	403	214205_x_at	FLJ12069	AK022131	FLJ12069
	404	205008_s_at	KIP2	NM_006383	proteína de interacción con la subunidad catalítica de la proteína quinasa dependiente de ADN, 2
55	405	211272_s_at	DAGK1	AF064771	diacilglicerol quinasa α del clon 24
	406	206316_s_at	KIAA0166	NM_014708	producto del gen KIAA0166
	407	209737_at	KIAA0705	AB014605	proteína KIAA0705
	408	213117_at	FLJ10262	AW138594	proteína hipotética FLJ10262
	409	217161_x_at		X17406	proteoglicano específico de cartilago
60	410	209436_at	KIAA0762	AB018305	proteína KIAA0762
	411	213993_at		AI885290	espondina 1, proteína de la matriz extracelular (f-espondina)
	412	214354_x_at		T91506	similar a N-acilesfingosina amidohidrolasa (ceramidasa ácida)
65	413	205799_s_at		M95548	proteína transportadora de aminoácidos

ES 2 504 242 T3

	414	213522_s_at		AA527578	familia de transportadores de solutos 16 (transportadores de ácido monocarboxílico), miembro 3
5	415	212393_at		AL096767	secuencia de ADN del clon 579N16 en el cromosoma 22
	416	219179_at	LOC51339	NM_016651	proteína del nuevo gen del carcinoma hepatocelular 3
	417	201502_s_at		AI078167	factor nuclear del inhibidor del potenciador del gen del polipéptido ligero kappa en linfocitos B, α
10	418	201512_s_at		BC003633	homólogo A de translocasa de la membrana mitocondrial externa 70 (levadura)
	419	201576_s_at	GLB1	NM_000404	galactosidasa, beta 1
	420	201765_s_at		AL523158	
	421	215155_at	HEXA	J04178	cadena α de β -hexosaminidasa anómala
	422	203518_at	CHS1	NM_000081	síndrome de Chediak-Higashi 1
15	423	220801_s_at	HAO2	NM_016527	hidroxiácido oxidasa 2 (cadena larga)
	424	204690_at	STX8	NM_004853	sintaxina 8
	425	201126_s_at	MGAT1	NM_002406	manosil (α -,3-)-glicoproteína β -,2-N-acetilglucosaminiltransferasa
	426	206925_at	SIAT8D	NM_005668	sialiltransferasa 8(α -2 8-polisialiltransferasa)D
20	427	205319_at	PSCA	NM_005672	antígeno de células madre de próstata
	428	206199_at	CEACAM7	NM_006890	molécula de adhesión celular relacionada con el antígeno carcinoembrionario, 7
	429	207695_s_at	IGSF1	NM_001555	superfamilia de inmunoglobulinas, miembro 1
	430	219249_s_at	FLJ22041	NM_021939	proteína hipotética FLJ22041
25	431	213413_at	FLJ13555	BG434174	FLJ13555
	432	219793_at	SNX16	NM_022133	nexina de direccionamiento 16
	433	212035_s_at	KIAA1067	AI817079	proteína KIAA1067
	434	202730_s_at	PDCD4	NM_014456	muerte celular programada 4
	435	202130_at	sudD	AA725102	supresor de bimD6, homólogo de <i>Aspergillus nidulans</i>
30	436	206623_at	PDE6A	NM_000440	fosfodiesterasa 6A, específica de cGMP, bastón, α
	437	222201_s_at	KIAA1315	AB037736	proteína KIAA1315
	438	221752_at	KIAA1298	AL041728	proteína KIAA1298
	439	205003_at	KIAA0716	NM_014705	producto del gen KIAA0716
35	441	205006_s_at	NMT2	NM_004808	N-miristoiltransferasa 2
	442	205677_s_at	DLEU1	NM_005887	deleciónado en la leucemia linfocítica, 1
	443	204947_at	E2F1	NM_005225	factor de transcripción E2F, 1
	444	207505_at	PRKG2	NM_006259	proteína quinasa, dependiente de GMPc, tipo II
40	445	211036_x_at		BC006301	subunidad del complejo promotor de la anafase 5, clon MGC:13295,
	446	211269_s_at		K03122	ARNm del receptor de interleucina 2 (forma corta)
	447	202610_s_at	TRAP170	AF135802	componente del complejo proteico asociada al receptor de hormona tiroidea
45	448	204762_s_at		BE670563	proteína de unión a nucleótidos de guanina (proteína G), polipéptido O con acción activadora de alfa
	449	201313_at	ENO2	NM_001975	enolasa 2, (γ , neuronal)
	450	203225_s_at	FLJ11149	NM_018339	proteína hipotética FLJ11149
	451	221567_at	NOP	AF064599	proteína nucleolar Nop30
	452	201518_at	CBX1	NM_006807	homólogo 1 de cromobox (HP1 beta de <i>Drosophila</i>)
50	453	204970_s_at		NM_002359	familia de oncogenes del fibrosarcoma musculoaponeurótico v-maf (aviar), proteína G (MAFG),
	454	214615_at	P2Y10	NM_014499	supuesto receptor purinérgico
	455	221386_at	OR3A2	NM_002551	familia de receptores olfativos 3, subfamilia A, miembro 2
55	456	200014_s_at	HNRPC	NM_004500	ribonucleoproteína nuclear heterogénea C (C1C2)
	457	200053_at	SPAG7	NM_004890	antígeno asociado a esperma 7
	458	203462_x_at	EIF3S9	NM_003751	factor de iniciación de la traducción 3 sub 9 (eta 116kD)
	459	205917_at	ZNF264	NM_003417	proteína con dedos de zinc 264
	460	207753_at	ZNF304	NM_020657	proteína con dedos de zinc 304
60	461	209751_s_at		AF291676	proteína-2A de interacción con MBP-1
	462	217403_s_at	BC228680	AC074331	cromosoma 19, BAC CIT-HSPC_204F22
	463	219571_s_at	GIOT-3	NM_016265	GIOT-3 para represor de la transcripción inducible por gonadotropina-3

