



19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 366 178**

51 Int. Cl.:  
**C12Q 1/68** (2006.01)  
**G01N 33/574** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **02746538 .4**  
96 Fecha de presentación : **14.06.2002**  
97 Número de publicación de la solicitud: **1410011**  
97 Fecha de publicación de la solicitud: **21.04.2004**

54 Título: **Diagnóstico y pronóstico del cáncer de mama en pacientes.**

30 Prioridad: **18.06.2001 US 298918 P**  
**14.05.2002 US 380710 P**

45 Fecha de publicación de la mención BOPI:  
**18.10.2011**

45 Fecha de la publicación del folleto de la patente:  
**18.10.2011**

73 Titular/es: **MERCK SHARP & DOHME Corp.**  
**126 East Lincoln Avenue**  
**Rahway, New Jersey 07065, US**  
**The Netherlands Cancer Institute**

72 Inventor/es: **Dai, Hongyue;**  
**He, Yudong;**  
**Linsley, Peter, S.;**  
**Mao, Mao;**  
**Roberts, Christopher, J.;**  
**Van'T Veer, Laura, Johanna;**  
**Van de Vijver, Marc, J.;**  
**Bernards, Rene y**  
**Hart, A., A., M.**

74 Agente: **Molinero Zofío, Félix**

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

## DESCRIPCIÓN

Diagnóstico y pronóstico del cáncer de mama en pacientes.

## 1. CAMPO DE APLICACIÓN DE LA INVENCION

**[0001]** La presente invención se refiere a la identificación de genes marcadores que sirven para el diagnóstico y el pronóstico del cáncer de mama. Más concretamente, la invención se refiere a la identificación de un conjunto de genes marcadores asociados al cáncer de mama, un conjunto de genes marcadores expresados diferenciadamente en tumores de receptor de estrógeno (+) frente a receptor de estrógeno (-), un conjunto de genes marcadores expresados diferenciadamente en el *BRCA1* frente a tumores esporádicos y un conjunto de genes marcadores expresados diferenciadamente en tumores esporádicos procedentes de pacientes con buen pronóstico clínico (esto es, sin metástasis ni dolencia durante más de 5 años) frente a pacientes con pronóstico clínico deficiente (esto es, sin metástasis ni dolencia durante menos de 5 años). Para cada uno de los conjuntos marcadores mencionados más arriba la invención se refiere además a métodos para distinguir las afecciones que se dan en el cáncer de mama. También se describen métodos para determinar el curso del tratamiento de las pacientes con cáncer de mama.

## 2. ANTECEDENTES DE LA INVENCION

**[0002]** El número creciente de casos de cáncer declarados en los Estados Unidos, y de hecho en todo el mundo, constituye un grave problema. Actualmente se dispone sólo de unos cuantos tratamientos para tipos específicos de cáncer, pero que no ofrecen garantía alguna de éxito. Para aumentar su eficacia, estos tratamientos tendrían que facilitar no sólo una detección más precoz del carácter maligno del tumor, sino también una valoración fiable de la gravedad de dicho tumor.

**[0003]** La incidencia del cáncer de mama, una de las causas principales de muerte entre las mujeres, ha ido aumentando gradualmente en los Estados Unidos en los treinta últimos años. Su riesgo acumulativo es relativamente alto; se prevé que, en los Estados Unidos, 1 de cada 8 mujeres desarrollará cáncer de mama antes de que cumpla 85 años. De hecho, el cáncer de mama es el cáncer más corriente entre las mujeres y la segunda causa más frecuente de muerte por cáncer en los Estados Unidos. En 1997 se estimó en 181.000 el número de nuevos casos declarados en los Estados Unidos, así como que 44.000 personas morirían de cáncer de mama (Parker et al., CA Cancer J. Clin. 47:5-27 (1997); Chu et al., J. Nat. Cancer Inst. 88:1571-1579 (1996)). Aunque se conoce en gran medida el mecanismo de oncogénesis de la mayoría de los carcinomas de mama, existen factores genéticos que pueden predisponer a algunas mujeres a desarrollar cáncer de mama ((Miki et al., Science, 266:66-71(1994)). El descubrimiento y la caracterización del *BRCA1* y *BRCA2* ha ampliado recientemente nuestro conocimiento de los factores genéticos que pueden contribuir al cáncer de mama familiar. Las mutaciones germinales en estos dos loci suponen de un 50 a un 85% del riesgo de cáncer de mama o de ovarios a lo largo de la vida (Casey, Curr. Opin. Oncol. 9:88-93 (1997); Marcus et al., Cancer 77:697-709 (1996)). Solo de un 5% a un 10% de casos de cáncer están relacionados con los genes de susceptibilidad al cáncer de mama, *BRCA1* y *BRCA2*. El riesgo de cáncer acumulado a lo largo de la vida para las mujeres portadoras del *BRCA1* mutante se calcula en aproximadamente el 92%, mientras que el riesgo acumulado a lo largo de la vida para la mayoría de no portadoras se estima en aproximadamente un 10%. El *BRCA1* es un gen supresor de tumores que interviene en la reparación del ADN y en el control del ciclo celular, ambos de suma importancia para el mantenimiento de la estabilidad genómica. Más del 90% de todas las mutaciones reveladas hasta ahora desembocan en un truncamiento prematuro del producto proteínico con función anormal o suprimida. La histología del cáncer de mama en los portadores de la mutación *BRCA1* varía en casos esporádicos, pero el análisis de la mutación es la única manera de hallar a la portadora.

Al igual que el *BRCA1*, el *BRCA2* interviene en el desarrollo del cáncer de mama y, al igual que el *BRC41*, juega su papel en la reparación del ADN. Sin embargo, a diferencia del *BRCA1*, no está presente en el cáncer de ovarios

**[0004]** En la técnica se conocen marcadores moleculares para diferenciar los tipos de tumores.

Perou CM, et al. (Nature 2000 406:747-752) efectúan retratos moleculares de tumores del cáncer de mama humano. Se identificó un subconjunto de genes en el que la variación en expresión era mayor entre tumores diferentes que entre muestras emparejadas del mismo tumor. Alizadeh AA, et al. (Nature 2000 403:503-511) describen dos formas diferentes de linfoma difuso de células B grandes (DLBCL) identificadas sobre la base del patrón de expresión de genes. Se identificó un subconjunto de genes al hallarse expresado de forma selectiva en una de las dos formas de DLBCL. Perou CM, et al. (PNAS 1999 96:9212-9217) describen patrones de expresión de genes diferenciados en las células epiteliales mamarias humanas y cánceres de mama. En respuesta a una serie de perturbaciones experimentales, se identificó un subconjunto de genes en los que se apreciaba claramente una expresión diferenciada en las células epiteliales mamarias humanas. Khan J, et al. (Nature Medicine 2001 7:673-679) describen un criterio de clasificación y predicción diagnóstica de cánceres mediante

perfiles de expresión de genes y redes neuronales artificiales. Se usó un subconjunto de genes para clasificar pequeñas muestras de tumores de células redondas y azules en categorías diagnósticas. Hedenfalk I, et al. (New Eng J Med 2001 344:539.548) describen perfiles de expresión de genes en el cáncer de mama hereditario e identificaron un subconjunto de genes que se dividía en tumores mutantes de BRCA1, tumores mutantes de BRCA2 y tumores esporádicos de tejido de cáncer de mama. A diferencia de estos documentos donde se da a conocer métodos para distinguir los tipos de tumores, la presente invención ofrece métodos para determinar si los individuos con cáncer de mama han tenido un pronóstico bueno o deficiente.

**[0005]** Con el cáncer de mama se ha relacionado también otros genes, por ejemplo el c-erb-2 (*HER2*) y el p53 (Beenken et al., Ann. Surg. 233(5):630-638 (2001)). La superexpresión del c-erb-2 (*HER2*) y del p53 se ha relacionado con pronósticos deficientes (Rudolph et al., Hum. Pathol. 32(3):311-319 (2001)), pues ha sido una expresión aberrante de productos de *mdm2* (Lukas et al., Cancer Res. 61(7):3212-3219 (2001)) y ciclina 1 y p27 (Porter & Roberts, International Publication WO98/33450, publicado el 6 de agosto de 1998). No obstante, no se han identificado otros marcadores clínicamente útiles y estrechamente asociados al cáncer de mama.

**[0006]** Los tumores esporádicos, o sea, aquellos no asociados comúnmente a una mutación germinal conocida, constituyen el cáncer de mama en la mayoría de los casos. También es probable que otros factores, no genéticos, influyan de modo significativo en la etiología de la enfermedad. Independientemente del origen del cáncer, la morbilidad y mortalidad del cáncer de mama aumentan de modo significativo si no se detectan en la fase inicial de su progreso. Por consiguiente, se han dedicado muchos esfuerzos a la detección precoz de la transformación celular y la formación de tumores en el tejido mamario.

**[0007]** La identificación y caracterización del tumor basándose en marcadores hace esperar una mayor fiabilidad del diagnóstico y del pronóstico. Típicamente, para el diagnóstico del cáncer de mama se requiere una prueba histopatológica de la presencia del tumor. Además del diagnóstico, los exámenes histopatológicos también suministran información sobre el pronóstico y la selección de regímenes de tratamiento. El pronóstico se puede basar también en parámetros clínicos, como el tamaño del tumor, el grado del tumor, la edad de la paciente y la metástasis del ganglio linfático.

**[0008]** El diagnóstico y/o el pronóstico se puede determinar en diversos grados de eficacia por observación directa del exterior de la mama, mediante mamografía o por otros métodos de imagen por rayos X (Jatoi, Am. J. Surg. 177:518-524 (1999)). Estos últimos, no obstante, tienen un coste elevado. Cada vez que se le hace una mamografía, la paciente corre un leve riesgo de tener un tumor de mama inducido por las propiedades ionizantes de la radiación usada durante la prueba. Además, el proceso es caro y las interpretaciones subjetivas de un técnico pueden llevar a imprecisión. Por ejemplo, un estudio mostró discrepancias clínicas importantes en aproximadamente un tercio de un conjunto de mamografías que fueron interpretadas individualmente por un grupo de radiólogos a que se encuestó. Asimismo, muchas mujeres opinan que someterse a una mamografía es una experiencia dolorosa. En consecuencia, el National Cancer Institute desaconseja las mamografías para mujeres que no hayan cumplido los cincuenta años, ya que este colectivo no es tan propenso a desarrollar cáncer de mama como el de las mujeres de más edad. Hay que tener en cuenta, no obstante, que, aunque sólo el 22% de los casos de cáncer de mama se da en mujeres menores de cincuenta años, los datos muestran que el cáncer de mama es más agresivo en mujeres premenopáusicas.

**[0009]** En la práctica clínica es importante un diagnóstico exacto de los diversos tipos de cáncer de mama, ya que las opciones de tratamiento, el pronóstico y la probabilidad de la respuesta terapéutica varían todos en función del diagnóstico. Con un pronóstico acertado, o con una determinación de la supervivencia libre de metástasis distante, el oncólogo podrá personalizar la administración de quimioterapia adyuvante, mientras que a las mujeres con peores pronósticos se les aplican los tratamientos más agresivos. Además, la predicción exacta de pronósticos deficientes tendría gran impacto en los ensayos clínicos de nuevas terapias contra el cáncer de mama, ya que entonces las potenciales pacientes de estudio podrían estratificarse según el pronóstico. Así, los ensayos se limitarían a las pacientes con pronósticos deficientes, con lo que a su vez sería más fácil discernir si una terapia experimental es eficaz.

**[0010]** Hasta la fecha no se ha identificado ningún conjunto de predictores de pronóstico satisfactorios basados exclusivamente en información clínica. La detección de las mutaciones del *BRCA1* o del *BRCA2* constituye un paso hacia el diseño de terapias para controlar mejor e impedir la aparición de estos tumores. Sin embargo, no existen medios equivalentes para el diagnóstico de pacientes con tumores esporádicos, que es el tipo más corriente de tumor de cáncer de mama, ni tampoco se dispone de medios para diferenciar los subtipos de cáncer de mama.

### 3. RESUMEN DE LA INVENCION

**[0011]** La invención suministra conjuntos de marcadores de genes que distinguen entre diversos tipos y subtipos de cáncer de mama, así como métodos para su uso. La invención suministra un método para determinar si un individuo aquejado de cáncer de mama tiene un pronóstico bueno o deficiente, en donde dicho

individuo es un humano, en donde dicho buen pronóstico indica que se espera que el individuo en cuestión no tenga metástasis distantes dentro de los cinco años siguientes al diagnóstico inicial de cáncer de mama, y en donde dicho pronóstico deficiente indica que se espera que el individuo en cuestión tenga metástasis distantes dentro de los cinco años siguientes al diagnóstico inicial de cáncer de mama, y que supone: (ia) calcular un primer parámetro clasificador entre un primer perfil de expresión y una plantilla de buenos pronósticos, o (ib) calcular un segundo parámetro clasificador entre dicho primer perfil de expresión y dicha plantilla de buenos pronósticos y un tercer parámetro clasificador entre dicho primer perfil de expresión y una plantilla de pronósticos deficientes; dicho primer perfil de expresión comprende los niveles de expresión de una primera serie de genes en una muestra de células tomada del individuo, dicha plantilla de buenos pronósticos comprende para cada gen de dicha primera serie de genes el nivel medio de expresión de dicho gen en una serie de pacientes que tengan metástasis distante dentro de los cinco años siguientes al diagnóstico inicial de cáncer de mama; dicha primera serie de genes consta de al menos 5 de los genes para los que se incluye una lista de marcadores en la Tabla 5, y (iia) determinar si dicho individuo tiene buen pronóstico si dicho primer parámetro de clasificación está por encima de un umbral establecido o si dicho primer perfil de expresión es más parecido a dicha plantilla de buenos pronósticos que a dicha plantilla de pronósticos deficientes, o (iib) determinar si dicho individuo tiene dicho pronóstico deficiente si dicho primer parámetro de clasificación está por debajo de dicho umbral establecido o si dicho primer perfil de expresión es más parecido a dicha plantilla de pronósticos deficientes que a dicha plantilla de buenos pronósticos. En una de las materializaciones se describe un método para clasificar una muestra de células como ER (+) o ER (-) que supone detectar una diferencia en la expresión de una primera serie de genes con respecto a un control, constanding dicha primera serie de genes de al menos 5 de los genes que corresponden a los marcadores enumerados en la Tabla 1. En materializaciones específicas, dicha serie de genes consta de al menos 50, 100, 200, 500, 1.000, hasta 2.460 de los genes enumerados en la Tabla 1. En otra materialización específica, dicha serie de genes consta de cada uno de los genes correspondientes a los 2.460 marcadores enumerados en la Tabla 1. En otra materialización específica, dicha serie consta de los 550 marcadores enumerados en la Tabla 2. En otra materialización específica, dicho control comprende ácidos nucleicos obtenidos de una reserva de tumores de pacientes esporádicas concretas. En otra materialización específica, dicha detección comprende los pasos de: (a) generar una plantilla de ER (+) mediante la hibridación de ácidos nucleicos obtenidos de una serie de pacientes con E (+) dentro de una serie de pacientes esporádicas contra ácidos nucleicos obtenidos de una reserva de tumores de pacientes esporádicas concretas; (b) generar una plantilla de ER (-) mediante la hibridación de ácidos nucleicos obtenidos de una serie de pacientes con ER (-) dentro de dicha serie de pacientes esporádicas contra ácidos nucleicos obtenidos de dicha reserva de tumores de pacientes esporádicas concretas dentro de dicha serie; (c) hibridar ácidos nucleicos obtenidos de una muestra específica contra dicha reserva; y (d) determinar la semejanza de la expresión de gen marcador en la muestra específica con la plantilla de ER (+) y la plantilla de ER (-), en donde si dicha expresión es más parecida a la plantilla de ER (+), se clasifica la muestra como ER (+), mientras que si dicha expresión es más parecida a la plantilla de ER (-), se clasifica la muestra como ER (-).

**[0012]** Ya se han descrito más arriba métodos aplicados a la clasificación como *BRCA1* o esporádico. La invención suministra los métodos de más arriba aplicados a la clasificación de pacientes como poseedoras de buen pronóstico o pronóstico deficiente. Para los marcadores de *BRCA1*/gen esporádico, se puede usar un método en el que la serie de genes es al menos de 5, 20, 50, 100, 200 or 300 de los *BRCA1*/marcadores esporádicos enumerados en la Tabla 3. En una materialización específica, se usan los 100 marcadores óptimos enumerados en la Tabla 4. Para los marcadores de pronóstico, la invención implica que se pueden usar al menos 5, 20, 50, 100 o 200 marcadores de genes enumerados en la Tabla 5. En una materialización específica, se usan los 70 marcadores óptimos enumerados en la Tabla 6.

**[0013]** La invención posibilita además la combinación de los marcadores. En otra materialización se usan al menos 5 marcadores de la Tabla 5 en conjunción con al menos 5 marcadores de la Tabla 3. En otra materialización se usan al menos 5 marcadores de la Tabla 1 en conjunción con al menos 5 marcadores de la Tabla 5. En otra materialización se usan simultáneamente al menos 5 marcadores de cada una de las Tablas 1, 3 y 5.

**[0014]** También se describe un método para clasificar una muestra como ER(+) or ER(-) calculando la semejanza entre la expresión de al menos 5 de los marcadores enumerados en la Tabla 1 en la muestra con la expresión de los mismos marcadores en una reserva de ácido nucleico de ER (-) y una reserva de ácido nucleico de ER (+), que comprende los pasos de (a) etiquetar ácidos nucleicos obtenidos de una muestra con un primer fluoróforo para obtener una primera reserva de ácidos nucleicos etiquetados con fluoróforo; (b) etiquetar con un segundo fluoróforo una primera reserva de ácidos nucleicos obtenidos de dos o más muestras de ER(+) y una segunda reserva de ácidos nucleicos obtenidos de dos o más muestras de ER(-); (c) poner en contacto dicho primer ácido nucleico etiquetado con fluoróforo y dicha primera reserva de segundo ácido nucleico etiquetado con fluoróforo con dicho primer biochip en condiciones en que pueda darse dicha hibridación, y poner en contacto dicho primer ácido nucleico etiquetado con fluoróforo y dicha segunda reserva de segundo ácido nucleico etiquetado con fluoróforo con dicho segundo biochip en condiciones en que pueda darse dicha hibridación, detectando en cada uno de una serie de loci discretos en el primer biochip una primera señal de emisión fluorescente de dicho primer ácido nucleico etiquetado con fluoróforo y una segunda señal de emisión

fluorescente de dicha primera reserva de segunda materia genética etiquetada con fluoróforo cuyo destino es dicho primer biochip en dichas condiciones, y detectando en cada uno de los marcadores de loci de dicho segundo biochip dicha primera señal de emisión fluorescente de dicho primer ácido nucleico etiquetado con fluoróforo y una tercera señal de emisión fluorescente de dicha segunda reserva de segundo ácido nucleico etiquetado con fluoróforo; (d) determinar la semejanza de la muestra con las reservas de ER(-) y ER(+) comparando dichas primeras señales de emisión de fluorescencia y dichas primeras señales de emisión de fluorescencia por una parte, y por otra dichas primeras señales de emisión de fluorescencia y dichas terceras señales de emisión de fluorescencia, y clasificar como ER(+) las muestras en que las primeras señales de emisión de fluorescencia sean más parecidas a dichas segundas señales de emisión de fluorescencia que a dichas terceras señales de emisión de fluorescencia y clasificar como ER(-) las muestras en que las primeras señales de emisión sean más parecidas a dichas terceras señales de emisión de fluorescencia que a dichas segundas señales de emisión de fluorescencia, en donde dicha semejanza se define como criterio estadístico. La invención señala además que los demás conjuntos de marcadores que da a conocer pueden usarse en el método de más arriba para distinguir el *BRCA1* de los tumores esporádicos, así como las pacientes con diagnóstico deficiente de las pacientes con buen diagnóstico.

**[0015]** En una materialización específica, dicha semejanza se calcula determinando una primera suma de las diferencias de niveles de expresión para cada marcador entre dicho primer ácido nucleico etiquetado con fluoróforo y dicha primera reserva de segundo ácido nucleico etiquetado con fluoróforo y una segunda suma de las diferencias de niveles de expresión para cada marcador entre dicho primer ácido nucleico etiquetado con fluoróforo y dicha segunda reserva de segundo ácido nucleico etiquetado con fluoróforo, en donde si dicha primera suma es mayor que dicha segunda suma, la muestra se clasifica como ER(-), y si dicha segunda suma es mayor que dicha primera suma, la mezcla se clasifica como ER(+). En otra materialización específica, dicha semejanza se calcula computando un primer parámetro clasificador P1 entre una plantilla de ER(+) y la expresión de dichos marcadores en dicha muestra, y un segundo parámetro clasificador P2 entre una plantilla de ER(-) y la expresión de dichos marcadores en dicha muestra, en donde P1 y P2 se calculan según la fórmula:

$$P_i = (\bar{z}_i \cdot \bar{y}) / (\|\bar{z}_i\| \cdot \|\bar{y}\|),$$

**Ecuación (1)**

en donde  $\bar{z}_1$  y  $\bar{z}_2$  son plantillas de ER(-) y ER(+), respectivamente, y se calculan hallando la media de dicha segunda señal de emisión de fluorescencia para cada uno de dichos marcadores en dicha primera reserva de segundo ácido nucleico etiquetado con fluoróforo y dicha tercera señal de emisión de fluorescencia para cada uno de dichos marcadores en dicha segunda reserva de segundo ácido nucleico etiquetado con fluoróforo, respectivamente, y donde  $\bar{y}$  indica la primera señal de emisión de fluorescencia de cada uno de dichos marcadores en la muestra que se va a clasificar como ER(+) o ER(-), en donde la expresión de los marcadores en la muestra es parecida a ER(+) si  $P_1 < P_2$ , y parecida a ER (-) si  $P_1 > P_2$ .

**[0016]** Asimismo se describe un método para identificar genes marcadores cuya expresión esté asociada a un fenotipo en concreto. También se describe un método para determinar un conjunto de genes marcadores cuya expresión esté asociada a un fenotipo en concreto y que comprende los pasos de: (a) seleccionar el fenotipo que tenga dos o más categorías de fenotipo; (b) identificar una serie de genes en donde la expresión de dichos genes esté en correlación o anticorrelación con una de las categorías de fenotipo y en donde el coeficiente de correlación para cada gene se calcula según la ecuación:

$$\rho = (\bar{c} \cdot \bar{r}) / (\|\bar{c}\| \cdot \|\bar{r}\|)$$

**Ecuación (2)**

en donde  $\bar{c}$  es un número que representa dicha categoría de fenotipo y  $\bar{r}$  es el cociente de expresión logarítmica en todas las muestras para cada gen en concreto, en donde si el coeficiente de correlación tiene el valor absoluto de un valor umbral o mayor, dicha expresión de dicho gen está asociada a la categoría de fenotipo y en donde dicha serie de genes es un conjunto de genes marcadores cuya expresión está asociada a un fenotipo en concreto. El umbral depende del número de muestras empleadas; el umbral se puede calcular como  $3 \times 1/\sqrt{n-3}$ , en donde  $1/\sqrt{n-3}$  es la anchura de distribución y n = el número de muestras. En una materialización específica donde n = 98, dicho valor umbral es 0,3. En una materialización específica, dicho conjunto de genes marcadores se valida mediante: (a) el uso de un criterio estadístico para aleatorizar la asociación entre dichos genes marcadores y dicha categoría de fenotipo, creando de este modo un coeficiente de correlación de control para cada gen marcador; (b) la repetición del paso (a) cien veces o más para desarrollar una distribución de frecuencia de dichos coeficientes de correlación de control para cada gen marcador; (c) la determinación del número de genes marcadores que tengan con coeficiente de correlación de control de un valor umbral o mayor, creando de este modo un conjunto de genes marcadores de control, y (d) la comparación del número de genes marcadores de control identificados como tales con el número de genes marcadores, en donde si el valor p es la diferencia entre el número de genes marcadores y el número de genes de control es menor de 0,01, se valida dicho conjunto de genes marcadores. En otra materialización específica, dicho conjunto de genes marcadores se optimiza con un método que consiste en: (a) ordenación jerarquizada de genes por amplitud de correlación o por

relevancia de los coeficientes de correlación, y (b) selección de un número arbitrario de genes marcadores de los puestos más altos de la lista jerarquizada. El valor umbral depende del número de muestras sometidas a prueba.

**[0017]** Asimismo se describe un método para asignar a una persona una serie de categorías en un ensayo clínico, que consiste en determinar para cada persona el nivel de expresión de al menos cinco de los marcadores de pronóstico enumerados en la Tabla 6, determinar a partir de los mismos si la persona tiene un patrón de expresión que se halla en correlación con un buen pronóstico o un pronóstico deficiente, y clasificar a dicha persona en una categoría de un ensayo clínico si se establece que dicha persona tiene un buen pronóstico, o en una categoría diferente si se establece que dicha persona tiene un pronóstico deficiente. También se describe un método para clasificar a una persona en una de una serie de categorías de un ensayo clínico, donde cada una de dichas categorías está asociada a un fenotipo diferente, y que consiste en determinar para cada una de dichas personas el nivel de expresión de al menos cinco marcadores de un conjunto de marcadores, en donde dicho conjunto de marcadores comprende marcadores asociados a cada una de dichas categorías clínicas, determinando a partir de dichos marcadores si la persona tiene un patrón de expresión que se halla en correlación con una de las categorías clínicas y clasificando a dicha persona en una de dichas categorías si se establece que dicha persona tiene un fenotipo asociado a dicha categoría.

**[0018]** Asimismo se describe un método para determinar si la primera célula u organismo tiene uno de al menos dos diferentes fenotipos, consistiendo dichos al menos dos diferentes fenotipos en un primer fenotipo y un segundo fenotipo, método que consiste en: (a) comparar el nivel de expresión de cada uno de una serie de genes en una primera muestra de la primera célula u organismo con el nivel de expresión de cada uno de dichos genes, respectivamente, en una muestra colectiva de una serie de células u organismos, comprendiendo dicha serie de células u organismos diferentes células u organismos cada uno de los cuales muestra los al menos dos diferentes fenotipos, respectivamente, para producir un primer valor comparado; (b) comparar dicho primer valor comparado con un segundo valor comparado, en donde dicho segundo valor comparado es el producto de un método que consiste en comparar el nivel de expresión de cada uno de dichos genes en una muestra de una célula u organismo caracterizado por tener dicho primer fenotipo al nivel de expresión de cada uno de dichos genes, respectivamente, en dicha muestra colectiva; (c) comparar dicho primer valor comparado con un tercer valor comparado, en donde dicho tercer valor comparado es el producto de un método que consiste en comparar el nivel de expresión de cada uno de dichos genes en una muestra de una célula u organismo caracterizado por tener dicho segundo fenotipo al nivel de expresión de cada uno de dichos genes, respectivamente, en dicha muestra colectiva; (d) opcionalmente, llevar a cabo una o más veces el paso de comparar dicho primer valor comparado con uno o más valores comparados adicionales, respectivamente, siendo cada uno de dichos valores comparados adicionales el producto de un método que consiste en comparar el nivel de expresión de cada uno de dichos genes en una muestra de una célula u organismo caracterizado por tener un fenotipo diferente de dichos primer y segundo fenotipos, pero incluido en los dichos al menos dos diferentes fenotipos, al nivel de expresión de cada uno de dichos genes, respectivamente, en dicha muestra colectiva, y (e) determinar a cuál de dichos segundo, tercero y, si están presentes, a uno o más valores comparados adicionales, es más parecido dicho primer valor comparado adicional, en donde se determina que dicha primera célula u organismo tiene el fenotipo de la célula u organismo usados para producir dicho valor comparado más parecido al primer valor comparado.

**[0019]** En una materialización específica del método de más arriba, cada uno de dichos valores comparados es el cociente de los niveles de expresión de cada uno de estos genes. En otra materialización específica, cada uno de dichos niveles de expresión de cada uno de dichos genes en dicha muestra colectiva se normaliza antes de proceder a cualquiera de dichos pasos de comparación. En otra materialización específica, la normalización de dichos niveles de expresión se lleva a cabo dividiendo cada uno de dichos niveles de expresión por la mediana o el nivel medio de expresión de uno o más genes constitutivos en dicha muestra colectiva. En una materialización más específica, dichos niveles de expresión normalizados se someten a una transformación logarítmica y dichos pasos de comparación consisten en sustraer dicha transformación logarítmica del logaritmo de dichos niveles de expresión de cada uno de dichos genes en dicha muestra de dicha célula u organismo. En otra materialización específica, dichos al menos dos fenotipos diferentes son etapas diferentes de una enfermedad o desorden. En otra materialización específica, dichos al menos dos fenotipos diferentes son diferentes pronósticos de una enfermedad o desorden. En una materialización específica más, dichos niveles de expresión de cada uno de dichos genes, respectivamente, en dicha muestra colectiva o dichos niveles de expresión de cada uno de dichos genes en una muestra de dicha célula u organismo caracterizada por tener dicho primer fenotipo, dicho segundo fenotipo o dicho fenotipo diferente de dicho primer y segundo fenotipos, respectivamente, se almacenan en un ordenador.

**[0020]** Asimismo se describen biochips que comprenden los conjuntos de marcadores que se dan a conocer. Se describe un biochip que comprende al menos 5 marcadores derivados de cualquiera de los de las Tablas 1-6, en donde al menos un 50% de las sondas del biochip están presentes en alguna de las Tablas 1-6. En materializaciones más específicas, al menos un 60%, 70%, 80%, 90%, 95% o 98% de las sondas de dicho biochip están presentes en alguna de las Tablas 1- 6.

**[0021]** Asimismo se describe un biochip para distinguir entre muestras de células ER (+) y ER (-) que comprenden un despliegue de posición direccionable de sondas polinucleótidas conectadas a un soporte, comprendiendo dichas sondas polinucleótidas una serie de sondas polinucleótidas de diferentes secuencias polinucleótidas que comprenden una secuencia complementaria e hibridable con una serie de genes, consistiendo dicha serie en al menos 5 de los genes que corresponden a los marcadores enumerados en la Tabla 1 o en la Tabla 2, en donde al menos un 50% de las sondas del biochip están presentes en la Tabla 1 o en la Tabla 2. También se describe una selección para distinguir muestras de células afines a Brai y de tipo tumor esporádico que comprenden un despliegue de posición direccionable de sondas polinucleótidas conectadas a un soporte, comprendiendo dichas sondas polinucleótidas una serie de sondas polinucleótidas de diferentes secuencias polinucleótidas, comprendiendo cada una de dichas secuencias polinucleótidas una secuencia complementaria e hibridable con una serie de genes, consistiendo dicha serie en al menos 5 de los genes que corresponden a los marcadores enumerados en la Tabla 3 o en la Tabla 4, en donde al menos un 50% de las sondas del biochip están presente en la Tabla 3 o en la Tabla 4. Asimismo se describe un biochip para distinguir muestras de células de pacientes que tienen un buen pronóstico y muestras de células de pacientes que tienen un pronóstico deficiente, el cual comprende un despliegue de posición direccionable de sondas polinucleótidas conectadas a un soporte, comprendiendo dichas sondas polinucleótidas una serie de sondas polinucleótidas de diferentes secuencias polinucleótidas, consistiendo cada una de dichas diferentes secuencias polinucleótidas una secuencia complementaria e hibridable con una serie de genes, consistiendo dicha serie en al menos 5 de los genes que corresponden a los marcadores enumerados en la Tabla 5 o en la Tabla 6, en donde al menos un 50% de las sondas del biochip están presentes en la Tabla 5 o en la Tabla 6. También se describen biochips que comprenden al menos 5, 20, 50, 100, 200, 500, 100, 1.250, 1.500, 1.750, o 2.000 de los genes marcadores de categoría ER- enumerados en la Tabla 1, al menos 5, 20, 50, 100, 200 o 300 de los genes marcadores esporádicos del *BRCA1* enumerados en la Tabla 3, o al menos 5, 20, 50, 100 o 200 de los genes marcadores de pronóstico enumerados en la Tabla 5, en cualquier combinación, en donde al menos un 50%, 60%, 70%, 80%, 90%, 95% o 98% de las sondas de dichos biochips están presentes en la Tabla 1, la Tabla 3 y/o la Tabla 5.

**[0022]** Se describe un kit para determinar la categoría de ER de una muestra que comprende al menos dos biochips, de los que cada uno comprende 5 de los marcadores enumerados en la Tabla 1, y un sistema informático para determinar la semejanza del nivel de ácido nucleico derivado de los marcadores enumerados en la Tabla 1 en una muestra con el de una reserva de ER (-) y una reserva de ER (+), sistema informático que comprende un procesador y una memoria que lleva codificados uno o más programas acoplados al procesador, en donde dichos uno o más programas hacen que el procesador ejecute un método consistente en computar las diferencias agregadas en expresión de cada marcador entre la muestra y la reserva de ER (-) y las diferencias agregadas en expresión de cada marcador entre la muestra y la reserva de ER (+), o un método consistente en determinar la correlación calculada según la Ecuación (4). También se describen kits capaces de distinguir entre el *BRCA1* y los tumores esporádicos y muestras de pacientes con buen pronóstico de muestras de pacientes con pronóstico deficiente, mediante la inclusión de los conjuntos de genes marcadores apropiados. También se describe un kit para determinar si una muestra procede de una paciente con buen pronóstico o de una con pronóstico deficiente, que comprende al menos un biochip que comprende sondas para al menos 5 de los genes correspondientes a los marcadores enumerados en la Tabla 5, y un soporte informático legible en el que se hayan grabado uno o más programas para determinar la semejanza del nivel de ácido nucleico derivado de los marcadores enumerados en la Tabla 5 en una muestra con el de una reserva de muestras procedentes de individuos con buen pronóstico y una reserva de muestras procedentes de individuos con pronóstico deficiente, en donde dichos uno o más programas hace que un ordenador ejecute un método consistente en computar las diferencias agregadas en expresión de cada marcador entre la muestra y la reserva de buen pronóstico y las diferencias agregadas en expresión de cada marcador entre la muestra y la reserva de pronóstico deficiente, o un método consistente en determinar la correlación de expresión de los marcadores de la muestra con la expresión en las reservas de buen pronóstico y pronóstico deficiente, correlación que se calcula según la Ecuación (3).

#### 4. BREVE DESCRIPCIÓN DE LAS FIGURAS

##### **[0023]**

La FIG. 1 es un diagrama de tipo Venn que muestra la superposición entre los conjuntos de marcadores que aquí se dan a conocer, incluidos los 2.460 marcadores de ER, los 430 marcadores de *BRCA1*/esporádicos y los 231 indicadores de pronóstico.

La FIG. 2 muestra los procedimientos experimentales para medir diferentes cambios en la densidad de transcripción de ARNm en los tumores de cáncer de mama usados en este estudio. En cada experimento, el ARNc etiquetado como Cy5 de un tumor X se hibrida en un biochip humano de 25k junto con una reserva de ARNc etiquetada como Cy3 hecha de muestras de ARNc de los tumores 1, 2, ... N. Los datos de expresión digital se obtuvieron por escaneo y procesamiento de imágenes. El error de modelado nos permitió asignar un valor-p a cada medición del cociente de transcripción.

La FIG. 3 El agrupamiento bidimensional revela dos tipos diferenciados de tumores. El agrupamiento se basó en los datos de expresión de genes de 98 tumores de cáncer de mama sobre 4.986 genes relevantes. El gris oscuro (rojo) representa regulación hacia arriba, el gris claro (verde) representa regulación hacia abajo, el negro indica sin cambios en la expresión y el gris indica que los datos no están disponibles. Se seleccionaron 4.986 genes que mostraron un cambio de cociente de más del doble en más de cinco experimentos. Los datos clínicos seleccionados para los resultados de las pruebas de mutaciones del *BRCA1*, receptor de estrógeno (ER) y receptor de proestrógeno (PR), grado del tumor, infiltrado linfocítico y angioinvasión se muestran a la derecha. El negro indica negativo y el blanco indica positivo. El patrón predominante en la parte inferior consta de 36 pacientes, de las cuales 34 son ER-negativas (total 39), y 16 son portadoras de mutación de *BRCA1* (total 18).

FIG. 4 Una porción de resultados agrupados no supervisados, como los de la FIG. 3. *ESR1* (el gen receptor de estrógenos) está regulado conjuntamente con un grupo de genes fuertemente regulados entre sí para formar un patrón dominante.