ES 2 504 242 T3

	464	203119_at	MGC2574	NM_024098	proteína hipotética MGC2574
	465	203721_s_at	LOC51096	NM_016001	proteína CGI-48
	466	204880_at	MGMT	NM_002412	O-6-metilguanina-ADN metiltransferasa
5	467	205664_at	KIN	NM_012311	determinante antigénico de homólogo de proteína recA (ratón)
	468	212597_s_at		AL079310	mapeo del nuevo gen humano al cromosoma 22
	469	212759_s_at		AI703074	similar a factor de transcripción 7, 2 (específico de linfocitos T, caja HMG)
	470	213824_at		AF221520	proteína 2 de unión a proteína quinasa C
10	471	218137_s_at	FLJ13159	NM_021940	proteína hipotética FLJ13159
	472	220445_s_at	TRAG3	NM_004909	gen asociado a resistencia a taxol 3
	473	203720_s_at		NM_001983	deficiencia para la reparación en roedores de la reparación por escisión por complementación cruzada, grupo de complementación 1
15	474	201046_s_at	RAD23A	NM_005053	homólogo A de RAD23 (<i>S. cerevisiae</i>)
	475	202344_at	HSF1	NM_005526	factor de transcripción de choque térmico 1
	476	202580_x_at	FOX M1	NM_021953	caja forkhead M1
	477	205690_s_at	G10	NM_003910	transcrito G10 materno
20	478	206307_s_at	FOXD1	NM_004472	caja forkhead D1
	479	213090_s_at		AI744029	factor asociado a proteína de unión a caja TATA (TBP)
	480	218215_s_at	NR1H2	NM_007121	subfamilia de receptores nucleares 1, grupo H, miembro 2
	481	207469_s_at	PIR	NM_003662	pirina
	482	209062_x_at	RAC3	AF010227	coactivador asociado al receptor 3
25	483	200823_x_at	RPL29	NM_000992	proteína ribosómica L29
	484	202868_s_at	POP4	NM_006627	homólogo de POP4 (procesamiento del precursor, <i>S. cerevisiae</i>)
	485	217747_s_at	RPS9	NM_001013	proteína ribosómica S9
30	486	32723_at		L02547	subunidad de 50 kDa del factor de estimulación de escisión (clon pZ50-19)
	487	212105_s_at		BE910323	polipéptido de la caja DEADH (Asp-Glu-Ala-AspHis) 9
	488	207320_x_at		NM_004602	staufer (<i>Drosophila</i> , proteína de unión a ARN) (STAU), variante de transcrito T4,
	489	200081_s_at		BE741754	proteína ribosómica S6
35	490	217559_at		AI001784	muy similar a proteína ribosómica A42735, L10, citosólica
	491	217907_at	HSPC071	NM_014161	proteína HSPC071
	492	213504_at		W63732	subunidad 6 de COP9 (homólogo de MOV34, 34 kD)
	493	204102_s_at	EEF2	NM_001961	factor de elongación de la traducción eucariota 2
40	494	208696_at		AF275798	PNAS-102
	495	209275_s_at	CLN3	AF015593	proteína CLN3
	496	214591_at		BF215673	proteína similar a Kelch de <i>Drosophila</i>
	497	201642_at	IFNGR2	NM_005534	receptor de interferón gamma 2 (transductor de interferón gamma 1)
45	498	219080_s_at	CTPS2	NM_019857	CTP sintasa II
	499	202282_at	HADH2	NM_004493	hidroxiacil-coenzima A deshidrogenasa, tipo II
	500	201453_x_at	RHEB2	NM_005614	homólogo de Ras enriquecido en el cerebro 2
	501	208733_at		NM_002865	RAB2, miembro de la familia de oncogenes RAS
	502	211004_s_at		BC002553	similar a aldehído deshidrogenasa 7
50	503	201400_at	PSMB3	NM_002795	proteasoma (prosome, macropaína) sub β tipo 3
	504	208799_at		BC004146	proteasoma (prosome, macropaína) sub β tipo 5
	505	216088_s_at		AL078633	secuencia de ADN del clon RP5-1005F21 en el cromosoma 20
	506	213912_at	KIAA0984	AW134976	proteína KIAA0984
	507	213913_s_at	KIAA0984	AW134976	proteína KIAA0984
55	508	205356_at	USP13	NM_003940	proteasa específica de ubiquitina 13 (isopeptidasa T-3)
	509	208166_at	MMP16	NM_022564	metaloproteínasa de matriz 16 (insertada en membrana) variante de transcrito 2,
	510	202018_s_at	LTF	NM_002343	lactotransferrina
60	511	213829_x_at	FLJ20478	AK000485	FLJ20478
	512	212766_s_at		AW294587	similar a proteína hipotética FLJ12484,
	513	202969_at		Y09216	quinasa regulada por fosforilación en tirosina (Y) de doble especificidad, 2
	514	203218_at		W37431	proteína quinasa activada por mitógeno 9

65

ES 2 504 242 T3

	515	204171_at	RPS6KB1	NM_003161	proteína S6 quinasa ribosómica,
	516	204906_at		BC002363	proteína S6 quinasa ribosómica, 90 kD, polipéptido 2,
	517	205126_at	VRK2	NM_006296	quinasa relacionada con vaccinia 2
5	518	214716_at	DKFZp434P0116	AW504018	proteína hipotética DKFZp434P0116
	519	37170_at	HRIHFB2017	AB015331	HRIHFB2017
	520	211474_s_at		BC004948	MGC:10846,
	521	204237_at	CED-6	NM_016315	proteína CED-6
10	522	203005_at	LTBR	NM_002342	receptor de linfotóxina β (superfamilia de TNFR, miembro 3)
	523	211000_s_at	gp130-RAPS	AB015706	gp130 de la forma soluble portadora del péptido antigénico de la artritis reumatoide
	524	202679_at	NPC1	NM_000271	enfermedad de Niemann-Pick, tipo C1
15	525	205282_at	LRP8	NM_004631	proteína relacionada con el receptor de lipoproteína de baja densidad 8, receptor de apolipoproteína E
	526	207270_x_at	CMRF35	NM_006678	receptor similar a inmunoglobulina de leucocitos CMRF35
	527	211846_s_at		AF110314	receptor similar a inmunoglobulina de herpesvirus HlgR
20	528	201810_s_at		AL562152	proteína de unión a dominio SH3 5 (asociada a BTK)
	529	213684_s_at		BF671400	proteína LIM (similar a enigma de unión a proteína quinasa C de rata)
	530	206636_at	RASA2	NM_006506	activador de proteína p21 de RAS, 2
25	531	219907_at	SNT-2	NM_006653	diana de factor neurotrófico asociado a suc1, 2 (adaptador de señalización de FGFR)
	532	201281_at	GP110	NM_007002	glicoproteína de membrana celular, 110000M(r) (antígeno de superficie)
	533	209462_at		U48437	proteína similar a precursor de amiloide 1
30	534	210880_s_at		AB001467	Efs2,
	535	212567_s_at		AL523310	CS0DC001YN06 (3 prima)
	536	202222_s_at	DES	NM_001927	desmina
	537	219615_s_at	KCNK5	NM_003740	canal de potasio, subfamilia K, miembro 5 (TASK-2)
	538	214210_at		AL049764	clon RP3-362J20 en cromosoma 22q13.1-13.31
35	539	52078_at		AI828080	
	540	203765_at	GCL	NM_012198	gran calcina
	541	213308_at		AB028945	proteína de unión a dominio SH3 de cortactina
	542	203082_at	KIAA0187	NM_014753	producto del gen KIAA0187
	543	212543_at	AIM1	U83115	proteína similar a β gamma-cristalina no lente
40	544	209709_s_at	RHAMM	U29343	receptor de ácido hialurónico
	545	205807_s_at	TUFT1	NM_020127	tuftelina 1
	546	213895_at		BF445047	proteína de membrana epitelial 1
	547	201565_s_at	ID2	NM_002166	inhibidor de la unión a ADN 2, proteína hélice-bucle-hélice dominante negativa
45	548	202639_s_at		AI689052	proteína de unión a RAN 3
	549	209230_s_at	COM1	AF135266	homólogo de proteína p8
	550	218626_at	4E-T	NM_019843	transportador de eIF4E
	551	221506_s_at		BG258639	carioferina beta 2b, transportina
50	552	221521_s_at		BC003186	proteína HSPC037, clon MGC:673,
	553	202069_s_at		AI826060	isocitrato deshidrogenasa 3 (NAD+) α
	554	203524_s_at	MPST	NM_021126	mercaptopiruvato sulfurtransferasa
	555	201916_s_at	SEC63L	NM_007214	SEC63, similar a componente de translocon de retículo endoplasmático (<i>S. cerevisiae</i>)
55	556	217776_at	ARSDR1	AF167438	deshidrogenasa/reductasa de cadena corta regulada por andrógenos 1
	557	218623_at	LOC51617	NM_015980	proteína HMP19
	558	208777_s_at		AF001212	subunidad de proteasoma 26S, 9
60	559	203971_at	SLC31A1	NM_001859	familia de transportadores de solutos 31 (transportadores de cobre), miembro 1
	560	215243_s_at		AF099730	gen de conexina 31 (GJB3),
	561	201591_s_at		NM_007184	receptor candidato de imidazolina (I-1),
	562	205466_s_at	HS3ST1	NM_005114	heparán sulfato (glucosamina) 3-O-sulfotransferasa 1
65	563	207829_s_at	BNIP1	NM_013978	proteína 1 de interacción con Bcl-2 adenovirus E1B 19kD, variante de transcrito BNIP1-a,

ES 2 504 242 T3

	564	205742_at	TNNI3	NM_000363	troponina I, cardíaca
	565	211178_s_at		AF038602	proteína 1 de unión a CD2, forma corta
5	566	219113_x_at	LOC51171	NM_016246	deshidrogenasa/reductasa de cadena corta retiniana retSDR3
	567	217807_s_at	GLTSCR2	NM_015710	gen de región candidato supresor de tumores glioma 2
	568	206133_at	HSXIAPAF1	NM_017523	factor 1 asociado a XIAP
	569	201379_s_at	TPD52L2	NM_003288	similar a proteína tumoral D52, 2
	570	209373_at		BC003179	MGC:4419,
10	571	210142_x_at		AF117234	flotilina
	572	218942_at	FLJ22055	NM_024779	proteína hipotética FLJ22055
	573	216716_at		U15197	proteína ABO de grupo histo-sanguíneo, secuencia 3 UTR parcial.
	574	218185_s_at	FLJ10511	NM_018120	proteína hipotética FLJ10511
15	575	219215_s_at	FLJ20327	NM_017767	proteína hipotética FLJ20327
	576	214683_s_at		AI251890	quinasa 1 similar a CDC
	577	202612_s_at	CRSP2	NM_004229	cofactor necesario para la activación de la transcripción de Sp1, subunidad 2 (150 kD)
	578	208932_at		BC001416	proteína fosfatasa 4 (anteriormente X), subunidad catalítica,
20	579	206766_at	ITGA10	AF112345	subunidad de integrina α 10
	580	216213_at		AF155113	antígeno NY-REN-55
	581	205609_at	ANGPT1	NM_001146	angiopoyetina 1
	582	204644_at	tNOX	AF207881	hidroquinona (NADH) oxidasa asociada a tumores
25	583	206812_at	ADRB3	NM_000025	adrenérgico, β -3-, receptor
	584	202176_at	ERCC3	NM_000122	deficiencia para la reparación en roedores de la reparación por escisión por complementación cruzada, grupo de complementación 3 (grupo B de complementación de xeroderma pigmentosum)
	585	219816_s_at	FLJ10482	NM_018107	proteína hipotética FLJ10482
30	586	200934_at	DEK	NM_003472	oncogén DEK (unión a ADN)
	587	206098_at	ZID	NM_006626	proteína con dedos de zinc con dominio de interacción
	588	206987_x_at	FGF18	NM_003862	factor de crecimiento de fibroblastos 18
	589	208392_x_at	IFI75	NM_004510	proteína inducida por interferón 75, 52 kD
35	590	213048_s_at		W26593	translocación de SET (asociado a leucemia mieloide)
	591	220861_at		AF118067	PRO1578
	592	222115_x_at		BC003693	similar al gen 3930401K13 de ADNc de RIKEN
	593	222236_s_at	FLJ20246	AK000253	FLJ20246
	594	41644_at	KIAA0790	AB018333	proteína KIAA0790,
40	595	204107_at		NM_002505	factor de transcripción nuclear Y, α
	596	209706_at		AF247704	proteína con caja homeótica Nkx3.1
	597	210174_at		AF228413	factor de transcripción de hepatocitos
	598	213844_at	HOXA5	NM_019102	caja homeótica A5
	599	214108_at		AI346181	proteína MAX
45	600	212064_x_at		AI471665	Proteína con dedos de zinc asociada a MYC (factor de transcripción de unión a purina)
	601	214316_x_at		AI348935	calreticulina
	602	201514_s_at	G3BP	NM_005754	proteína de unión a dominio SH3 de la proteína activadora de Ras-GTPasa
50	603	203092_at	hTIM44	AF026030	supuesto receptor de importación de proteínas de la membrana interna mitocondrial
	604	208725_at		AL031668	secuencia de ADN del clon RP1-64K7 en el cromosoma 20q11.21-11.23.
	605	219819_s_at	HSPC007	NM_014018	proteína HSPC007
55	606	200724_at		BC003358	proteína ribosómica L10,
	607	214350_at		AI762021	sintofina, beta 2 (proteína A1 asociada adistrofina, 59kD, componente básico 2)
	608	210197_at		BC003622	similar a inositol 1,3,4-trifosfato 56 quinasa,
60	609	218150_at	ARL5	NM_012097	similar a factor de ribosilación de ADP 5
	610	214241_at		AA723057	NADH deshidrogenasa (ubiquinona) subcomplejo 1 beta, 8 (19 kD, ASH1)

65

ES 2 504 242 T3

	611	206253_at	DLG2	NM_001364	homólogo de discos grandes (<i>Dros.</i>) 2 (chapsina-110)
	612	205435_s_at	KIAA1048	NM_014911	proteína KIAA1048
	613	215154_at	DKFZp434G043	AL080134	DKFZp434G043
5	614	215771_x_at		X15786	gen ret-II.
	615	213795_s_at		BF740139	clon RP4-534B8, cromosoma 20
	616	209042_s_at		BC001738	Similar a enzima conjugadora de ubiquitina E2G 2 (homóloga a UBC7 de levadura),
10	617	201602_s_at		NM_002480	miosina fosfatasa, subunidad diana 1
	618	211133_x_at		AF009643	proteína transcrito 5 similar a inmunoglobulina de clon 6
	619	211245_x_at		AF002256	homólogo del receptor inhibidor de linfocitos citolíticos cl-9
	620	210713_at		U61166	proteína que contiene dominio SH3, SH3P17
15	621	210840_s_at	KIAA0051	D29640	KIAA0051
	622	218157_x_at	SPEC1	NM_020239	efector de proteína pequeña 1 de Cdc42
	623	208292_at	BMP10	NM_014482	proteína morfogenética ósea 10
	624	210518_at		AB035305	ARNm de CDH8 para cadherina-8,
	625	205064_at	SPRR1B	NM_003125	proteína pequeña rica en prolina 1B (cornifina)
20	626	206990_at	TNR	NM_003285	tenascina R (restrictina, janusina)
	627	211304_x_at		D50134	canal de K de rectificación interna
	628	203950_s_at	CIC-6a	NM_001286	canal de cloruro 6 (CLCN6), variante de transcrito
	629	214211_at		AA083483	ferritina, polipéptido pesado 1
	630	211056_s_at		BC006373	clon MGC:12762
25	631	212185_x_at	MT2A	NM_005953	metalotioneína 2A
	632	203415_at	PDCD6	NM_013232	muerte celular programada 6
	633	205388_at	TNNC2	NM_003279	troponina C2, rápida
	634	208871_at		D31840	DRPLA
	635	218840_s_at	FLJ10631	NM_018161	proteína hipotética FLJ10631
30	636	206400_at	LGALS7	NM_002307	lectina de unión a galactósido, soluble, 7 (galectina 7)
	637	207733_x_at	PSG9	NM_002784	glicoproteína beta-1 específica del embarazo 9
	638	207850_at	GRO3	NM_002090	oncogén GRO3
	639	210884_s_at		AF168619	HE2 gamma1
35	640	220162_s_at	LOC64170	NM_022352	proteína con dominio de reclutamiento de caspasa 9
	641	209749_s_at		AI623989	enzima convertidora de angiotensina I (peptidil-dipeptidasa A) 1
	642	216010_x_at		D89324	ADN para alfa (1,31,4) fucosiltransferasa,
	643	203970_s_at	PEX3	NM_003630	factor de biogénesis de peroxisomas 3
40	644	203972_s_at		AB035307	ARNm para Pex3p,
	645	217749_at	LOC51137	NM_016128	proteína de la cubierta gamma-cop
	646	207022_s_at	LDHC1	NM_002301	lactato deshidrogenasa C, variante de transcrito 1
	647	209905_at		AI246769	caja homeótica A9
	648	209099_x_at	HJ1	U73936	Jagged 1
45	649	222201_s_at	KIAA1315	AB037736	proteína KIAA1315,
	650	201335_s_at	KIAA0382	NM_015313	proteína KIAA0382;
	651	221748_s_at	tensin	AL046979	tensina
	652	205139_s_at	UST	NM_005715	uronil 2-sulfotransferasa
50	653	210678_s_at		U56418	ácido lisofosfatídico aciltransferasa-beta
	654	215293_s_at	DKFZp564E053	AL049261	DKFZp564E053
	655	41858_at	DKFZp564E053	AL049261	DKFZp564E053
	656	218047_at	FLJ12492	NM_024586	proteína hipotética FLJ12492
	657	215069_at	FLJ21412	AK025065	FLJ21412
55	658	201095_at	DAP	NM_004394	proteína asociada a la muerte
	659	220197_at	ATP6N1B	NM_020632	proteína accesoria 1B, no catalítica, lisosómica, transportadora de H(+), ATPasa,
	660	213351_s_at	KIAA0779	AI934469	KIAA0779
	661	203285_s_at	HS2ST1	NM_012262	sulfato de heparán 2-O-sulfotransferasa
60	662	204326_x_at	MT1L	NM_002450	metalotioneína 1L
	663	208581_x_at	MT1X	NM_005952	metalotioneína 1X
	660	212690_at	KIAA0725	AB018268	proteína KIAA0725,