FIG. 5A Histograma de coeficientes de correlación de genes relevantes entre sus cocientes de expresión y la categoría del receptor de estrógenos (ER) (esto es, el nivel de ER). El histograma para datos experimentales aparece en línea gris. Los resultados de un ensayo Monte-Carlo aparecen en línea negra gruesa. Hay 2.460 genes cuyos datos de expresión están en correlación con la categoría de ER a un nivel superior a 0,3 o en anticorrelación con la categoría de ER a un nivel inferior a -0,3.

FIG. 5B La distribución del número de genes que se ajustó a los mismos criterios de selección (amplitud de correlación por encima de 0,3) de entre 10.000 ensayos Monte-Carlo. Se estima que este conjunto de 2.460 genes informa de la categoría de ER a un nivel de fiabilidad de  $p > 99.99\%$ .

FIG. 6 Tasas de error de clasificación para Tipo 1 y Tipo 2 como función del número (excepto 2,460) de genes marcadores usados en el clasificador. La tasa de error combinada más baja se da cuando se usan aproximadamente 550 genes marcadores.

FIG. 7 Clasificación de 98 muestras de tumores como ER(+) o ER(-) basándose en niveles de expresión de los 550 genes marcadores óptimos. Las muestras de ER(+) (por encima de la línea blanca) muestran un patrón de expresión claramente diferente al de las muestras de ER(-) (por debajo de la línea blanca).

FIG. 8 Correlación entre los niveles de expresión en muestras de cada paciente y el perfil medio del grupo de ER(-) con respecto a la correlación con el grupo de ER(+). Los cuadrados representan muestras de pacientes clínicamente ER(-); los redondeles representan muestras de pacientes clínicamente ER(-).

FIG. 9A El histograma de coeficientes de correlación de expresión de genes sobre el cociente de cada gen relevante con la clase de mutación del *BRCA1* aparece en línea continua. La línea discontinua indica una distribución de frecuencia a partir de un ensayo Monte-Carlo. 430 genes mostraron una amplitud de correlación o anticorrelación superior a 0,35.

FIG. 9B Distribución de frecuencia del número de genes que muestran una amplitud de correlación o anticorrelación superior a 0,35 para el control de los 10.000 ensayos Monte-Carlo. Media=115.  $p(n > 430) = 0.48\%$  y  $p(> 430/2) = 9.0\%$ .

FIG. 10 Tasas de error de clasificación para Tipo 1 y Tipo 2 como función del número de genes discriminantes usados en el clasificador (plantilla). La tasa de error combinada más baja se da cuando se usan aproximadamente 100 genes marcadores discriminantes.

FIG. 11A La clasificación de 38 tumores del grupo ER(-) en dos subgrupos, *BRCA1* y esporádicos, usando el conjunto óptimo de 100 genes marcadores discriminantes. Las pacientes por encima de la línea blanca se caracterizan por patrones asociados al *BRCA1*.

FIG. 11B Correlación entre niveles de expresión en muestras de cada paciente ER(-) y el perfil medio del grupo *BRCA1* frente a correlación con el grupo esporádico. Los cuadrados representan muestras de pacientes con tumores de tipo esporádico; los redondeles representan muestras de pacientes portadoras de la mutación de *BRCA1*.

FIG. 12A El histograma de los coeficientes de correlación de cociente de expresión de genes de cada gen relevante con la clase de pronóstico (grupo de metástasis distantes y grupo de metástasis no distantes) aparece en línea continua. La distribución procedente de un ensayo Monte-Carlo aparece en línea discontinua. La amplitud de correlación o anticorrelación de 231 genes marcadores es superior a 0,3.

FIG. 12B Distribución de frecuencia del número de genes cuya amplitud de correlación o anticorrelación fue superior a 0,3 en 10,000 ensayos Monte-Carlo.



FIG. 13 Tasa de error de clasificación del grupo de metástasis distante para Tipo 1 y Tipo 2 como función del número de genes discriminantes usados en el clasificador. La tasa de error combinada más baja se da cuando se usan aproximadamente 70 genes marcadores discriminantes.

FIG. 14 Clasificación de 78 tumores esporádicos en dos grupos de pronóstico, metástasis distante (pronóstico deficiente) y metástasis no distante (buen pronóstico) mediante el conjunto óptimo de 70 genes marcadores discriminantes. Las pacientes por encima de la línea blanca se caracterizan por tener buen pronóstico. Las pacientes por debajo de la línea blanca se caracterizan por tener pronóstico deficiente.

FIG. 15 Correlación entre los niveles de expresión en muestras de cada paciente y el perfil medio del grupo de buen pronóstico frente a la correlación con el grupo de pronóstico deficiente. Los cuadrados representan muestras de pacientes que tienen pronóstico deficiente, los redondeles representan muestras de pacientes que tienen pronóstico deficiente. Los cuadrados rojos representan las "recaídas", mientras que los redondeles azules representan las "no recaídas". Se clasificó erróneamente a 13 de un total de 78.

FIG. 16 Probabilidad de recaída como función de tiempo desde el diagnóstico. La predicción del Grupo A y del grupo B se realizó según el método de "dejar uno fuera" basándose en el conjunto óptimo de 70 genes marcadores discriminantes. Las 43 pacientes del grupo A se dividen en 37 pacientes del grupo de metástasis no distante y 6 pacientes del grupo de metástasis distante. Las 35 pacientes del grupo B se dividen en 28 pacientes del grupo de metástasis distante y 7 pacientes del grupo de metástasis no distante.

FIG. 17 Probabilidad de metástasis distante como función de tiempo desde el diagnóstico para individuos con receptor de progesterona PR(+) (sí) o PR(-) (no).

FIG. 18 Probabilidad de metástasis distante como función de tiempo desde el diagnóstico para individuos de ER(+) (sí) o ER(-) (no).

FIG. 19A, B Probabilidad de metástasis distante como función de tiempo desde el diagnóstico. Los grupos se definieron por grados de tumor.

FIG. 20A Clasificación de 19 tumores esporádicos independientes en dos grupos de pronóstico, metástasis distante y metástasis no distante, mediante los 70 genes marcadores óptimos. Las pacientes por encima de la línea blanca tienen buen pronóstico. Las pacientes por debajo de la línea blanca tienen pronóstico deficiente.

FIG. 20B La correlación entre los cocientes de expresión de cada paciente y el cociente medio de expresión del grupo de buen pronóstico se define mediante el conjunto de ensayo con respecto a la correlación entre los cocientes de expresión de cada paciente y el cociente medio de expresión del conjunto de ensayo de pronóstico deficiente. De nueve pacientes en el grupo de buen pronóstico, tres son del "grupo de metástasis distante"; de diez pacientes en el grupo de buen pronóstico, una paciente es del "grupo de metástasis no distante". Esta tasa de error de 4 sobre 19 se corresponde con la de 13 sobre 78 en las 78 pacientes iniciales.

FIG. 20C Probabilidad de recaída como función de tiempo desde el diagnóstico para dos grupos cuya predicción se hizo basándose en la expresión de los 70 genes marcadores óptimos.

FIG. 21A Sensibilidad frente a 1-especificidad para la clasificación de buen pronóstico.

FIG. 21B Sensibilidad frente a 1-especificidad para la clasificación de pronóstico deficiente.

FIG. 21C Tasa total de error como función de umbral en la probabilidad modelada. Se usaron seis parámetros clínicos (categoría de ER, categoría de PR, grado de tumor, tamaño de tumor, edad de la paciente y presencia o ausencia de angioinvasión) para realizar el modelado clínico.

FIG. 22 Comparación del cociente logarítmico de muestras individuales mediante la "reserva de muestras de material" frente a la intensidad logarítmica sustraída media mediante la "reserva de muestras matemáticas" para 70 genes indicadores en las 78 muestras de tumores esporádicos. La "reserva de muestras de material" estaba constituida por las 78 muestras de tumores esporádicos.

FIG. 23A Los resultados de la validación cruzada dejando un elemento fuera basándose en los datos de un solo canal. Las muestras se agrupan según el coeficiente de correlación de cada muestra con el perfil medio de "buen pronóstico" y "pronóstico deficiente" en los 70 genes examinados. La línea blanca separa las muestras de los pacientes con pronóstico deficiente (abajo) de las de aquellos con buen pronóstico (arriba).

FIG. 23B Diagrama de dispersión de coeficientes de correlación con la expresión media en muestras de “buen pronóstico” y “pronóstico deficiente”. La falsa tasa positiva (esto es, la tasa de clasificaciones incorrectas de una muestra de una paciente con buen pronóstico como perteneciente a una con pronóstico deficiente) fue de 10 sobre 44, mientras que la falsa tasa negativa fue de 6 sobre 34.

5 FIG. 24A Hibridación de datos de un solo canal para muestras ordenadas jerárquicamente por coeficientes de correlación con el clasificador de buen pronóstico. Las muestras clasificadas como “buen pronóstico” se hallan por encima de la línea blanca, mientras que las clasificadas como “pronóstico deficiente” se hallan por debajo de la misma.

10 FIG. 24B Diagrama de dispersión de coeficientes de correlación con tres muestras clasificadas incorrectamente que se hallan a la derecha del valor umbral de coeficiente de correlación. Dicho valor umbral de correlación se fijó en 0,2727 para limitar las falsas muestras negativas a aproximadamente el 10% de las mismas.

## DESCRIPCIÓN DETALLADA DE LA INVENCION

### 5.1 INTRODUCCIÓN

15 **[0024]** La invención se refiere a conjuntos de marcadores genéticos cuyos patrones de expresión guardan correlación con importantes características del cáncer de mama, esto es, la categoría del receptor de estrógeno (ER), la categoría del *BRCA1* y la probabilidad de recaída (esto es, metástasis distante o pronóstico deficiente). Se describen conjuntos marcadores genéticos que pueden distinguir las tres siguientes categorías clínicas. En  
20 primer lugar, la invención se refiere a conjuntos de marcadores cuya expresión guarda correlación con la categoría de ER de un paciente y que puede usarse para distinguir los pacientes ER (+) de los pacientes ER (-). La categoría de ER es un indicador de pronóstico muy útil y un indicador de la probabilidad de que un paciente responda a ciertas terapias como el tamoxifeno.

**[0025]** Igualmente, entre las mujeres que son ER positivas la tasa de respuesta (superior al 50%) a la terapia hormonal es mucho mayor que la tasa de respuesta (inferior al 10%) en pacientes con categoría negativa de ER.  
25 En pacientes con tumores de ER positivo la posibilidad de conseguir una respuesta hormonal es directamente proporcional al nivel de ER (P. Calabresi y P. S. Schein, MEDICAL ONCOLOGY (2ND ED.), McGraw-Hill, Inc., New York (1993)). En segundo lugar, la invención se refiere además a conjuntos de marcadores cuya expresión guarda correlación con la presencia de mutaciones de *BRCA1* y que pueden usarse para distinguir los tumores de tipo *BRCA1* de los tumores esporádicos. En tercer lugar, la invención se refiere a marcadores genéticos cuya  
30 expresión guarda correlación con el pronóstico clínico y que puede usarse para distinguir entre las pacientes que tienen buen pronóstico (esto es, que no tengan metástasis distante de un tumor dentro de los cinco años) y las de pronóstico deficiente (esto es, que tengan metástasis distante de un tumor durante cinco años). En cuanto al uso de estos marcadores, se suministran métodos para distinguir entre estos grupos de pacientes y para determinar pautas generales de tratamiento. También se suministran biochips que contienen estos marcadores,  
35 así como métodos para elaborar dichos biochips. Cada uno de los marcadores corresponde a un gen del genoma humano, es decir, dicho marcador es identificable como el todo o una porción de un gen. Por último, como cada uno de los marcadores de más arriba está en correlación con ciertas afecciones relacionadas con el cáncer de mama, los marcadores, o las proteínas codificadas en éstos, serán muy probablemente el destino de medicamentos contra el cáncer de mama.

### 40 5.2 DEFINICIONES

**[0026]** Tal como se usa aquí, “tumor BRCA1” significa tumor que tiene células que contienen una mutación del locus *BRCA1*.

**[0027]** La “amplitud absoluta” de expresiones de correlación significa la distancia, tanto si es positiva como negativa, a un valor cero, es decir, ambos coeficientes, -0,35 y 0,35, tienen una amplitud absoluta de 0,35.

45 **[0028]** “Categoría” significa un estado de expresión de genes de un conjunto de marcadores genéticos cuya expresión está en estrecha correlación con un fenotipo concreto. Por ejemplo, “categoría de ER” significa un estado de expresión de genes de un conjunto de marcadores genéticos cuya expresión está en estrecha correlación con la del ESR1 (gen receptor de estrógeno), en donde el patrón de expresión de estos genes varía significativamente entre los tumores que expresan el receptor y los tumores que no expresan el receptor.

50 **[0029]** “Buen pronóstico” significa que se espera que una paciente no tenga metástasis distantes de un tumor de mama dentro de los cinco años siguientes al diagnóstico inicial de cáncer de mama.

**[0030]** “Pronóstico deficiente” significa que se espera que una paciente tenga metástasis distantes de un tumor de mama dentro de los cinco años siguientes al diagnóstico inicial de cáncer de mama.

**[0031]** “Marcador” significa un gen entero, o un EST derivado de dicho gen, cuyo nivel de expresión cambia en ciertas condiciones. Allí donde la expresión del gen esté en correlación con un estado en concreto, el gen será un marcador para ese estado.

5 **[0032]** “Polinucleótidos derivados de marcador” significa el ARN transcrito de un gen marcador, cualquier ADNc o ARNc producido del mismo y cualquier ácido nucleico derivado del mismo, como el ácido nucleico sintético que tenga una secuencia derivada del gen correspondiente al gen marcador.

### 5.3 MARCADORES ÚTILES EN EL DIAGNÓSTICO Y PRONÓSTICO DEL CÁNCER DE MAMA

#### 5.3.1 CONJUNTOS DE MARCADORES

10 **[0033]** Se describe un conjunto de 4.986 marcadores genéticos cuya expresión está en correlación con la existencia de cáncer de mama mediante análisis de agrupación. En la SEQ ID Nos: 1-2,699 se enumera un subconjunto de dichos marcadores identificados como útiles para el diagnóstico o el pronóstico. La invención también se refiere a un método de uso de dichos marcadores para distinguir los tipos de tumores en el diagnóstico o el pronóstico.

15 **[0034]** También se describe un conjunto de 2.460 marcadores genéticos que pueden clasificar a las pacientes de cáncer de mama por la categoría del receptor de estrógeno (ER), esto es, distinguir entre pacientes o tumores con ER (+) y ER (-) obtenidos de dichas pacientes. La categoría de ER es un importante indicador de la probabilidad de la respuesta de un paciente a algunas quimioterapias (esto es, al tamoxifeno). Dichos marcadores se enumeran en la Tabla 1. La invención también se refiere a subconjuntos de al menos 5, 10, 25, 50, 100, 200, 300, 400, 500, 750, 1.000, 1.250, 1.500, 1.750 o 2.000 marcadores genéticos, sacados del conjunto de 2.460 marcadores, que también pueden distinguir pacientes o tumores con ER (+) y ER (-). Preferiblemente, el número de marcadores será de 550. Se describe además un conjunto de 550 de los 2.460 marcadores que son óptimos para distinguir la categoría de ER (Tabla 2). También se suministra un método de uso de dichos marcadores para distinguir entre pacientes o tumores ER (+) y ER (-) obtenidos de los mismos.

25 **[0035]** En otra materialización se describe un conjunto de 430 marcadores genéticos que pueden clasificar a pacientes de cáncer de mama con ER (-) por la categoría del *BRCA1*, es decir, distinguir entre tumores que contienen una mutación de *BRCA1* y tumores esporádicos. Dichos marcadores se enumeran en la Tabla 3. Además, se proporcionan subconjuntos de al menos 5, 10, 20, 30, 40, 50, 75, 100, 150, 200, 250, 300 o 350 marcadores, sacados del conjunto de 430 marcadores, que también distinguen entre tumores que contienen una mutación de *BRCA1* y tumores esporádicos. Preferiblemente, el número de marcadores será de 100. En la Tabla 30 4 se suministra un conjunto preferido de 100 marcadores. También se suministra un método de uso de dichos marcadores para distinguir entre pacientes o tumores *BRCA1* y esporádicos obtenidos de los mismos.

35 **[0036]** La invención se refiere a un conjunto de 231 marcadores genéticos que pueden distinguir entre las pacientes con un buen pronóstico de cáncer de mama (ausencia de metástasis distantes de tumores de cáncer de mama dentro de los cinco años siguientes) y las pacientes con un pronóstico deficiente de cáncer de mama (presencia de metástasis distantes de tumores de cáncer de mama dentro de los cinco años siguientes). Dichos marcadores se enumeran en la Tabla 5. Se suministran subconjuntos de al menos 5, 10, 20, 30, 40, 50, 75, 100, 150 o 200 marcadores, sacados del conjunto de 231, que también distinguen entre pacientes con buen y con pronóstico deficiente. En la Tabla 6 se suministra un conjunto preferido de 70 marcadores. En una materialización específica, el conjunto de marcadores consta de los doce marcadores relacionados con la 40 quinasas y los siete marcadores relacionados con la división de células o mitosis. La invención también suministra un método de uso de los marcadores de más arriba para distinguir entre pacientes con buen pronóstico y con pronóstico deficiente.

**Table 1. 2.460 marcadores de genes que distinguen entre muestras de células ER(+) y ER(-).**

Banco de Genes		Bando de Genes	
Número de Acceso	SEQ ID NO	Número de Acceso	SEQ ID NO
AA555029_RC	SEQ ID NO 1	NM_006984	SEQ ID NO 1344
AB000509	SEQ ID NO 2	NM_007005	SEQ ID NO 1345
AB001451	SEQ ID NO 3	NM_007006	SEQ ID NO 1346
AB002301	SEQ ID NO 4	NM_007019	SEQ ID NO 1347
AB002308	SEQ ID NO 5	NM_007027	SEQ ID NO 1348
AB002351	SEQ ID NO 6	NM_007044	SEQ ID NO 1350
AB002448	SEQ ID NO 7	NM_007050	SEQ ID NO 1351

AB006628	SEQ ID NO 9	NM_007057	SEQ ID NO 1352
AB006630	SEQ ID NO 10	NM_007069	SEQ ID NO 1353
AB006746	SEQ ID NO 11	NM_007074	SEQ ID NO 1355
AB007458	SEQ ID NO 12	NM_007088	SEQ ID NO 1356
AB007855	SEQ ID NO 13	NMB_007111	SEQ ID NO 1357
AB007857	SEQ ID NO 14	NMB_007146	SEQ ID NO 1358
AB007863	SEQ ID NO 15	NM_007173	SEQ ID NO 1359
AB007883	SEQ ID NO 16	NMB_007177	SEQ ID NO 1360
AB007896	SEQ ID NO 17	NMB_007196	SEQ ID NO 1361
AB007899	SEQ ID NO 18	NM_007203	SEQ ID NO 1362
AB007916	SEQ ID NO 19	NM_007214	SEQ ID NO 1363
AB007950	SEQ ID NO 20	NMB_007217	SEQ ID NO 1364
AB011087	SEQ ID NO 21	NM_007231	SEQ ID NO 1365
AB011089	SEQ ID NO 22	NM_007268	SEQ ID NO 1367
AB011104	SEQ ID NO 23	NM_007274	SEQ ID NO 1368
AB011105	SEQ ID NO 24	NM_007275	SEQ ID NO 1369
AB011121	SEQ ID NO 25	NM_007281	SEQ ID NO 1370
AB011132	SEQ ID NO 26	NM_007309	SEQ ID NO 1371
AB011152	SEQ ID NO 27	NMB_007315	SEQ ID NO 1372
AB011179	SEQ ID NO 28	NM_007334	SEQ ID NO 1373
AB014534	SEQ ID NO 29	NM_007358	SEQ ID NO 1374
AB014568	SEQ ID NO 30	NM_009585	SEQ ID NO 1375
AB018260	SEQ ID NO 31	NM_009587	SEQ ID NO 1376
AB018268	SEQ ID NO 32	NM_009588	SEQ ID NO 1377
AB018289	SEQ ID NO 33	NMB_012062	SEQ ID NO 1378
AB018345	SEQ ID NO 35	NMB_012067	SEQ ID NO 1379
AB020677	SEQ ID NO 36	NMB_012101	SEQ ID NO 1380
AB020689	SEQ ID NO 37	NMB_012105	SEQ ID NO 1381
AB020695	SEQ ID NO 38	NMB_012108	SEQ ID NO 1382
AB020710	SEQ ID NO 39	NM_012110	SEQ ID NO 1383
AB023139	SEQ ID NO 40	NMB_012124	SEQ ID NO 1384
AB023151	SEQ ID NO 41	NM_012142	SEQ ID NO 1386
AB023152	SEQ ID NO 42	NM_012155	SEQ ID NO 1388
AB023163	SEQ ID NO 43	NM_012175	SEQ ID NO 1389
AB023173	SEQ ID NO 44	NMB_012177	SEQ ID NO 1390
AB023211	SEQ ID NO 45	NM_012205	SEQ ID NO 1391
AB024704	SEQ ID NO 46	NM_012219	SEQ ID NO 1393
AB028985	SEQ ID NO 47	NMB_012242	SEQ ID NO 1394
AB028986	SEQ ID NO 48	NMB_012250	SEQ ID NO 1395
AB028998	SEQ ID NO 49	NM_012261	SEQ ID NO 1397
AB029031	SEQ ID NO 51	NM_012286	SEQ ID NO 1398
AB032951	SEQ ID NO 52	NM_012319	SEQ ID NO 1400
AB032966	SEQ ID NO 53	NM_012332	SEQ ID NO 1401
AB032969	SEQ ID NO 54	NM_012336	SEQ ID NO 1402
AB032977	SEQ ID NO 56	NMB_012339	SEQ ID NO 1404
AB033007	SEQ ID NO 58	NMB_012341	SEQ ID NO 1405
AB033034	SEQ ID NO 59	NMB_012391	SEQ ID NO 1406

AB033035	SEQ ID NO 60	NM_012394	SEQ ID NO 1407
AB033040	SEQ ID NO 61	NM_012413	SEQ ID NO 1408
AB033049	SEQ ID NO 63	NM_012421	SEQ ID NO 1409
AB033050	SEQ ID NO 64	NMB_012425	SEQ ID NO 1410
AB033053	SEQ ID NO 65	NM_012427	SEQ ID NO 1411
AB033055	SEQ ID NO 66	NM_012429	SEQ ID NO 1413
AB033058	SEQ ID NO 67	NM_012446	SEQ ID NO 1414
AB033073	SEQ ID NO 68	NM_012463	SEQ ID NO 1415
AB033092	SEQ ID NO 69	NM_012474	SEQ ID NO 1416
AB033111	SEQ ID NO 70	NM_013230	SEQ ID NO 1417
AB036063	SEQ ID NO 71	NM_013233	SEQ ID NO 1418
AB037720	SEQ ID NO 72	NM_013238	SEQ ID NO 1419
AB037743	SEQ ID NO 74	NM_013239	SEQ ID NO 1420
AB037745	SEQ ID NO 75	NM_013242	SEQ ID NO 1421
AB037756	SEQ ID NO 76	NM_01 3257	SEQ ID NO 1423
AB037765	SEQ ID NO 77	NM_013261	SEQ ID NO 1424
AB037778	SEQ ID NO 78	NM_013262	SEQ ID NO 1425
AB037791	SEQ ID NO 79	NM_013277	SEQ ID NO 1426
AB037793	SEQ ID NO 80	NM_013296	SEQ ID NO 1427
AB037802	SEQ ID NO 81	NM_013301	SEQ ID NO 1428
AB037806	SEQ ID NO 82	NM_013324	SEQ ID NO 1429
AB037809	SEQ ID NO 83	NM_013327	SEQ ID NO 1430
AB037836	SEQ ID NO 84	NM_013336	SEQ ID NO 1431
AB037844	SEQ ID NO 85	NM_013339	SEQ ID NO 1432
AB037845	SEQ ID NO 86	NM_013363	SEQ ID NO 1433
AB037848	SEQ ID NO 87	NM_013378	SEQ ID NO 1435
AB037863	SEQ ID NO 88	NM_013384	SEQ ID NO 1436
AB037864	SEQ ID NO 89	NM_013385	SEQ ID NO 1437
AB040881	SEQ ID NO 90	NM_013406	SEQ ID NO 1438
AB040900	SEQ ID NO 91	NM_013437	SEQ ID NO 1439
AB040914	SEQ ID NO 92	NM_013451	SEQ ID NO 1440
AB040926	SEQ ID NO 93	NM_013943	SEQ ID NO 1441
AB040955	SEQ ID NO 94	NMB_013994	SEQ ID NO 1442
AB040961	SEQ ID NO 95	NMB_013995	SEQ ID NO 1443
AF000974	SEQ ID NO 97	NMB_014026	SEQ ID NO 1444
AF005487	SEQ ID NO 98	NM_014029	SEQ ID NO 1445
AF007153	SEQ ID NO 99	NM_014036	SEQ ID NO 1446
AF007155	SEQ ID NO 100	NM_014062	SEQ ID NO 1447
AF015041	SEQ ID NO 101	NMB_014074	SEQ ID NO 1448
AF016004	SEQ ID NO 102	NM_014096	SEQ ID NO 1450
AF016495	SEQ ID NO 103	NM_014109	SEQ ID NO 1451
AF020919	SEQ ID NO 104	NM_014112	SEQ ID NO 1452
AF026941	SEQ ID NO 105	NM_014147	SEQ ID NO 1453
AF035191	SEQ ID NO 106	NM_014149	SEQ ID NO 1454
AF035284	SEQ ID NO 107	NMB_014164	SEQ ID NO 1455
AF035318	SEQ ID NO 108	NMB_014172	SEQ ID NO 1456
AF038182	SEQ ID NO 109	NM_014175	SEQ ID NO 1457

AF038193	SEQ ID NO 110	NM_014181	SEQ ID NO 1458
AF042838	SEQ ID NO 111	NM_014184	SEQ ID NO 1459
AF044127	SEQ ID NO 112	NM_014211	SEQ ID NO 1460
AF045229	SEQ ID NO 113	NM_014214	SEQ ID NO 1461
AF047002	SEQ ID NO 114	NM_014216	SEQ ID NO 1462
AF047826	SEQ ID NO 115	NM_014241	SEQ ID NO 1463
AF049460	SEQ ID NO 116	NM_014246	SEQ ID NO 1465
AF052101	SEQ ID NO 117	NM_014268	SEQ ID NO 1466
AF052117	SEQ ID NO 118	NM_014272	SEQ ID NO 1467
AF052155	SEQ ID NO 119	NM_014274	SEQ ID NO 1468
AF052159	SEQ ID NO 120	NM_014289	SEQ ID NO 1469
AF052176	SEQ ID NO 122	NM_014298	SEQ ID NO 1470
AF052185	SEQ ID NO 123	NM_014302	SEQ ID NO 1471
AF055270	SEQ ID NO 126	NM_014315	SEQ ID NO 1473
AF058075	SEQ ID NO 127	NM_014316	SEQ ID NO 1474
AF061034	SEQ ID NO 128	NM_014317	SEQ ID NO 1475
AF063725	SEQ ID NO 129	NM_014320	SEQ ID NO 1476
AF063936	SEQ ID NO 130	NM_014321	SEQ ID NO 1477
AF065241	SEQ ID NO 131	NM_014325	SEQ ID NO 1478
AF067972	SEQ ID NO 132	NM_014335	SEQ ID NO 1479
AF070536	SEQ ID NO 133	NM_014363	SEQ ID NO 1480
AF070552	SEQ ID NO 134	NM_014364	SEQ ID NO 1481
AF070617	SEQ ID NO 135	NM_014365	SEQ ID NO 1482
AF073770	SEQ ID NO 138	NM_014373	SEQ ID NO 1483
AF076612	SEQ ID NO 139	NM_014382	SEQ ID NO 1484
AF079529	SEQ ID NO 140	NM_014395	SEQ ID NO 1485
AF090913	SEQ ID NO 142	NM_014398	SEQ ID NO 1486
AF095719	SEQ ID NO 143	NM_014399	SEQ ID NO 1487
AF098641	SEQ ID NO 144	NM_014402	SEQ ID NO 1488
AF099032	SEQ ID NO 145	NM_014428	SEQ ID NO 1489
AF100756	SEQ ID NO 146	NM_014448	SEQ ID NO 1490
AF101051	SEQ ID NO 147	NM_014449	SEQ ID NO 1491
AF103375	SEQ ID NO 148	NM_014450	SEQ ID NO 1492
AF103458	SEQ ID NO 149	NM_014452	SEQ ID NO 1493
AF103530	SEQ ID NO 150	NM_014453	SEQ ID NO 1494
AF103804	SEQ ID NO 151	NM_014456	SEQ ID NO 1495
AF111849	SEQ ID NO 152	NM_014479	SEQ ID NO 1497
AF112213	SEQ ID NO 153	NM_014501	SEQ ID NO 1498
AF113132	SEQ ID NO 154	NM_014552	SEQ ID NO 1500
AF116682	SEQ ID NO 156	NM_014553	SEQ ID NO 1501
AF118224	SEQ ID NO 157	NM_014570	SEQ ID NO 1502
AF118274	SEQ ID NO 158	NM_014575	SEQ ID NO 1503
AF119256	SEQ ID NO 159	NM_014585	SEQ ID NO 1504
AF119665	SEQ ID NO 160	NM_014595	SEQ ID NO 1505
AF121255	SEQ ID NO 161	NM_014624	SEQ ID NO 1507
AF131748	SEQ ID NO 162	NM_014633	SEQ ID NO 1508
AF131753	SEQ ID NO 163	NM_014640	SEQ ID NO 1509

AF131760	SEQ ID NO 164	NM_014642	SEQ ID NO 1510
AF131784	SEQ ID NO 165	NM_014643	SEQ ID NO 1511
AF131828	SEQ ID NO 166	NM_014656	SEQ ID NO 1512
AF135168	SEQ ID NO 167	NM_014668	SEQ ID NO 1513
AF141882	SEQ ID NO 168	NM_014669	SEQ ID NO 1514
AF148505	SEQ ID NO 169	NM_014673	SEQ ID NO 1515
AF149785	SEQ ID NO 170	NM_014675	SEQ ID NO 1516
AF151810	SEQ ID NO 171	NM_014679	SEQ ID NO 1517
AF152502	SEQ ID NO 172	NM_014680	SEQ ID NO 1518
AF155120	SEQ ID NO 174	NM_014696	SEQ ID NO 1519
AF159092	SEQ ID NO 175	NM_014700	SEQ ID NO 1520
AF161407	SEQ ID NO 176	NM_014715	SEQ ID NO 1521
AF161553	SEQ ID NO 177	NM_014721	SEQ ID NO 1522
AF164104	SEQ ID NO 178	NM_014737	SEQ ID NO 1524
AF167706	SEQ ID NO 179	NM_014738	SEQ ID NO 1525
AF175387	SEQ ID NO 180	NM_014747	SEQ ID NO 1526
AF176012	SEQ ID NO 181	NM_014750	SEQ ID NO 1527
AF186780	SEQ ID NO 182	NM_014754	SEQ ID NO 1528
AF217508	SEQ ID NO 184	NM_014767	SEQ ID NO 1529
AF220492	SEQ ID NO 185	NM_014770	SEQ ID NO 1530
AF224266	SEQ ID NO 186	NM_014773	SEQ ID NO 1531
AF230904	SEQ ID NO 187	NM_014776	SEQ ID NO 1532
AF234532	SEQ ID NO 188	NM_014782	SEQ ID NO 1533
AF257175	SEQ ID NO 189	NM_014785	SEQ ID NO 1534
AF257659	SEQ ID NO 190	NM_014791	SEQ ID NO 1535
AF272357	SEQ ID NO 191	NM_014808	SEQ ID NO 1536
AF279865	SEQ ID NO 192	NM_014811	SEQ ID NO 1537
AI497657_RC	SEQ ID NO 193	NM_014812	SEQ ID NO 1538
AJ012755	SEQ ID NO 194	NM_014838	SEQ ID NO 1540
AJ223353	SEQ ID NO 195	NM_014862	SEQ ID NO 1542
AJ224741	SEQ ID NO 196	NM_014865	SEQ ID NO 1543
AJ224864	SEQ ID NO 197	NM_014870	SEQ ID NO 1544
AJ225092	SEQ ID NO 198	NM_014875	SEQ ID NO 1545
AJ225093	SEQ ID NO 199	NM_014886	SEQ ID NO 1547
AJ249377	SEQ ID NO 200	NM_014889	SEQ ID NO 1548
AJ270996	SEQ ID NO 202	NM_014905	SEQ ID NO 1549
AJ272057	SEQ ID NO 203	NM_014935	SEQ ID NO 1550
AJ275978	SEQ ID NO 204	NM_014945	SEQ ID NO 1551
AJ276429	SEQ ID NO 205	NM_014965	SEQ ID NO 1552
AK000004	SEQ ID NO 206	NM_014967	SEQ ID NO 1553
AK000005	SEQ ID NO 207	NM_014968	SEQ ID NO 1554
AK000106	SEQ ID NO 208	NM_015032	SEQ ID NO 1555
AK000142	SEQ ID NO 209	NM_015239	SEQ ID NO 1556
AK000168	SEQ ID NO 210	NM_015383	SEQ ID NO 1557
AK000345	SEQ ID NO 212	NM_015392	SEQ ID NO 1558
AK000543	SEQ ID NO 213	NM_015416	SEQ ID NO 1559
AK000552	SEQ ID NO 214	NM_015417	SEQ ID NO 1560

AK000643	SEQ ID NO 216	NM_015420	SEQ ID NO 1561
AK000660	SEQ ID NO 217	NM_015434	SEQ ID NO 1562
AK000689	SEQ ID NO 218	NM_015474	SEQ ID NO 1563
AK000770	SEQ ID NO 220	NM_015507	SEQ ID NO 1565
AK000933	SEQ ID NO 221	NM_015513	SEQ ID NO 1566
AK001100	SEQ ID NO 223	NM_015515	SEQ ID NO 1567
AK001164	SEQ ID NO 224	NM_015523	SEQ ID NO 1568
AK001166	SEQ ID NO 225	NM_015524	SEQ ID NO 1569
AK001295	SEQ ID NO 226	NM_015599	SEQ ID NO 1571
AK001380	SEQ ID NO 227	NM_015623	SEQ ID NO 1572
AK001423	SEQ ID NO 228	NM_015640	SEQ ID NO 1573
AK001438	SEQ ID NO 229	NM_015641	SEQ ID NO 1574
AK001492	SEQ ID NO 230	NM_015678	SEQ ID NO 1575
AK001499	SEQ ID NO 231	NM_015721	SEQ ID NO 1576
AK001630	SEQ ID NO 232	NM_015892	SEQ ID NO 1578
AK001872	SEQ ID NO 234	NM_015895	SEQ ID NO 1579
AK001890	SEQ ID NO 235	NM_015907	SEQ ID NO 1580
AK002016	SEQ ID NO 236	NM_015925	SEQ ID NO 1581
AK002088	SEQ ID NO 237	NM_015937	SEQ ID NO 1582
AK002206	SEQ ID NO 240	NM_015954	SEQ ID NO 1583
AL035297	SEQ ID NO 241	NM_015955	SEQ ID NO 1584
AL049265	SEQ ID NO 242	NM_015961	SEQ ID NO 1585
AL049365	SEQ ID NO 244	NM_015984	SEQ ID NO 1587
AL049370	SEQ ID NO 245	NM_015986	SEQ ID NO 1588
AL049381	SEQ ID NO 246	NM_015987	SEQ ID NO 1589
AL049397	SEQ ID NO 247	NM_015991	SEQ ID NO 1590
AL049415	SEQ ID NO 248	NM_016002	SEQ ID NO 1592
AL049667	SEQ ID NO 249	NM_016028	SEQ ID NO 1594
AL049801	SEQ ID NO 250	NM_016029	SEQ ID NO 1595
AL049932	SEQ ID NO 251	NM_016047	SEQ ID NO 1596
AL049935	SEQ ID NO 252	NM_016048	SEQ ID NO 1597
AL049943	SEQ ID NO 253	NM_016050	SEQ ID NO 1598
AL049949	SEQ ID NO 254	NM_016056	SEQ ID NO 1599
AL049963	SEQ ID NO 255	NM_016058	SEQ ID NO 1600
AL049987	SEQ ID NO 256	NM_016066	SEQ ID NO 1601
AL050021	SEQ ID NO 257	NM_016072	SEQ ID NO 1602
AL050024	SEQ ID NO 258	NM_016073	SEQ ID NO 1603
AL050090	SEQ ID NO 259	NM_016108	SEQ ID NO 1605
AL050148	SEQ ID NO 260	NM_016109	SEQ ID NO 1606
AL050151	SEQ ID NO 261	NM_016121	SEQ ID NO 1607
AL050227	SEQ ID NO 262	NM_016126	SEQ ID NO 1608
AL050367	SEQ ID NO 263	NM_016127	SEQ ID NO 1609
AL050370	SEQ ID NO 264	NM_016135	SEQ ID NO 1610
AL050371	SEQ ID NO 265	NM_016142	SEQ ID NO 1612
AL050372	SEQ ID NO 266	NM_016153	SEQ ID NO 1613
AL050388	SEQ ID NO 267	NM_016171	SEQ ID NO 1614
AL079276	SEQ ID NO 268	NM_016175	SEQ ID NO 1615