65

Referencias citadas

5	5.242.974	20030069180	WO01053312	WO02094988	WO04030615
	5.384.261	20030073623	WO01054472	WO02095010	WO04031413
	5.405.783	20030087818	WO01055322	WO02102235	WO04031414
	5.412.087	20030096982	WO01057182	WO02102993	WO04033637
	5.424.186	20030104366	WO01057188	WO02102994	WO04037996
	5.429.807	20030124128	WO01059063	WO03000735	WO04039943
	5.436.327	20030124148	WO01060860	WO03003906	WO04039956
10	5.445.934	20030128884	WO01070979	WO03004622	WO04041170
	5.472.672	20030166064	WO01073027	WO03004989	WO04042022
	5.527.681	20030194704	WO01075067	WO03010336	WO04043361
	5.529.756	20030219744	WO01077168	WO03016475	WO04047728
15	5.532.128	20030219767	WO01077288	WO03016476	WO04048938
	5.545.531	20030236392	WO01077291	WO03016500	WO04053081
	5.554.501	20040005560	WO01088088	WO03018807	WO04060270
	5.556.752	20040009478	WO01090154	WO03025138	WO04063355
	5.561.071	20040009491	WO01090304	WO03027258	WO04070062
20	5.571.639	20040013663	WO01094629	WO03038063	WO04072265
	5.593.839	20040034196	WO02002623	WO03042661	WO04074301
	5.599.695	20040037842	WO02004514	WO03042661	WO04076613
	5.624.711	20040058340	WO02012235	WO03042661	WO04079014
	5.658.734	20040110194	WO02012328	WO03046152	WO91001999
25	5.700.637	20040115636	WO02024956	WO03054152	WO92019745
	5874283	DE 432977	WO02028999	WO03062379	WO94023039
	6.004.755	EP 1104808	WO02029086	WO03064589	WO95014772
	6060283	WO00021991	WO02031198	WO03072035	WO97012967
	6.136.182	WO00026244	WO02036766	WO03078572	WO97024441
30	6171787	WO00032776	WO02050279	WO03081201	WO97033902
	6.218.114	WO00036107	WO02057414	WO03091388	WO98010067
	6.218.122	WO00055320	WO02059260	WO03094848	WO98045437
	6.271.002	WO00055350	WO02059271	WO03101283	WO99026976
	6440694	WO00056880	WO02059377	WO04003162	WO990033980
35	200200131971	WO00058473	WO02060317	WO04018641	WO99033981
	20020055474	WO00060078	WO02068652	WO04020497	WO99035158
	20020064872	WO00061610	WO02070737	WO04020593	WO99035170
	20030065157	WO00063438	WO02074237	WO04022059	WO99038972
	20020072089	WO00078961	WO02076488	WO04023973	WO99047540
40	20020081659	WO01000828	WO02081498	WO04024097	WO99050416
	20020110547	WO01025256	WO02081517	WO04024892	WO99054460
	20020123619	WO01027158	WO02081731	WO04028479	WO99058642
	20020150581	WO0104046	WO02083876		WO99064576
	20020151681	WO01042451	WO02083898		
45	20020168637	WO01042467	WO02086443		
	20020192678	WO01044448	WO02090526		
	20030194734	WO01046697	WO02094629		
	20030022239	WO01049716			

50 Ahr *et al.* (2002) "Identification of high risk breast-cancer patients by gene-expression profiling" *Lancet* 359:131-132

Chang *et al.* (2003) "Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer" *Lancet* 362:362-9

Early Breast Cancer Trialists' Collaborative Group (1995) "Effects of radiotherapy and surgery in early breast cancer. An overview of the randomized trials" *N Engl J Med* 333:1444-1455

55 Early Breast Cancer Trialists' Collaborative Group (1998a) "Polychemotherapy for early breast cancer: an overview of the randomized trials" *Lancet* 352:930-942

Early Breast Cancer Trialists' Collaborative Group (1998b) "Tamoxifen for early breast cancer: an overview of randomized trials" *Lancet* 351:1451-1467

Efron (1981) "Censored data and the bootstrap" *J Am Stat Assoc* 76:312-319

60 Eifel *et al.* (2001) "National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, noviembre 1-3, 2000" *J Natl Cancer Inst* 93:979-989

Foekens *et al.* (1989b) "Prognostic value of estrogen and progesterone receptors measured by enzyme immunoassays in human breast tumor cytosols" *Cancer Res* 49:5823-5828

Foekens *et al.* (1989a) "Prognostic value of receptors for insulin-like growth factor 1, somatostatin, and epidermal growth factor in human breast cancer" *Cancer Res* 49:7002-7009

65 Goldhirsch *et al.* (2003) "Meeting highlights: Updated International Expert Consensus on the Primary Therapy of Early Breast Cancer" *J Clin Oncol* 21:3357-3365

5 Golub *et al.* (1999) "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring" *Science* 286:531-537

6 Gruvberger *et al.* (2001) "Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns" *Cancer Res* 61:5979-5984

7 Hedenfalk *et al.* (2001) "Gene-expression profiles in hereditary breast cancer" *N Engl J Med* 344:539-548

8 Herrera-Gayol *et al.* (1999) "Adhesion proteins in the biology of breast cancer: contribution of CD44" *Exp Mol Pathol* 66:149-156

9 Huang *et al.* (2003) "Gene expression predictors of breast cancer outcomes" *Lancet* 361:1590-1596

10 Kaplan *et al.* (1958) "Non-parametric estimation of incomplete observations" *J Am Stat Assoc* 53:457-481

11 Keyomarsi *et al.* (2002) "Cyclin E and survival in patients with breast cancer" *N Engl J Med* 347:1566-1575

12 Lipshutz *et al.* (1999) "High density synthetic oligonucleotide arrays" *Nat Genet* 21:20-24

13 Ma *et al.* (2003) "Gene expression profiles of human breast cancer progression" *Proc Natl Acad Sci USA* 100:5974-5979

14 Ntzani *et al.* (2003) "Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment" *Lancet* 362:1439-1444

15 Perou *et al.* (2000) "Molecular portraits of human breast tumors" *Nature* 406:747-752

16 Ramaswamy *et al.* (2001) "Multiclass cancer diagnosis using tumor gene expression signatures" *Proc Natl Acad Sci USA* 98:15149-15154

17 Ramaswamy *et al.* (2003) "A molecular signature of metastasis in primary solid tumors" *Nat Genet* 33:1-6

18 Ransohoff (2004) "Rules of evidence for cancer molecular-marker discovery and validation" *Nat Rev Cancer* 4:309-314

19 Sørlie *et al.* (2001) "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications" *Proc Natl Acad Sci USA* 98:10869-10874

20 Sørlie *et al.* (2003) "Repeated observation of breast tumor subtypes in independent gene expression data sets" *Proc Natl Acad Sci USA* 100:8418-8423

21 Sotiriou *et al.* (2003) "Gene expression profiles derived from fine needle aspiration correlate with response to systemic chemotherapy in breast cancer" *Breast Cancer Res* 4:R3

22 Sotiriou *et al.* (2003) "Breast cancer classification and prognosis based on gene expression profiles from a populationbased study" *Proc Natl Acad Sci USA* 100:10393-10398

23 Su *et al.* (2001) "Molecular classification of human carcinomas by use of gene expression signatures" *Cancer Res* 61:7388-7393

24 van de Vijver *et al.* (2002) "A gene expression signature as a predictor of survival in breast cancer" *N Engl J Med* 347:1999-2009

25 van't Veer *et al.* (2002) "Gene expression profiling predicts clinical outcome of breast cancer" *Nature* 415:530-536

26 Wang *et al.* (2004) "Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer" *J Clin Oncol* 22:1564-1571

27 Woelfle *et al.* (2003) "Molecular signature associated with bone marrow micrometastasis in human breast cancer" *Cancer Res* 63:5679-5684

40

45

50

55

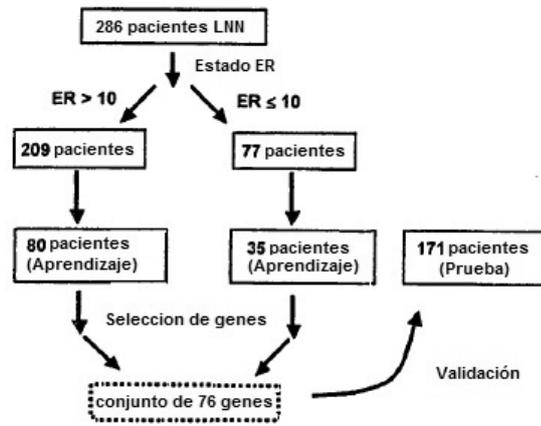
60

65

REIVINDICACIONES

- 5 1. Método para evaluar el estado de un cáncer de mama, estadificar a los pacientes de cáncer de mama o determinar un protocolo de tratamiento del paciente en una muestra biológica obtenida de un paciente, que comprende las etapas de medir los niveles de expresión en la muestra de los genes que codifican un ARNm identificado por las SEQ ID NO: 96-111 en la Tabla 10, en el que los niveles de expresión génica por encima o por debajo de los niveles límite predeterminados son indicativos del estado de un cáncer de mama o del estadio de un cáncer de mama.
- 10 2. Método según la reivindicación 1 en el que la etapa corresponde a la clasificación mediante el sistema TNM o a pacientes con perfiles de expresión génica similares.
- 15 3. Método según la reivindicación 1 que comprende adicionalmente medir los niveles de expresión de genes seleccionados del grupo que consiste en aquellos que codifican un ARNm identificado por las SEQ ID NO: 36-95 en la Tabla 10, en el que los niveles de expresión génica por encima o por debajo de los niveles límite predeterminados son indicativos del estado de un cáncer de mama o del estadio de un cáncer de mama; particularmente en el que el método proporciona un pronóstico para los pacientes ER positivo o ER negativo.
- 20 4. Método según cualquiera de las reivindicaciones 1-3, en el que la muestra se prepara mediante un método seleccionado del grupo que consiste en la preparación de tejido en bruto (preferentemente obtenido a partir de una pieza quirúrgica o de una biopsia) y la microdissección por captura con láser.
- 25 5. Método según cualquiera de las reivindicaciones 1-3 que comprende adicionalmente:
- (a) medir el nivel de expresión de al menos un gen que codifica un ARNm identificado por las SEQ ID NO: 112-132 en la Tabla 10; o
- (b) medir el nivel de expresión de al menos un gen expresado constitutivamente en la muestra; o
- (c) determinar el estado del receptor de estrógeno (ER) de la muestra; particularmente en el que el estado ER se determina (i) midiendo el nivel de expresión de al menos un gen indicativo del estado ER o midiendo la presencia de ER en la muestra o en el que la presencia de ER se mide mediante inmunohistoquímica o (ii) midiendo el nivel de expresión de al menos un gen indicativo del estado ER o midiendo la presencia de ER en la muestra o en el que la presencia de ER se mide mediante inmunohistoquímica.
- 30 6. Método según cualquiera de las reivindicaciones 1-3, en el que
- (a) la muestra se obtiene de un tumor primario; o
- (b) el patrón de expresión de los genes se compara con un patrón de expresión indicativo de un paciente recidivante; particularmente en el que la comparación de los patrones de expresión se lleva a cabo con los métodos de reconocimiento de patrones; más particularmente en el que los métodos de reconocimiento de patrones incluyen el uso de un análisis de riesgos proporcionales de Cox.
- 35 7. Método según cualquiera de las reivindicaciones 1-3 en el que
- (a) los niveles límite predeterminados son una sobreexpresión o subexpresión al menos 1,5 veces mayor en la muestra con respecto a las células benignas o al tejido normal; o
- (b) los niveles límite predeterminados tienen al menos una sobreexpresión con un valor p estadísticamente significativo en la muestra que tiene células metastásicas con respecto a las células benignas o al tejido normal, particularmente en el que el valor p es inferior a 0,05; o
- (c) la expresión génica se mide en una micromatriz o matriz génica, particularmente en el que (i) la micromatriz es una matriz de ADNc o una matriz de oligonucleótidos o (ii) la micromatriz o matriz génica comprende adicionalmente uno o más reactivos de control interno; o
- (d) en el que la expresión génica se determina mediante la amplificación de ácidos nucleicos llevada a cabo mediante una reacción en cadena de la polimerasa (PCR) de ARN extraído de la muestra; particularmente en el que dicha PCR es una reacción en cadena de la polimerasa con transcriptasa inversa (RT-PCR), que opcionalmente comprende además uno o más reactivos de control interno.
- 40 8. Método según cualquiera de las reivindicaciones 1-3, en el que la expresión génica se detecta
- (a) midiendo o detectando una proteína codificada por el gen, particularmente en el que la proteína es detectada por un anticuerpo específico para la proteína; o
- (b) midiendo una característica del gen, particularmente en el que la característica medida está seleccionada del grupo que consiste en la amplificación, metilación, mutación y variación alélica de ADN.
- 45 9. Uso de una composición que comprende un conjunto de sondas que consiste en sondas para detectar los genes identificados por las SEQ ID NO: 96-111 en la Tabla 10 en un método según cualquiera de las reivindicaciones 1 ó 3.
- 50
- 55
- 60
- 65

- 5 10. Uso de un kit para llevar a cabo un ensayo para determinar el pronóstico de cáncer de mama en una muestra biológica en el que el kit comprende materiales para detectar secuencias aisladas de ácido nucleico, sus complementos, o porciones de los mismos de una combinación de genes que codifican un ARNm identificado por las SEQ ID NO: 96-111 en la Tabla 10, en el que el kit opcionalmente comprende además materiales para detectar
- 10 i. reactivos para llevar a cabo un análisis de micromatrices; o
ii. un medio a través del cual se ensayan dichas secuencias de ácido nucleico, sus complementos, o porciones de los mismos.
- 15 11. Micromatriz o matriz génica para llevar a cabo el método según las reivindicaciones 1 ó 3; que consiste en secuencias aisladas de ácido nucleico, sus complementos, o porciones de los mismos de una combinación de genes que codifican un ARNm identificado por las SEQ ID NO: 96-111 en la Tabla 10 y opcionalmente uno o más reactivos de control interno; en el que la combinación es suficiente para caracterizar el estado de un cáncer de mama o el riesgo de recidiva en una muestra biológica;
- 20 12. Micromatriz o matriz génica según la reivindicación 11; en la que (i) la medición o caracterización es una sobreexpresión o subexpresión al menos 1,5 veces mayor; o (ii) la medición proporciona una sobreexpresión o subexpresión con un valor p estadísticamente significativo.
- 25 13. Micromatriz o matriz génica según la reivindicación 11 ó 12
(a) en la que el valor p es inferior a 0,05 o en la que la micromatriz comprende una matriz de ADNc o una matriz de oligonucleótidos; y/o
(b) que comprende una matriz de ADNc o una matriz de oligonucleótidos.
- 30
- 35
- 40
- 45
- 50
- 55
- 60
- 65