AL079298	SEQ ID NO 269	NM_016184	SEQ ID NO 1616
AL080079	SEQ ID NO 271	NM_016185	SEQ ID NO 1617
AL080192	SEQ ID NO 273	NM_016187	SEQ ID NO 1618
AL080199	SEQ ID NO 274	NM_016199	SEQ ID NO 1619
AL080209	SEQ ID NO 275	NM_016210	SEQ ID NO 1620
AL080234	SEQ ID NO 277	NM_016217	SEQ ID NO 1621
AL080235	SEQ ID NO 278	NM_016228	SEQ ID NO 1623
AL096737	SEQ ID NO 279	NM_016229	SEQ ID NO 1624
AL110126	SEQ ID NO 280	NM_016235	SEQ ID NO 1625
AL110139	SEQ ID NO 281	NM_016240	SEQ ID NO 1626
AL110202	SEQ ID NO 283	NM_016243	SEQ ID NO 1627
AL110212	SEQ ID NO 284	NM_016250	SEQ ID NO 1628
AL110260	SEQ ID NO 285	NM_016267	SEQ ID NO 1629
AL117441	SEQ ID NO 286	NM_016271	SEQ ID NO 1630
AL117452	SEQ ID NO 287	NM_016299	SEQ ID NO 1631
AL117477	SEQ ID NO 288	NM_016306	SEQ ID NO 1632
AL117502	SEQ ID NO 289	NM_016308	SEQ ID NO 1634
AL117523	SEQ ID NO 290	NM_016321	SEQ ID NO 1635
AL117595	SEQ ID NO 291	NM_016337	SEQ ID NO 1636
AL117599	SEQ ID NO 292	NM_016352	SEQ ID NO 1637
AL117600	SEQ ID NO 293	NM_016359	SEQ ID NO 1638
AL117609	SEQ ID NO 294	NM_016401	SEQ ID NO 1641
AL117617	SEQ ID NO 295	NM_016403	SEQ ID NO 1642
AL117666	SEQ ID NO 296	NM_016411	SEQ ID NO 1643
AL122055	SEQ ID NO 297	NM_016423	SEQ ID NO 1644
AL133033	SEQ ID NO 298	NM_016463	SEQ ID NO 1647
AL133035	SEQ ID NO 299	NM_016475	SEQ ID NO 1649
AL133074	SEQ ID NO 301	NM_016477	SEQ ID NO 1650
AL133096	SEQ ID NO 302	NM_016491	SEQ ID NO 1651
AL133105	SEQ ID NO 303	NM_016495	SEQ ID NO 1652
AL133108	SEQ ID NO 304	NM_016542	SEQ ID NO 1653
AL133572	SEQ ID NO 305	NM_016548	SEQ ID NO 1654
AL133619	SEQ ID NO 307	NM_016569	SEQ ID NO 1655
AL133622	SEQ ID NO 308	NM_016577	SEQ ID NO 1656
AL133623	SEQ ID NO 309	NM_016582	SEQ ID NO 1657
AL133624	SEQ ID NO 310	NM_016593	SEQ ID NO 1658
AL133632	SEQ ID NO 311	NM_016603	SEQ ID NO 1659
AL133644	SEQ ID NO 312	NM_016612	SEQ ID NO 1660
AL133645	SEQ ID NO 313	NM_016619	SEQ ID NO 1661
AL133651	SEQ ID NO 314	NM_016623	SEQ ID NO 1663
AL137310	SEQ ID NO 316	NM_016625	SEQ ID NO 1664
AL137316	SEQ ID NO 317	NM_016629	SEQ ID NO 1665
AL137332	SEQ ID NO 318	NM_016640	SEQ ID NO 1666
AL137342	SEQ ID NO 319	NM_016645	SEQ ID NO 1667
AL137362	SEQ ID NO 321	NM_016650	SEQ ID NO 1668
AL137381	SEQ ID NO 322	NM_016657	SEQ ID NO 1669
AL137407	SEQ ID NO 323	NM_016733	SEQ ID NO 1670

AL137448	SEQ ID NO 324	NM_016815	SEQ ID NO 1671
AL137502	SEQ ID NO 326	NM_016817	SEQ ID NO 1672
AL137514	SEQ ID NO 327	NM_016818	SEQ ID NO 1673
AL137540	SEQ ID NO 328	NM_016839	SEQ ID NO 1675
AL137566	SEQ ID NO 330	NM_017414	SEQ ID NO 1676
AL137615	SEQ ID NO 331	NM_017422	SEQ ID NO 1677
AL137673	SEQ ID NO 335	NM_017423	SEQ ID NO 1678
AL137718	SEQ ID NO 336	NM_017447	SEQ ID NO 1679
AL137736	SEQ ID NO 337	NM_017518	SEQ ID NO 1680
AL137751	SEQ ID NO 338	NM_017522	SEQ ID NO 1681
AL137761	SEQ ID NO 339	NM_017540	SEQ ID NO 1682
AL157431	SEQ ID NO 340	NM_017555	SEQ ID NO 1683
AL157432	SEQ ID NO 341	NM_017572	SEQ ID NO 1684
AL157454	SEQ ID NO 342	NM_017585	SEQ ID NO 1685
AL157476	SEQ ID NO 343	NM_017586	SEQ ID NO 1686
AL157480	SEQ ID NO 344	NM_017596	SEQ ID NO 1687
AL157482	SEQ ID NO 345	NM_017606	SEQ ID NO 1688
AL157484	SEQ ID NO 346	NM_017617	SEQ ID NO 1689
AL157492	SEQ ID NO 347	NM_017633	SEQ ID NO 1690
AL157505	SEQ ID NO 348	NM_017634	SEQ ID NO 1691
AL157851	SEQ ID NO 349	NM_017646	SEQ ID NO 1692
AL160131	SEQ ID NO 350	NM_017660	SEQ ID NO 1693
AL161960	SEQ ID NO 351	NM_017680	SEQ ID NO 1694
AL162049	SEQ ID NO 352	NM_017691	SEQ ID NO 1695
AL355708	SEQ ID NO 353	NM_017698	SEQ ID NO 1696
D13643	SEQ ID NO 355	NM_017702	SEQ ID NO 1697
D14678	SEQ ID NO 356	NM_017731	SEQ ID NO 1699
D25328	SEQ ID NO 357	NM_017732	SEQ ID NO 1700
D26070	SEQ ID NO 358	NM_017733	SEQ ID NO 1701
D26488	SEQ ID NO 359	NM_017734	SEQ ID NO 1702
D31887	SEQ ID NO 360	NM_017746	SEQ ID NO 1703
D38521	SEQ ID NO 361	NM_017750	SEQ ID NO 1704
D38553	SEQ ID NO 362	NM_017761	SEQ ID NO 1705
D42043	SEQ ID NO 363	NM_017763	SEQ ID NO 1706
D42047	SEQ ID NO 364	NM_017770	SEQ ID NO 1707
D43950	SEQ ID NO 365	NM_017779	SEQ ID NO 1708
D50402	SEQ ID NO 366	NM_017780	SEQ ID NO 1709
D50914	SEQ ID NO 367	NM_017782	SEQ ID NO 1710
D55716	SEQ ID NO 368	NM_017786	SEQ ID NO 1711
D80001	SEQ ID NO 369	NM_017791	SEQ ID NO 1712
D80010	SEQ ID NO 370	NM_017805	SEQ ID NO 1713
D82345	SEQ ID NO 371	NM_017816	SEQ ID NO 1714
D83781	SEQ ID NO 372	NM_017821	SEQ ID NO 1715
D86964	SEQ ID NO 373	NM_017835	SEQ ID NO 1716
D86978	SEQ ID NO 374	NM_017843	SEQ ID NO 1717
D86985	SEQ ID NO 375	NM_017857	SEQ ID NO 1718
D87076	SEQ ID NO 376	NM_017901	SEQ ID NO 1719

# ES 2 366 178 T3

D87453	SEQ ID NO 377	NM_017906	SEQ ID NO 1720
D87469	SEQ ID NO 378	NM_017918	SEQ ID NO 1721
D87682	SEQ ID NO 379	NM_017961	SEQ ID NO 1722
G26403	SEQ ID NO 380	NM_017996	SEQ ID NO 1723
J02639	SEQ ID NO 381	NM_018000	SEQ ID NO 1724
J04162	SEQ ID NO 382	NM_018004	SEQ ID NO 1725
K02403	SEQ ID NO 384	NM_018011	SEQ ID NO 1726
L05096	SEQ ID NO 385	NM_018014	SEQ ID NO 1727
L10333	SEQ ID NO 386	NM_018022	SEQ ID NO 1728
L11645	SEQ ID NO 387	NM_018031	SEQ ID NO 1729
L21934	SEQ ID NO 388	NM_018043	SEQ ID NO 1730
L22005	SEQ ID NO 389	NM_018048	SEQ ID NO 1731
L48692	SEQ ID NO 391	NM_018062	SEQ ID NO 1732
M12758	SEQ ID NO 392	NM_018069	SEQ ID NO 1733
M15178	SEQ ID NO 393	NM_018072	SEQ ID NO 1734
M21551	SEQ ID NO 394	NM_018077	SEQ ID NO 1735
M24895	SEQ ID NO 395	NM_018086	SEQ ID NO 1736
M26383	SEQ ID NO 396	NM_018087	SEQ ID NO 1737
M27749	SEQ ID NO 397	NM_018093	SEQ ID NO 1738
M28170	SEQ ID NO 398	NM_018098	SEQ ID NO 1739
M29873	SEQ ID NO 399	NM_018099	SEQ ID NO 1740
M29874	SEQ ID NO 400	NM_018101	SEQ ID NO 1741
M30448	SEQ ID NO 401	NM_018103	SEQ ID NO 1742
M30818	SEQ ID NO 402	NM_018109	SEQ ID NO 1744
M31932	SEQ ID NO 403	NM_018123	SEQ ID NO 1746
M37033	SEQ ID NO 404	NM_018131	SEQ ID NO 1747
M55914	SEQ ID NO 405	NM_018136	SEQ ID NO 1748
M63438	SEQ ID NO 406	NM_018138	SEQ ID NO 1749
M65254	SEQ ID NO 407	NM_018166	SEQ ID NO 1750
M68874	SEQ ID NO 408	NM_018171	SEQ ID NO 1751
M73547	SEQ ID NO 409	NM_018178	SEQ ID NO 1752
M77142	SEQ ID NO 410	NM_018181	SEQ ID NO 1753
M80899	SEQ ID NO 411	NM_018186	SEQ ID NO 1754
M90657	SEQ ID NO 413	NM_018194	SEQ ID NO 1757
M93718	SEQ ID NO 414	NM_018204	SEQ ID NO 1758
M96577	SEQ ID NO 415	NM_018208	SEQ ID NO 1759
NM_000022	SEQ ID NO 417	NM_018212	SEQ ID NO 1760
NM_000044	SEQ ID NO 418	NM_018234	SEQ ID NO 1763
NM_000050	SEQ ID NO 419	NM_018255	SEQ ID NO 1764
NM_000057	SEQ ID NO 420	NM_018257	SEQ ID NO 1765
NM_000060	SEQ ID NO 421	NM_018265	SEQ ID NO 1766
NM_000064	SEQ ID NO 422	NM_018271	SEQ ID NO 1767
NM_000073	SEQ ID NO 424	NM_018290	SEQ ID NO 1768
NM_000077	SEQ ID NO 425	NM_018295	SEQ ID NO 1769
NM_000086	SEQ ID NO 426	NM_018304	SEQ ID NO 1770
NM_000087	SEQ ID NO 427	NM_018306	SEQ ID NO 1771
NM_000095	SEQ ID NO 429	NM_018326	SEQ ID NO 1772

NM_000096	SEQ ID NO 430	NM_018346	SEQ ID NO 1773
NM_000100	SEQ ID NO 431	NM_018366	SEQ ID NO 1775
NM_000101	SEQ ID NO 432	NM_018370	SEQ ID NO 1776
NM_000104	SEQ ID NO 433	NM_018373	SEQ ID NO 1777
NM_000109	SEQ ID NO 434	NM_018379	SEQ ID NO 1778
NM_000125	SEQ ID NO 435	NM_018384	SEQ ID NO 1779
NM_000127	SEQ ID NO 436	NM_018389	SEQ ID NO 1780
NM_000135	SEQ ID NO 437	NM_018410	SEQ ID NO 1783
NM_000137	SEQ ID NO 438	NM_018439	SEQ ID NO 1785
NM_000146	SEQ ID NO 439	NM_018454	SEQ ID NO 1786
NM_000149	SEQ ID NO 440	NM_018455	SEQ ID NO 1787
NM_000154	SEQ ID NO 441	NM_018465	SEQ ID NO 9788
NM_000161	SEQ ID NO 443	NM_018471	SEQ ID NO 1789
NM_000165	SEQ ID NO 444	NM_018478	SEQ ID NO 1790
NM_000168	SEQ ID NO 445	NM_018479	SEQ ID NO 1791
NM_000169	SEQ ID NO 446	NM_018529	SEQ ID NO 1793
NM_000175	SEQ ID NO 447	NM_018556	SEQ ID NO 1794
NM_000191	SEQ ID NO 448	NM_018569	SEQ ID NO 1795
NM_000201	SEQ ID NO 450	NM_018584	SEQ ID NO 1796
NM_000211	SEQ ID NO 451	NM_018653	SEQ ID NO 1797
NM_000213	SEQ ID NO 452	NM_018660	SEQ ID NO 1798
NM_000224	SEQ ID NO 453	NM_018683	SEQ ID NO 1799
NM_000239	SEQ ID NO 454	NM_018685	SEQ ID NO 1800
NM_000251	SEQ ID NO 455	NM_018686	SEQ ID NO 1801
NM_000268	SEQ ID NO 456	NM_018695	SEQ ID NO 1802
NM_000270	SEQ ID NO 458	NM_018728	SEQ ID NO 1803
NM_000271	SEQ ID NO 459	NM_018840	SEQ ID NO 1804
NM_000283	SEQ ID NO 460	NM_018842	SEQ ID NO 1805
NM_000284	SEQ ID NO 461	NM_018950	SEQ ID NO 1806
NM_000286	SEQ ID NO 462	NM_018988	SEQ ID NO 1807
NM_000291	SEQ ID NO 463	NM_019000	SEQ ID NO 1808
NM_000299	SEQ ID NO 464	NM_019013	SEQ ID NO 1809
NM_000300	SEQ ID NO 465	NM_019025	SEQ ID NO 1810
NM_000310	SEQ ID NO 466	NM_019027	SEQ ID NO 1811
NM_000311	SEQ ID NO 467	NM_019041	SEQ ID NO 1812
NM_000317	SEQ ID NO 468	NM_019044	SEQ ID NO 1813
NM_000320	SEQ ID NO 469	NM_019063	SEQ ID NO 1815
NM_000342	SEQ ID NO 470	NM_019084	SEQ ID NO 1816
NM_000346	SEQ ID NO 471	NM_019554	SEQ ID NO 1817
NM_000352	SEQ ID NO 472	NM_019845	SEQ ID NO 1818
NM_000355	SEQ ID NO 473	NM_019858	SEQ ID NO 1819
NM_000358	SEQ ID NO 474	NM_020130	SEQ ID NO 1820
NM_000359	SEQ ID NO 475	NM_020133	SEQ ID NO 1821
NM_000362	SEQ ID NO 476	NM_020143	SEQ ID NO 1822
NM_000365	SEQ ID NO 477	NM_020150	SEQ ID NO 1823
NM_000381	SEQ ID NO 478	NM_020163	SEQ ID NO 1824
NM_000397	SEQ ID NO 480	NM_020166	SEQ ID NO 1825

NM_000399	SEQ ID NO 481	NM_020169	SEQ ID NO 1826
NM_000414	SEQ ID NO 482	NM_020179	SEQ ID NO 1827
NM_000416	SEQ ID NO 483	NM_020184	SEQ ID NO 1828
NM_000422	SEQ ID NO 484	NM_020186	SEQ ID NO 1829
NM_000424	SEQ ID NO 485	NM_020188	SEQ ID NO 1830
NM_000433	SEQ ID NO 486	NM_020189	SEQ ID NO 1831
NM_000436	SEQ ID NO 487	NM_020197	SEQ ID NO 1832
NM_000450	SEQ ID NO 488	NM_020199	SEQ ID NO 1833
NM_000462	SEQ ID NO 489	NM_020215	SEQ ID NO 1834
NM_000495	SEQ ID NO 490	NM_020347	SEQ ID NO 1836
NM_000507	SEQ ID NO 491	NM_020365	SEQ ID NO 1837
NM_000526	SEQ ID NO 492	NM_020386	SEQ ID NO 1838
NM_000557	SEQ ID NO 493	NM_020445	SEQ ID NO 1839
NM_000560	SEQ ID NO 494	NM_020639	SEQ ID NO 1840
NM_000576	SEQ ID NO 495	NM_020659	SEQ ID NO 1841
NM_000579	SEQ ID NO 496	NM_020675	SEQ ID NO 1842
NM_000584	SEQ ID NO 497	NM_020686	SEQ ID NO 1843
NM_000591	SEQ ID NO 498	NM_020974	SEQ ID NO 1844
NM_000592	SEQ ID NO 499	NM_020978	SEQ ID NO 1845
NM_000593	SEQ ID NO 500	NM_020979	SEQ ID NO 1846
NM_000594	SEQ ID NO 501	NM_020980	SEQ ID NO 1847
NM_000597	SEQ ID NO 502	NM_021000	SEQ ID NO 1849
NM_000600	SEQ ID NO 504	NM_021004	SEQ ID NO 1850
NM_000607	SEQ ID NO 505	NM_021025	SEQ ID NO 1851
NM_000612	SEQ ID NO 506	NM_021063	SEQ ID NO 1852
NM_000627	SEQ ID NO 507	NM_021065	SEQ ID NO 1853
NM_000633	SEQ ID NO 508	NM_021077	SEQ ID NO 1854
NM_000636	SEQ ID NO 509	NM_021095	SEQ ID NO 1855
NM_000639	SEQ ID NO 510	NM_021101	SEQ ID NO 1856
NM_000647	SEQ ID NO 511	NM_021103	SEQ ID NO 1857
NM_000655	SEQ ID NO 512	NM_021128	SEQ ID NO 1858
NM_000662	SEQ ID NO 513	NM_021147	SEQ ID NO 1859
NM_000663	SEQ ID NO 514	NM_021151	SEQ ID NO 1860
NM_000666	SEQ ID NO 515	NM_021181	SEQ ID NO 1861
NM_000676	SEQ ID NO 516	NM_021190	SEQ ID NO 1862
NM_000685	SEQ ID NO 517	NM_021198	SEQ ID NO 1863
NM_000693	SEQ ID NO 518	NM_021200	SEQ ID NO 1864
NM_000699	SEQ ID NO 519	NM_021203	SEQ ID NO 1865
NM_000700	SEQ ID NO 520	NM_021238	SEQ ID NO 1866
NM_000712	SEQ ID NO 521	NM_021242	SEQ ID NO 1867
NM_000727	SEQ ID NO 522	S40706	SEQ ID NO 1869
NM_000732	SEQ ID NO 523	S53354	SEQ ID NO 1870
NM_000734	SEQ ID NO 524	S59184	SEQ ID NO 1871
NM_000767	SEQ ID NO 525	S62138	SEQ ID NO 1872
NM_000784	SEQ ID NO 526	U09848	SEQ ID NO 1873
NM_000802	SEQ ID NO 528	U10991	SEQ ID NO 1874
NM_000824	SEQ ID NO 529	U17077	SEQ ID NO 1875

NM_000849	SEQ ID NO 530	U18919	SEQ ID NO 1876
NM_000852	SEQ ID NO 531	U41387	SEQ ID NO 1877
NM_000874	SEQ ID NO 532	U45975	SEQ ID NO 1878
NM_000878	SEQ ID NO 533	U49835	SEQ ID NO 1879
NM_000884	SEQ ID NO 534	U56725	SEQ ID NO 1880
NM_000908	SEQ ID NO 537	U58033	SEQ ID NO 1881
NM_000909	SEQ ID NO 538	U61167	SEQ ID NO 1882
NM_000926	SEQ ID NO 539	U66042	SEQ ID NO 1883
NM_000930	SEQ ID NO 540	U68385	SEQ ID NO 1885
NM_000931	SEQ ID NO 541	U68494	SEQ ID NO 1886
NM_000947	SEQ ID NO 542	U74612	SEQ ID NO 1887
NM_000949	SEQ ID NO 543	U75968	SEQ ID NO 1888
NM_000950	SEQ ID NO 544	U79293	SEQ ID NO 1889
NM_000954	SEQ ID NO 545	U80736	SEQ ID NO 1890
NM_000964	SEQ ID NO 546	U82987	SEQ ID NO 1891
3 NM_001003	SEQ ID NO 549	U83115	SEQ ID NO 1892
NM_001016	SEQ ID NO 551	U89715	SEQ ID NO 1893
NM_001047	SEQ ID NO 553	U90916	SEQ ID NO 1894
NM_001066	SEQ ID NO 555	U92544	SEQ ID NO 1895
NM_001071	SEQ ID NO 556	U96131	SEQ ID NO 1896
NM_001078	SEQ ID NO 557	U96394	SEQ ID NO 1897
NM_001085	SEQ ID NO 558	W61000_RC	SEQ ID NO 1898
NM_001089	SEQ ID NO 559	X00437	SEQ ID NO 1899
NM_001109	SEQ ID NO 560	X00497	SEQ ID NO 1900
NM_001122	SEQ ID NO 561	X01394	SEQ ID NO 1901
NM_001124	SEQ ID NO 562	X03084	SEQ ID NO 1902
NM_001161	SEQ ID NO 563	X07834	SEQ ID NO 1905
NM_001165	SEQ ID NO 564	X14356	SEQ ID NO 1906
NM_001166	SEQ ID NO 565	X16302	SEQ ID NO 1907
NM_001168	SEQ ID NO 566	X52486	SEQ ID NO 1909
NM_001179	SEQ ID NO 567	X52882	SEQ ID NO 1910
NM_001185	SEQ ID NO 569	X56807	SEQ ID NO 1911
NM_001203	SEQ ID NO 570	X57809	SEQ ID NO 1912
NM_001207	SEQ ID NO 573	X57819	SEQ ID NO 1913
NM_001216	SEQ ID NO 574	X58529	SEQ ID NO 1914
NM_001218	SEQ ID NO 575	X59405	SEQ ID NO 1915
NM_001223	SEQ ID NO 576	X72475	SEQ ID NO 1918
NM_001225	SEQ ID NO 577	X73617	SEQ ID NO 1919
NM_001233	SEQ ID NO 578	X74794	SEQ ID NO 1920
NM_001236	SEQ ID NO 579	X75315	SEQ ID NO 1921
NM_001237	SEQ ID NO 580	X79782	SEQ ID NO 1922
NM_001251	SEQ ID NO 581	X82693	SEQ ID NO 1923
NM_001255	SEQ ID NO 582	X83301	SEQ ID NO 1924
NM_001262	SEQ ID NO 583	X93006	SEQ ID NO 1926
NM_001263	SEQ ID NO 584	X94232	SEQ ID NO 1927
NM_001267	SEQ ID NO 585	X98834	SEQ ID NO 1929
NM_001276	SEQ ID NO 587	X99142	SEQ ID NO 1930

NM_001280	SEQ ID NO 588	Y14737	SEQ ID NO 1932
NM_001282	SEQ ID NO 589	Z11887	SEQ ID NO 1933
NM_001295	SEQ ID NO 590	Z48633	SEQ ID NO 1935
NM_001305	SEQ ID NO 591	NM_004222	SEQ ID NO 1936
NM_001310	SEQ ID NO 592	NM_016405	SEQ ID NO 1937
NM_001312	SEQ ID NO 593	NM_017690	SEQ ID NO 1938
NM_001321	SEQ ID NO 594	Contig29_RC	SEQ ID NO 1939
NM_001327	SEQ ID NO 595	Contig237_RC	SEQ ID NO 1940
NM_001329	SEQ ID NO 596	Contig263_RC	SEQ ID NO 1941
NM_001333	SEQ ID NO 597	Contig292_RC	SEQ ID NO 1942
NM_001338	SEQ ID NO 598	Contig382_RC	SEQ ID NO 1944
NM_001360	SEQ ID NO 599	Contig399_RC	SEQ ID NO 1945
NM_001363	SEQ ID NO 600	Contig448_RC	SEQ ID NO 1946
NM_001381	SEQ ID NO 601	Contig569_RC	SEQ ID NO 1947
NM_001394	SEQ ID NO 602	Contig580_RC	SEQ ID NO 1948
NM_001395	SEQ ID NO 603	Contig678_RC	SEQ ID NO 1949
NM_001419	SEQ ID NO 604	Contig706_RC	SEQ ID NO 1950
NM_001424	SEQ ID NO 605	Contig718_RC	SEQ ID NO 1951
NM_001428	SEQ ID NO 606	Contig719_RC	SEQ ID NO 1952
NM_001436	SEQ ID NO 607	Contig742_RC	SEQ ID NO 1953
NM_001444	SEQ ID NO 608	Contig753_RC	SEQ ID NO 1954
NM_001446	SEQ ID NO 609	Contig758_RC	SEQ ID NO 1956
NM_001453	SEQ ID NO 611	Contig760_RC	SEQ ID NO 1957
NM_001456	SEQ ID NO 612	Contig842_RC	SEQ ID NO 1958
NM_001457	SEQ ID NO 613	Contig848_RC	SEQ ID NO 1959
NM_001463	SEQ ID NO 614	Contig924_RC	SEQ ID NO 1960
NM_001465	SEQ ID NO 615	Contig974_RC	SEQ ID NO 1961
NM_001481	SEQ ID NO 616	Contig1018_RC	SEQ ID NO 1962
NM_001493	SEQ ID NO 617	Contig1056_RC	SEQ ID NO 1963
NM_001494	SEQ ID NO 618	Contig1061_RC	SEQ ID NO 1964
NM_001500	SEQ ID NO 619	Contig1129_RC	SEQ ID NO 1965
NM_001504	SEQ ID NO 620	Contig1148	SEQ ID NO 1966
NM_001511	SEQ ID NO 621	Contig1239_RC	SEQ ID NO 1967
NM_001513	SEQ ID NO 622	Contig1277	SEQ ID NO 1968
NM_001527	SEQ ID NO 623	Contig1333_RC	SEQ ID NO 1969
NM_001529	SEQ ID NO 624	Contig1386_RC	SEQ ID NO 1970
NM_001530	SEQ ID NO 625	Contig1389_RC	SEQ ID NO 1971
NM_001540	SEQ ID NO 626	Contig1418_RC	SEQ ID NO 1972
NM_001550	SEQ ID NO 627	Contig1462_RC	SEQ ID NO 1973
NM_001551	SEQ ID NO 628	Contig1505_RC	SEQ ID NO 1974
NM_001552	SEQ ID NO 629	Contig1540_RC	SEQ ID NO 1975
NM_001554	SEQ ID NO 631	Contig1584_RC	SEQ ID NO 1976
NM_001558	SEQ ID NO 632	Contig1632_RC	SEQ ID NO 1977
NM_001560	SEQ ID NO 633	Contig1682_RC	SEQ ID NO 1978
NM_001565	SEQ ID NO 634	Contig1778_RC	SEQ ID NO 1979
NM_001569	SEQ ID, NO 635	Contig1829	SEQ ID NO 1981
NM_001605	SEQ ID NO 636	Contig1838_RC	SEQ ID NO 1982

NM_001609	SEQ ID NO 637	Contig1938_RC	SEQ ID NO 1983
NM_001615	SEQ ID NO 638	Contig1970_RC	SEQ ID NO 1984
NM_001623	SEQ ID NO 639	Contig1998_RC	SEQ ID NO 1985
NM_001627	SEQ ID NO 640	Contig2099_RC	SEQ ID NO 1986
NM_001628	SEQ ID NO 641	Contig2143_RC	SEQ ID NO 1987
NM_001630	SEQ ID NO 642	Contig2237_RC	SEQ ID NO 1988
NM_001634	SEQ ID NO 643	Contig2429_RC	SEQ ID NO 1990
NM_001656	SEQ ID NO 644	Contig2504_RC	SEQ ID NO 1991
NM_001673	SEQ ID NO 645	Contig2512_RC	SEQ ID NO 1992
NM_001675	SEQ ID NO 647	Contig2575_RC	SEQ ID NO 1993
NM_001679	SEQ ID NO 648	Contig2578_RC	SEQ ID NO 1994
NM_001689	SEQ ID NO 649	Contig2639_RC	SEQ ID NO 1995
NM_001703	SEQ ID NO 650	Contig2647_RC	SEQ ID NO 1996
NM_001710	SEQ ID NO 651	Contig2657_RC	SEQ ID NO 1997
NM_001725	SEQ ID NO 652	Contig2728_RC	SEQ ID NO 1998
NM_001730	SEQ ID NO 653	Contig2745_RC	SEQ ID NO 1999
NM_001733	SEQ ID NO 654	Contig2811_RC	SEQ ID NO 2000
NM_001734	SEQ ID NO 655	Contig2873_RC	SEQ ID NO 2001
NM_001740	SEQ ID NO 656	Contig2883_RC	SEQ ID NO 2002
NM_001745	SEQ ID NO 657	Contig2915_RC	SEQ ID NO 2003
NM_001747	SEQ ID NO 658	Contig2928_RC	SEQ ID NO 2004
NM_001756	SEQ ID NO 659	Contig3024_RC	SEQ ID NO 2005
NM_001757	SEQ ID NO 660	Contig3094_RC	SEQ ID NO 2006
NM_001758	SEQ ID NO 661	Contig3164_RC	SEQ ID NO 2007
NM_001762	SEQ ID NO 662	Contig3495_RC	SEQ ID NO 2009
NM_001767	SEQ ID NO 663	Contig3607_RC	SEQ ID NO 2010
NM_001770	SEQ ID NO 664	Contig3659_RC	SEQ ID NO 2011
NM_001777	SEQ ID NO 665	Contig3677_RC	SEQ ID NO 2012
NM_001778	SEQ ID NO 666	Contig3682_RC	SEQ ID NO 2013
NM_001781	SEQ ID NO 667	Contig3734_RC	SEQ ID NO 2014
NM_001786	SEQ ID NO 668	Contig3834_RC	SEQ ID NO 2015
NM_001793	SEQ ID NO 669	Contig3876_RC	SEQ ID NO 2016
NM_001803	SEQ ID NO 671	Contig3902_RC	SEQ ID NO 2017
NM_001806	SEQ ID NO 672	Contig3940_RC	SEQ ID NO 2018
NM_001809	SEQ ID NO 673	Contig4380_RC	SEQ ID NO 2019
NM_001814	SEQ ID NO 674	Contig4388_RC	SEQ ID NO 2020
NM_001826	SEQ ID NO 675	Contig4467_RC	SEQ ID NO 2021
NM_001830	SEQ ID NO 677	Contig4949_RC	SEQ ID NO 2023
NM_001838	SEQ ID NO 678	Contig5348_RC	SEQ ID NO 2024
NM_001839	SEQ ID NO 679	Contig5403_RC	SEQ ID NO 2025
NM_001853	SEQ ID NO 681	Contig5716_RC	SEQ ID NO 2026
NM_001859	SEQ ID NO 682	Contig6118_RC	SEQ ID NO 2027
NM_001861	SEQ ID NO 683	Contig6164_RC	SEQ ID NO 2028
NM_001874	SEQ ID NO 685	Contig6181_RC	SEQ ID NO 2029
NM_001885	SEQ ID NO 686	Contig6514_RC	SEQ ID NO 2030
NM_001892	SEQ ID NO 688	Contig6612_RC	SEQ ID NO 2031
NM_001897	SEQ ID NO 689	Contig6881_RC	SEQ ID NO 2032



NM_001899	SEQ ID NO 690	Contig8165_RC	SEQ ID NO 2033
NM_001905	SEQ ID NO 691	Contig8221_RC	SEQ ID NO 2034
NM_001912	SEQ ID NO 692	Contig8347_RC	SEQ ID NO 2035
NM_001914	SEQ ID NO 693	Contig8364_RC	SEQ ID NO 2036
NM_001919	SEQ ID NO 694	Contig8888_RC	SEQ ID NO 2038
NM_001941	SEQ ID NO 695	Contig9259_RC	SEQ ID NO 2039
NM_001943	SEQ ID NO 696	Contig9541_RC	SEQ ID NO 2040
NM_001944	SEQ ID NO 697	Contig10268_RC	SEQ ID NO 2041
NM_001953	SEQ ID NO 699	Contig10363_RC	SEQ ID NO 2042
NM_001954	SEQ ID NO 700	Contig10437_RC	SEQ ID NO 2043
NMB_001955	SEQ ID NO 701	Contig11086_RC	SEQ ID NO 2045
NMB_001956	SEQ ID NO 702	Contig11275_RC	SEQ ID NO 2046
NM_001958	SEQ ID NO 703	Contig11648_RC	SEQ ID NO 2047
NMB_001961	SEQ ID NO 705	Contig12216_RC	SEQ ID NO 2048
NMB_001970	SEQ ID NO 706	Contig12369_RC	SEQ ID NO 2049
NMB_001979	SEQ ID NO 707	Contig12814_RC	SEQ ID NO 2050
NM_001982	SEQ ID NO 708	Contig12951_RC	SEQ ID NO 2051
NMB_002017	SEQ ID NO 710	Contig13480_RC	SEQ ID NO 2052
NM_002033	SEQ ID NO 713	Contig14284_RC	SEQ ID NO 2053
NM_002046	SEQ ID NO 714	Contig14390_RC	SEQ ID NO 2054
NM_002047	SEQ ID NO 715	Contig14780_RC	SEQ ID NO 2055
NM_002051	SEQ ID NO 716	Contig14954_RC	SEQ ID NO 2056
NM_002053	SEQ ID NO 717	Contig14981_RC	SEQ ID NO 2057
NM_002061	SEQ ID NO 718	Contig15692_RC	SEQ ID NO 2058
NM_002065	SEQ ID NO 719	Contig16192_RC	SEQ ID NO 2059
NM_002068	SEQ ID NO 720	Contig16759_RC	SEQ ID NO 2061
NM_002077	SEQ ID NO 722	Contig16786_RC	SEQ ID NO 2062
NM_002091	SEQ ID NO 723	Contig16905_RC	SEQ ID NO 2063
NMB_002101	SEQ ID NO 724	Contig17103_RC	SEQ ID NO 2064
NM_002106	SEQ ID NO 725	Contig17105_RC	SEQ ID NO 2065
NMB_002110	SEQ ID NO 726	Contig17248_RC	SEQ ID NO 2066
NM_002111	SEQ ID NO 727	Contig17345_RC	SEQ ID NO 2067
NMB_002115	SEQ ID NO 728	Contig18502_RC	SEQ ID NO 2069
NMB_002118	SEQ ID NO 729	Contig20156_RC	SEQ ID NO 2071
NMB_002123	SEQ ID NO 730	Contig20302_RC	SEQ ID NO 2073
NMB_002131	SEQ ID NO 731	Contig20600_RC	SEQ ID NO 2074
NMB_002136	SEQ ID NO 732	Contig20617_RC	SEQ ID NO 2075
NM_002145	SEQ ID NO 733	Contig20629_RC	SEQ ID NO 2076
NMB_002164	SEQ ID NO 734	Contig20651_RC	SEQ ID NO 2077
NMB_002168	SEQ ID NO 735	Contig21130_RC	SEQ ID NO 2078
NM_002184	SEQ ID NO 736	Contig21185_RC	SEQ ID NO 2079
NM_002185	SEQ ID NO 737	Contig21421_RC	SEQ ID NO 2080
NM_002189	SEQ ID NO 738	Contig21787_RC	SEQ ID NO 2081
NM_002200	SEQ ID NO 739	Contig21812_RC	SEQ ID NO 2082
NM_002201	SEQ ID NO 740	Contig22418_RC	SEQ ID NO 2083
NM_002213	SEQ ID NO 741	Contig23085_RC	SEQ ID NO 2084
NMB_002219	SEQ ID NO 742	Contig23454_RC	SEQ ID NO 2085

NM_002222	SEQ ID NO 743	Contig24138_RC	SEQ ID NO 2086
NM_002239	SEQ ID NO 744	Contig24252_RC	SEQ ID NO 2087
NM_002243	SEQ ID NO 745	Contig24655_RC	SEQ ID NO 2089
NM_002245	SEQ ID NO 746	Contig25055_RC	SEQ ID NO 2090
NM_002250	SEQ ID NO 747	Contig25290_RC	SEQ ID NO 2091
NM_002254	SEQ ID NO 748	Contig25343_RC	SEQ ID NO 2092
NM_002266	SEQ ID NO 749	Contig25362_RC	SEQ ID NO 2093
NM_002273	SEQ ID NO 750	Contig25617_RC	SEQ ID NO 2094
NM_002281	SEQ ID NO 751	Contig25659_RC	SEQ ID NO 2095
NM_002292	SEQ ID NO 752	Contig25722_RC	SEQ ID NO 2096
NM_002298	SEQ ID NO 753	Contig25809_RC	SEQ ID NO 2097
NM_002300	SEQ ID NO 754	Contig25991	SEQ ID NO 2098
NM_002308	SEQ ID NO 755	Contig26022_RC	SEQ ID NO 2099
NMB_002314	SEQ ID NO 756	Contig26077_RC	SEQ ID NO 2100
NM_002337	SEQ ID NO 757	Contig26310_RC	SEQ ID NO 2101
NM_002341	SEQ ID NO 758	Contig26371_RC	SEQ ID NO 2102
NM_002342	SEQ ID NO 759	Contig26438_RC	SEQ ID NO 2103
NM_002346	SEQ ID NO 760	Contig26706_RC	SEQ ID NO 2104
NM_002349	SEQ ID NO 761	Contig27088_RC	SEQ ID NO 2105
NM_002350	SEQ ID NO 762	Contig27186_RC	SEQ ID NO 2106
NM_002356	SEQ ID NO 763	Contig27228_RC	SEQ ID NO 2107
NM_002358	SEQ ID NO 764	Contig27344_RC	SEQ ID NO 2109
NM_002370	SEQ ID NO 765	Contig27386_RC	SEQ ID NO 2110
NM_002395	SEQ ID NO 766	Contig27624_RC	SEQ ID NO 2111
NMB_002416	SEQ ID NO 767	Contig27749_RC	SEQ ID NO 2112
NM_002421	SEQ ID NO 768	Contig27882_RC	SEQ ID NO 2113
NM_002426	SEQ ID NO 769	Contig27915_RC	SEQ ID NO 2114
NM_002435	SEQ ID NO 770	Contig28030_RC	SEQ ID NO 2115
NM_002438	SEQ ID NO 771	Contig28081_RC	SEQ ID NO 2116
NM_002444	SEQ ID NO 772	Contig28152_RC	SEQ ID NO 2117
NM_002449	SEQ ID NO 773	Contig28550_RC	SEQ ID NO 2119
NM_002450	SEQ ID NO 774	Contig28552_RC	SEQ ID NO 2120
NM_002456	SEQ ID NO 775	Contig28712_RC	SEQ ID NO 2121
NM_002466	SEQ ID NO 776	Contig28888_RC	SEQ ID NO 2122
NM_002482	SEQ ID NO 777	Contig28947_RC	SEQ ID NO 2123
NM_002497	SEQ ID NO 778	Contig29126_RC	SEQ ID NO 2124
NM_002510	SEQ ID NO 779	Contig29193_RC	SEQ ID NO 2125
NMB_002515	SEQ ID NO 781	Contig29369_RC	SEQ ID NO 2126
NM_002524	SEQ ID NO 782	Contig29639_RC	SEQ ID NO 2127
NM_002539	SEQ ID NO 783	Contig30047_RC	SEQ ID NO 2129
NM_002555	SEQ ID NO 785	Contig30154_RC	SEQ ID NO 2131
NM_002570	SEQ ID NO 787	Contig30209_RC	SEQ ID NO 2132
NM_002579	SEQ ID NO 788	Contig30213_RC	SEQ ID NO 2133
NM_002587	SEQ ID NO 789	Contig30230_RC	SEQ ID NO 2134
NM_002590	SEQ ID NO 790	Contig30267_RC	SEQ ID NO 2135
NM_002600	SEQ ID NO 791	Contig30390_RC	SEQ ID NO 2136
NMB_002614	SEQ ID NO 792	Contig30480_RC	SEQ ID NO 2137

NMB_002618	SEQ ID NO 794	Contig30609_RC	SEQ ID NO 2138
NM_002626	SEQ ID NO 795	Contig30934_RC	SEQ ID NO 2139
NM_002633	SEQ ID NO 796	Contig31150_RC	SEQ ID NO 2140
NM_002639	SEQ ID NO 797	Contig31186_RC	SEQ ID NO 2141
NM_002648	SEQ ID NO 798	Contig31251_RC	SEQ ID NO 2142
NM_002659	SEQ ID NO 799	Contig31288_RC	SEQ ID NO 2143
NM_002661	SEQ ID NO 800	Contig31291_RC	SEQ ID NO 2144
NM_002662	SEQ ID NO 801	Contig31295_RC	SEQ ID NO 2145
NM_002664	SEQ ID NO 802	Contig31424_RC	SEQ ID NO 2146
NM_002689	SEQ ID NO 804	Contig31449_RC	SEQ ID NO 2147
NM_002690	SEQ ID NO 805	Contig31596_RC	SEQ ID NO 2148
NM_002709	SEQ ID NO 806	Contig31864_RC	SEQ ID NO 2149
NM_002727	SEQ ID NO 807	Contig31928_RC	SEQ ID NO 2150
NM_002729	SEQ ID NO 808	Contig31966_RC	SEQ ID NO 2151
NM_002734	SEQ ID NO 809	Contig31986_RC	SEQ ID NO 2152
NM_002736	SEQ ID NO 810	Contig32084_RC	SEQ ID NO 2153
NM_002740	SEQ ID NO 811	Contig32105_RC	SEQ ID NO 2154
NM_002748	SEQ ID NO 813	Contig32185_RC	SEQ ID NO 2156
NM_002774	SEQ ID NO 814	Contig32242_RC	SEQ ID NO 2157
NM_002775	SEQ ID NO 815	Contig32322_RC	SEQ ID NO 2158
NM_002776	SEQ ID NO 816	Contig32336_RC	SEQ ID NO 2159
NM_002789	SEQ ID NO 817	Contig32558_RC	SEQ ID NO 2160
NM_002794	SEQ ID NO 818	Contig32798_RC	SEQ ID NO 2161
NM_002796	SEQ ID NO 819	Contig33005_RC	SEQ ID NO 2162
NM_002800	SEQ ID NO 820	Contig33230_RC	SEQ ID NO 2163
NM_002801	SEQ ID NO 821	Contig33260_RC	SEQ ID NO 2164
NM_002808	SEQ ID NO 822	Contig33654_RC	SEQ ID NO 2166
NM_002821	SEQ ID NO 824	Contig33741_RC	SEQ ID NO 2167
NM_002826	SEQ ID NO 825	Contig33771_RC	SEQ ID NO 2168
NM_002827	SEQ ID NO 826	Contig33814_RC	SEQ ID NO 2169
NM_002838	SEQ ID NO 827	Contig33815_RC	SEQ ID NO 2170
NM_002852	SEQ ID NO 828	Contig33833	SEQ ID NO 2171
NM_002854	SEQ ID NO 829	Contig33998_RC	SEQ ID NO 2172
NM_002856	SEQ ID NO 830	Contig34079	SEQ ID NO 2173
NM_002857	SEQ ID NO 831	Contig34080_RC	SEQ ID NO 2174
NM_002858	SEQ ID NO 832	Contig34222_RC	SEQ ID NO 2175
NM_002888	SEQ ID NO 833	Contig34233_RC	SEQ ID NO 2176
NM_002890	SEQ ID NO 834	Contig34303_RC	SEQ ID NO 2177
NM_002901	SEQ ID NO 836	Contig34393_RC	SEQ ID NO 2178
NM_002906	SEQ ID NO 837	Contig34477_RC	SEQ ID NO 2179
NMB_002916	SEQ ID NO 838	Contig34766_RC	SEQ ID NO 2181
NM_002923	SEQ ID NO 839	Contig34952	SEQ ID NO 2182
NM_002933	SEQ ID NO 840	Contig34989_RC	SEQ ID NO 2183
NM_002936	SEQ ID NO 841	Contig35030_RC	SEQ ID NO 2184
NM_002937	SEQ ID NO 842	Contig35251_RC	SEQ ID NO 2185
NM_002950	SEQ ID NO 843	Contig35629_RC	SEQ ID NO 2186 .
NM_002961	SEQ ID NO 844	Contig35635_RC	SEQ ID NO 2187

NM_002964	SEQ ID NO 845	Contig35763_RC	SEQ ID NO 2188
NM_002965	SEQ ID NO 846	Contig35814_RC	SEQ ID NO 2189
NM_002966	SEQ ID NO 847	Contig35896_RC	SEQ ID NO 2190
NM_002982	SEQ ID NO 849	Contig35976_RC	SEQ ID NO 2191
NM_002983	SEQ ID NO 850	Contig36042_RC	SEQ ID NO 2192
NM_002984	SEQ ID NO 851	Contig36081_RC	SEQ ID NO 2193
NM_002985	SEQ ID NO 852	Contig36152_RC	SEQ ID NO 2194
NM_002988	SEQ ID NO 853	Contig36193_RC	SEQ ID NO 2195
NM_002996	SEQ ID NO 854	Contig36312_RC	SEQ ID NO 2196
NM_002997	SEQ ID NO 855	Contig36323_RC	SEQ ID NO 2197
NM_002999	SEQ ID NO 856	Contig36339_RC	SEQ ID NO 2198
NMB_003012	SEQ ID NO 857	Contig36647_RC	SEQ ID NO 2199
NM_003022	SEQ ID NO 858	Contig36744_RC	SEQ ID NO 2200
NM_003034	SEQ ID NO 859	Contig36761_RC	SEQ ID NO 2201
NM_003035	SEQ ID NO 860	Contig36879_RC	SEQ ID NO 2202
NM_003039	SEQ ID NO 861	Contig36900_RC	SEQ ID NO 2203
NM_003051	SEQ ID NO 862	Contig37015_RC	SEQ ID NO 2204
NM_003064	SEQ ID NO 863	Contig37024_RC	SEQ ID NO 2205
NM_003066	SEQ ID NO 864	Contig37072_RC	SEQ ID NO 2207
NM_003088	SEQ ID NO 865	Contig37140_RC	SEQ ID NO 2208
NM_003090	SEQ ID NO 866	Contig37141_RC	SEQ ID NO 2209
NM_003096	SEQ ID NO 867	Contig37204_RC	SEQ ID NO 2210
NM_003099	SEQ ID NO 868	Contig37281_RC	SEQ ID NO 2211
NMB_003102	SEQ ID NO 869	Contig37287_RC	SEQ ID NO 2212
NMB_003104	SEQ ID NO 870	Contig37439_RC	SEQ ID NO 2213
NMB_003108	SEQ ID NO 871	Contig37562_RC	SEQ ID NO 2214
NMB_003121	SEQ ID NO 873	Contig37571_RC	SEQ ID NO 2215
NMB_003134	SEQ ID NO 874	Contig37598	SEQ ID NO 2216
NM_003137	SEQ ID NO 875	Contig37758_RC	SEQ ID NO 2217
NM_003144	SEQ ID NO 876	Contig37778_RC	SEQ ID NO 2218
NM_003146	SEQ ID NO 877	Contig37884_RC	SEQ ID NO 2219
NM_003149	SEQ ID NO 878	Contig37946_RC	SEQ ID NO 2220
NM_003151	SEQ ID NO 879	Contig38170_RC	SEQ ID NO 2221
NM_003157	SEQ ID NO 880	Contig38288_RC	SEQ ID NO 2223
NM_003158	SEQ ID NO 881	Contig38398_RC	SEQ ID NO 2224
NM_003165	SEQ ID NO 882	Contig38580_RC	SEQ ID NO 2226
NM_003172	SEQ ID NO 883	Contig38630_RC	SEQ ID NO 2227
NMB_003177	SEQ ID NO 884	Contig38652_RC	SEQ ID NO 2228
NM_003197	SEQ ID NO 885	Contig38683_RC	SEQ ID NO 2229
NM_003202	SEQ ID NO 886	Contig38726_RC	SEQ ID NO 2230
NM_003213	SEQ ID NO 887	Contig38791_RC	SEQ ID NO 2231
NMB_003217	SEQ ID NO 888	Contig38901_RC	SEQ ID NO 2232
NM_003225	SEQ ID NO 889	Contig38983_RC	SEQ ID NO 2233
NM_003226	SEQ ID NO 890	Contig39090_RC	SEQ ID NO 2234
NM_003236	SEQ ID NO 892	Contig39132_RC	SEQ ID NO 2235
NM_003239	SEQ ID NO 893	Contig39157_RC	SEQ ID NO 2236
NM_003248	SEQ ID NO 894	Contig39226_RC	SEQ ID NO 2237

NM_003255	SEQ ID NO 895	Contig39285_RC	SEQ ID NO 2238
NM_003258	SEQ ID NO 896	Contig39556_RC	SEQ ID NO 2239
NM_003264	SEQ ID NO 897	Contig39591_RC	SEQ ID NO 2240
NM_003283	SEQ ID NO 898	Contig39826_RC	SEQ ID NO 2241
NM_003318	SEQ ID NO 899	Contig39845_RC	SEQ ID NO 2242
NM_003329	SEQ ID NO 900	Contig39891_RC	SEQ ID NO 2243
NM_003332	SEQ ID NO 901	Contig39922_RC	SEQ ID NO 2244
NM_003358	SEQ ID NO 902	Contig39960_RC	SEQ ID NO 2245
NM_003359	SEQ ID NO 903	Contig40026_RC	SEQ ID NO 2246
NM_003360	SEQ ID NO 904	Contig40121_RC	SEQ ID NO 2247
NM_003368	SEQ ID NO 905	Contig40128_RC	SEQ ID NO 2248
NM_003376	SEQ ID NO 906	Contig40146	SEQ ID NO 2249
NM_003380	SEQ ID NO 907	Contig40208_RC	SEQ ID NO 2250
NM_003392	SEQ ID NO 908	Contig40212_RC	SEQ ID NO 2251
NM_003412	SEQ ID NO 909	Contig40238_RC	SEQ ID NO 2252
NM_003430	SEQ ID NO 910	Contig40434_RC	SEQ ID NO 2253
NM_003462	SEQ ID NO 911	Contig40446_RC	SEQ ID NO 2254
NM_003467	SEQ ID NO 912	Contig40500_RC	SEQ ID NO 2255
NM_003472	SEQ ID NO 913	Contig40573_RC	SEQ ID NO 2256
NM_003479	SEQ ID NO 914	Contig40813_RC	SEQ ID NO 2258
NM_003489	SEQ ID NO 915	Contig40816_RC	SEQ ID NO 2259
NM_003494	SEQ ID NO 918	Contig40845_RC	SEQ ID NO 2261
NM_003498	SEQ ID NO 917	Contig40889_RC	SEQ ID NO 2262
NM_003504	SEQ ID NO 919	Contig41035	SEQ ID NO 2263
NM_003508	SEQ ID NO 920	Contig41234_RC	SEQ ID NO 2264
NMB_003510	SEQ ID NO 921	Contig41413_RC	SEQ ID NO 2266
NMB_003512	SEQ ID NO 922	Contig41521_RC	SEQ ID NO 2267
NM_003528	SEQ ID NO 923	Contig41530_RC	SEQ ID NO 2268
NM_003544	SEQ ID NO 924	Contig41590	SEQ ID NO 2269
NM_003561	SEQ ID NO 925	Contig41618_RC	SEQ ID NO 2270
NM_003563	SEQ ID NO 926	Contig41624_RC	SEQ ID NO 2271
NM_003568	SEQ ID NO 927	Contig41635_RC	SEQ ID NO 2272
NM_003579	SEQ ID NO 928	Contig41676_RC	SEQ ID NO 2273
NM_003600	SEQ ID NO 929	Contig41689_RC	SEQ ID NO 2274
NM_003615	SEQ ID NO 931	Contig41804_RC	SEQ ID NO 2275
NM_003627	SEQ ID NO 932	Contig41887_RC	SEQ ID NO 2276
NM_003645	SEQ ID NO 935	Contig41905_RC	SEQ ID NO 2277
NM_003651	SEQ ID NO 936	Contig41954_RC	SEQ ID NO 2278
NM_003657	SEQ ID NO 937	Contig41983_RC	SEQ ID NO 2279
NM_003662	SEQ ID NO 938	Contig42006_RC	SEQ ID NO 2280
NM_003670	SEQ ID NO 939	Contig42014_RC	SEQ ID NO 2281
NM_003675	SEQ ID NO 940	Contig42036_RC	SEQ ID NO 2282
NM_003676	SEQ ID NO 941	Contig42041_RC	SEQ ID NO 2283
NM_003681	SEQ ID NO 942	Contig42139	SEQ ID NO 2284
NM_003683	SEQ ID NO 943	Contig42161_RC	SEQ ID NO 2285
NM_003686	SEQ ID NO 944	Contig42220_RC	SEQ ID NO 2286
NM_003689	SEQ ID NO 945	Contig42306_RC	SEQ ID NO 2287

NM_003714	SEQ ID NO 946	Contig42311_RC	SEQ ID NO 2288
NM_003720	SEQ ID NO 947	Contig42313_RC	SEQ ID NO 2289
NM_003726	SEQ ID NO 948	Contig42402_RC	SEQ ID NO 2290
NM_003729	SEQ ID NO 949	Contig42421_RC	SEQ ID NO 2291
NM_003740	SEQ ID NO 950	Contig42430_RC	SEQ ID NO 2292
NM_003772	SEQ ID NO 952	Contig42431_RC	SEQ ID NO 2293
NM_003791	SEQ ID NO 953	Contig42542_RC	SEQ ID NO 2294
NM_003793	SEQ ID NO 954	Contig42582	SEQ ID NO 2295
NM_003795	SEQ ID NO 955	Contig42631_RC	SEQ ID NO 2296
NM_003806	SEQ ID NO 956	Contig42751_RC	SEQ ID NO 2297
NM_003821	SEQ ID NO 957	Contig42759_RC	SEQ ID NO 2298
NM_003829	SEQ ID NO 958	Contig43054	SEQ ID NO 2299
NM_003831	SEQ ID NO 959	Contig43079_RC	SEQ ID NO 2300
NM_003862	SEQ ID NO 960	Contig43195_RC	SEQ ID NO 2301
NM_003866	SEQ ID NO 961	Contig43368_RC	SEQ ID NO 2302
NM_003875	SEQ ID NO 962	Contig43410_RC	SEQ ID NO 2303
NM_003878	SEQ ID NO 963	Contig43476_RC	SEQ ID NO 2304
NM_003894	SEQ ID NO 965	Contig43549_RC	SEQ ID NO 2305
NM_003897	SEQ ID NO 966	Contig43645_RC	SEQ ID NO 2306
NM_003904	SEQ ID NO 967	Contig43648_RC	SEQ ID NO 2307
NM_003929	SEQ ID NO 968	Contig43673_RC	SEQ ID NO 2308
NM_003933	SEQ ID NO 969	Contig43679_RC	SEQ ID NO 2309
NM_003937	SEQ ID NO 970	Contig43694_RC	SEQ ID NO 2310
NM_003940	SEQ ID NO 971	Contig43747_RC	SEQ ID NO 2311
NM_003942	SEQ ID NO 972	Contig43918_RC	SEQ ID NO 2312
NM_003944	SEQ ID NO 973	Contig43983_RC	SEQ ID NO 2313
NM_003953	SEQ ID NO 974	Contig44040_RC	SEQ ID NO 2314
NM_003954	SEQ ID NO 975	Contig44064_RC	SEQ ID NO 2315
NM_003975	SEQ ID NO 976	Contig44195_RC	SEQ ID NO 2316
NM_003981	SEQ ID NO 977	Contig44226_RC	SEQ ID NO 2317
NM_003982	SEQ ID NO 978	Contig44289_RC	SEQ ID NO 2320
NM_003986	SEQ ID NO 979	Contig44310_RC	SEQ ID NO 2321
NM_004003	SEQ ID NO 980	Contig44409	SEQ ID NO 2322
NM_004010	SEQ ID NO 981	Contig44413_RC	SEQ ID NO 2323
NM_004024	SEQ ID NO 982	Contig44451_RC	SEQ ID NO 2324
NM_004038	SEQ ID NO 983	Contig44585_RC	SEQ ID NO 2325
NM_004049	SEQ ID NO 984	Contig44656_RC	SEQ ID NO 2326
NM_004052	SEQ ID NO 985	Contig44703_RC	SEQ ID NO 2327
NM_004053	SEQ ID NO 986	Contig44708_RC	SEQ ID NO 2328
NM_004079	SEQ ID NO 987	Contig44757_RC	SEQ ID NO 2329
NM_004104	SEQ ID NO 988	Contig44829_RC	SEQ ID NO 2331
NM_004109	SEQ ID NO 989	Contig44870	SEQ ID NO 2332
NM_004110	SEQ ID NO 990	Contig44893_RC	SEQ ID NO 2333
NM_004120	SEQ ID NO 991	Contig44909_RC	SEQ ID NO 2334
NM_004131	SEQ ID NO 992	Contig44939_RC	SEQ ID NO 2335
NM_004143	SEQ ID NO 993	Contig45022_RC	SEQ ID NO 2336
NM_004154	SEQ ID NO 994	Contig45032_RC	SEQ ID NO 2337
NM_004170	SEQ ID NO 996	Contig45041_RC	SEQ ID NO 2338

NM_004172	SEQ ID NO 997	Contig45049_RC	SEQ ID NO 2339
NM_004176	SEQ ID NO 998	Contig45090_RC	SEQ ID NO 2340
NM_004180	SEQ ID NO 999	Contig45156_RC	SEQ ID NO 2341
NM_004181	SEQ ID NO 1000	Contig45316_RC	SEQ ID NO 2342
NM_004184	SEQ ID NO 1001	Contig45321	SEQ ID NO 2343
NM_004203	SEQ ID NO 1002	Contig45375_RC	SEQ ID NO 2345
NM_004207	SEQ ID NO 1003	Contig45443_RC	SEQ ID NO 2346
NM_004217	SEQ ID NO 1004	Contig45454_RC	SEQ ID NO 2347
NM_004219	SEQ ID NO 1005	Contig45537_RC	SEQ ID NO 2348
NM_004221	SEQ ID NO 1006	Contig45588_RC	SEQ ID NO 2349
NM_004233	SEQ ID NO 1007	Contig45708_RC	SEQ ID NO 2350
NM_004244	SEQ ID NO 1008	Contig45816_RC	SEQ ID NO 2351
NM_004252	SEQ ID NO 1009	Contig45847_RC	SEQ ID NO 2352
NM_004265	SEQ ID NO 1010	Contig45891_RC	SEQ ID NO 2353
NM_004267	SEQ ID NO 1011	Contig46056_RC	SEQ ID NO 2354
NM_004281	SEQ ID NO 1012	Contig46062_RC	SEQ ID NO 2355
NM_004289	SEQ ID NO 1013	Contig46075_RC	SEQ ID NO 2356
NM_004298	SEQ ID NO 1015	Contig46164_RC	SEQ ID NO 2357
NM_004301	SEQ ID NO 1016	Contig46218_RC	SEQ ID NO 2358
NM_004305	SEQ ID NO 1017	Contig46223_RC	SEQ ID NO 2359
NM_004311	SEQ ID NO 1018	Contig46244_RC	SEQ ID NO 2360
NM_004315	SEQ ID NO 1019	Contig46262_RC	SEQ ID NO 2361
NM_004323	SEQ ID NO 1020	Contig46362_RC	SEQ ID NO 2364
NM_004330	SEQ ID NO 1021	Contig46443_RC	SEQ ID NO 2365
NM_004336	SEQ ID NO 1022	Contig46553_RC	SEQ ID NO 2367
NM_004338	SEQ ID NO 1023	Contig46597_RC	SEQ ID NO 2368
NM_004350	SEQ ID NO 1024	Contig46653_RC	SEQ ID NO 2369
NM_004354	SEQ ID NO 1025	Contig46709_RC	SEQ ID NO 2370
NM_004358	SEQ ID NO 1026	Contig46777_RC	SEQ ID NO 2371
NM_004360	SEQ ID NO 1027	Contig46802_RC	SEQ ID NO 2372
NM_004362	SEQ ID NO 1028	Contig46890_RC	SEQ ID NO 2374
NM_004374	SEQ ID NO 1029	Contig46922_RC	SEQ ID NO 2375
NM_004378	SEQ ID NO 1030	Contig46934_RC	SEQ ID NO 2376
NM_004392	SEQ ID NO 1031	Contig46937_RC	SEQ ID NO 2377
NM_004395	SEQ ID NO 1032	Contig46991_RC	SEQ ID NO 2378
NM_004414	SEQ ID NO 1033	Contig47016_RC	SEQ ID NO 2379
NM_004418	SEQ ID NO 1034	Contig47045_RC	SEQ ID NO 2380
NM_004425	SEQ ID NO 1035	Contig47106_RC	SEQ ID NO 2381
NM_004431	SEQ ID NO 1036	Contig47146_RC	SEQ ID NO 2382
NM_004436	SEQ ID NO 1037	Contig47230_RC	SEQ ID NO 2383
NM_004438	SEQ ID NO 1038	Contig47405_RC	SEQ ID NO 2384
NM_004443	SEQ ID NO 1039	Contig47456_RC	SEQ ID NO 2385
NM_004446	SEQ ID NO 1040	Contig47465_RC	SEQ ID NO 2386
NM_004451	SEQ ID NO 1041	Contig47498_RC	SEQ ID NO 2387
NM_004454	SEQ ID NO 1042	Contig47578_RC	SEQ ID NO 2388
NM_004456	SEQ ID NO 1043	Contig47645_RC	SEQ ID NO 2389
NM_004458	SEQ ID NO 1044	Contig47680_RC	SEQ ID NO 2390

NM_004472	SEQ ID NO 1045	Contig47781_RC	SEQ ID NO 2391
NM_004480	SEQ ID NO 1046	Contig47814_RC	SEQ ID NO 2392
NM_004482	SEQ ID NO 1047	Contig48004_RC	SEQ ID NO 2393
NM_004494	SEQ ID NO 1048	Contig48043_RC	SEQ ID NO 2394
NM_004496	SEQ ID NO 1049	Contig48057_RC	SEQ ID NO 2395
NM_004503	SEQ ID NO 1050	Contig48076_RC	SEQ ID NO 2396
NM_004504	SEQ ID NO 1051	Contig48249_RC	SEQ ID NO 2397
NM_004515	SEQ ID NO 1052	Contig48263_RC	SEQ ID NO 2398
NM_004522	SEQ ID NO 1053	Contig48270_RC	SEQ ID NO 2399
NM_004523	SEQ ID NO 1054	Contig48328_RC	SEQ ID NO 2400
NM_004525	SEQ ID NO 1055	Contig48518_RC	SEQ ID NO 2401
NM_004556	SEQ ID NO 1056	Contig48572_RC	SEQ ID NO 2402
NM_004559	SEQ ID NO 1057	Contig48659_RC	SEQ ID NO 2403
NM_004569	SEQ ID NO 1058	Contig48722_RC	SEQ ID NO 2404
NM_004577	SEQ ID NO 1059	Contig48774_RC	SEQ ID NO 2405
NM_004585	SEQ ID NO 1060	Contig48776_RC	SEQ ID NO 2406
NM_004587	SEQ ID NO 1061	Contig48800_RC	SEQ ID NO 2407
NM_004594	SEQ ID NO 1062	Contig48806_RC	SEQ ID NO 2408
NM_004599	SEQ ID NO 1063	Contig48852_RC	SEQ ID NO 2409
NM_004633	SEQ ID NO 1066	Contig48900_RC	SEQ ID NO 2410
NM_004642	SEQ ID NO 1067	Contig48913_RC	SEQ ID NO 2411
NM_004648	SEQ ID NO 1068	Contig48970_RC	SEQ ID NO 2413
NM_004663	SEQ ID NO 1069	Contig49058_RC	SEQ ID NO 2414
NM_004664	SEQ ID NO 1070	Contig49063_RC	SEQ ID NO 2415
NM_004684	SEQ ID NO 1071	Contig49093	SEQ ID NO 2416
NM_004688	SEQ ID NO 1072	Contig49098_RC	SEQ ID NO 2417
NM_004694	SEQ ID NO 1073	Contig49169_RC	SEQ ID NO 2418
NM_004695	SEQ ID NO 1074	Contig49233_RC	SEQ ID NO 2419
NM_004701	SEQ ID NO 1075	Contig49270_RC	SEQ ID NO 2420
NM_004708	SEQ ID NO 1077	Contig49282_RC	SEQ ID NO 2421
NM_004711	SEQ ID NO 1078	Contig49289_RC	SEQ ID NO 2422
NM_004726	SEQ ID NO 1079	Contig49342_RC	SEQ ID NO 2423
NM_004750	SEQ ID NO 1081	Contig49344	SEQ ID NO 2424
NM_004761	SEQ ID NO 1082	Contig49388_RC	SEQ ID NO 2425
NM_004762	SEQ ID NO 1083	Contig49405_RC	SEQ ID NO 2426
NM_004780	SEQ ID NO 1085	Contig49445_RC	SEQ ID NO 2427
NM_004791	SEQ ID NO 1086	Contig49468_RC	SEQ ID NO 2428
NM_004798	SEQ ID NO 1087	Contig49509_RC	SEQ ID NO 2429
NM_004808	SEQ ID NO 1088	Contig49578_RC	SEQ ID NO 2431
NM_004811	SEQ ID NO 1089	Contig49581_RC	SEQ ID NO 2432
NM_004833	SEQ ID NO 1090	Contig49631_RC	SEQ ID NO 2433
NM_004835	SEQ ID NO 1091	Contig49673_RC	SEQ ID NO 2435
NM_004843	SEQ ID NO 1092	Contig49743_RC	SEQ ID NO 2436
NM_004847	SEQ ID NO 1093	Contig49790_RC	SEQ ID NO 2437
NM_004848	SEQ ID NO 1094	Contig49818_RC	SEQ ID NO 2438
NM_004864	SEQ ID NO 1095	Contig49849_RC	SEQ ID NO 2439
NM_004865	SEQ ID NO 1096	Contig49855	SEQ ID NO 2440



NM_004866	SEQ ID NO 1097	Contig49910_RC	SEQ ID NO 2441
NM_004877	SEQ ID NO 1098	Contig49948_RC	SEQ ID NO 2442
NM_004900	SEQ ID NO 1099	Contig50004_RC	SEQ ID NO 2443
NM_004906	SEQ ID NO 1100	Contig50094	SEQ ID NO 2444
NM_004910	SEQ ID NO 1101	Contig50120_RC	SEQ ID NO 2446
NM_004918	SEQ ID NO 1103	Contig50153_RC	SEQ ID NO 2447
NM_004923	SEQ ID NO 1104	Contig50189_RC	SEQ ID NO 2448
NM_004938	SEQ ID NO 1105	Contig50276_RC	SEQ ID NO 2449
NM_004951	SEQ ID NO 1106	Contig50288_RC	SEQ ID NO 2450
NM_004968	SEQ ID NO 1107	Contig50297_RC	SEQ ID NO 2451
NM_004994	SEQ ID NO 1108	Contig50391_RC	SEQ ID NO 2452
NM_004999	SEQ ID NO 1109	Contig50410	SEQ ID NO 2453
NM_005001	SEQ ID NO 1110	Contig50523_RC	SEQ ID NO 2454
NM_005002	SEQ ID NO 1111	Contig50529	SEQ ID NO 2455
NM_005012	SEQ ID NO 1112	Contig50588_RC	SEQ ID NO 2456
NM_005032	SEQ ID NO 1113	Contig50592	SEQ ID NO 2457
NM_005044	SEQ ID NO 1114	Contig50669_RC	SEQ ID NO 2458
NM_005046	SEQ ID NO 1115	Contig50719_RC	SEQ ID NO 2460
NM_005049	SEQ ID NO 1116	Contig50728_RC	SEQ ID NO 2461
NM_005067	SEQ ID NO 1117	Contig50731_RC	SEQ ID NO 2462
NM_005077	SEQ ID NO 1118	Contig50802_RC	SEQ ID NO 2463
NM_005080	SEQ ID NO 1119	Contig50822_RC	SEQ ID NO 2464
NM_005084	SEQ ID NO 1120	Contig50850_RC	SEQ ID NO 2466
NM_005130	SEQ ID NO 1122	Contig50860_RC	SEQ ID NO 2467
NM_005139	SEQ ID NO 1123	Contig50913_RC	SEQ ID NO 2468
NM_005168	SEQ ID NO 1125	Contig50950_RC	SEQ ID NO 2469
NM_005190	SEQ ID NO 1126	Contig51066_RC	SEQ ID NO 2470
NM_005196	SEQ ID NO 1127	Contig51105_RC	SEQ ID NO 2472
NM_005213	SEQ ID NO 1128	Contig51117_RC	SEQ ID NO 2473
NM_005218	SEQ ID NO 1129	Contig51196_RC	SEQ ID NO 2474
NM_005235	SEQ ID NO 1130	Contig51235_RC	SEQ ID NO 2475
NM_005245	SEQ ID NO 1131	Contig51254_RC	SEQ ID NO 2476
NM_005249	SEQ ID NO 1132	Contig51352_RC	SEQ ID NO 2477
NM_005257	SEQ ID NO 1133	Contig51369_RC	SEQ ID NO 2478
NM_005264	SEQ ID NO 1134	Contig51392_RC	SEQ ID NO 2479
NM_005271	SEQ ID NO 1135	Contig51403_RC	SEQ ID NO 2480
NM_005314	SEQ ID NO 1136	Contig51685_RC	SEQ ID NO 2483
NM_005321	SEQ ID NO 1137	Contig51726_RC	SEQ ID NO 2484
NM_005322	SEQ ID NO 1138	Contig51742_RC	SEQ ID NO 2485
NM_005325	SEQ ID NO 1139	Contig51749_RC	SEQ ID NO 2486
NM_005326	SEQ ID NO 1140	Contig51775_RC	SEQ ID NO 2487
NM_005335	SEQ ID NO 1141	Contig51800	SEQ ID NO 2488
NM_005337	SEQ ID NO 1142	Contig51809_RC	SEQ ID NO 2489
NM_005342	SEQ ID NO 1143	Contig51821_RC	SEQ ID NO 2490
NM_005345	SEQ ID NO 1144	Contig51888_RC	SEQ ID NO 2491
NM_005357	SEQ ID NO 1145	Contig51953_RC	SEQ ID NO 2493
NM_005375	SEQ ID NO 1146	Contig51967_RC	SEQ ID NO 2495