1A

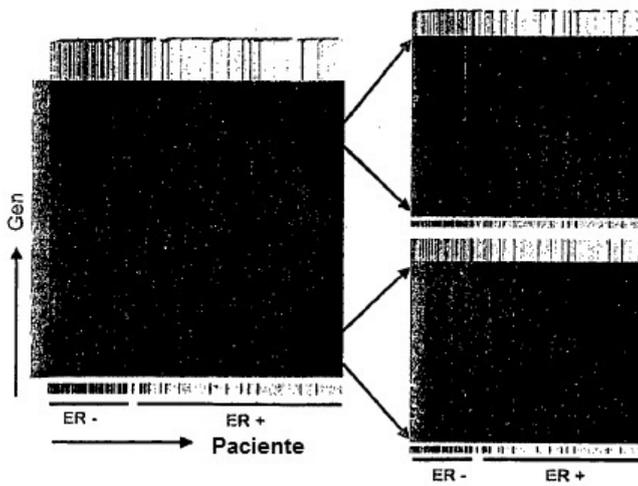
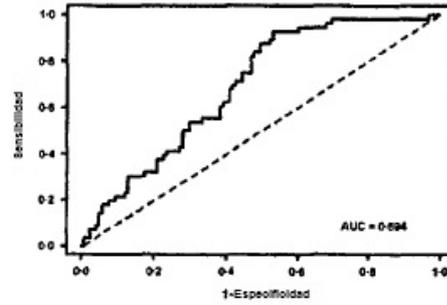
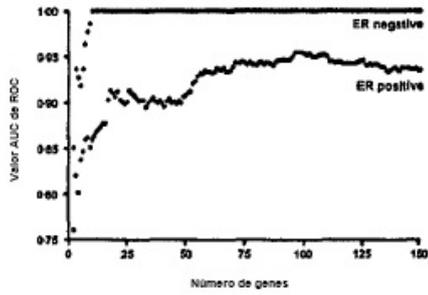


Figura 1B

A. Selección de genes



B. Conjunto de validación 171 pacientes LNN

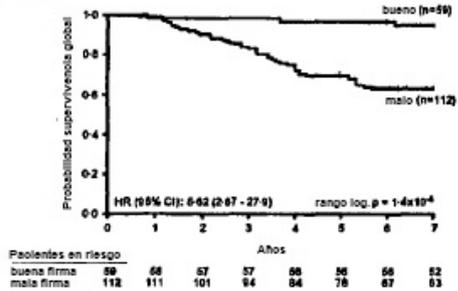
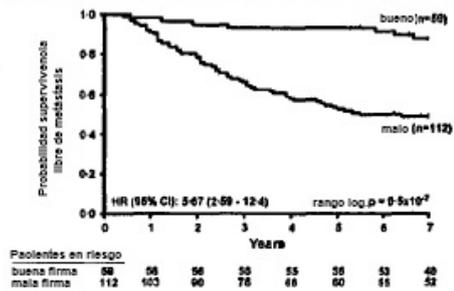
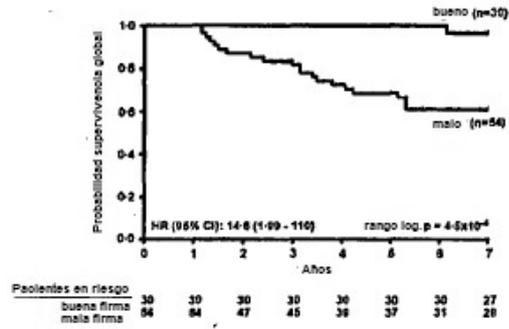
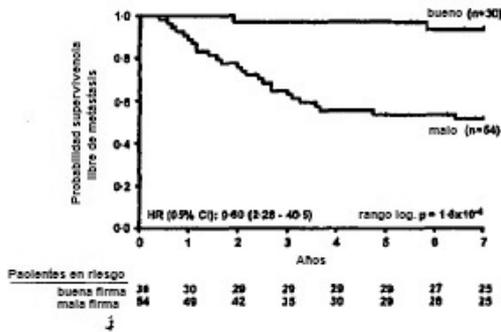
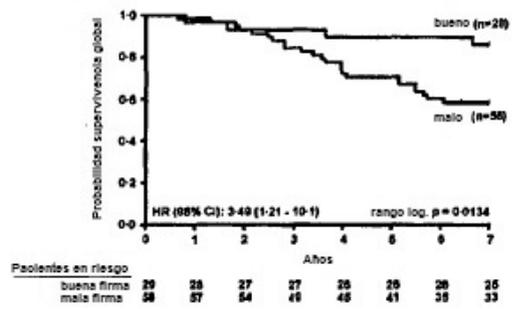
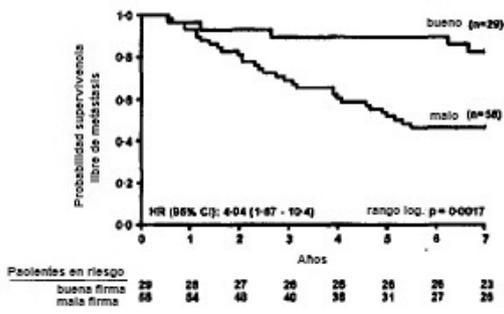


Figura 2

A. Premenopáusica



B. Posmenopáusica



C. Tumores 10-20 mm

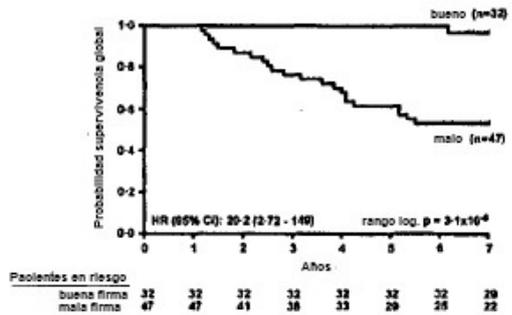
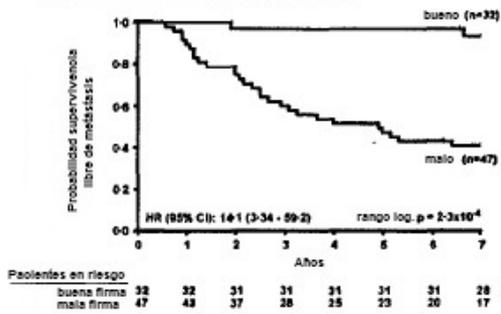