NM_005391	SEQ ID NO 1147	Contig51981_RC	SEQ ID NO 2496
NM_005408	SEQ ID NO 1148	Contig51994_RC	SEQ ID NO 2497
NM_005409	SEQ ID NO 1149	Contig52082_RC	SEQ ID NO 2498
NM_005410	SEQ ID NO 1150	Contig52094_RC	SEQ ID NO 2499
NM_005426	SEQ ID NO 1151	Contig52320	SEQ ID NO 2500
NM_005433	SEQ ID NO 1152	Contig52398_RC	SEQ ID NO 2501
NM_005441	SEQ ID NO 1153	Contig52425_RC	SEQ ID NO 2503
NM_005443	SEQ ID NO 1154	Contig52482_RC	SEQ ID NO 2504
NM_005483	SEQ ID NO 1155	Contig52543_RC	SEQ ID NO 2505
NM_005486	SEQ ID NO 1156	Contig52553_RC	SEQ ID NO 2506
NM_005496	SEQ ID NO 1157	Contig52579_RC	SEQ ID NO 2507
NM_005498	SEQ ID NO 1158	Contig52603_RC	SEQ ID NO 2508
NM_005499	SEQ ID NO 1159	Contig52639_RC	SEQ ID NO 2509
NM_005514	SEQ ID NO 1160	Contig52641_RC	SEQ ID NO 2510
NM_005531	SEQ ID NO 1162	Contig52684	SEQ ID NO 2511
NM_005538	SEQ ID NO 1163	Contig52705_RC	SEQ ID NO 2512
NM_005541	SEQ ID NO 1164	Contig52720_RC	SEQ ID NO 2513
NM_005544	SEQ ID NO 1165	Contig52722_RC	SEQ ID NO 2514
NM_005548	SEQ ID NO 1166	Contig52723_RC	SEQ ID NO 2515
NM_005554	SEQ ID NO 1167	Contig52740_RC	SEQ ID NO 2516
NM_005555	SEQ ID NO 1168	Contig52779_RC	SEQ ID NO 2517
NM_005556	SEQ ID NO 1169	Contig52957_RC	SEQ ID NO 2518
NM_005557	SEQ ID NO 1170	Contig52994_RC	SEQ ID NO 2519
NM_005558	SEQ ID NO 1171	Contig53022_RC	SEQ ID NO 2520
NM_005562	SEQ ID NO 1172	Contig53038_RC	SEQ ID NO 2521
NM_005563	SEQ ID NO 1173	Contig53047_RC	SEQ ID NO 2522
NM_005565	SEQ ID NO 1174	Contig53130	SEQ ID NO 2523
NM_005566	SEQ ID NO 1175	Contig53183_RC	SEQ ID NO 2524
NM_005572	SEQ ID NO 1176	Contig53242_RC	SEQ ID NO 2526
NM_005582	SEQ ID NO 1177	Contig53248_RC	SEQ ID NO 2527
NM_005608	SEQ ID NO 1178	Contig53260_RC	SEQ ID NO 2528
NM_005614	SEQ ID NO 1179	Contig53296_RC	SEQ ID NO 2531
NM_005617	SEQ ID NO 1180	Contig53307_RC	SEQ ID NO 2532
NM_005620	SEQ ID NO 1181	Contig53314_RC	SEQ ID NO 2533
NM_005625	SEQ ID NO 1182	Contig53401_RC	SEQ ID NO 2534
NM_005651	SEQ ID NO 1183	Contig53550_RC	SEQ ID NO 2535
NM_005658	SEQ ID NO 1184	Contig53551_RC	SEQ ID NO 2536
NM_005659	SEQ ID NO 1185	Contig53598_RC	SEQ ID NO 2537
NM_005667	SEQ ID NO 1186	Contig53646_RC	SEQ ID NO 2538
NM_005686	SEQ ID NO 1187	Contig53658_RC	SEQ ID NO 2539
NM_005690	SEQ ID NO 1188	Contig53698_RC	SEQ ID NO 2540
NM_005720	SEQ ID NO 1190	Contig53719_RC	SEQ ID NO 2541
NM_005727	SEQ ID NO 1191	Contig53742_RC	SEQ ID NO 2542
NM_005733	SEQ ID NO 1192	Contig53757_RC	SEQ ID NO 2543
NM_005737	SEQ ID NO 1193	Contig53870_RC	SEQ ID NO 2544
NM_005742	SEQ ID NO 1194	Contig53952_RC	SEQ ID NO 2546
NM_005746	SEQ ID NO 1195	Contig53962_RC	SEQ ID NO 2547

NM_005749	SEQ ID NO 1196	Contig53968_RC	SEQ ID NO 2548
NM_005760	SEQ ID NO 1197	Contig54113_RC	SEQ ID NO 2549
NM_005764	SEQ ID NO 1198	Contig54142_RC	SEQ ID NO 2550
NM_005794	SEQ ID NO 1199	Contig54232_RC	SEQ ID NO 2551
NM_005796	SEQ ID NO 1200	Contig54242_RC	SEQ ID NO 2552
NM_005804	SEQ ID NO 1201	Contig54260_RC	SEQ ID NO 2553
NM_005813	SEQ ID NO 1202	Contig54263_RC	SEQ ID NO 2554
NM_005824	SEQ ID NO 1203	Contig54295_RC	SEQ ID NO 2555
NM_005825	SEQ ID NO 1204	Contig54318_RC	SEQ ID NO 2556
NM_005849	SEQ ID NO 1205	Contig54325_RC	SEQ ID NO 2557
NM_005853	SEQ ID NO 1206	Contig54389_RC	SEQ ID NO 2558
NM_005855	SEQ ID NO 1207	Contig54394_RC	SEQ ID NO 2559
NM_005864	SEQ ID NO 1208	Contig54414_RC	SEQ ID NO 2560
NM_005874	SEQ ID NO 1209	Contig54425	SEQ ID NO 2561
NM_005876	SEQ ID NO 1210	Contig54477_RC	SEQ ID NO 2562
NM_005880	SEQ ID NO 1211	Contig54503_RC	SEQ ID NO 2563
NM_005891	SEQ ID NO 1212	Contig54534_RC	SEQ ID NO 2564
NM_005892	SEQ ID NO 1213	Contig54560_RC	SEQ ID NO 2566
NM_005899	SEQ ID NO 1214	Contig54581_RC	SEQ ID NO 2567
NM_005915	SEQ ID NO 1215	Contig54609_RC	SEQ ID NO 2568
NM_005919	SEQ ID NO 1216	Contig54666_RC	SEQ ID NO 2569
NM_005923	SEQ ID NO 1217	Contig54667_RC	SEQ ID NO 2570
NM_005928	SEQ ID NO 1218	Contig54726_RC	SEQ ID NO 2571
NM_005932	SEQ ID NO 1219	Contig54742_RC	SEQ ID NO 2572
NM_005935	SEQ ID NO 1220	Contig54745_RC	SEQ ID NO 2573
NM_005945	SEQ ID NO 1221	Contig54757_RC	SEQ ID NO 2574
NM_005953	SEQ ID NO 1222	Contig54761_RC	SEQ ID NO 2575
NM_005978	SEQ ID NO 1223	Contig54813_RC	SEQ ID NO 2576
NM_005990	SEQ ID NO 1224	Contig54867_RC	SEQ ID NO 2577
NM_006002	SEQ ID NO 1225	Contig54895_RC	SEQ ID NO 2578
NM_006004	SEQ ID NO 1226	Contig54898_RC	SEQ ID NO 2579
NM_006005	SEQ ID NO 1227	Contig54913_RC	SEQ ID NO 2580
NM_006006	SEQ ID NO 1228	Contig54965_RC	SEQ ID NO 2582
NM_006017	SEQ ID NO 1229	Contig54968_RC	SEQ ID NO 2583
NM_006018	SEQ ID NO 1230	Contig55069_RC	SEQ ID NO 2584
NM_006023	SEQ ID NO 1231	Contig55181_RC	SEQ ID NO 2585
NM_006027	SEQ ID NO 1232	Contig55188_RC	SEQ ID NO 2586
NM_006029	SEQ ID NO 1233	Contig55221_RC	SEQ ID NO 2587
NM_006033	SEQ ID NO 1234	Contig55254_RC	SEQ ID NO 2588
NM_006051	SEQ ID NO 1235	Contig55265_RC	SEQ ID NO 2589
NM_006055	SEQ ID NO 1236	Contig55377_RC	SEQ ID NO 2591
NM_006074	SEQ ID NO 1237	Contig55397_RC	SEQ ID NO 2592
NM_006086	SEQ ID NO 1238	Contig55448_RC	SEQ ID NO 2593
NM_006087	SEQ ID NO 1239	Contig55468_RC	SEQ ID NO 2594
NM_006096	SEQ ID NO 1240	Contig55500_RC	SEQ ID NO 2595
NM_006101	SEQ ID NO 1241	Contig55538_RC	SEQ ID NO 2596
NM_006103	SEQ ID NO 1242	Contig55558_RC	SEQ ID NO 2597

NM_006111	SEQ ID NO 1243	Contig55606_RC	SEQ ID NO 2598
NM_006113	SEQ ID NO 1244	Contig55674_RC	SEQ ID NO 2599
NM_006115	SEQ ID NO 1245	Contig55725_RC	SEQ ID NO 2600
NM_006117	SEQ ID NO 1246	Contig55728_RC	SEQ ID NO 2601
NM_006142	SEQ ID NO 1247	Contig55756_RC	SEQ ID NO 2602
NM_006144	SEQ ID NO 1248	Contig55769_RC	SEQ ID NO 2603
NM_006148	SEQ ID NO 1249	Contig55771_RC	SEQ ID NO 2605
NM_006153	SEQ ID NO 1250	Contig55813_RC	SEQ ID NO 2607
NM_006159	SEQ ID NO 1251	Contig55829_RC	SEQ ID NO 2608
NM_006170	SEQ ID NO 1252	Contig55852_RC	SEQ ID NO 2609
NM_006197	SEQ ID NO 1253	Contig55883_RC	SEQ ID NO 2610
NM_006224	SEQ ID NO 1255	Contig55920_RC	SEQ ID NO 2611
NM_006227	SEQ ID NO 1256	Contig55940_RC	SEQ ID NO 2612
NM_006235	SEQ ID NO 1257	Contig55950_RC	SEQ ID NO 2613
NM_006243	SEQ ID NO 1258	Contig55991_RC	SEQ ID NO 2614
NM_006264	SEQ ID NO 1259	Contig55997_RC	SEQ ID NO 2615
NM_006271	SEQ ID NO 1261	Contig56023_RC	SEQ ID NO 2616
NM_006274	SEQ ID NO 1262	Contig56030_RC	SEQ ID NO 2617
NM_006290	SEQ ID NO 1265	Contig56093_RC	SEQ ID NO 2618
NM_006291	SEQ ID NO 1266	Contig56205_RC	SEQ ID NO 2621
NM_006296	SEQ ID NO 1267	Contig56270_RC	SEQ ID NO 2622
NM_006304	SEQ ID NO 1268	Contig56276_RC	SEQ ID NO 2623
NM_006314	SEQ ID NO 1269	Contig56291_RC	SEQ ID NO 2624
NM_006332	SEQ ID NO 1270	Contig56298_RC	SEQ ID NO 2625
NM_006357	SEQ ID NO 1271	Contig56307	SEQ ID NO 2627
NM_006366	SEQ ID NO 1272	Contig56390_RC	SEQ ID NO 2628
NM_006372	SEQ ID NO 1273	Contig56434_RC	SEQ ID NO 2629
NM_006377	SEQ ID NO 1274	Contig56457_RC	SEQ ID NO 2630
NM_006378	SEQ ID NO 1275	Contig56534_RC	SEQ ID NO 2631
NM_006383	SEQ ID NO 1276	Contig56670_RC	SEQ ID NO 2632
NM_006389	SEQ ID NO 1277	Contig56678_RC	SEQ ID NO 2633
NM_006393	SEQ ID NO 1278	Contig56742_RC	SEQ ID NO 2634
NM_006398	SEQ ID NO 1279	Contig56759_RC	SEQ ID NO 2635
NM_006406	SEQ ID NO 1280	Contig56765_RC	SEQ ID NO 2636
NM_006408	SEQ ID NO 1281	Contig56843_RC	SEQ ID NO 2637
NM_006410	SEQ ID NO 1282	Contig57011_RC	SEQ ID NO 2638
NM_006414	SEQ ID NO 1283	Contig57023_RC	SEQ ID NO 2639
NM_006417	SEQ ID NO 1284	Contig57057_RC	SEQ ID NO 2640
NM_006430	SEQ ID NO 1285	Contig57076_RC	SEQ ID NO 2641
NM_006460	SEQ ID NO 1286	Contig57081_RC	SEQ ID NO 2642
NM_006461	SEQ ID NO 1287	Contig57091_RC	SEQ ID NO 2643
NM_006469	SEQ ID NO 1288	Contig57138_RC	SEQ ID NO 2644
NM_006470	SEQ ID NO 1289	Contig57173_RC	SEQ ID NO 2645
NM_006491	SEQ ID NO 1290	Contig57230_RC	SEQ ID NO 2646
NM_006495	SEQ ID NO 1291	Contig57258_RC	SEQ ID NO 2647
NM_006500	SEQ ID NO 1292	Contig57270_RC	SEQ ID NO 2648
NM_006509	SEQ ID NO 1293	Contig57272_RC	SEQ ID NO 2649

NMB_006516	SEQ ID NO 1294	Contig57344_RC	SEQ ID NO 2650
NM_006533	SEQ ID NO 1295	Contig57430_RC	SEQ ID NO 2651
NM_006551	SEQ ID NO 1296	Contig57458_RC	SEQ ID NO 2652
NM_006556	SEQ ID NO 1297	Contig57493_RC	SEQ ID NO 2653
NM_006558	SEQ ID NO 1298	Contig57584_RC	SEQ ID NO 2654
NM_006564	SEQ ID NO 1299	Contig57595	SEQ ID NO 2655
NM_006573	SEQ ID NO 1300	Contig57602_RC	SEQ ID NO 2656
NM_006607	SEQ ID NO 1301	Contig57609_RC	SEQ ID NO 2657
NM_006622	SEQ ID NO 1302	Contig57610_RC	SEQ ID NO 2658
NM_006623	SEQ ID NO 1303	Contig57644_RC	SEQ ID NO 2659
NM_006636	SEQ ID NO 1304	Contig57725_RC	SEQ ID NO 2660
NM_006670	SEQ ID NO 1305	Contig57739_RC	SEQ ID NO 2661
NM_006681	SEQ ID NO 1306	Contig57825_RC	SEQ ID NO 2662
NM_006682	SEQ ID NO 1307	Contig57864_RC	SEQ ID NO 2663
NM_006696	SEQ ID NO 1308	Contig57940_RC	SEQ ID NO 2664
NM_006698	SEQ ID NO 1309	Contig58260_RC	SEQ ID NO 2665
NM_006705	SEQ ID NO 1310	Contig58272_RC	SEQ ID NO 2666
NM_006739	SEQ ID NO 1311	Contig58301_RC	SEQ ID NO 2667
NM_006748	SEQ ID NO 1312	Contig58368_RC	SEQ ID NO 2668
NM_006759	SEQ ID NO 1313	Contig58471_RC	SEQ ID NO 2669
NM_006762	SEQ ID NO 1314	Contig58755_RC	SEQ ID NO 2671
NM_006763	SEQ ID NO 1315	Contig59120-RC	SEQ ID NO 2672
NM_006769	SEQ ID NO 1316	Contig60157_RC	SEQ ID NO 2673
NM_006770	SEQ ID NO 1317	Contig60864_RC	SEQ ID NO 2676
NM_006780	SEQ ID NO 1318	Contig61254_RC	SEQ ID NO 2677
NM_006787	SEQ ID NO 1319	Contig61815	SEQ ID NO 2678
NM_006806	SEQ ID NO 1320	Contig61975	SEQ ID NO 2679
NMB_006813	SEQ ID NO 1321	Contig62306	SEQ ID NO 2680
NM_006825	SEQ ID NO 1322	Contig62568_RC	SEQ ID NO 2681
NM_006826	SEQ ID NO 1323	Contig62922_RC	SEQ ID NO 2682
NM_006829	SEQ ID NO 1324	Contig62964_RC	SEQ ID NO 2683
NM_006834	SEQ ID NO 1325	Contig63520_RC	SEQ ID NO 2685
NM_006835	SEQ ID NO 1326	Contig63649_RC	SEQ ID NO 2686
NM_006840	SEQ ID NO 1327	Contig63683_RC	SEQ ID NO 2687
NM_006845	SEQ ID NO 1328	Contig63748_RC	SEQ ID NO 2688
NM_006847	SEQ ID NO 1329	Contig64502	SEQ ID NO 2689
NM_006851	SEQ ID NO 1330	Contig64688	SEQ ID NO 2690
NM_006855	SEQ ID NO 1331	Contig64775_RC	SEQ ID NO 2691
NM_006864	SEQ ID NO 1332	Contig65227	SEQ ID NO 2692
NM_006868	SEQ ID NO 1333	Contig65663	SEQ ID NO 2693
NM_006875	SEQ ID NO 1334	Contig65785_RC	SEQ ID NO 2694
NM_006889	SEQ ID NO 1336	Contig65900	SEQ ID NO 2695
NM_006892	SEQ ID NO 1337	Contig66219_RC	SEQ ID NO 2696
NMB_006912	SEQ ID NO 1338	Contig66705_RC	SEQ ID NO 2697
NM_006931	SEQ ID NO 1341	Contig66759_RC	SEQ ID NO 2698
NM_006941	SEQ ID NO 1342	Contig67182_RC	SEQ ID NO 2699
NM_006943	SEQ ID NO 1343		

Tabla 2. 550 marcadores de categoría de ER preferidos sacados de la Tabla 1

Identificador	correlación	Nombre	Descripción
NM_002051	0.763977	GATA3	Proteína 3 de enlace con GATA
AB020689	0.753592	KIAA0882	Proteína KIAA0882
NM_001218	0.753225	CA12	Anhidrasa carbónica XII
NM_000125	0.748421	ESR1	Receptor de estrógeno 1
Contig56678_RC	0.747816		ESTs
NM_004496	0.729116	HNF3A	factor nuclear 3 de hepatocitos, alfa
NM_017732	0.713398	FLJ20262	Proteína hipotética FLJ20262
NM_006806	-0.712678	BTG3	Familia BTG, miembro 3
Contig56390_RC	0.705940		ESTs
Contig37571_RC	0.704468		ESTs
NM_004559	-0.701617	NSEP1	Proteína 1 de enlace con elemento sensible a la nucleasa
Contig50153_RC	-0.696652		ESTs, ligeramente parecidos al precursor de proteína de enlace proteoglicano [H.sapiens]
NMB_012155	0.694332	EMAP-2	Proteína asociada a microtubulo como EMAP equinodermo
Contig237_RC	0.687485	FLJ21127	Proteína hipotética FLJ21127
NMB_01 9063	-0.686064	C2ORF2	Marco de lectura abierta 2 del cromosoma 2
NMB_012219	-0.680900	MRAS	Homólogo de oncógeno RAS de músculo
NM_001982	0.676114	ERBB3	Homólogo 3 de oncógeno viral de leucemia eritoblástica aviar v-erb-b 2
NM_006623	-0.675090	PHGDH	Fosfogliceratodehidrogenasa
NMB_000636	-0.674282	SOD2	Superóxido dismutasa 2, mitocondrial
NMB_006017	-0.670353	PROML1	Prominina 1 similar a (ratón)
Contig57940_RC	0.667915	MAP-1	Proteína MAP-1
Contig46934_RC	0.666908		ESTs, ligeramente parecidos a gradiente 2 Anterior JE0350 [H.sapiens]
NM_005080	0.665772	XBP1	Proteína 1 de enlace con X-caja
NM_014246	0.665725	CELSR1	Cadherina, receptor 1 de tipo G de siete pases EGF LAG, homólogo de flamingo (Drosophila)
Contig54667_RC	-0.663727		Secuencia de ADN humano del clon RP1-187J11 en cromosoma 6q11.1-22.33. Contiene el gen para una nueva proteína parecida a las proteínas previstas S. pombe y S. cerevisiae, el gen para una nueva proteína parecida a los inhibidores C de quinasa proteínica, el extremo de 3' del gen para una nueva proteína parecida a Drosophila L82 y proteínas de gusanos previstas, ESTs, STSs, GSSs y dos islas CpG putativas
Contig51994_RC	0.663715		ESTs, ligeramente parecido a B0416.1 [C.elegans]
NM_016337	0.663006	RNB6	RNB6
NMB_015640	-0.660165	PAI-RBP1	Proteína de enlace con PAI-1 ARNm
X07834	-0.657798	SOD2	Superóxido dismutasa 2, mitocondrial
NMB_012319	0.657666	LIV-1	Proteína LIV-1, regulada por estrógeno
Contig41887_RC	0.656042		ESTs, ligeramente parecidos al homólogo de la proteína de la membrana del corpúsculo de zimógeno de rata [H.sapiens]
NM_003462	0.655349	P28	dineína, axonemal, polipéptido intermedio ligero

Contig58301_RC	0.654268		ARNm de Homo sapiens; ADNc DKFZp667D095 (del clon DKFZp667D095)
NM_005375	0.653783	MYB	Homólogo de oncógeno viral de mieloblastosis aviar v-myb
NMB_017447	-0.652445	YG81	Proteína hipotética LOC54149
Contig924_RC	-0.650658		ESTs
M55914	-0.650181	MPB1	Proteína 1 de enlace con el promotor MYC
NM_006004	-0.649819	UQCRH	Proteína bisagra con ubiquinol-citocromo c reductasa
NM_000964	0.649072	RARA	Receptor de ácido retinoico, alfa
NM_013301	0.647583	HSU79303	Proteína prevista por el clon 23882
AB023211	-0.647403	PDI2	Peptidil arginina deiminasa, tipo II
NM_016629	-0.646412	LOC51323	Proteína hipotética
K02403	0.645532	C4A	Componente de complemento 4A
NM_016405	-0.642201	HSU93243	Homólogo de Ubc6p
Contig46597_RC	0.641733		ESTs
Contig55377_RC	0.640310		ESTs
NM_001207	0.637800	BTF3	Factor 3 de transcripción básica
NM_018166	0.636422	FLJ10647	Proteína hipotética FLJ10647
AL110202	-0.635398		ARNm de Homo sapiens; ADNc DKFZp586I2022 (del clon DKFZp586I2022)
AL133105	-0.635201	DKFZp434F 2322	Proteína hipotética DKFZp434F2322
NMB_016839	-0.635169	RBMS1	Motivo de enlace con ARN, proteína 1 interactuante monocatenaria
Contig53130	-0.634812		ESTs, ligeramente parecidos al canal controlado por nucleótido cíclico activado por hiperpolarización hHCN2 [H.sapiens]
NM_018014	-0.634460	BCL11A	CLUlinfoma 11A de célula B (proteína de dedo de cinc)
NM_006769	-0.632197	LMO4	Dominio LIM solo 4
U92544	0.631170	JCL-1	Proteína asociada al carcinoma hepatocelular; gen 1 asociado al cáncer de mama
Contig49233_RC	-0.631047		Homo sapiens, parecido al factor 2 de enlace con receptor nuclear, clon IMAGE:3463191, ARNm, cds parcial
AL133033	0.629690	KIAA1025	KIAA1025 protein
AL049265	0.629414		ARN de Homo sapiens; ADNc DKFZp564F053 (del clon DKFZp564F053)
NM_018728	0.627989	MYO5C	Miosina 5C
NM_004780	0.627856	TCEAL1	Factor A de alargamiento de transcripción de tipo 1(SII)
Contig760_RC	0.627132		ESTs
Contig399_RC	0.626543	FLJ12538	Proteína hipotética FLJ12538 parecida a la proteína relacionada con ras RAB17
M83822	0.625092	CDC4L	Ciclo de división de célula de tipo 4
NM_001255	-0.625089	CDC20	CDC20 (ciclo de división de célula 20, S. cerevisiae, homólogo)
NM_006739	-0.624903	MCM5	Deficiente mantenimiento de minicromosomas (S. cerevisiae) 5 (ciclo de división celular 46)
NM_002888	-0.624664	RARRES1	Respondedor 1 a receptor de ácido retinoico (inducido por tazaroteno)

NM_003197	0.623850	TCEB1L	Factor B de alargamiento de transcripción (SIII), afín a polipéptido 1
NM_006787	0.623625	JCL-1	Proteína asociada al carcinoma hepatocelular; gen 1 asociado al cáncer de mama
Contig49342_RC	0.622179		ESTs
AL133619	0.621719		ARNm de Homo sapiens: ADNc DKFZp434E2321 (del clon DKFZp434E2321); cds parcial
AL133622	0.621577	KIAA0876	Proteína KIAA0876
NM_004648	-0.621532	PTPNS1	Tirosina fosfatasa proteínica, sustrato 1 de tipo no- receptor
NM_001793	-0.621530	CDH3	Cadherina 3, tipo 1, P-caderina (placentaria)
NM_003217	0.620915	TEGT	Trascricpción mejorada de genes de testículos (inhibidor 1 de BAX)
NM_001551	0.620832	IGBP1	Proteína 1 asociada a la inmunoglobulina (CD79A)
NM_002539	-0.620683	ODC1	Ornitina decarboxilasa 1
Contig55997_RC	-0.619932		ESTs
NM_000633	0.619547	BCL2	CLL/linfoma 2 de célula B
NMB_016267	-0.619096	TONDU	TONDU
Contig3659_RC	0.618048	FLJ21174	Proteína hipotética FLJ21174
NM_000191	0.617250	HMGCL	3-hidroximetil-3-metilglutaril-Coenzima A liasa (hydroximetilglutaricaciduria)
NM_001267	0.616890	CHAD	Condroadherina
Contig39090_RC	0.616385		ESTs
AF055270	-0.616268	HSSG1	Proteína 1 suprimida de choque de calor
Contig43054	0.616015	FLJ21603	Proteína hipotética FLJ21603
NM_001428	-0.615855	ENO1	Enolasa 1, (alfa)
Contig51369_RC	0.615466		ESTs
Contig36647_RC	0.615310	GFRA1	Alfa 1 de receptor de familia GDNF
NM_014096	-0.614832	PRO1659	Proteína PRO1659
NM_015937	0.614735	LOC51604	Proteína CGI-06
Contig49790_RC	-0.614463		ESTs
NM_006759	-0.614279	UGP2	UDP-glucosa pirofosforilasa 2
Contig53598_RC	-0.613787	FLJ11413	Proteína hipotética FLJ11413
AF113132	-0.613561	PSA	Fosfoserina aminotransferasa
AK000004	0.613001		ARNm de Homo sapiens para proteína FLJ00004, cds parcial
Contig52543_RC	0.612960		ADNc de Homo sapiens, fis FLJ13945, clon Y79AA1000969
AB032966	-0.611917	KIAA1140	Proteína KIAA1140
AL080192	0.611544		ADNc de Homo sapiens; FLJ21238 fis, clon COL01115
X56807	-0.610654	DSC2	Desmocolina 2
Contig30390_RC	0.609614		ESTs
AL137362	0.609121	FLJ22237	Proteína hipotética FLJ22237
NM_014211	-0.608585	GABRP	Ácido gamma-aminobutírico acid (GABA) receptor A, pi
NM_006696	0.608474	SMAP	Proteína coactivadora del receptor de hormonas de la tiroides
Contig45588_RC	-0.608273		ADN de Homo sapiens; FLJ22610 fis, clon HSI04930
NM_003358	0.608244	UGCG	UDP-glucosa ceramida glucosiltransferasa
NMB_006153	-0.608129	NCK1	Proteína 1 del adaptador de NCK



NMB_001453	-0.606939	FOXC1	Forkhead box C1
Contig54666_RC	0.606475		oy65e02.x1 NCI_CGAP_CLL1 ADNC de Homo sapiens clon IMAGE: 1670714 3' parecido a TR:Q29168 PROTEINA DESCONOCIDA Q29168; secuencia de ARNm.
NM_005945	-0.605945	MPB1	Proteína 1 de enlace con promotor MYC
Contig55725_RC	-0.605841		ESTs, ligeramente parecido a T50635 proteína hipotética DKFZp762L0311.1 [H.sapiens]
Contig37015_RC	-0.605780		ESTs, ligeramente parecidos a UAS3_PROTEÍNA HUMANA UBASH3A PROTEIN [H.sapiens]
AL157480	-0.604362	SH3BP1	Proteína 1 de enlace con dominio SH3
NM_005325	-0.604310	H1F1	Familia de histona H1, miembro 1
NM_001446	-0.604061	FABP7	Proteína 7 de enlace con el ácido graso, cerebro
Contig263_RC	0.603318		ADNC de Homo sapiens cDNA: fis FLJ23000, clon LNG00194
Contig8347_RC	-0.603311		ESTs
NM_002988	-0.603279	SCYA18	Pequeña subfamilia A inducible por citoquina (Cys- Cys), miembro 18, pulmonar y regulada por activación
AF111849	0.603157	HELO1	Enzima 2 de alargamiento de ácido graso poliinsaturado de cadena larga homólogo de levadura
NM_014700	0.603042	KIAA0665	Producto del gen KIAA0665
NM_001814	-0.602988	CTSC	Catepsia C
AF116682	-0.602350	PRO2013	Proteína hipotética PRO2013
AB037836	0.602024	KIAA1415	Proteína KIAA1415
AB002301	0.602005	KIAA0303	Proteína KIAA0303
NM_002996	-0.601841	SCYD1	Pequeña subfamilia D (Cys-X3-Cys), miembro 1, (fractalina, neurotactina)
NM_018410	-0.601765	KFZp762E1312	Proteína hipotética DKFZp762E1312
Contig49581_RC	-0.601571	KIAA1350	Proteína KIAA1350
NM_003088	-0.601458	SNL	Gaseado (afin a Drosophila) (tipo de homólogo de fascina de erizo de mar)
Contig47045_RC	0.601088		ESTs, ligeramente parecido a DP1 PROTEÍNA HUMANA 1 DE POLIPOSIS DE LOCUS [H.sapiens]
NM_001806	-0.600954	CEBPG	Proteína (C/EBP) de enlace con CCAAT/potenciador, gamma
NM_004374	0.600766	COX6C	Subunidad Vlc de citocromo c oxidasa
Contig52641_RC	0.600132	MOUSE	ESTs, ligeramente parecidos a AUTOANTIGENO B CENTRÓMERO CENB PRINCIPAL [M.musculus]
NM_000100	-0.600127	CSTB	Cistatina B (estefina B)
NM_002250	-0.600004	KCNN4	Intermediario de potasio/pequeño canal de conductancia activado por calcio, subfamilia N, miembro 4
AB033035	-0.599423	KIAA1209	Proteína KIAA1209
Contig53968_RC	0.599077		ESTs
NM_002300	-0.598246	LDHB	Lactato dehidrogenasa B
NM_000507	0.598110	FBP1	Fructosa-1,6-bisfosfatasa 1
NM_002053	-0.597756	GBP1	Proteína 1 de enlace con guanilato, inducible por interferón, 67kD
AB007883	0.597043	KIAA0423	Proteína KIAA0423

NM_004900	-0.597010	DJ742C19.2	Forbolina (parecida a la proteína editora de ARNm apolipoproteína B)
NM_004480	0.596321	FUT8	Fucosiltransferasa 8 (alfa (1,6) fucosiltransferasa)
Contig35896_RC	0.596281		ESTs
NM_020974	0.595173	CEGP1	Proteína CEGP1
NM_000662	0.595114	NAT1	N-acetiltransferasa 1 (arilamina N-acetiltransferasa)
NMB_006113	0.595017	VAV3	oncógeno vav 3
NM_014865	-0.594928	KIAA0159	Proteína 1 asociada a SMC de condensación de cromosoma
Contig55538_RC	-0.594573	BA395L14.2	Proteína hipotética bA395L14.2
NM_016056	0.594084	LOC51643	Proteína CGI-119
NM_003579	-0.594063	RAD54L	RAD54 (afín a S.cerevisiae)
NM_014214	-0.593860	IMPA2	Inositol(mio)-1 (o 4)- monofosfatasa 2
U79293	0.593793		Secuencia de ARNm del clon humano 23948
NM_005557	-0.593746	KRT16	queratina 16 (queratoderma palmoplantar focal no epidermolítico)
NM_002444	-0.592405	MSN	Moesina
NM_003681	-0.592155	PDXK	piridoxal (pyridoxina, vitamina B6) quinasa
NM_006372	-0.591711	NSAP1	Proteína 1 asociada a NS1
NM_005218	-0.591192	DEFB1	Defensina, beta 1
NM_004642	-0.591081	DOC1	Borrado en cáncer oral (ratón, homólogo) 1
AL133074	0.590359		ADNC de Homo sapiens; fis FLJ22139, clon HEP20959
M73547	0.590317	D5S346	Segmento de ADN, sonda de una sola copia LNS-CAI/LNS-CAII (borrada en poliposis)
Contig65663	0.590312		ESTs
AL035297	-0.589728		Gen de H.sapiens de PAC 747L4
Contig35629_RC	0.589383		ESTs
NM_019027	0.588862	FLJ20273	Proteína hipotética
NM_012425	-0.588804		Proteína 1 del supresor del Ras de Homo sapiens (RSU1), ARNm
NM_020179	-0.588326	FN5	Proteína FN5
AF090913	-0.587275	TMSB10	Timosina, beta 10
NM_004176	0.587190	SREBF1	Factor 1 de transcripción de enlace con el elemento regulatorio de esterol
NM_016121	0.586941	LOC51133	Antígeno NY-REN-45
NM_014773	0.586871	KIAA0141	Producto del gen KIAA0141
NM_019000	0.586677	FLJ20152	Proteína hipotética
NM_016243	0.585942	LOC51706	Citocromo b5 reductasa 1 (B5R.1)
NM_014274	-0.585815	ABP/ZF	Proteína de enlace Alu con dominio de dedo de cinc
NM_018379	0.585497	FLJ11280	Proteína hipotética FLJ11280
AL157431	-0.585077	DKFZp762A227	Proteína hipotética DKFZp762A227
D38521	-0.584684	KIAA0077	Proteína KIAA0077
NM_002570	0.584272	PACE4	Sistema 4 apareado de exfoliado de aminoácido básico

NM_001809	-0.584252	CENPA	Proteína centrómera A (17kD)
NM_003318	-0.583556	TTK	Proteína quinasa TTK
NM_014325	-0.583555	CORO1C	Coronina, proteína de enlace con actina, 1C
NM_005667	0.583376	ZFP103	Proteína homóloga de cinc homóloga de Zfp103 en ratón
NM_004354	0.582420	CCNG2	Ciclina G2
NM_003670	0.582235	BHLHB2	Contenido de helix básico-bucle-dominio de helix domain, class B, 2
NM_001673	-0.581902	ASNS	Asparagina sintetasa
NM_001333	-0.581402	CTSL2	Catepsina L2
Contig54295_RC	0.581256		ESTs
Contig33998_RC	0.581018		ESTs
NM_006002	-0.580592	UCHL3	Ubiquitina carboxyl-terminal esterasa L3 (ubiquitin tiolesterasa)
NM_015392	0.580568	NPDC1	Proliferación, diferenciación y control neural, 1
NM_004866	0.580138	SCAMP1	Proteína 1 de la membrana portadora secretora 1
Contig50391_RC	0.580071		ESTs
NM_000592	0.579965	C4B	Componente de complemento 4B
Contig50802_RC	0.579881		ESTs
Contig41635_RC	-0.579468		ESTs
NM_006845	-0.579339	KNSL6	Tipo 6 de quinesina (quinesina asociada a centrómero mitótico)
NM_003720	-0.579296	DSCR2	Gen 2 de la región crítica del síndrome de Down
NM_000060	0.578967	BTD	Biotinidasa
AL050388	-0.578736		ARNm de Homo sapiens; ADNc DKFZp564M2422 (de clon DKFZp564M2422); cds parcial
NM_003772	-0.578395	JRKL	Homólogo afín a jerky (ratón)
NM_014398	-0.578388	TSC403	Parecido a glicoproteína de membrana asociada a lisosoma
NM_001280	0.578213	CIRBP	Proteína de enlace con ARN inducible por frío
NM_001395	-0.577369	DUSP9	Fosfatasa 9 de especificidad dual
NM_016229	-0.576290	LOC51700	Citocromo b5 reductasa b5R.2
NM_006096	-0.575615	NDRG1	N-myc regulado corriente abajo
NM_001552	0.575438	IGFBP4	Proteína 4 de enlace con el factor de crecimiento afín a insulina
NM_005558	-0.574818	LAD1	Ladinina 1
Contig54534_RC	0.574784		Pseudogen transportador de glucosa humana
Contig1239_RC	0.573822		Cromosoma humano 16 BAC clon CIT987SK-A-362G6
Contig57173_RC	0.573807		ARNm de Homo sapiens para proteína KIAA1737, cds parcial
NM_004414	-0.573538	DSCR1	Gen1 de la región crítica del síndrome de Down
NM_021103	-0.572722	TMSB10	Timosina, beta 10
NM_002350	-0.571917	LYN	Homólogo del oncógeno relacionado con el sarcoma viral Yamaguchi v-sí-1
Contig51235_RC	0.571049		ADNc de Homo sapiens: fis FLJ23388, clon HEP17008
NM_013384	0.570987	TMSG1	Supresor de metástasis tumoral
NM_014399	0.570936	NET-6	Proteína tetraspan NET-6
Contig26022_RC	-0.570851		ESTs

AB023152	0.570561	KIAA0935	Proteína KIAA0935
NM_021077	-0.569944	NMB	Neuromedina B
NM_003498	-0.569129	SNN	Estannina
U17077	-0.568979	BENE	Proteína BENE
D86985	0.567698	KIAA0232	Producto del gen KIAA0232
NM_006357	-0.567513	UBE2E3	Enzima E2E 3 conjugadora con ubiquitina (homólogo de levadura UBC4/5)
AL049397	-0.567434		ARNm de Homo sapiens; ADNc DKFZp586C1019 (de clon DKFZp586C1019)
Contig64502	0.567433		ESTs, ligeramente parecidos a [M.musculus] desconocido
Contig56298_RC	-0.566892	FLJ13154	Proteína hipotética FLJ13154
Contig46056_RC	0.566634		ESTs, ligeramente parecidos a YZ28 PROTEÍNA HIPOTÉTICA HUMANA ZAP128 [H.sapiens]
AF007153	0.566044		Clon de Homo sapiens 23736 de secuencia de ARNm
Contig1778_RC	-0.565789		ESTs
NM_017702	-0.565789	FLJ20186	Proteína hipotética FLJ20186
Contig39226_RC	0.565761		Fis FLJ12187 de ADNc de Homo sapiens, clon MAMMA1000831
NM_000168	0.564879	GLI3	Miembro GL13 de familia GLI-Kruppel (Síndrome de Greig)
Contig57609_RC	0.564751		ESTs, ligeramente parecido a SUBUNIDAD TFIID 135 KDA T2D3_DE FACTOR de INICIACIÓN DE TRANSCRIPCIÓN HUMANA [H.sapiens]
U45975	0.564602	PIB5PA	Fosfatidilinositol (4,5) bisfosfato 5-fosfatasa, A
AF038182	0.564596		Clon 23860 de Homo sapiens secuencia de ARNm
Contig5348_RC	0.564480		ESTs, ligeramente parecidos al factor de transcripción 1607338A BTF3a [H.sapiens]
NM_001321	-0.564459	CSRP2	Proteína 2 rica en cisteína y glicina
Contig25362_RC	-0.563801		ESTs
NM_001609	0.563782	ACADSB	Acil-Coenzima A dehidrogenasa, cadena ramificada corta
Contig40146	0.563731		ADNc de Homo sapiens wi84e12.x1 NCI_CGAP_Kid12 clon IMAGE: 2400046 3' parecido a PROTEÍNA RAS DE TIPO RAS SW:RASD_DICDI P03967; secuencia de ARNm
NMB_016002	0.563403	LOC51097	Proteína CGI-49
Contig34303_RC	0.563157		ADNc de Homo sapiens; fis FLJ21517, clon COL05829
Contig55883_RC	0.563141		ESTs
NM_017961	0.562479	FLJ20813	Proteína hipotética FLJ20813
M21551	-0.562340	NMB	Neuromedina B
Contig3940_RC	-0.561956	YWHAH	Proteína de activación de tirosina 3- Monooxigenasa-sa/triptofano 5-monooxigenasa, polipéptido eta
AB033111	-0.561746	KIAA1285	Proteína KIAA1285
Contig43410_RC	0.561678		ESTs
Contig42006_RC	-0.561677		ESTs
Contig57272_RC	0.561228		ESTs
G26403	-0.561068	YWHAH	Proteína de activación de tirosina 3-monooxigenasa/triptofano 5-monooxygenasa, polipéptido eta

NM_005915	-0.560813	MCM6	Deficiente en mantenimiento de minicromosoma (mis5, S. pombe) 6
NM_003875	-0.560668	GMPS	Monofosfato sintetasa de guanina
AK000142	0.559651	AK000142	ADNc de Homo sapiens fis FLJ20135 fis, clon COL06818.
NM_002709	-0.559621	PPP1CB	Fosfatasa proteínica 1, subunidad catalítica, isoforma beta
NM_001276	-0.558868	CHI3L1	Tipo 1 de quitinasa 3 (glicoproteína de cartilago-39)
NM_002857	0.558862	PXF	Proteína farnesilatada de peroxisomal
Contig33815_RC	-0.558741	FLJ22833	Proteína hipotética FLJ22833
NM_003740	-0.558491	KCNK5	Canal de potasio, subfamilia K, miembro 5 (TASK-2)
Contig53646_RC	0.558455		ESTs
NM_005538	-0.558350	INHBC	Inhibina, beta C
NM_002111	0.557860	HD	Huntingtina (enfermedad de Huntington)
NM_003683	-0.557807	D21S2056E	Segmento de ADN en cromosoma 21 (único) secuencia expresada 2056
NM_003035	-0.557380	SIL	TAL1 (SCL) interrupting locus
Contig4388_RC	-0.557216		Homo sapiens, parecida a proteína integral 3 de membrana, clon MGC:3011, ARNm, cds completo
Contig38288_RC	-0.556426		ESTs, ligeramente parecido a proteína ISHUSS disulfuro-isomerasa [H.sapiens]
NM_015417	0.556184	DKFZP434I114	Proteína DKFZP434I114
NM_015507	-0.556138	EGFL6	Dominio afín a EGF, múltiple 6
AF279865	0.555951	KIF13B	Miembro de la familia de la quinesia 13B
Contig31288_RC	-0.555754		ESTs
NM_002966	-0.555620	S100A10	Proteína A10 de enlace con calcio S100 (ligando de anexina II, calpactina I, polipéptido ligero (p11))
NM_017585	-0.555476	SLC2A6	Familia 2 del portador de soluto 2 (transportador de glucosa facilitado), miembro 6
NM_013296	-0.555367	HSU54999	Proteína LGN
NM_000224	0.554838	KRT18	Queratina 18
Contig49270_RC	-0.554593	KIAA1553	Proteína KIAA1553
NM_004848	-0.554538	ICB-1	Gen inducido por membrana basal
NM_007275	0.554278	FUS1	Candidato de cáncer de pulmón
NM_007044	-0.553550	KATNA1	Katanina p60 (contenedora de ATPasa) subunidad A 1
Contig1829	0.553317		ESTs
AF272357	0.553286	NPDC1	Proliferación, diferenciación y control neural, 1
Contig57584_RC	-0.553080		Homo sapiens, parecido a agrupamiento rico en, gen C8, clon MGC:2577, ARNm, cds completo
NM_003039	-0.552747	SLC2A5	Familia 2 del portador de soluto 2 (transportador de glucosa facilitado), miembro 5
NM_014216	0.552321	ITPK1	inositol 1,3,4-trifosfato 5/6 quinasa
NM_007027	-0.552064	TOPBP1	Proteína de enlace con topoisomerasa (ADN) II
AF118224	-0.551916	ST14	Supresión de tumorigenicidad 14 (carcinoma de colon, matriptasa, epitina)
X75315	-0.551853	HSRNASEB	Seb4D
NM_012101	-0.551824	ATDC	Proteína asociada al grupo D de ataxia-telangiectasia

AL157482	-0.551329	FLJ23399	Proteína hipotética FLJ23399
NM_012474	-0.551150	UMPK	Uridina monofosfato quinasa
Contig57081_RC	0.551103		ESTs
NM_006941	-0.551069	SOX10	SRY (región Y de determinación de sexo)-caja 10
NM_004694	0.550932	SLC16A6	Familia 2 del portador de soluto 16 (transportadores de ácido monocarboxílico), miembro 6
Contig9541_RC	0.550680		ESTs
Contig20617_RC	0.550546		ESTs
NM_004252	0.550365	SLC9A3R1	Familia 9 del portador intercambiador de sodio/hidrógeno), isoforma 3 factor regulador 1
NM_015641	-0.550200	KFZP586B2022	Testina
NM_004336	-0.550164	BUB1	Injerto desinhibido por benzimidazoles 1 (homólogo de levadura)
Contig39960_RC	-0.549951	FLJ21079	Proteína hipotética FLJ21079
NM_020686	0.549659	NPD009	Proteína NPD009
NM_002633	-0.549647	PGM1	Fosfoglucomutasa 1
Contig30480_RC	0.548932		ESTs
NM_003479	0.548896	PTP4A2	proteína tirosina fosfatasa tipo IVA, miembro 2
NM_001679	-0.548768	ATP1 B3	ATPasa, Na <sup>+</sup> /K <sup>+</sup> transportador, polipéptido beta 3
NM_001124	-0.548601	ADM	Adrenomedulina
NM_001216	-0.548375	CA9	Anhidrasa carbónica IX
U58033	-0.548354	MTMR2	Proteína 2 relacionada con miotubularina
NM_018389	-0.547875	FLJ11320	Proteína hipotética FLJ11320
AF176012	0.547867	JDP1	Dominio J que contiene proteína 1
Contig66705_RC	-0.546926	ST5	Supresión de tumorigenicidad 5
NMB_018194	0.546878	FLJ10724	Proteína hipotética FLJ10724
NM_006851	-0.546823	RTVP1	Proteína relacionada con la patogénesis de glioma
Contig53870_RC	0.546756		ESTs
NM_002482	-0.546012	NASP	Proteína del esperma autoanígeno nuclear (de enlace con histona)
NM_002292	0.545949	LAMB2	Laminina, beta 2 (laminina S)
NMB_014696	-0.545758	KIAA0514	Producto del gen KIAA0514
Contig49855	0.545517		ESTs
AL117666	0.545203	DKFZP586	DKFZP58601624 Proteína O1624
NM_004701	-0.545185	CCNB2	Ciclina B2
NM_007050	0.544890	PTPRT	Proteína tirosina fosfatasa, tipo receptor, T
NMB_000414	0.544778	HSD17B4	Hidroxisteroide (17-beta) dehidrogenasa 4
Contig52398_RC	-0.544775		ADNc de Homo sapiens: fis FLJ21950, clon HEP04949
AB007916	0.544496	KIAA0447	Producto del gen KIAA0447
Contig66219_RC	0.544467	FLJ22402	Proteína hipotética FLJ22402
D87453	0.544145	KIAA0264	Proteína KIAA0264
NM_015515	-0.543929	DKFZP434G032	Proteína DKFZP434G032
NM_001530	-0.543898	HIF1A	Factor 1 inducible por hipoxia, subunidad alfa (factor de transcripción helix básico-bucle-helix)
NM_004109	-0.543893	FDX1	Ferredoxina 1
NM_000381	-0.543871	MID1	Midlina 1 (síndrome de Opitz/BBB)
Contig43983_RC	0.543523	CS2	Calsinténina-2

AL137761	0.543371		ARNm de Homo sapiens; ADNc DKFZp586L2424 (del clon DKFZp586L2424)
NM_005764	-0.543175	DD96	Proteína epitelial aumentada en carcinoma, proteína 17 asociada a membrana
Contig1838_RC	0.542996		ADNc de Homo sapiens: fis FLJ22722, clon HSI14444
NM_006670	0.542932	5T4	Oncofetal trofoblasto glicoproteína
Contig28552_RC	-0.542617		ARN de Homo sapiens; ADNc DKFZp434C0931 (del clon DKFZp434C0931); cds parcial
Contig14284_RC	0.542224		ESTs
NM_006290	-0.542115	TNFAIP3	Factor de necrosis tumoral, proteína 3 inducida por alfa
AL050372	0.541463		ARNm de Homo sapiens; ADNc DKFZp434A091 (del clon DKFZp434A091); cds parcial
NM_014181	-0.541095	HSPC159	Proteína HSPC159
Contig37141_RC	0.540990		ADNc de Homo sapiens: fis FLJ23582, clon LNG13759
NM_000947	-0.540621	PRIM2A	Primasa, polipéptido 2A (58kD)
NMB_002136	0.540572	HNRPA1	Ribonucleoproteína A1 nuclear heterogénea
NM_004494	-0.540543	HDGF	Factor del crecimiento derivado de hepatoma (proteína de tipo 1 del grupo de alta movilidad)
Contig38983_RC	0.540526		ESTs
Contig27882_RC	-0.540506		ESTs
Z11887	-0.540020	MMP7	Metaloproteínasa 7 de matriz (matrilisina, uterina)
NM_014575	-0.539725	SCHIP-1	Proteína 1 interactuante con schwannomina
Contig38170_RC	0.539708		ESTs
Contig44064_RC	0.539403		ESTs
U68385	0.539395	MEIS3	Homólogo 3 de Meis (ratón)
Contig51967_RC	0.538952		ESTs
Contig37562_RC	0.538657		ESTs, ligeramente parecidos a la proteína relacionada con transformación [H.sapiens]
Contig40500_RC	0.538582		ESTs, ligeramente parecidos a producto proteínico sin nombre [H.sapiens]
Contig1129_RC	0.538339		ESTs
NM_002184	0.538185	IL6ST	Transductor de señal de interleuquina 6 (gp130, receptor de oncostatina M)
AL049381	0.538041		ADNc de Homo sapiens fis FLJ12900, clon NT2RP2004321
NM_002189	-0.537867	IL15RA	Receptor de interleuquina 15, alfa
NM_012110	-0.537562	CHIC2	Dominio hidrófobo 2 rico en cisteína
AB040881	-0.537473	KIAA1448	Proteína KIAA1448
NM_016577	-0.537430	RAB6B	RAB6B, miembro de la familia del oncógeno RAS
NM_001745	0.536940	CAMLG	Ligando modulador de calcio
NM_005742	-0.536738	P5	Proteína relacionada con proteína disulfuro isomerasa
AB011132	0.536345	KIAA0560	Producto del gen KIAA0560
Contig54898_RC	0.536094	PNN protein	Pinina, asociada a desmosoma
Contig45049_RC	-0.536043	FUT4	Fucosiltransferasa 4 (alfa (1,3) fucosiltransferasa, específico de mieloides)

NM_006864	-0.535924	LILRB3	Receptor afín a inmunoglobulina de leucocitos, subfamilia B (con dominios TM e ITIM), miembro 3
Contig53242_RC	-0.535909		ADNc de Homo sapiens fis FLJ11436, clon HEMBA1001213
NM_005544	0.535712	IRS1	Sustrato 1 del receptor de insulina
Contig47456_RC	0.535493	CACNA1D	Canal de calcio, dependiente del voltaje, tipo L, subunidad alfa 1 D
Contig42751_RC	-0.535469		ESTs
Contig29126_RC	-0.535186		ESTs
NM_012391	0.535067	PDEF	Factor de transcripción de Ets específico del epitelio de la próstata
NMB_012429	0.534974	SEC14L2	SEC14 tipo 2 de (S. cerevisiae)
NMB_018171	0.534898	FLJ10659	Proteína hipotética FLJ10659
Contig53047_RC	-0.534773	TTYH1	Homólogo 1 de tweety (Drosophila)
Contig54968_RC	0.534754		ADNc de Homo sapiens fis FLJ13558, clon PLACE1007743
Contig2099_RC	-0.534694	KIAA1691	Proteína KIAA9691
NM_005264	0.534057	GFRA1	Alfa 1 de receptor de familia GDNF
NM_014036	-0.533638	SBBI42	Precursor de proteína de membrana de tipo BCM
NMB_018101	-0.533473	FLJ10468	Proteína hipotética FLJ10468
Contig56765_RC	0.533442	K02E10.2	ESTs, moderadamente parecidos a [C.elegans]
AB006746	-0.533400	PLSCR1	Fosfolípido escramblasa 1
NMB_001089	0.533350	ABCA3	Casete de enlace con ATP, subfamilia A (ABC1), miembro 3
NMB_018188	-0.533132	FLJ10709	Proteína hipotética FLJ10709
X94232	-0.532925	MAPRE2	Proteína asociada a microtúbulo, familia RP/EB, miembro 2
AF234532	-0.532910	MYO10	Miosina X
Contig292_RC	0.532853	FLJ22386	Proteína hipotética FLJ22386
NMB_000101	-0.532767	CYBA	Citocromo b-245, polipéptido alfa
Contig47814_RC	-0.532656	HHGP	Proteína HHGP
NM_014320	-0.532430	SOUL	Proteína de enlace con heme putativo
NM_020347	0.531976	LZTFL1	Tipo 1 de factor de transcripción de cremallera de Leucina
NM_004323	0.531936	BAG1	Atanógeno asociado a BCL2
Contig50850_RC	-0.531914		ESTs
Contig11648_RC	0.531704		ESTs
NMB_018131	-0.531559	FLJ10540	Proteína hipotética FLJ10540
NM_004688	-0.531329	NMI	Interactor N-myc (y STAT)
NM_014870	0.531101	KIAA0478	Producto del gen KIAA0478
Contig31424_RC	0.530720		ESTs
NM_000874	-0.530545	IFNAR2	Receptor 2 de interferón (alfa, beta y omega)
Contig50588_RC	0.530145		ESTs
NMB_016463	0.529998	HSPC195	Proteína hipotética
NMB_013324	0.529966	CISH	Proteína contenedora de SH2 inducible por citoquina
NM_006705	0.529840	GADD45G	Paro de crecimiento e inducible a daño de ADN, gamma
Contig38901_RC	-0.529747		ESTs
NM_004184	-0.529635	WARS	Triptofanil-ARNt sintetasa
NM_015955	-0.529538	LOC51072	Proteína CGI-27
AF151810	0.529416	CGI-52	Parecido a proteína 2 de transferencia de



			fosfatidilcolina
NMB_002164	-0.529117	INDO	Indoleamina-pirrol 2,3 dioxigenasa
NM_004267	-0.528679	CHST2	Carbohidrato (condroitina 6/queratano) sulfotransferasa 2
Contig32185_RC	-0.528529		ADNc de Homo sapiens fis FJ13997, clon Y79AA1002220
NM_004154	-0.528343	P2RY6	Receptor pirimidinergico P2Y, acoplado a proteína G, 6
NM_005235	0.528294	ERBB4	Tipo 4 de homólogo de oncógeno viral de leucemia eritoblástica aviar v-erb-a
Contig40208_RC	-0.528062	LOC56938	Factor de transcripción BMAL2
NMB_013262	0.527297	MIR	Proteína interactuante con la cadena ligera reguladora de miosina
NM_003034	-0.527148	SIAT8A	Sialiltransferasa 8 (alfa-N-acetilneuraminato: alfa-2,8- sialiltransferasa, sintasa GD3) A
NM_004556	-0.527146	NFKBIE	Factor nuclear de potenciador del gen polipéptido ligero kappa en inhibidor de células B, epsilon
NM_002046	-0.527051	GAPD	Gliceraldehído-3-fosfato dehidrogenasa
NMB_001905	-0.526986	CTPS	Sintasa CTP
Contig42402_RC	0.526852		ESTs
NM_014272	-0.526283	ADAMTS7	Tipo A de disintegrina y metaloproteasa (afín a repolisina) con motivo de tipo 1 de trombospondina, 7
AF076612	0.526205	CHRD	Cordina
Contig57725_RC	-0.526122		ARNm de Homo sapiens para factor de transcripción TCF-3 de caja HMG, cds completo
Contig42041_RC	-0.525877		ESTs
Contig44656_RC	-0.525868		ESTs, muy parecido al precursor del receptor de S02392 alfa-2-macroglobulina [H.sapiens]
NMB_018004	-0.525610	FLJ10134	Proteína hipotética FLJ10134
Contig56434_RC	0.525510		ADNc de Homo sapiens fis FLJ13603, clon PLACE1010270
D25328	-0.525504	PFKP	Fosfofructoquinasa, plaqueta
Contig55950_RC	-0.525358	FLJ22329	Proteína hipotética FLJ22329
NM_002648	-0.525211	PIM1	Oncógeno pim-1
AL157505	0.525186		ARNm de Homo sapiens; ADNc DKFZp586P1124 (del clon DKFZp586P1124)
AF061034	-0.525185	FIP2	ARNm de Homo sapiens alternativamente convertido a FIP2, cds completo
NMB_014721	-0.525102	KIAA0680	Producto del gen KIAA0680
NMB_001634	-0.525030	AMD1	S-adenosilmetionina decarboxilasa 1
NM_006304	-0.524911	DSS1	Borrado en región 1 de mano hendida-pie hendido
Contig37778_RC	0.524667		ESTs, muy parecidos a HLHUSB MHC clase II antígeno de histocompatibilidad precursor de cadena HLA-DP alfa-1 [H.sapiens]
NM_003099	0.524339	SNX1	Nexina 1 de clasificación
AL079298	0.523774	MCCC2	Metilcrotonoil-Coenzima A carboxilasa 2 (beta)
NM_019013	-0.523663	FLJ10156	Proteína hipotética
NM_000397	-0.523293	CYBB	Citocromo b-245, polipéptido beta (enfermedad granulomatosa crónica)

NM_014811	0.523132	KIAA0649	Producto del gen KIAA0649
Contig20600_RC	0.523072		ESTs
NMB_005190	-0.522710	CCNC	Ciclina C
AL161960	-0.522574	FLJ21324	Proteína hipotética FLJ21324
AL117502	0.522280		ARNm de Homo sapiens; ADNc DKFZp434D0935 (del clon DKFZp434D0935)
AF131753	-0.522245		Secuencia de ARNm del clon 24859 de Homo sapiens
NM_000320	0.521974	QDPR	Dihidropteridina reductasa quinoide
NMB_002115	-0.521870	HK3	Hexoquinasa 3 (célula blanca)
NM_006460	0.521696	HIS1	Inducible por HMBA
NMB_018683	-0.521679	ZNF313	Proteína de dedo de cinc 313
NM_004305	-0.521539	BIN1	Integrador puente 1
NM_006770	-0.521538	MARCO	Receptor macrófago con estructura colágena
NM_001166	-0.521530	BIRC2	Contenedor de reptición 2 del baculoviral IAP
D42047	0.521522	KIAA0089	Proteína KIAA0089
NMB_016235	-0.521298	GPRC5B	Receptor acoplado a la proteína G, familia C, grupo 5, miembro B
NM_004504	-0.521189	HRB	Proteína de enlace con VIH-1 Rev
NM_002727	-0.521146	PRG1	Proteoglicano 1, gránulo secretor
AB029031	-0.520761	KIAA1108	Proteína KIAA1108
NM_005556	-0.520692	KRT7	Queratina 7
NMB_018031	0.520600	WDR6	Dominio 6 de repetición de WD
AL117523	-0.520579	KIAA1053	Proteína KIAA1053
NMB_004515	-0.520363	ILF2	Factor 2 de enlace con el potenciador de interleuquina, 45kD
NM_004708	-0.519935	PDCD5	Muerte programada de célula 5
NM_005935	0.519765	MLLT2	Leucemia mielóide/linfóide o de trazado mixto (homólogo de tritórax (Drosophila); trasladado a 2
Contig49289_RC	-0.519546		ARNm de Homo sapiens; ADNc DKFZp586J1119 (del clon DKFZp586J1119); cds completo
NMB_000211	-0.519342	ITGB2	Integrina, beta 2 (antígeno CD18 (p95), antígeno 1 asociado a la función de linfocito; antígeno macrófago 1 (mac-1) subunidad beta)
AL079276	0.519207	LOC58495	Proteína putativa de dedo de cinc procedente de EUROIMAGE 566589
Contig57825_RC	0.519041		ESTs
NM_002466	-0.518911	MYBL2	Tipo 2 de homólogo de oncógeno viral de leucemia eritoblástica aviar v-erb-a
NMB_016072	-0.518802	LOC51026	Proteína CGI-141
AB007950	-0.518699	KIAA0481	Producto del gen KIAA0481
NMB_001550	-0.518549	IFRD1	Regulador de desarrollo 1 relacionado con el interferón
AF155120	-0.518221	UBE2V1	Variante 1 de la enzima E que conjuga con ubiquitina
Contig49849_RC	0.517983		ESTs, ligeramente parecidos a la proteína AF188706 1 g20 [H.sapiens]
NMB_016625	-0.517936	LOC51319	Proteína hipotética
NM_004049	-0.517862	BCL2A1	Proteína A1 relacionada con BCL2
Contig50719_RC	0.517740		ESTs
D80010	-0.517620	LPIN1	Lipina 1

NM_000299	-0.517405	PKP1	Placofilina 1 (síndrome de displasia ectodérmica/fragilidad epidérmica)
AL049365	0.517080	FTL	Ferritina, polipéptido ligero
Contig65227	0.517003		ESTs
NM_004865	-0.516808	TBPL1	Tipo 1 de TBP
Contig54813_RC	0.516246	FLJ13962	Proteína hipotética FLJ13962
NM_003494	-0.516221	DYSF	Disferlina, distrofia de miopatía muscular de miembros 2B (autosómico recesivo)
NM_004431	-0.516212	EPHA2	EphA2
AL117600	-0.516067	DKFZP564 J0863	Proteína DKFZP564J0863
AL080209	-0.516037	DKFZP586 F2423	Proteína hipotética DKFZp586F2423
NM_000135	-0.515613	FANCA	Anemia de Fanconi, grupo de complementación A
NM_000050	-0.515494	ASS	Argininosuccinato sintetasa
NMB_001830	-0.515439	CLCN4	Canal de cloruro 4
NMB_018234	-0.515365	FLJ10829	Proteína hipotética FLJ10829
Contig53307_RC	0.515328		ESTs, ligeramente parecidos a la proteína KIAA1437 [H.sapiens]
AL117617	-0.515141		ARNm de Homo sapiens; ADNc DKFZp564H0764 (del clon DKFZp564H0764)
NM_002906	-0.515098	RDX	Radixina
NMB_003360	-0.514427	UGT8	UDP glicosiltransferasa 8 (UDP-galactosa ceramida galactosiltransferasa)
NM_018478	0.514332	HSMNP1	Proteína del hipotálamo no caracterizada HSMNP1
M90657	-0.513908	TM4SF1	Miembro 1 de la superfamilia de la transmembrana 4
NM_014967	0.513793	KIAA1018	Proteína KIAA1018
Contig1462_RC	0.513604	C11ORF15	Marco de lectura abierta 15 del cromosoma 11
ontig37287_RC	-0.513324		ESTs
NM_000355	-0.513225	TCN2	Transcobalamina II; anemia macrocítica anemia
AB037756	0.512914	KIAA1335	hypothetical protein KIAA1335
Contig842_RC	-0.512880		ESTs
NMB_018186	-0.512878	FLJ10706	Proteína hipotética FLJ10706
NM_014668	0.512746	KIAA0575	Producto del gen KIAA0575
NM_003226	0.512611	TFF3	Factor de trébol 3 (intestinal)
Contig56457_RC	-0.512548	TMEFF1	Proteína de transmembrana con afín a EGF y dos dominios 1 afines a la folistatina
AL050367	-0.511999		ARNm de Homo sapiens; ADNc DKFZp564A026 (del clon DKFZp564A026)
NM_014791	-0.511963	KIAA0175	Producto del gen KIAA0175
Contig36312_RC	0.511794		ESTs
NM_004811	-0.511447	LPXN	Leupaxina
Contig67182_RC	-0.511416		ESTs, muy parecidos al precursor del antígeno epitelial de tipo V [H.sapiens]
Contig52723_RC	-0.511134		ESTs
Contigl 7105_RC	-0.511072		ARNm de Homo sapiens para proteína citoplasmática putativa (ORF1-FL21)
NMB_014449	0.511023	A	Proteína "A"

Contig52957_RC	0.510815		ESTs
Contig49388_RC	0.510582	FLJ13322	Proteína hipotética FLJ13322
NM_017786	0.510557	FLJ20366	Proteína hipotética FLJ20366
AL157476	0.510478		ARNm de Homo sapiens; ADNc DKFZp761 C082 (del clon DKFZp761 C082)
NMB_001919	0.510242	DCI	Dodecenoil-Coenzima A delta isomerasa (3,2 transenoil-Coenzima A isomerasa)
NM_000268	-0.510165	NF2	Neurofibromina 2 (neuroma acústico bilateral)
NMB_016210	0.510018	LOC51161	Proteína g20
Contig45816_RC	-0.509977		ESTs
NM_003953	-0.509969	MPZL1	Proteína de la mielina 1 de tipo cero
NM_000057	-0.509669	BLM	Síndrome de Bloom
NM_014452	-0.509473	DR6	Receptor de muerte 6
Contig45156_RC	0.509284		ESTs, moderadamente parecidos al dominio motor de KIF12 [M.musculus]
NM_006943	0.509149	SOX22	SRY (región determinante del sexo Y)-caja 22
NM_000594	-0.509012	TNF	Factor de necrosis tumoral (superfamilia TNF, miembro 2)
AL137316	-0.508353	KIAA1609	Proteína KIAA1609
NM_000557	-0.508325	GDF5	Factor de diferenciación del crecimiento 5 (proteína morfogénica 1 derivada de cartílago)
NMB_018685	-0.508307	ANLN	Anilina (homólogo de Detritus de Drosophila), proteína de enlace con actina
Contig53401_RC	0.508189		ESTs
NM_014364	-0.508170	GAPDS	Gliceraldehído-3-fosfato dehidrogenasa, específica de los testículos
Contig50297_RC	0.508137		ESTs, moderadamente parecidos a ALU8_ALU HUMANO SUBFAMILIA SX ENTRADA DE AVISO DE CONTAMINACIÓN DE SECUENCIA [H.sapiens]
Contig51800	0.507891		ESTs, moderadamente parecidos a ALU6_ALU HUMANO SUBFAMILIA SP ENTRADA DE AVISO DE CONTAMINACIÓN DE SECUENCIA [H.sapiens]
Contig49098_RC	-0.507716	MGC4090	Proteína hipotética MGC4090
NM_002985	-0.507554	SCYA5	Citoquina inducible pequeña A5 (RANTES)
AB007899	0.507439	KIAA0439	Proteína KIAA0439; homólogo de ubiquitina-proteína ligasa de levadura Rsp5
AL110139	0.507145		ARNm de Homo sapiens; ADNc DKFZp56401763 (del clon DKFZp56401763)
Contig51117_RC	0.507001		ESTs
NMB_017660	-0.506768	FLJ20085	Proteína hipotética FLJ20085
NM_018000	0.506686	FLJ10116	Proteína hipotética FLJ10116
NM_005555	-0.506516	KRT6B	Queratina 6B
NM_005582	-0.506462	LY64	Homólogo de antígeno de linfocito 64 (ratón homolog, radioprotector, 105kD)
Contig47405_RC	0.506202		ESTs
NM_014808	0.506173	KIAA0793	Producto del gen KIAA0793
NM_004938	-0.506121	DAPK1	Proteína quinasa 1 asociada a la muerte
NM_020659	-0.505793	TTYH1	Homólogo 1 de tweety (Drosophila)
NM_006227	-0.505604	PLTP	Proteína de transferencia de fosfolípidos

NMB_014268	-0.505412	MAPRE2	Proteína asociada a microtúbulo, familia RP/EB, miembro 2
NM_004711	0.504849	SYNGR1	Sinaptogirina 1
NMB_004418	-0.504497	DUSP2	Fosfatasa 2 de especificidad dual
NM_003508	-0.504475	FZD9	Homólogo 9 (Drosophila) rizado

**Tabla 3. 430 marcadores de genes que distinguen entre muestras tumorales relacionadas con BRCA1-1 y muestras tumorales esporádicas**

5

Bando de Genes Número de Acceso	SEQ ID NO	Bando de Genes Número de Acceso	SEQ ID NO
AB002301	SEQ ID NO 4	NM_012391	SEQ ID NO 1406
AB004857	SEQ ID NO 8	NM_012428	SEQ ID NO 1412
AB007458	SEQ ID NO 12	NM_013233	SEQ ID NO 1418
AB014534	SEQ ID NO 29	NM_013253	SEQ ID NO 1422
AB018305	SEQ ID NO 34	NM_013262	SEQ ID NO 1425
AB020677	SEQ ID NO 36	NM_013372	SEQ ID NO 1434
AB020689	SEQ ID NO 37	NM_013378	SEQ ID NO 1435
AB023151	SEQ ID NO 41	NM_014096	SEQ ID NO 1450
AB023163	SEQ ID NO 43	NM_014242	SEQ ID NO 1464
AB028986	SEQ ID NO 48	NM_014314	SEQ ID NO 1472
AB029025	SEQ ID NO 50	NM_014398	SEQ ID NO 1486
AB032966	SEQ ID NO 53	NM_014402	SEQ ID NO 1488
AB032988	SEQ ID NO 57	NM_014476	SEQ ID NO 1496
AB033049	SEQ ID NO 63	NM_014521	SEQ ID NO 1499
AB033055	SEQ ID NO 66	NM_014585	SEQ ID NO 1504
AB037742	SEQ ID NO 73	NM_014597	SEQ ID NO 1506
AB041269	SEQ ID NO 96	NM_014642	SEQ ID NO 1510
AF000974	SEQ ID NO 97	NM_014679	SEQ ID NO 1517
AF042838	SEQ ID NO 111	NM_014680	SEQ ID NO 1518
AF052155	SEQ ID NO 119	NM_014700	SEQ ID NO 1520
AF055084	SEQ ID NO 125	NM_014723	SEQ ID NO 1523
AF063725	SEQ ID NO 129	NM_014770	SEQ ID NO 1530
AF070536	SEQ ID NO 133	NM_014785	SEQ ID NO 1534
AF070617	SEQ ID NO 135	NM_014817	SEQ ID NO 1539
AF073299	SEQ ID NO 136	NM_014840	SEQ ID NO 1541
AF079529	SEQ ID NO 140	NM_014878	SEQ ID NO 1546
AF090353	SEQ ID NO 141	NM_015493	SEQ ID NO 1564
AF116238	SEQ ID NO 155	NM_015523	SEQ ID NO 1568
AF151810	SEQ ID NO 171	NM_015544	SEQ ID NO 1570
AF220492	SEQ ID NO 185	NM_015623	SEQ ID NO 1572
AJ224741	SEQ ID NO 196	NM_015640	SEQ ID NO 1573
AJ250475	SEQ ID NO 201	NM_015721	SEQ ID NO 1576
AJ270996	SEQ ID NO 202	NM_015881	SEQ ID NO 1577
AJ272057	SEQ ID NO 203	NM_015937	SEQ ID NO 1582
AK000174	SEQ ID NO 211	NM_015964	SEQ ID NO 1586
AK000617	SEQ ID NO 215	NM_015984	SEQ ID NO 1587