Figura 3

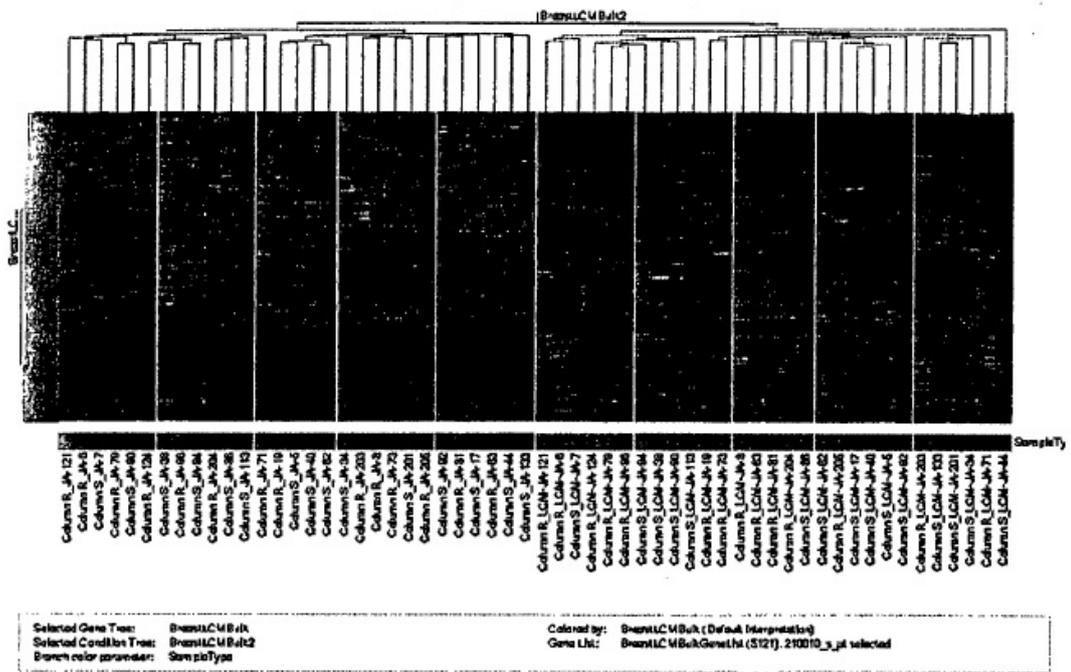


Figura 4

Comparación de la intensidad de expresión de los genes de control con una sola ronda y con dos rondas de amplificación de ARN

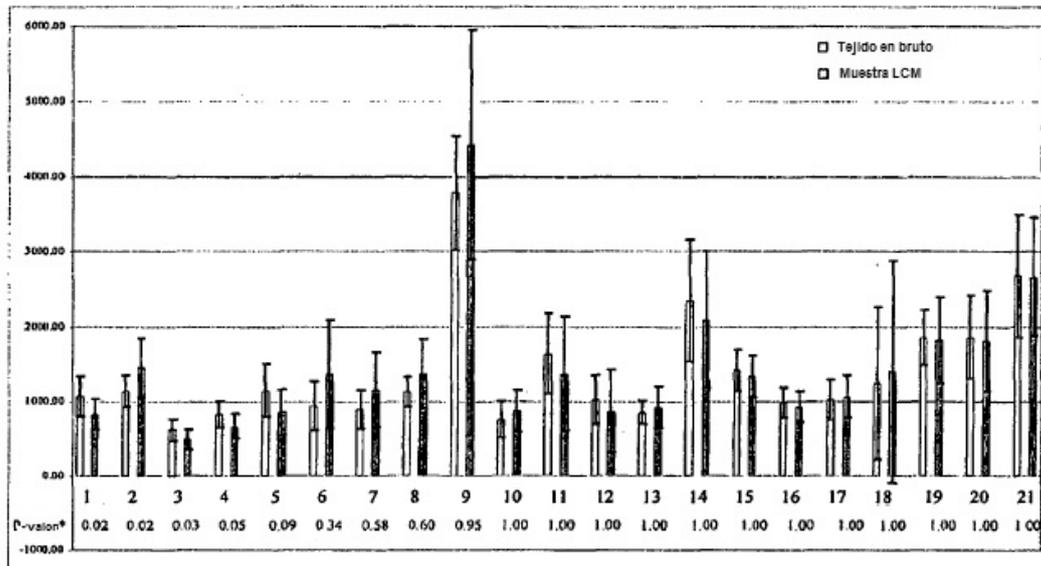
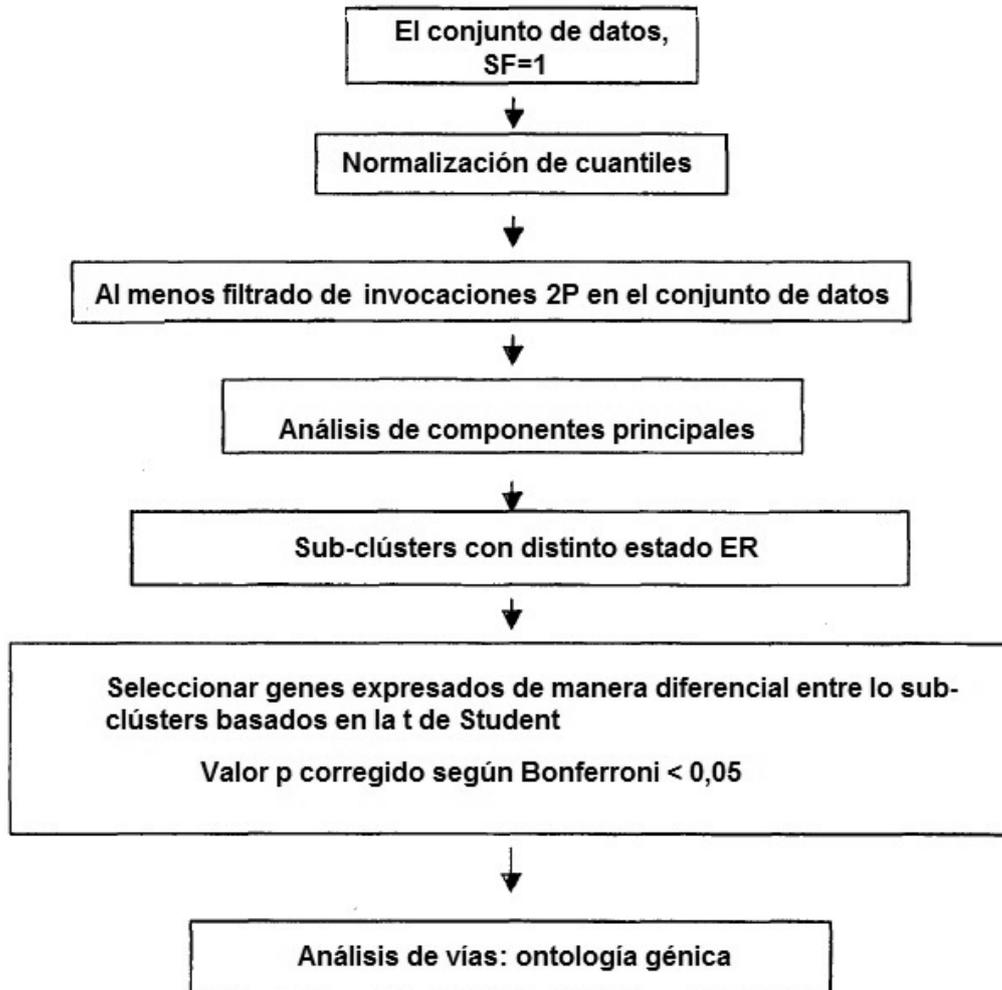