AK000959	SEQ ID NO 222	NM_016000	SEQ ID NO 1591
AK001438	SEQ ID NO 229	NM_016018	SEQ ID NO 1593
AK001838	SEQ ID NO 233	NM_016066	SEQ ID NO 1601
AK002107	SEQ ID NO 238	NM_016073	SEQ ID NO 1603
AK002197	SEQ ID NO 239	NM_016081	SEQ ID NO 1604
AL035297	SEQ ID NO 241	NM_016140	SEQ ID NO 1611
AL049346	SEQ ID NO 243	NM_016223	SEQ ID NO 1622
AL049370	SEQ ID NO 245	NM_016267	SEQ ID NO 1629
AL049667	SEQ ID NO 249	NM_016307	SEQ ID NO 1633
AL080222	SEQ ID NO 276	NM_016364	SEQ ID NO 1639
AL096737	SEQ ID NO 279	NM_016373	SEQ ID NO 1640
AL110163	SEQ ID NO 282	NM_016459	SEQ ID NO 1646
AL133057	SEQ ID NO 300	NM_016471	SEQ ID NO 1648
AL133096	SEQ ID NO 302	NM_016548	SEQ ID NO 1654
AL133572	SEQ ID NO 305	NM_016620	SEQ ID NO 1662
AL133619	SEQ ID NO 307	NM_016820	SEQ ID NO 1674
AL133623	SEQ ID NO 309	NM_017423	SEQ ID NO 1678
AL137347	SEQ ID NO 320	NM_017709	SEQ ID NO 1698
AL137381	SEQ ID NO 322	NM_017732	SEQ ID NO 1700
AL137461	SEQ ID NO 325	NM_017734	SEQ ID NO 1702
AL137540	SEQ ID NO 328	NM_017750	SEQ ID NO 1704
AL137555	SEQ ID NO 329	NM_017763	SEQ ID NO 1706
AL137638	SEQ ID NO 332	NM_017782	SEQ ID NO 1710
AL137639	SEQ ID NO 333	NM_017816	SEQ ID NO 1714
AL137663	SEQ ID NO 334	NM_018043	SEQ ID NO 1730
AL137761	SEQ ID NO 339	NM_018072	SEQ ID NO 1734
AL157431	SEQ ID NO 340	NM_018093	SEQ ID NO 1738
AL161960	SEQ ID NO 351	NM_018103	SEQ ID NO 1742
AL355708	SEQ ID NO 353	NM_018171	SEQ ID NO 1751
AL359053	SEQ ID NO 354	NM_018187	SEQ ID NO 1755
D26488	SEQ ID NO 359	NM_018188	SEQ ID NO 1756
D38521	SEQ ID NO 361	NM_018222	SEQ ID NO 1761
D50914	SEQ ID NO 367	NM_018228	SEQ ID NO 1762
D80001	SEQ ID NO 369	NM_018373	SEQ ID NO 1777
G26403	SEQ ID NO 380	NM_018390	SEQ ID NO 1781
K02276	SEQ ID NO 383	NM_018422	SEQ ID NO 1784
M21551	SEQ ID NO 394	NM_018509	SEQ ID NO 1792
M27749	SEQ ID NO 397	NM_018584	SEQ ID NO 1796
M28170	SEQ ID NO 398	NM_018653	SEQ ID NO 1797
M73547	SEQ ID NO 409	NM_018660	SEQ ID NO 1798
M80899	SEQ ID NO 411	NM_018683	SEQ ID NO 1799
NM_000067	SEQ ID NO 423	NM_019049	SEQ ID NO 1814
NM_000087	SEQ ID NO 427	NM_019063	SEQ ID NO 1815
NM_000090	SEQ ID NO 428	NM_020150	SEQ ID NO 1823
NMB_000165	SEQ ID NO 444	NM_020987	SEQ ID NO 1848
NM_000168	SEQ ID NO 445	NM_021095	SEQ ID NO 1855
NM_000196	SEQ ID NO 449	NM_021242	SEQ ID NO 1867

NM_000269	SEQ ID NO 457	U41387	SEQ ID NO 1877
NM_000310	SEQ ID NO 466	U45975	SEQ ID NO 1878
NM_000396	SEQ ID NO 479	U58033	SEQ ID NO 1881
NM_000397	SEQ ID NO 480	U67784	SEQ ID NO 1884
NM_000597	SEQ ID NO 502	U68385	SEQ ID NO 1885
NM_000636	SEQ ID NO 509	U80736	SEQ ID NO 1890
NM_000888	SEQ ID NO 535	X00437	SEQ ID NO 1899
NM_000903	SEQ ID NO 536	X07203	SEQ ID NO 1904
NM_000930	SEQ ID NO 540	X16302	SEQ ID NO 1907
NM_000931	SEQ ID NO 541	X51630	SEQ ID NO 1908
NM_000969	SEQ ID NO 547	X57809	SEQ ID NO 1912
NM_000984	SEQ ID NO 548	X57819	SEQ ID NO 1913
NM_001026	SEQ ID NO 552	X58529	SEQ ID NO 1914
NM_001054	SEQ ID NO 554	X66087	SEQ ID NO 1916
NM_001179	SEQ ID NO 567	X69150	SEQ ID NO 1917
NM_001184	SEQ ID NO 568	X72475	SEQ ID NO 1918
NM_001204	SEQ ID NO 571	X74794	SEQ ID NO 1920
NM_001206	SEQ ID NO 572	X75315	SEQ ID NO 1921
NM_001218	SEQ ID NO 575	X84340	SEQ ID NO 1925
NM_001275	SEQ ID NO 586	X98260	SEQ ID NO 1928
NM_001394	SEQ ID NO 602	Y07512	SEQ ID NO 1931
NM_001424	SEQ ID NO 605	Y14737	SEQ ID NO 1932
NM_001448	SEQ ID NO 610	Z34893	SEQ ID NO 1934
NM_001504	SEQ ID NO 620	Contig237_RC	SEQ ID NO 1940
NM_001553	SEQ ID NO 630	Contig292_RC	SEQ ID NO 1942
NM_001674	SEQ ID NO 646	Contig372_RC	SEQ ID NO 1943
NM_001675	SEQ ID NO 647	Contig756_RC	SEQ ID NO 1955
NM_001725	SEQ ID NO 652	Contig842_RC	SEQ ID NO 1958
NM_001740	SEQ ID NO 656	Contig1632_RC	SEQ ID NO 1977
NM_001756	SEQ ID NO 659	Contig1826_RC	SEQ ID NO 1980
NM_001770	SEQ ID NO 664	Contig2237_RC	SEQ ID NO 1988
NM_001797	SEQ ID NO 670	Contig2915_RC	SEQ ID NO 2003
NM_001845	SEQ ID NO 680	Contig3164_RC	SEQ ID NO 2007
NM_001873	SEQ ID NO 684	Contig3252_RC	SEQ ID NO 2008
NM_001888	SEQ ID NO 687	Contig3940_RC	SEQ ID NO 2018
NM_001892	SEQ ID NO 688	Contig9259_RC	SEQ ID NO 2039
NM_001919	SEQ ID NO 694	Contig10268_RC	SEQ ID NO 2041
NM_001946	SEQ ID NO 698	Contig10437_RC	SEQ ID NO 2043
NM_001953	SEQ ID NO 699	Contig10973_RC	SEQ ID NO 2044
NM_001960	SEQ ID NO 704	Contig14390_RC	SEQ ID NO 2054
NM_001985	SEQ ID NO 709	Contig16453_RC	SEQ ID NO 2060
NM_002023	SEQ ID NO 712	Contig16759_RC	SEQ ID NO 2061
NM_002051	SEQ ID NO 716	Contig19551	SEQ ID NO 2070
NM_002053	SEQ ID NO 717	Contig24541_RC	SEQ ID NO 2088
NM_002164	SEQ ID NO 734	Contig25362_RC	SEQ ID NO 2093
NM_002200	SEQ ID NO 739	Contig25617_RC	SEQ ID NO 2094
NM_002201	SEQ ID NO 740	Contig25722_RC	SEQ ID NO 2096

NM_002213	SEQ ID NO 741	Contig26022_RC	SEQ ID NO 2099
NM_002250	SEQ ID NO 747	Contig27915_RC	SEQ ID NO 2114
NM_002512	SEQ ID NO 780	Contig28081_RC	SEQ ID NO 2116
NM_002542	SEQ ID NO 784	Contig28179_RC	SEQ ID NO 2118
NM_002561	SEQ ID NO 786	Contig28550_RC	SEQ ID NO 2119 D
NM_002615	SEQ ID NO 793	Contig29639_RC	SEQ ID NO 2127
NM_002686	SEQ ID NO 803	Contig29647_RC	SEQ ID NO 2128
NM_002709	SEQ ID NO 806	Contig30092_RC	SEQ ID NO 2130
NM_002742	SEQ ID NO 812	Contig30209_RC	SEQ ID NO 2132
NM_002775	SEQ ID NO 815	Contig32185_RC	SEQ ID NO 2156
NM_002975	SEQ ID NO 848	Contig32798_RC	SEQ ID NO 2161
NM_002982	SEQ ID NO 849	Contig33230_RC	SEQ ID NO 2163
NM_003104	SEQ ID NO 870	Contig33394_RC	SEQ ID NO 2165
NM_003118	SEQ ID NO 872	Contig36323_RC	SEQ ID NO 2197
NM_003144	SEQ ID NO 876	Contig36761_RC	SEQ ID NO 2201
NM_003165	SEQ ID NO 882	Contig37141_RC	SEQ ID NO 2209
NM_003197	SEQ ID NO 885	Contig37778_RC	SEQ ID NO 2218
NM_003202	SEQ ID NO 886	Contig38285_RC	SEQ ID NO 2222
NM_003462	SEQ ID NO 911	Contig39826_RC	SEQ ID NO 2241
NM_003500	SEQ ID NO 918	Contig40212_RC	SEQ ID NO 2251
NM_003561	SEQ ID NO 925	Contig40712_RC	SEQ ID NO 2257
NM_003607	SEQ ID NO 930	Contig41402_RC	SEQ ID NO 2265
NM_003633	SEQ ID NO 933	Contig41635_RC	SEQ ID NO 2272
NM_003641	SEQ ID NO 934	Contig42006_RC	SEQ ID NO 2280
NM_003683	SEQ ID NO 943	Contig42220_RC	SEQ ID NO 2286
NM_003729	SEQ ID NO 949	Contig42306_RC	SEQ ID NO 2287
NM_003793	SEQ ID NO 954	Contig43918_RC	SEQ ID NO 2312
NM_003829	SEQ ID NO 958	Contig44195_RC	SEQ ID NO 2316
NM_003866	SEQ ID NO 961	Contig44265_RC	SEQ ID NO 2318
NM_003904	SEQ ID NO 967	Contig44278_RC	SEQ ID NO 2319
NM_003953	SEQ ID NO 974	Contig44757_RC	SEQ ID NO 2329
NM_004024	SEQ ID NO 982	Contig45588_RC	SEQ ID NO 2349
NM_004053	SEQ ID NO 986	Contig46262_RC	SEQ ID NO 2361
NM_004295	SEQ ID NO 1014	Contig46288_RC	SEQ ID NO 2362
NM_004438	SEQ ID NO 1038	Contig46343_RC	SEQ ID NO 2363
NM_004559	SEQ ID NO 1057	Contig46452_RC	SEQ ID NO 2366
NM_004616	SEQ ID NO 1065	Contig46868_RC	SEQ ID NO 2373
NM_004741	SEQ ID NO 1080	Contig46937_RC	SEQ ID NO 2377
NM_004772	SEQ ID NO 1084	Contig48004_RC	SEQ ID NO 2393
NM_004791	SEQ ID NO 1086	Contig48249_RC	SEQ ID NO 2397
NM_004848	SEQ ID NO 1094	Contig48774_RC	SEQ ID NO 2405
NM_004866	SEQ ID NO 1097	Contig48913_RC	SEQ ID NO 2411
NM_005128	SEQ ID NO 1121	Contig48945_RC	SEQ ID NO 2412
NM_005148	SEQ ID NO 1124	Contig48970_RC	SEQ ID NO 2413
NM_005196	SEQ ID NO 1127	Contig49233_RC	SEQ ID NO 2419
NM_005326	SEQ ID NO 1140	Contig49289_RC	SEQ ID NO 2422
NM_005518	SEQ ID NO 1161	Contig49342_RC	SEQ ID NO 2423



NM_005538	SEQ ID NO 1163	Contig49510_RC	SEQ ID NO 2430
NM_005557	SEQ ID NO 1170	Contig49855_RC	SEQ ID NO 2440
NM_005718	SEQ ID NO 1189	Contig49948_RC	SEQ ID NO 2442
NM_005804	SEQ ID NO 1201	Contig50297_RC	SEQ ID NO 2451
NM_005824	SEQ ID NO 1203	Contig50669_RC	SEQ ID NO 2458
NM_005935	SEQ ID NO 1220	Contig50673_RC	SEQ ID NO 2459
NM_006002	SEQ ID NO 1225	Contig50838_RC	SEQ ID NO 2465
NM_006148	SEQ ID NO 1249	Contig51068_RC	SEQ ID NO 2471
NM_006235	SEQ ID NO 1257	Contig51929	SEQ ID NO 2492
NM_006271	SEQ ID NO 1261	Contig51953_RC	SEQ ID NO 2493
NM_006287	SEQ ID NO 1264	Contig52405_RC	SEQ ID NO 2502
NM_006296	SEQ ID NO 1267	Contig52543_RC	SEQ ID NO 2505
NM_006378	SEQ ID NO 1275	Contig52720_RC	SEQ ID NO 2513
NM_006461	SEQ ID NO 1287	Contig53281_RC	SEQ ID NO 2530
NM_006573	SEQ ID NO 1300	Contig53598_RC	SEQ ID NO 2537
NM_006622	SEQ ID NO 1302	Contig53757_RC	SEQ ID NO 2543
NM_006696	SEQ ID NO 1308	Contig53944_RC	SEQ ID NO 2545
NM_006769	SEQ ID NO 1316	Contig54425	SEQ ID NO 2561
NM_006787	SEQ ID NO 1319	Contig54547_RC	SEQ ID NO 2565
NM_006875	SEQ ID NO 1334	Contig54757_RC	SEQ ID NO 2574
NM_006885	SEQ ID NO 1335	Contig54916_RC	SEQ ID NO 2581
NM_006918	SEQ ID NO 1339	Contig55770_RC	SEQ ID NO 2604
NM_006923	SEQ ID NO 1340	Contig55801_RC	SEQ ID NO 2606
NM_006941	SEQ ID NO 1342	Contig56143_RC	SEQ ID NO 2619
NM_007070	SEQ ID NO 1354	Contig56160_RC	SEQ ID NO 2620
NM_007088	SEQ ID NO 1356	Contig56303_RC	SEQ ID NO 2626
NM_007146	SEQ ID NO 1358	Contig57023_RC	SEQ ID NO 2639
NM_007173	SEQ ID NO 1359	Contig57138_RC	SEQ ID NO 2644
NM_007246	SEQ ID NO 1366	Contig57609_RC	SEQ ID NO 2657
NM_007358	SEQ ID NO 1374	Contig58301_RC	SEQ ID NO 2667
NM_012135	SEQ ID NO 1385	Contig58512_RC	SEQ ID NO 2670
NM_012151	SEQ ID NO 1387	Contig60393	SEQ ID NO 2674
NM_012258	SEQ ID NO 1396	Contig60509_RC	SEQ ID NO 2675
NM_012317	SEQ ID NO 1399	Contig61254_RC	SEQ ID NO 2677
NM_012337	SEQ ID NO 1403	Contig62306	SEQ ID NO 2680
NM_012339	SEQ ID NO 1404	Contig64502	SEQ ID NO 2689

**Tabla 4. 100 marcadores preferidos de la Tabla 3 que distinguen entre tumores relacionados con *BRCA1* y tumores esporádicos.**

Identificador	Correlación	Nombre de secuencia	Descripción
NM_001892	-0.651689	CSNK1A1	Caseína quinasa 1, alfa 1
NM_018171	-0.637696	FLJ10659	Proteína hipotética FLJ10659
Contig40712_RC	-0.612509		ESTs
NM_001204	-0.608470	BMPR2	Receptor de proteína morfogenética de los huesos, tipo II (serina/treonina quinasa)

NM_005148	-0.598612	UNC119	Homólogo de unc119 (C.elegans)
G26403	0.585054	YWHAH	Proteína de activación de tirosina 3- monooxigenasa/triptofano 5-monooxigenasa, polipéptido eta
NM_015640	0.583397	PAI-RBP1	Proteína de enlace con PAI-1 ARNm
Contig9259_RC	0.581362		ESTs
AB033049	-0.578750	KIAA1223	Proteína KIAA1223
NM_015523	0.576029	DKFZP566E144	Pequeño fragmento de nucleasa
Contig41402_RC	-0.571650		Secuencia de ADN humano del clon RP11-16L21 en el cromosoma 9. Contiene el gen para leucotrieno B4 12-hidroxidehidrogenasa dependiente de NDAP, el gen para una nueva proteína de dominio DnaJ parecida a Drosophila, C. elegans y a las proteínas previstas por Arabidopsis, el gen GNG10 para la proteína gene de enlace con nucleótido de guanina 10, un nuevo gen, ESTs, STSs, GSSs y seis islas CpG
NM_004791	-0.564819	ITGBL1	Integrina, tipo 1 beta (con dominios de repetición afines a EGF)
NM_007070	0.561173	FAP48	Proteína asociada a FKBP
NM_014597	0.555907	HSU15552	ARNm ácido de proteína 82 kDa
AF000974	0.547194	TRIP6	Interactor 6 del receptor de hormonas de la tiroides
NM_016073	-0.547072	CGI-142	CGI-142
Contig3940_RC	0.544073	YWHAH	Proteína de activación de tirosina 3-mono-oxygenasa/triptofano 5-monooxigenasa, polipéptido eta
NM_003683	0.542219	D2152056E	Segmento de AND en cromosoma 21 (único) secuencia expresada 2056
Contig58512_RC	-0.528458		ARNm de proteína relacionada con el tumor de páncreas de Homo sapiens (FKSG12), cds 21completo
NM_003904	0.521223	ZNF259	Proteína de dedo de cinc 259
Contig26022_RC	0.517351		ESTs
Contig48970_RC	-0.516953	KIAA0892	Proteína KIAA0892
NM_016307	-0.515398	PRX2	Proteína de caja homeo relacionada apareada
AL137761	-0.514891		ARNm de Homo sapiens; ADNc DKFZp586L2424 (del clon DKFZp586L2424)
NM_001919	-0.514799	DCI	Dodecenol-Coenzima A delta isomerasa (3,2 transenol-Coenzima A isomerasa)
NM_000196	-0.514004	HSD11B2	Hidroxiesteroide (11-beta) dehidrogenasa 2
NM_002200	0.513149	IRF5	Factor 5 regulador del interferón
AL133572	0.511340		ARNm de Homo sapiens; ADNc DKFZp43410535 (del clon DKFZp43410535); cds parcial
NM_019063	0.511127	C2ORF2	Marco de lectura abierta 2 del cromosoma 2
Contig25617_RC	0.509506		ESTs
NM_007358	0.508145	M96	Proteína putativa de enlace con ADN
NM_014785	-0.507114	KIAA0258	Producto del gen KIAA0258
NM_006235	0.506585	POU2AF1	Dominio POU, clase 2, factor asociador 1
NM_014680	-0.505779	KIAA0100	Producto del gen KIAA0100
X66087	0.500842	MYBL1	Tipo 1 del homólogo del oncógeno viral de la mieloblastosis aviar v-myb
Y07512	-0.500686	PRKG1	Proteína quinasa, dependiente de cGMP, tipo I

NM_006296	0.500344	VRK2	Quinasa 2 relacionada con la vaccinia
Contig44278_RC	0.498260	DKFZP434K114	Proteína DKFZP434K114
Contig56160_RC	-0.497695		ESTs
NM_002023	-0.497570	FMOD	Fibromodulina
M28170	0.497095	CD19	Antígeno CD19
D26488	0.496511	KIAA0007	Proteína KIAA0007
X72475	0.496125		ARNm de H.sapiens para región variable de cadena ligera kappa Ig reorganizada (I.114)
K02276	0.496068	MYC	Homólogo del oncógeno viral de la mieloblastosis aviar v-myc
NM_013378	0.495648	VPREB3	Gen 3 de pre-B linfocito
X58529	0.495608	IGHM	Mu constante pesado de inmunoglobulina
NM_000168	-0.494260	GLI3	Miembro GL13 de la familia GLI-Kruppel (síndrome de Greig)
NM_004866	-0.492967	SCAMP1	Proteína 1 de la membrana portadora secretora
NM_003729	0.488971	RPC	Fosfato ciclasa del terminal 3' de ARN
NM_006875	0.487407	PIM2	Oncógeno pim-2
NM_018188	0.487126	FLJ10709	Proteína hipotética FLJ10709
NM_004848	0.485408	ICB-1	Gen inducido por la membrana basal
NM_001179	0.483253	ART3	ADP-ribosiltransferasa 3
NM_016548	-0.482329	LOC51280	Proteína GP73 de la membrana de Golgi
NM_007146	-0.481994	ZNF161	Proteína de dedo de cinc 161
NM_021242	-0.481754	STRAIT11499	Proteína hipotética STRAIT11499
NM_016223	0.481710	PACSIN3	Sustrato de proteína quinasa C y caseína quinasa en neuronas 3
NM_003197	-0.481526	TCEB1L	Factor B de alargamiento de transcripción (SIII), afin a polipéptido 1
NM_000067	-0.481003	CA2	Anhidrasa carbónica II
NM_006885	-0.479705	ATBF1	Factor 1 de transcripción de enlace con AT
NM_002542	0.478282	OGG1	8-oxoguanina ADN glicosilasa
AL133619	-0.476596		ARNm de Homo sapiens mRNA; ADNc DKFZp434E2321 (del clon DKFZp434E2321); cds parcial cds
D80001	0.476130	KIAA0179	Proteína KIAA0179
NM_018660	-0.475548	LOC55893	Factor regulador del virus del papiloma PRF-1
AB004857	0.473440	SLC11A2	Familia 11 del portador de soluto (transportadores de iones de metal divalente acoplados a protones), miembro 2
NM_002250	0.472900	KCNN4	Canal activado por calcio de intermedia/pequeña conductancia de potasio, subfamilia N, miembro 4
Contig56143_RC	-0.472611		ESTs, ligeramente parecidos al precursor de cadena del colágeno A54849 alfa 1 (VII) [H.sapiens]
NM_001960	0.471502	EEF1D	Factor 1 delta de alargamiento de conversión eucariótica (proteína de intercambio con nucleótido de guanina)
Contig52405_RC	-0.470705		ESTs, moderadamente parecidos a ALU8_ALU HUMANO SUBFAMILIA SX ENTRADA DE AVISO DE CONTAMINACIÓN DE SECUENCIA [H.sapiens]
Contig30092_RC	-0.469977		ARNm de isoforma B (PRDM6) de proteína 6 de dedo de cinc del dominio PR de Homo sapiens, cds parcial; alternativamente empalmado
NM_003462	-0.468753	P28	Dineína, axonemal, polipéptido intermedio ligero
Contig60393	0.468475		ESTs

Contig842_RC	0.468158		ESTs
NM_002982	0.466362	SCYA2	Citoquina inducible pequeña A2 (proteína 1 quemoatáctica de monocito, homólogo de ratón Sig-je)
Contig14390_RC	0.464150		ESTs
NM_001770	0.463847	CD19	Antígeno CD19
AK000617	-0.463158		ARNm de Homo sapiens; ADNc DKFZp434L235 (del clon DKFZp434L235)
AF073299	-0.463007	SLC9A2	Familia 9 del portador de soluto (intercambiador de sodio/hidrógeno), isoforma 2
NM_019049	0.461990	FLJ20054	Proteína hipotética
AL137347	-0.460778	DKFZP761M1511	Proteína hipotética
NM_000396	-0.460263	CTSK	Catepsina K (picnodisostosis)
NM_018373	-0.459268	FLJ11271	Proteína hipotética FLJ11271
NM_002709	0.458500	PPP1CB	Proteína fosfatasa 1, subunidad catalítica, isoforma beta
NM_016820	0.457516	OGG1	8-oxoguanina ADN glicosilasa
Contig10268_RC	0.456933		Secuencia de ADN humano del clon RP11-196N14 en el cromosoma 20. Contiene ESTs, STSs, GSSs e islas CpG. Contiene tres nuevos genes, parte de un gen para una nueva proteína parecida a la proteína serina/treonina fosfatasa 4 subunidad reguladora 1 (PP4R1) y un gen para una nueva proteína con una dominio de anquirina.
NM_014521	-0.456733	SH3BP4	Proteína 4 de enlace con el dominio SH3
AJ272057	-0.456548	STRAIT11499	Proteína hipotética STRAIT11499
NM_015964	-0.456187	LOC51673	Proteína específica del cerebro
Contig16759_RC	-0.456169		ESTs
NM_015937	-0.455954	LOC51604	Proteína CGI-06
NM_007246	-0.455500	KLHL2	Kelch 2 afín a (Drosophila) (Mayven)
NM_001985	-0.453024	ETFB	Flavoproteína de transferencia de electrones, polipéptido beta
NM_000984	-0.452935	RPL23A	Proteína ribosomal L23a
Contig51953_RC	-0.451695		ESTs
NM_015984	0.450491	UCH37	Ubiquitina C-terminal hidrolasa UCH37
NM_000903	-0.450371	DIA4	Diaforasa (NADH/NADPH) (citocromo b-5 reductasa)
NM_001797	-0.449862	CDH11	Cadherina 11, tipo 2, OB-cadherina (osteoblasto)
NM_014878	0.449818	KIAA0020	Producto del gen KIAA0020
NM_002742	-0.449590	PRKCM	Proteína quinasa C, mu

**Tabla 5. 231 marcadores de genes que distinguen entre pacientes con buen pronóstico y pacientes con pronóstico deficiente.**

Banco de Genes	SEQ ID NO	Banco de Genes	SEQ ID NO
Número de Acceso		Número de Acceso	
AA555029_RC	SEQ ID NO 1	NM_013296	SEQ ID NO 1427
AB020689	SEQ ID NO 37	NM_013437	SEQ ID NO 1439
AB032973	SEQ ID NO 55	NM_014078	SEQ ID NO 1449
AB033007	SEQ ID NO 58	NM_014109	SEQ ID NO 1451
AB033043	SEQ ID NO 62	NM_014321	SEQ ID NO 1477
AB037745	SEQ ID NO 75	NM_014363	SEQ ID NO 1480
AB037863	SEQ ID NO 88	NM_014750	SEQ ID NO 1527

# ES 2 366 178 T3

AF052159	SEQ ID NO 120	NM_014754	SEQ ID NO 1528
AF052162	SEQ ID NO 121	NM_014791	SEQ ID NO 1535
AF055033	SEQ ID NO 124	NM_014875	SEQ ID NO 1545
AF073519	SEQ ID NO 137	NM_014889	SEQ ID NO 1548
AF148505	SEQ ID NO 169	NM_014968	SEQ ID NO 1554
AF155117	SEQ ID NO 173	NM_015416	SEQ ID NO 1559
AF161553	SEQ ID NO 177	NM_015417	SEQ ID NO 1560
AF201951	SEQ ID NO 183	NM_015434	SEQ ID NO 1562
AF257175	SEQ ID NO 189	NM_015984 S	SEQ ID NO 1587
AJ224741	SEQ ID NO 196	NM_016337	SEQ ID NO 1636
AK000745	SEQ ID NO 219	NM_016359	SEQ ID NO 1638
AL050021	SEQ ID NO 257	NM_016448	SEQ ID NO 1645
AL050090	SEQ ID NO 259	NM_016569	SEQ ID NO 1655
AL080059	SEQ ID NO 270	NM_016577	SEQ ID NO 1656
AL080079	SEQ ID NO 271	NM_017779	SEQ ID NO 1708
AL080110	SEQ ID NO 272	NM_018004	SEQ ID NO 1725
AL133603	SEQ ID NO 306	NM_018098	SEQ ID NO 1739
AL133619	SEQ ID NO 307	NM_018104	SEQ ID NO 1743
AL137295	SEQ ID NO 315	NM_018120	SEQ ID NO 1745
AL137502	SEQ ID NO 326	NM_018136	SEQ ID NO 1748
AL137514	SEQ ID NO 327	NM_018265	SEQ ID NO 1766
AL137718	SEQ ID NO 336	NM_018354	SEQ ID NO 1774
AL355708	SEQ ID NO 353	NM_018401	SEQ ID NO 1782
D25328	SEQ ID NO 357	NM_018410	SEQ ID NO 1783
L27560	SEQ ID NO 390	NM_018454	SEQ ID NO 1786
M21551	SEQ 10 NO 394	NM_018455	SEQ ID NO 1787
NM_000017	SEQ ID NO 416	NM_019013	SEQ ID NO 1809
NM_000096	SEQ ID NO 430	NM_020166	SEQ ID NO 1825
NM_000127	SEQ ID NO 436	NM_020188	SEQ ID NO 1830
NM_000158	SEQ ID NO 442	NM_020244	SEQ ID NO 1835
NM_000224	SEQ ID NO 453	NM_020386	SEQ ID NO 1838
NM_000286	SEQ ID NO 462	NM_020675	SEQ ID NO 1842
NM_000291	SEQ ID NO 463	NM_020974	SEQ ID NO 1844
NM_000320	SEQ ID NO 469	R70506_RC	SEQ ID NO 1868
NM_000436	SEQ ID NO 487	U45975	SEQ ID NO 1878
NM_000507	SEQ ID NO 491	U58033	SEQ ID NO 1881
NM_000599	SEQ ID NO 503	U82987	SEQ ID NO 1891
NM_000788	SEQ ID NO 527	U96131	SEQ ID NO 1896
NM_000849	SEQ ID NO 530	X05610	SEQ ID NO 1903
NM_001007	SEQ ID NO 550	X94232	SEQ ID NO 1927
NM_001124	SEQ ID NO 562	Contig753_RC	SEQ ID NO 1954
NM_001168	SEQ ID NO 566	Contig1778_RC	SEQ ID NO 1979
NM_001216	SEQ ID NO 574	Contig2399_RC	SEQ ID NO 1989
NM_001280	SEQ ID NO 588	Contig2504_RC	SEQ ID NO 1991
NM_001282	SEQ ID NO 589	Contig3902_RC	SEQ ID NO 2017
NM_001333	SEQ ID NO 597	Contig4595	SEQ ID NO 2022
NM_001673	SEQ ID NO 645	Contig8581_RC	SEQ ID NO 2037

NM_001809	SEQ ID NO 673	Contig13480_RC	SEQ ID NO 2052
NM_001827	SEQ ID NO 676	Contig17359_RC	SEQ ID NO 2068
NM_001905	SEQ ID NO 691	Contig20217_RC	SEQ ID NO 2072
NM_002019	SEQ ID NO 711	Contig21812_RC	SEQ ID NO 2082
NM_002073	SEQ ID NO 721	Contig24252_RC	SEQ ID NO 2087
NM_002358	SEQ ID NO 764	Contig25055_RC	SEQ ID NO 2090
NM_002570	SEQ ID NO 787	Contig25343_RC	SEQ ID NO 2092
NM_002808	SEQ ID NO 822	Contig25991	SEQ ID NO 2098
NM_002811	SEQ ID NO 823	Contig27312_RC	SEQ ID NO 2108
NM_002900	SEQ ID NO 835	Contig28552_RC	SEQ ID NO 2120
NM_002916	SEQ ID NO 838	Contig32125_RC	SEQ ID NO 2155
NM_003158	SEQ ID NO 881	Contig32185_RC	SEQ ID NO 2156
NM_003234	SEQ ID NO 891	Contig33814_RC	SEQ ID NO 2169
NM_003239	SEQ ID NO 893	Contig34634_RC	SEQ ID NO 2180
NM_003258	SEQ ID NO 896	Contig35251_RC	SEQ ID NO 2185
NM_003376	SEQ ID NO 906	Contig37063_RC	SEQ ID NO 2206
NM_003600	SEQ ID NO 929	Contig37598	SEQ ID NO 2216
NM_003607	SEQ ID NO 930	Contig38288_RC	SEQ ID NO 2223
NM_003662	SEQ ID NO 938	Contig40128_RC	SEQ ID NO 2248
NM_003676	SEQ ID NO 941	Contig40831_RC	SEQ ID NO 2260
NM_003748	SEQ ID NO 951	Contig41413_RC	SEQ ID NO 2266
NM_003862	SEQ ID NO 960	Contig41887_RC	SEQ ID NO 2276
NM_003875	SEQ ID NO 962	Contig42421_RC	SEQ ID NO 2291
NM_003878	SEQ ID NO 963	Contig43747_RC	SEQ ID NO 2311
NM_003882	SEQ ID NO 964	Contig44064_RC	SEQ ID NO 2315
NM_003981	SEQ ID NO 977	Contig44289_RC	SEQ ID NO 2320
NM_004052	SEQ ID NO 985	Contig44799_RC	SEQ ID NO 2330
NM_004163	SEQ ID NO 995	Contig45347_RC	SEQ ID NO 2344
NM_004336	SEQ ID NO 1022	Contig45816_RC	SEQ ID NO 2351
NM_004358	SEQ ID NO 1026	Contig46218_RC	SEQ ID NO 2358
NM_004456	SEQ ID NO 1043	Contig46223_RC	SEQ ID NO 2359
NM_004480	SEQ ID NO 1046	Contig46653_RC	SEQ ID NO 2369
NM_004504	SEQ ID NO 1051	Contig46802_RC	SEQ ID NO 2372
NM_004603	SEQ ID NO 1064	Contig47405_RC	SEQ ID NO 2384
NM_004701	SEQ ID NO 1075	Contig48328_RC	SEQ ID NO 2400
NM_004702	SEQ ID NO 1076	Contig49670_RC	SEQ ID NO 2434
NM_004798	SEQ ID NO 1087	Contig50106_RC	SEQ ID NO 2445
NM_004911	SEQ ID NO 1102	Contig50410	SEQ ID NO 2453
NM_004994	SEQ ID NO 1108	Contig50802_RC	SEQ ID NO 2463
NM_005196	SEQ ID NO 1127	Contig51464_RC	SEQ ID NO 2481
NM_005342	SEQ ID NO 1143	Contig51519_RC	SEQ ID NO 2482
NM_005496	SEQ ID NO 1157	Contig51749_RC	SEQ ID NO 2486
NM_005563	SEQ ID NO 1173	Contig51963	SEQ ID NO 2494
NM_005915	SEQ ID NO 1215	Contig53226_RC	SEQ ID NO 2525
NM_006096	SEQ ID NO 1240	Contig53268_RC	SEQ ID NO 2529
NM_006101	SEQ ID NO 1241	Contig53646_RC	SEQ ID NO 2538
NM_006115	SEQ ID NO 1245	Contig53742_RC	SEQ ID NO 2542

NM_006117	SEQ ID NO 1246	Contig55188_RC	SEQ ID NO 2586
NM_006201	SEQ ID NO 1254	Contig55313_RC	SEQ ID NO 2590
NM_006265	SEQ ID NO 1260	Contig55377_RC	SEQ ID NO 2591
NM_006281	SEQ ID NO 1263	Contig55725_RC	SEQ ID NO 2600
NM_006372	SEQ ID NO 1273	Contig55813_RC	SEQ ID NO 2607
NM_006681	SEQ ID NO 1306	Contig55829_RC	SEQ ID NO 2608
NM_006763	SEQ ID NO 1315	Contig56457_RC	SEQ ID NO 2630
NM_006931	SEQ ID NO 1341	Contig57595	SEQ ID NO 2655
NM_007036	SEQ ID NO 1349	Contig57864_RC	SEQ ID NO 2663
NM_007203	SEQ ID NO 1362	Contig58368_RC	SEQ ID NO 2668
NM_012177	SEQ ID NO 1390	Contig60864_RC	SEQ ID NO 2676
NM_012214	SEQ ID NO 1392	Contig63102_RC	SEQ ID NO 2684
NM_012261	SEQ ID NO 1397	Contig63649_RC	SEQ ID NO 2686
NM_012429	SEQ ID NO 1413	Contig64688	SEQ ID NO 2690
NM_013262	SEQ ID NO 1425		