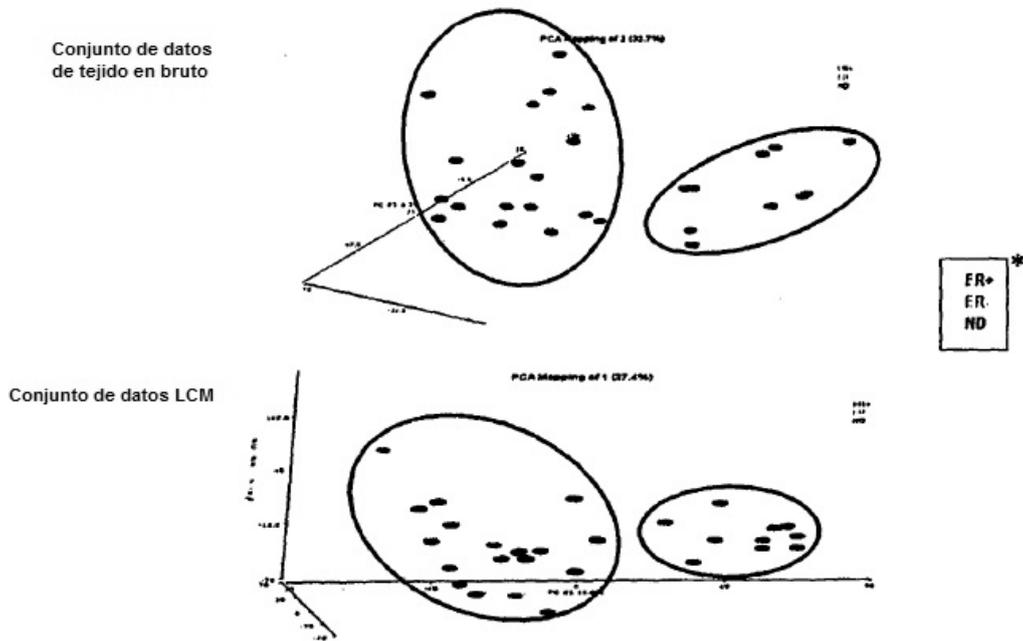
Figura 5

Figura 6

Flujo de trabajo de análisis de datos



Análisis de componentes principales con conjuntos de genes filtrados



*La clase de ER se basó en el inmunoensayo enzimático. Como límite se utilizaron 10 fmol/mg.

Figura 7

Figura 8

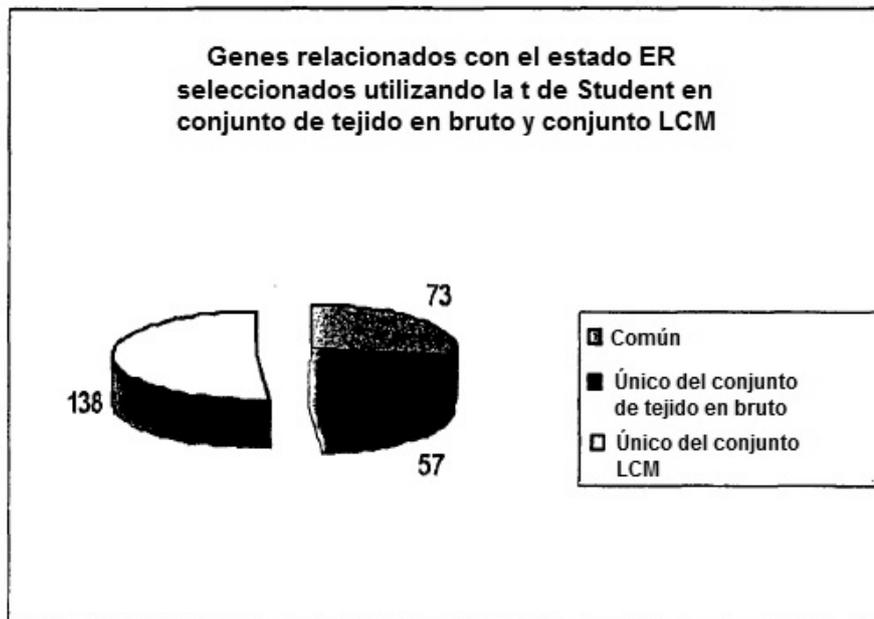
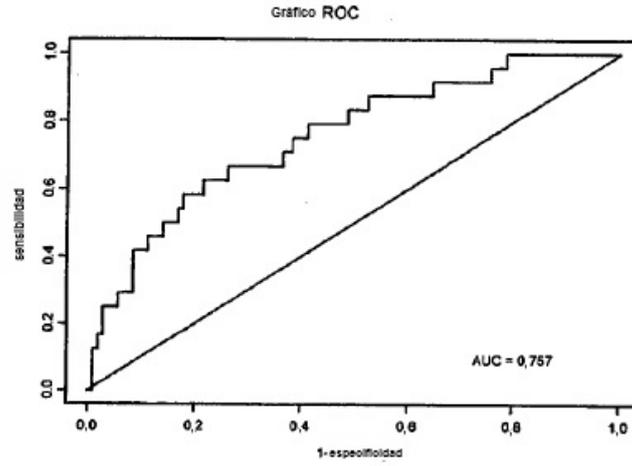


Figura 11

A.

Características	Conjunto de validación
Número	132
Edad (media±SD)	58±11
7-9 añ.	9 (7%)
41-55 añ.	42 (32%)
56-70 añ. (límite 55,1)	53 (40%)
>70 añ.	17 (13%)
Desconocida	1
Estado menopáusico	
Premenopáusico	31 (23%)
Postmenopáusico	101 (77%)
Estadio T	
T1	77 (58%)
T2	54 (41%)
T3/4	1
Grado de difer.	
Escaso	40 (30%)
Moderado	55 (42%)
Bueno	16 (12%)
Desconocido	21 (16%)
Metástasis <6 años	
Si	24 (18%)
No	108 (82%)

B.



132 pacientes		
	Recidivantes	No recidivantes
Positivo	21	63
Negativo	3	44
Sensibilidad:	88% (0,69 - 0,96)	
Especificidad:	41% (0,32 - 0,51)	
Freq. de recaída:	18%	
PPV:	26%	
NPV:	94%	
Oportunidad relativa:	4,9 (1,4-17)	

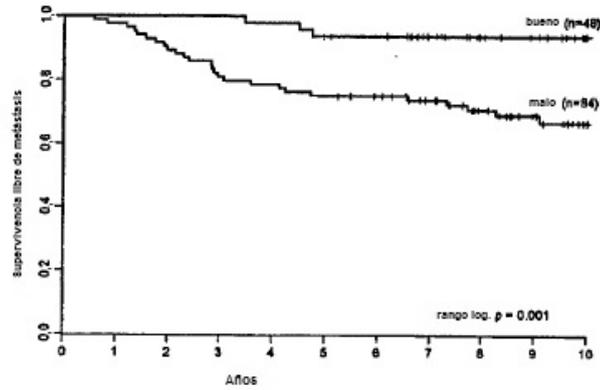


Figura 12

Figura 13

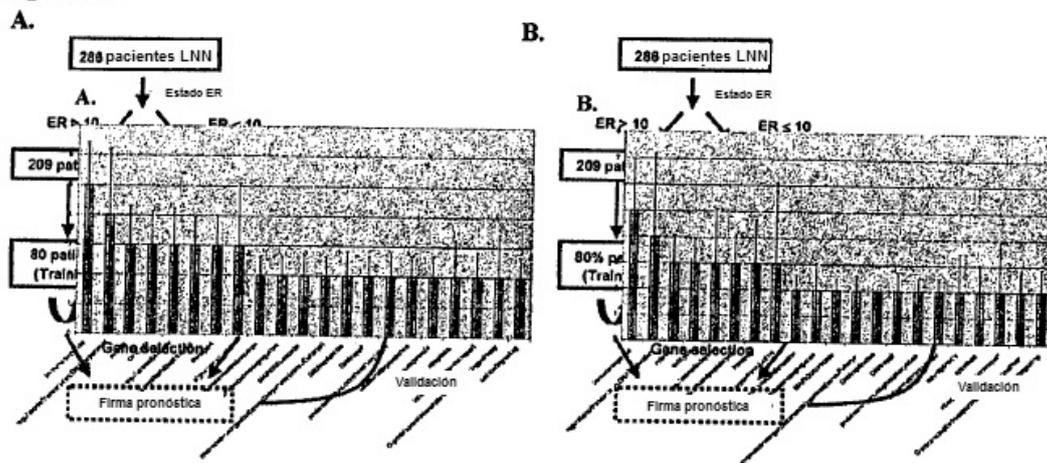


Figura 14