**Tabla 6. 70 marcadores de pronóstico preferidos sacados de la Tabla 5.**

AL080059	-0.527150		ARNm de Homo sapiens para proteína KIAA1750, cds parcial
Contig63649_RC	-0.468130		ESTs
Contig46218_RC	-0.432540		ESTs
NM_016359	-0.424930	LOC51203	Clon HQ0310 PRO0310p1
AA555029_RC	-0.424120		ESTs
NM_003748	0.420671	ALDH4	Aldehído dehidrogenasa 4 (glutamato gamma-semialdehído dehidrogenasa; pirrolina-5-carboxilato dehidrogenasa)
Contig38288_RC	-0.414970		ESTs, ligeramente parecidos a ISHUSS proteína disulfuro-isomerasa [H.sapiens]
NM_003862	0.410964	FGF18	Factor del crecimiento fibroblasto 18
Contig28552_RC	-0.409260		ARNm de Homo sapiens; ADNc DKFZp434C0931 (de clon DKFZp434C0931); cds parcial
Contig32125_RC	0.409054		ESTs
U82987	0.407002	BBC3	Componente 3 de enlace con Bcl-2
AL137718	-0.404980		ARN de Homo sapiens; ADNc DKFZp434C0931 (del clon DKFZp434C0931); cds parcial
AB037863	0.402335	KIAA1442	Proteína KIAA1442
NM_020188	-0.400070	DC13	Proteína DC13
NM_020974	0.399987	CEGP1	Proteína CEGP1
NM_000127	-0.399520	EXT1	Exostosis (múltiple) 1
NM_002019	-0.398070	FLT1	Tirosina quinasa 1 relacionada con fms (factor de crecimiento endotelial vascular/receptor de factor de permeabilidad vascular)
NM_002073	-0.395460	GNAZ	Proteína de enlace con nucleótido de guanina (proteína G), polipéptido alfa z
NM_000436	-0.392120	OXCT	3-oxoácido CoA transferasa
NM_004994	-0.391690	MMP9	Metaloproteínasa 9 de matriz (gelatinasa B, 92kD gelatinasa, 92kD

			tipo IV colagenasa)
Contig55377_RC	0.390600		ESTs
Contig35251_RC	-0.390410		ADNc de Homo sapiens: fis FLJ22719, clon HSI14307
Contig25991	-0.390370	ECT2	Oncógeno de secuencia transformadora 2 de célula epitelial
NM_003875	-0.386520	GMPS	Guanina monofosfato sintetasa
NM_006101	-0.385890	HEC	Altamente expresado en cáncer, rico en repeticiones de heptadas de leucina
NM_003882	0.384479	WISP1	WNT1 procedimiento de señalización inducible proteína 1
NM_003607	-0.384390	PK428	Proteína quinasa Ser-Thr relacionada con la proteína quinasa de distrofia miotónica
AF073519	-0.383340	SERF1A	Pequeño factor 1A rico en EDRK (telomérico)
AF052162	-0.380830	FLJ12443	Proteína hipotética FLJ12443
NM_000849	0.380831	GSTM3	Glutathione S-transferasa M3 (cerebro)
Contig32185_RC	-0.379170		ADNc de Homo sapiens fis FLJ13997, clon Y79AA1002220
NM_016577	-0.376230	RAB6B	RAB6B, miembro de familia de oncógeno RAS
Contig48328_RC	0.375252		ESTs, ligeramente parecidos a T17248 proteína hipotética DKFZp586G1122.1 [H.sapiens]
Contig46223_RC	0.374289		ESTs
NM_015984	-0.373880	UCH37	Ubiquitina C-terminal hidrolasa UCH37
NM_006117	0.373290	PECI	Peroxisomal D3,D2-enoil-CoA isomerasa
AK000745	-0.373060		ADNc de Homo sapiens fis FLJ20738, clon HEP08257
Contig40831_RC	-0.372930		ESTs
NM_003239	0.371524	TGFB3	Factor transformador del crecimiento, beta 3
NM_014791	-0.370860	KIAA0175	Producto del gen KIAA0175
X05610	-0.370860	COL4A2	Colágeno, tipo IV, alfa 2
NM_016448	-0.369420	L2DTL	Proteína L2DTL
NM_018401	0.368349	HSA250839	Gen para serina/treonina proteína quinasa
NM_000788	-0.367700	DCK	Deoxicitidina quinasa
Contig51464_RC	-0.367450	FLJ22477	Proteína hipotética FLJ22477
AL080079	-0.367390	DKFZP564D0462	Proteína hipotética DKFZp564D0462
NM_006931	-0.366490	SLC2A3	Familia 2 del portador de soluto (transportador de glucosa facilitado), miembro 3
AF257175	-0.365900		ARN (HCA64) de Homo sapiens del antígeno 64 asociado al carcinoma hepatocelular 64, cds completo
NM_014321	-0.365810	ORC6L	Complejo de reconocimiento de origen, subunidad 6 afín a (homólogo de levadura)
NM_002916	-0.365590	RFC4	Factor de replicación C (activador 1) 4 (37kD)
Contig55725_RC	-0.365350		ESTs, moderadamente parecidos a T50635 proteína hipotética DKFZp762L0311.1 [H.sapiens]
Contig24252_RC	-0.364990		ESTs
AF201951	0.363953	CFFM4	Subunidad beta de receptor epsilon de inmunoglobulina de alta afinidad
NM_005915	-0.363850	MCM6	Deficiente en mantenimiento de minicromosoma (mis5, S. pombe) 6
NM_001282	0.363326	AP2B1	Complejo 2 de proteína relacionada con adaptador, subunidad beta 1



Contig56457_RC	-0.361650	TMEFF1	Proteína de transmembrana con afin a EGF y dos dominios 1 afines a folistatina
NM_000599	-0.361290	IGFBP5	Proteína 5 de enlace con factor de crecimiento afin a insulina
NM_020386	-0.360780	LOC57110	Proteína relacionada con proteína H-REV107
NM_014889	-0.360040	MP1	Metaloproteasa 1 (familia de pitrilisina)
AF055033	-0.359940	IGFBP5	Proteína 5 de enlace con factor de crecimiento afin a insulina
NM_006681	-0.359700	NMU	Neuromedina U
NM_007203	-0.359570	AKAP2	A Proteína 2 de amarre a quinasa (PRKA)
Contig63102_RC	0.359255	FLJ11354	Proteína hipotética FLJ11354
NM_003981	-0.358260	PRC1	Regulador proteínico de la citoquinesis 1
Contig20217_RC	-0.357880		ESTs
NM_001809	-0.357720	CENPA	Proteína A (17kD) de centrómero
Contig2399_RC	-0.356600	SM-20	Similar a la proteína SM-20 de músculo liso de rata
NM_004702	-0.356600	CCNE2	Ciclina E2
NM_007036	-0.356540	ESM1	Molécula 1 específica de célula endotelial
NM_018354	-0.356000	FLJ11190	Proteína hipotética FLJ11190

**[0037]** Los conjuntos de marcadores enumerados en las Tablas 1-6 están parcialmente superpuestos; en otras palabras, algunos marcadores están presentes en múltiples conjuntos, mientras que otros son exclusivos de un conjunto en concreto (FIG. 1).

- 5 **[0038]** Así pues, en una materialización se describe un conjunto de 256 marcadores genéticos que pueden distinguir entre ER (+) y ER (-) y también entre tumores de *BRCA1* y tumores esporádicos (es decir, clasificar un tumor como ER (-) o ER (-) y relacionado con *BRCA1* o esporádico). En una materialización más específica, se describen subconjuntos de al menos 20, al menos 50, al menos 100 o al menos 150 del conjunto de 256 marcadores que pueden clasificar un tumor como ER (-) o ER (-) y relacionado con *BRCA1* o esporádico. En otra
- 10 materialización se describen 165 marcadores que pueden distinguir entre ER (+) y ER (-) y también entre pacientes con buen pronóstico frente a pronóstico deficiente (es decir, clasificar un tumor como ER (-) o como ER (+) y como que tenga que ser extirpado de una paciente con buen pronóstico o con pronóstico deficiente). Se describen asimismo subconjuntos de al menos 20, 50, 100 o 125 del conjunto completo de 165 marcadores, que también pueden clasificar un tumor como ER (-) o como ER (+) y como que tenga que ser extirpado de una
- 15 paciente con buen pronóstico o con pronóstico deficiente. Además, se describe un conjunto de doce marcadores que pueden distinguir entre tumores de *BRCA1* y tumores esporádicos, y entre pacientes con buen pronóstico frente a pronóstico deficiente. Por último, se describen once marcadores que son capaces de diferenciar las tres categorías. Y a la inversa, se describen 2.050 de los 2.460 marcadores con categoría de ER que pueden determinar solo la categoría de ER, 173 de los 430 marcadores ARC47 frente a marcadores esporádicos que solo pueden determinar el *BRCA1* frente a la categoría esporádica y 65 de los 231 marcadores de pronóstico que solo pueden determinar el pronóstico. En materializaciones más específicas se describen subconjuntos de al menos 20, 50, 100, 200, 500, 1.000, 1.500 o 2.000 de los 2.050 marcadores con categoría de ER que igualmente solo pueden determinar la categoría de ER. También se describen subconjuntos de al menos 20, 50, 100 o 150 de los 173 marcadores que también determinan solo el *BRCA1* frente a tumores esporádicos. Asimismo, se
- 20 describen subconjuntos de al menos 20, 30, 40, o 50 de los 65 marcadores de pronóstico que igualmente solo pueden determinar la categoría del pronóstico.

- 30 **[0039]** Cualquiera de los conjuntos de marcadores suministrados más arriba puede usarse solo específicamente o en combinación con otros marcadores que no sean del conjunto. Por ejemplo, los marcadores que distinguen la categoría de ER pueden usarse en combinación con los marcadores de *BRCA1* frente a los esporádicos, o con los marcadores de pronóstico, o ambos. Cualquiera de los conjuntos de marcadores suministrados más arriba puede usarse también en combinación con otros marcadores para el cáncer de mama, o para cualquier otra afección clínica o fisiológica.

**[0040]** La relación entre los conjuntos de marcadores se muestra en el diagrama de la FIG. 1.

### 5.3.2 IDENTIFICACIÓN DE MARCADORES

- 35 **[0041]** Se describen conjuntos de marcadores para la identificación de condiciones o indicaciones relacionadas con el cáncer de mama. De modo general, los conjuntos de marcadores se identificaron determinando cuáles de

entre -25.000 marcadores humanos tenían patrones de expresión que se hallaran en correlación con las condiciones o indicaciones.

**[0042]** En una materialización, el método para identificar conjuntos de marcadores es como sigue. Tras la extracción y el etiquetado de polinucleótidos destino, se compara la expresión de todos los marcadores (genes) en una muestra X con la expresión de todos los marcadores (genes) en un patrón o control. En una materialización, el patrón o control comprende moléculas polinucleótidas destino obtenidas de una muestra de un individuo normal (es decir, de un individuo no aquejado de cáncer de mama). En una materialización preferida, el patrón o control es una reserva de moléculas polinucleótidas destino. La reserva puede obtenerse de muestras recogidas de una serie de individuos normales. En una materialización preferida, la reserva comprende muestras tomadas de una serie de individuos que tienen tumores de tipo esporádico. En otra materialización preferida, la reserva comprende una población de ácidos nucleicos generados artificialmente y destinados a aproximar el nivel de ácido nucleico derivado de cada marcador hallado en una reserva de marcadores derivados de ácidos nucleicos obtenidos de muestras de tumores. En una materialización preferida más, la reserva se obtiene de líneas celulares o de muestras de líneas de células normales o de cáncer de mama.

**[0043]** La comparación puede realizarse por cualquiera de los medios conocidos en la técnica. Por ejemplo, los niveles de expresión de marcadores diversos pueden calcularse mediante la separación de moléculas polinucleótidas destino (por ej., ARN o ADNc), derivadas de los marcadores en geles de agarosa o poliacrilamida, seguida por la hibridación con sondas oligonucleótidas específicas de marcadores. Como alternativa, la comparación puede realizarse mediante el etiquetado de moléculas polinucleótidas destino seguidas por la separación en un gel secuenciador. Se colocan muestras polinucleótidas en el gel de modo que paciente y control o patrón se hallen en corredores adyacentes. La comparación de niveles de expresión se realiza visualmente o por medio de densitómetro. En una materialización preferida, la expresión de todos los marcadores se calcula simultáneamente por hibridación a un biochip. En cada aproximación, los marcadores que reúnen determinados requisitos se identifican como asociados al cáncer de mama.

**[0044]** El marcador se selecciona basándose en alguna diferencia significativa de expresión en una muestra al compararla con una condición de patrón o control. Se puede hacer la selección basándose bien en una regulación significativa hacia arriba o hacia abajo del marcador en la muestra de la paciente. También se puede hacer la selección calculando la relevancia estadística (esto es, el valor-p) de la correlación entre la expresión del marcador y la condición o indicación. A ser posible, deben usarse ambos criterios de selección. De este modo, en una materialización de la presente invención, los marcadores asociados al cáncer de mama se seleccionan si cumplen estas dos condiciones: que muestren un cambio de más del doble (aumento o disminución) en expresión en comparación con un patrón y que el valor-p para la correlación entre la existencia del cáncer de mama y el cambio en el marcador de expresión no sea superior a 0,01 (es decir, que sea significativo estadísticamente).

**[0045]** La expresión de los marcadores identificados en relación con el cáncer de mama se usa entonces para identificar marcadores que puedan diferenciar tumores en tipos clínicos. En una materialización específica que usa una serie de muestras de tumores, los marcadores se identifican mediante el cálculo de correlación de coeficientes entre la categoría clínica o parámetro(s) clínico(s) y el cociente de expresión lineal, logarítmico o cualquier transformación del mismo en todas las muestras de cada gen por separado. Concretamente, el coeficiente de correlación se calcula como

$$\rho = (\bar{c} \cdot \bar{r}) / (\|\bar{c}\| \cdot \|\bar{r}\|)$$

**Ecuación (2)**

en donde  $\bar{c}$  representa los parámetros o categorías clínicos y  $\bar{r}$  representa el cociente de expresión lineal, logarítmico o cualquier transformación del mismo entre muestra y control. Los marcadores en los que el coeficiente de correlación exceda de un límite se clasifican como marcadores en relación con el cáncer de mama específicos para un determinado tipo clínico. Dicho límite o umbral coincide con una cierta relevancia de los genes diferenciadores obtenidos mediante simulaciones Monte-Carlo. El umbral depende del número de muestras usadas; el umbral puede calcularse como  $3 \times 1/\sqrt{n-3}$ , donde  $1/\sqrt{n-3}$  es la anchura de distribución y  $n$  = el número de muestras. En una materialización específica, los marcadores se eligen si el coeficiente de correlación es superior a 0,3, aproximadamente, o inferior a 0,3, aproximadamente.

**[0046]** A continuación se calcula la relevancia de la correlación. Dicha relevancia puede calcularse por cualquier medio estadístico con el que se calcule dicha relevancia. En un ejemplo concreto, se genera un conjunto de datos ordenados mediante una técnica Monte-Carlo para aleatorizar la asociación entre la diferencia de expresión de un marcador en particular y la categoría clínica. La distribución de frecuencia de los marcadores que reúnen los requisitos mediante el cálculo de coeficientes de correlación se compara con el número de marcadores que reúnen los requisitos en los datos generados mediante la técnica Monte-Carlo. La distribución de frecuencia de los marcadores que reúnen los requisitos en los ensayos Monte-Carlo se usa para determinar si

el número de marcadores seleccionados por correlación con los datos clínicos es significativo. Véase el Ejemplo 4.

**[0047]** Una vez identificado un marcador, los marcadores pueden ordenarse jerárquicamente por su relevancia de diferenciación. Una forma de ordenación jerárquica es por la amplitud de correlación entre el cambio en la expresión y la condición específica que se está diferenciando. Otro medio también muy utilizado es usar una métrica estadística. En una materialización específica, la métrica es una estadística de tipo Fisher:

$$t = \frac{(\langle x_1 \rangle - \langle x_2 \rangle)}{\sqrt{(\sigma_1^2(n_1 - 1) + \sigma_2^2(n_2 - 1)) / (n_1 + n_2 - 1) \cdot (1/n_1 + 1/n_2)}}$$

**Ecuación (3)**

**[0048]** En esta ecuación, ( $x_1$ ) es la media ponderada de error del cociente logarítmico de las mediciones de expresión de transcripción dentro de un primer grupo de diagnóstico (por ej., ER(-),  $x_2$ ) es la media ponderada del cociente logarítmico dentro de un segundo grupo de diagnóstico relacionado con el primero (por ej., ER(+)),  $\sigma_1$  es la varianza del logaritmo del cociente dentro del grupo ER(-) y  $n_1$  es el número de muestras para las que se dispone de mediciones válidas de logaritmos del cociente.  $\sigma_2$  es la varianza del cociente logarítmico dentro de un segundo grupo de diagnóstico (por ej., ER(+)), y  $n_2$  es el número de muestras para las que se dispone de mediciones válidas de cocientes logarítmicos. El valor-t representa la diferencia de compensación de varianza entre dos medias.

**[0049]** El conjunto jerarquizado de marcadores puede usarse para optimizar el número de marcadores en el conjunto usado para la diferenciación. Esto se lleva a cabo generalmente en un método de "dejar uno fuera" de la forma siguiente: en un primer ensayo, se usa un subconjunto, por ejemplo 5, de los marcadores de los primeros puestos de la lista jerarquizada para generar una plantilla, donde de X muestras se usan X-1 para generar la plantilla y se predice la categoría de la muestra restante. Este proceso se repite para cada muestra hasta que cada una de las muestras X se haya previsto una vez. En un segundo ensayo, se añaden marcadores adicionales, por ejemplo 5, de modo que ahora se genera una plantilla a partir de 10 marcadores, y se predice el resultado de la muestra restante. Este proceso se repite hasta que se haya usado todo el conjunto de marcadores para generar una plantilla. Para cada uno de los ensayos se cuentan los errores de tipo 1 (falso positivo) y los errores de tipo 2 (falso negativo): el número óptimo de marcadores es aquel número donde la tasa de error 1 o la tasa de error 2, o preferiblemente la tasa total de error 1 y de error 2 sea más baja.

**[0050]** Para los marcadores de pronóstico, la validación del conjunto de marcadores puede llevarse a cabo mediante una estadística adicional, un modelo de supervivencia. Esta estadística genera la probabilidad de metástasis distantes de tumor como función de tiempo desde el diagnóstico inicial. Se puede usar una serie de modelos, como el Weibull, el normal, el normal logarítmico, el logístico logarítmico, el exponencial logarítmico o el Rayleigh logarítmico (Capítulo 12 de "Life Testing", S-PLUS 2000 GUIDE TO STATISTICS, Vol. 2, p. 368 (2000)). Para el modelo "normal", la probabilidad de metástasis distante P en un tiempo t se calcula como

$$P = \alpha \times \exp(-t^2/\tau^2)$$

**Ecuación (4)**

**[0051]** donde  $\alpha$  es fijo e igual a 1 y  $\tau$  es un parámetro que hay que ajustar y que mide la "previsión de vida".

**[0052]** Para todos aquellos experimentados en la técnica resultará evidente que los métodos de más arriba, especialmente los métodos estadísticos, descritos más arriba, no están limitados a la identificación de marcadores asociados al cáncer de mama, sino que pueden usarse para identificar conjuntos de genes marcadores asociados a cualquier fenotipo. El fenotipo puede ser la presencia o ausencia de una enfermedad como el cáncer, o la presencia o ausencia de cualquier categoría clínica de identificación asociado a ese cáncer. En el contexto de las enfermedades, el fenotipo puede ser un pronóstico como el tiempo de supervivencia, la probabilidad de metástasis distante de un estado de enfermedad o la probabilidad de una respuesta específica a un régimen terapéutico o profiláctico. El fenotipo no tiene por qué ser un cáncer o una enfermedad; el fenotipo puede ser una característica nominal asociada a un individuo sano.

### 5.3.3 RECOGIDA DE MUESTRAS

**[0053]** En la presente invención, se extraen moléculas polinucleótidas destino de una muestra tomada de un individuo aquejado de cáncer de mama. La muestra puede recogerse de cualquier manera clínicamente aceptable, pero siempre de forma que se conserven los polinucleótidos derivados de marcadores (por ej., ARN). A ser posible, el ANRm o los ácidos nucleicos derivados del mismo (esto es, el ADNc o el ADN amplificado) deben etiquetarse de forma que se distingan de las moléculas polinucleótidas de control o patrón y ambos se hibridarán simultánea o independientemente con un biochip que comprenda algunos o todos los marcadores o conjuntos o subconjuntos de marcadores descritos más arriba. Como alternativa, el ANRm o los ácidos nucleicos derivados del mismo pueden etiquetarse con la misma etiqueta que las moléculas polinucleótidas de control o

patrón, en las que se compara la intensidad de hibridación de cada uno en una sonda en concreto. La muestra puede consistir en cualquier muestra de tejido clínicamente relevante, como una biopsia tumoral o un aspirado con aguja fina, o una muestra de fluido corporal, como sangre, plasma, suero, linfa, fluido ascítico, fluido cístico, orina o exudado de pezón. La muestra puede tomarse de un ser humano o, en un contexto veterinario, de animales no humanos como rumiantes, equinos, porcinos u ovinos, o de animales domésticos de compañía como felinos y caninos.

**[0054]** Los métodos para preparar ARN total y poli(A) + son bien conocidos y se describen de modo general en Sambrook et al., *MOLECULAR CLONING - A LABORATORY MANUAL* (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989)) y Ausubel et al., *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY*, vol. 2, Current Protocols Publishing, New York (1994)).

**[0055]** El ARN puede aislarse de las células eucarióticas mediante procesos que implican la lisis de las células y la desnaturalización de las proteínas contenidas en las mismas. Las células de interés consisten en las células de tipo natural (esto es, no cancerosas), las células de tipo natural expuestas a medicamentos, las células tumorales u procedentes de tumores, las células modificadas, las células de línea celular normal o tumoral y las células modificadas expuestas a medicamentos.

**[0056]** Se pueden ejecutar pasos adicionales para eliminar el ADN. Se puede llevar a cabo una lisis celular con un detergente no iónico seguida de microcentrifugado para eliminar los núcleos y por ende la masa del ADN celular. En una materialización, el ARN se extrae de las células de los diversos tipos de interés mediante lisis de tiocianato de guanidina seguida de centrifugado de CsCl para separar el ARN del ADN (Chirgwin et al., *Biochemistry* 18:5294-5299 (1979)). El Poli(A)+ ARN se selecciona mediante selección con celulosa oligo-dT (véase Sambrook et al., *MOLECULAR CLONING - A LABORATORY MANUAL* (2<sup>ND</sup> ED.), Vols. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989). Como alternativa, la separación del ARN del ADN se puede llevar a cabo mediante extracción orgánica, por ejemplo, con fenol caliente o fenol/cloroformo/alcohol isoamílico.

**[0057]** Si se desea, se puede añadir inhibidores de RNasa al tampón de lisis. Igualmente, para ciertos tipos de células, puede ser aconsejable añadir al protocolo un paso de desnaturalización/digestión de proteína.

**[0058]** Para muchas aplicaciones, es deseable ante todo enriquecer el ARNm con respecto a otros ARNs celulares, como ARN de transferencia (ARNt) y ARN ribosómico (ARNr). La mayoría de los ARNm contienen una cola de poli(A) en su extremo 3'. Con esto es posible enriquecerlos por afinidad cromatográfica, por ejemplo, mediante oligo(dT) or poly(U) acoplado a un soporte sólido, como celulosa o Sephadex™ (véase Ausubel et al., *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY*, vol. 2, Current Protocols Publishing, New York (1994). Una vez ligado el poly(A)+ ARNm, se lo eluye de la columna de afinidad mediante 2 mM EDTA/0.1% SDS.

**[0059]** La muestra de ARN comprende una pluralidad de diferentes moléculas de ARNm, cada una de las cuales tiene una secuencia nucleótida diferente. En una materialización específica, las moléculas de ARNm en la muestra de ARN comprenden al menos 100 secuencias nucleótidas diferentes. Más preferiblemente, las moléculas de ARNm en la muestra de ARN comprenden moléculas de ARNm correspondientes a cada uno de los genes marcadores. En otra materialización específica, la muestra de ARN es una muestra de ARN de mamífero.

**[0060]** En una materialización específica, se usan ARN o ARNm totales de las células en los métodos de la invención. La fuente de ARN puede ser células de una planta o animal, humano, mamífero, primate, animal no humano, perro, gato, ratón, ave, levadura, eucariota, procariota, etc. En materializaciones específicas, el método de la invención se usa con una muestra que contiene ARNm o ARN total de 1x10<sup>6</sup> células o menos. En otra materialización se puede aislar las proteínas de las fuentes anteriores con métodos conocidos en la técnica para usarlas en análisis de expresión a nivel proteínico.

**[0061]** Las sondas a los homólogos de las secuencias de marcadores que aquí se dan a conocer pueden emplearse preferiblemente allí donde se analice ácido nucleico no humano.

## 5.4 MÉTODOS DE EMPLEO DE LOS CONJUNTOS DE MARCADORES DE CÁNCER DE MAMA

### 5.4.1 MÉTODOS DE DIAGNÓSTICO

**[0062]** A continuación se describen métodos de empleo de los conjuntos de marcadores para analizar una muestra de un individuo con el fin de determinar el tipo o subtipo de tumor del individuo a nivel molecular, tanto si es un tumor de tipo ER(+) o ER(-) y tanto si es un tumor asociado a *BRCA1* o esporádico. En realidad no es preciso que el individuo esté aquejado de cáncer de mama. Esencialmente, consiste en comparar la expresión de genes de marcadores específicos en el individuo, o una muestra tomada de los mismos, con un patrón o control. Por ejemplo, supongamos dos condiciones en relación con el cáncer de mama, X e Y. Podemos comparar el nivel de expresión de los marcadores de pronóstico de cáncer de mama para la condición X en un

individuo con el nivel de los polinucleótidos derivados del marcador en un control, en donde el nivel representa el nivel de expresión indicado por las muestras que tienen la condición X. En este caso, si la expresión de los marcadores en la muestra del individuo es sustancialmente (esto es, estadísticamente) diferente de la del control, entonces el individuo no tiene condición X. Donde, como aquí, la opción es bimodal (esto es, una muestra es o X o Y), se puede decir adicionalmente que el individuo tiene la condición Y. Por supuesto que también se puede realizar la comparación con un control que represente la condición Y. Preferiblemente, las dos se llevarán a cabo simultáneamente, de modo que cada control actúe a la vez como control positivo y como negativo. Así, el resultado típico puede ser bien una diferencia demostrable respecto a los niveles de expresión (esto es, la cantidad de ARN derivado del marcador, o los polinucleótidos derivados del mismo) representados por el control, o bien ninguna diferencia significativa.

**[0063]** Así pues, en una materialización, el método para determinar el estado concreto de un individuo en relación con un tumor comprende los pasos de (1) hibridar los nucleótidos destino etiquetados de un individuo con un biochip que contenga uno de los conjuntos de marcadores de más arriba; (2) hibridar las moléculas polinucleótidas de patrón o control con el biochip, en donde las moléculas de patrón o control se etiquetan de forma diferenciada a partir de las moléculas destino, y (3) determinar la diferencia en los niveles de transcripción, o la ausencia de la misma, entre el destino y el patrón o control, en donde la diferencia, o la ausencia de la misma, determina el estado del individuo en relación con el tumor. En una materialización más específica, las moléculas de patrón o control comprenden polinucleótidos obtenidos de marcadores a partir de una reserva de muestras de individuos normales, o una reserva de muestras de tumores de individuos con tumores de tipo esporádico. En una materialización preferida, el patrón o control es una reserva de polinucleótidos derivados de marcadores generada artificialmente, reserva que está destinada a mimetizar el nivel de expresión de marcadores apreciado en las muestras clínicas de tejido normal o de cáncer de mama que tenga una indicación clínica específica (esto es, canceroso o no canceroso); tumor de ER(+) o ER(-); tumor de *BRCA1*- o de tipo esporádico. En otra materialización específica, las moléculas de control comprenden una reserva obtenida de líneas celulares normales o de cáncer de mama.

**[0064]** A continuación se describen conjuntos de marcadores que sirven para distinguir tipos de tumores ER(+) de los de ER(-). Así, en una materialización del método de más arriba, el nivel de polinucleótidos (esto es, ARNm o polinucleótidos derivados del mismo) en una muestra de un individuo, expresada a partir de los marcadores suministrados en la Tabla 1, se compara con el nivel de expresión de los mismos marcadores a partir de un control, en donde el control comprende polinucleótidos en relación con marcadores obtenidos de muestras de ER(+), de ER (-) o de ambos. Preferiblemente, la comparación será tanto con ER(+) como con ER(-) y preferiblemente la comparación será con reservas de polinucleótidos procedentes de una serie de muestras de ER(+) y ER(-), respectivamente. Allí donde la expresión de marcadores del individuo se parezca más o esté en correspondencia más estrecha con el control de ER(+) y no se parezca o corresponda con el control de ER(-), al individuo se le clasificará como ER(+). Allí donde la reserva no sea ER(+) o ER(-) puros, por ejemplo, se usará una reserva esporádica. Se debe hibridar contra la reserva un conjunto de experimentos que usen individuos con categoría de ER conocida, con el fin de definir las plantillas de expresión para el grupo de ER(+) y de ER(-). A cada individuo con categoría de ER desconocida se lo hibridará contra la misma reserva y se comparará el perfil de expresión con las plantillas (s) para determinar la categoría de ER del individuo.

**[0065]** Lo que se describe a continuación son conjuntos de marcadores que sirven para distinguir los tumores relacionados con *BRCA1* de los tumores esporádicos. Por lo tanto, el método puede ejecutarse sustancialmente igual que para la determinación de ER(+/-), con la excepción de que los marcadores son los enumerados en las Tablas 3 y 4 y que los marcadores de control son una reserva de muestras de tumores de *BRCA1* de polinucleótidos derivados de marcadores y una reserva de polinucleótidos derivados de marcadores procedente de tumores esporádicos. Se considera que una paciente tiene una mutación germinal de *BRCA1* allí donde la expresión de los polinucleótidos derivados de marcador del individuo se parezca más, o donde esté en correspondencia más estrecha, a la del control del *BRCA1*. Cuando el control no sea *BRCA1* o esporádico puros, se pueden definir dos plantillas de manera parecida a la usada para hallar la categoría de ER, como se explica más arriba.

**[0066]** Para las dos anteriores materializaciones del método se puede usar todo el conjunto de marcadores (esto es, el conjunto completo de marcadores para las Tablas 1 o 3). En otras materializaciones se pueden usar subconjuntos de los marcadores. En una materialización preferida se usan los marcadores preferidos enumerados las Tablas 2 o 4.

**[0067]** La semejanza entre el perfil de expresión de marcadores de un individuo y el de un control puede calcularse de varias maneras. En el caso más simple, los perfiles pueden compararse visualmente en una lista impresa de datos de diferencia de expresión. Como alternativa, la semejanza puede calcularse matemáticamente.

**[0068]** En una materialización, la medida de la semejanza entre dos pacientes x e y, o una paciente x y una plantilla y, puede calcularse mediante la siguiente ecuación:

$$s = 1 - \left[ \frac{\sum_{i=1}^{N_p} \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sigma_{x_i} \sigma_{y_i}}}{\sqrt{\sum_{i=1}^{N_p} \left( \frac{x_i - \bar{x}}{\sigma_{x_i}} \right)^2 \sum_{i=1}^{N_p} \left( \frac{y_i - \bar{y}}{\sigma_{y_i}} \right)^2}} \right]$$

**Ecuación (5)**

En esa ecuación,  $x$  e  $y$  son dos pacientes con componentes de cociente logarítmico  $x_i$  e  $y_i$ ,  $i = 1, \dots, N = 4.986$ . A cada valor  $x_i$  va asociado el error  $\sigma_{x_i}$ . Cuanto menor es el valor  $\sigma_{x_i}$ , más fiable es la medición  $x_i$ .

$$\bar{x} = \frac{\sum_{i=1}^{N_p} \frac{x_i}{\sigma_{x_i}^2}}{\sum_{i=1}^{N_p} \frac{1}{\sigma_{x_i}^2}}$$

5 es la media aritmética ponderada de error.

[0069] En una materialización preferida, se desarrolla las plantillas por comparación de muestras. Se define la plantilla como la media ponderada de error del cociente logarítmico de la diferencia de expresión para el grupo de genes marcadores capaces de diferenciar la condición específica de la relacionada con el cáncer de mama. Por ejemplo, se definen plantillas para muestras de ER(+) y para muestras de ER(-). A continuación se calcula un parámetro clasificador. Este parámetro puede calcularse mediante diferencias de nivel de expresión entre la muestra y la plantilla o bien hallando un coeficiente de correlación. Dicho coeficiente,  $P_i$ , puede calcularse mediante la siguiente ecuación:

$$P_i = (\bar{z}_i \bullet \bar{y}) / (\|\bar{z}_i\| \cdot \|\bar{y}\|)$$

**Ecuación (1)**

donde  $Z_i$  es la plantilla de expresión  $i$ , e  $y$  es el perfil de expresión de una paciente.

[0070] De este modo, en una materialización más específica, el método de más arriba para determinar la categoría concreta de un individuo en relación con el tumor comprende los pasos de (1) hibridar los polinucleótidos destino etiquetados de un individuo con un biochip que contenga uno de los conjuntos de marcadores de más arriba; (2) hibridar moléculas polinucleótidas de patrón o control con el biochip, en donde las moléculas de patrón o control están etiquetadas de forma diferenciada con respecto a las moléculas destino; (3) determinar el cociente (o diferencia) de niveles de transcripción entre dos canales (individual y de control), o simplemente los niveles de transcripción del individuo, y (4) comparar los resultados de (3) con las plantillas predefinidas, en donde dicha determinación se lleva a cabo mediante la estadística de la Ecuación 1 o de la Ecuación 5 y en donde la diferencia, o la ausencia de la misma, determina la categoría del individuo en relación con el tumor.

#### 5.4.2 MÉTODOS DE PRONÓSTICO

[0071] La presente invención se refiere a conjuntos de marcadores que sirven para distinguir las muestras de pacientes con buen pronóstico de las muestras de pacientes con pronóstico deficiente. Así pues, la invención suministra un método de uso de estos marcadores para determinar si un individuo aquejado de cáncer de mama tendrá un pronóstico clínico bueno o deficiente. En una materialización, la invención suministra un método para determinar si un individuo aquejado de cáncer de mama tiene probabilidades de experimentar una recaída dentro de los cinco años siguientes al diagnóstico inicial (eso es, si un individuo tiene pronóstico deficiente) que comprende (1) comparar el nivel de expresión de los marcadores enumerados en la Tabla 5 en una muestra tomada del individuo al nivel de los mismos marcadores en un patrón o control, donde los niveles del patrón o control representan los hallados en un individuo con pronóstico deficiente, y (2) determinar si el nivel de los nucleótidos relacionados con marcadores en la muestra del individuo se diferencia significativamente del nivel del control, en donde si no se aprecia una diferencia sustancial, la paciente tiene pronóstico deficiente, mientras que si hay una diferencia sustancial, la paciente tiene buen pronóstico. Los experimentados en la técnica advertirán en seguida que los marcadores asociados al buen pronóstico también pueden usarse como controles. En una materialización más específica se prueban ambos controles. En caso de que la reserva no sea "buen pronóstico" o "pronóstico deficiente" puros, debe hibridarse contra la reserva un conjunto de experimentos con individuos con resultado conocido, para definir las plantillas de expresión para los grupos de buen diagnóstico y diagnóstico deficiente. A cada individuo con resultado desconocido se lo hibrida contra la misma reserva y el perfil de expresión resultante se compara con las plantillas para predecir el resultado del mismo.

[0072] El pronóstico deficiente del cáncer de mama puede indicar que un tumor es relativamente agresivo, mientras que el buen pronóstico puede indicar que un tumor es relativamente no agresivo. Existe un método para determinar el curso del tratamiento de una paciente de cáncer de mama, que consiste en determinar si el nivel de expresión de los 231 marcadores de la Tabla 5, o un subconjunto de los mismos, se corresponde con el nivel de dichos marcadores en una muestra que representa un patrón de expresión de buen pronóstico o un patrón de pronóstico deficiente, y determinar una duración de tratamiento, en donde si la expresión se corresponde con el patrón de pronóstico deficiente, el tumor es tratado como tumor agresivo.

**[0073]** Al igual que con los marcadores de diagnóstico, el método puede usar el conjunto completo de marcadores enumerados en la Tabla 5. Sin embargo, también pueden usarse subconjuntos de los marcadores. En una materialización preferida, se usa el subconjunto enumerado en la Tabla 6.

5 **[0074]** La clasificación de una muestra como “buen pronóstico” o “pronóstico deficiente” se lleva a cabo de manera sustancialmente igual que para los marcadores de diagnóstico descritos más arriba, en donde se genera una plantilla con la que se comparan los niveles de expresión de marcadores en la muestra.

10 **[0075]** El uso de conjuntos de marcadores no está limitado a los estados referentes al pronóstico del cáncer de mama, por lo que puede aplicarse en una variedad de fenotipos o estados, clínicos o experimentales, donde juegue algún papel la expresión de genes. Allí donde se haya identificado un conjunto de marcadores que corresponda a dos o más fenotipos puede usarse los conjuntos de marcadores para distinguir dichos fenotipos. Por ejemplo, los fenotipos pueden ser el diagnóstico y/o el pronóstico de categorías clínicas o fenotipos asociados a otros cánceres, otros estados de enfermedad u otros estados fisiológicos, en donde los datos de nivel de expresión proceden de un conjunto de genes en correlación con el estado de enfermedad o fisiológico concreto.

#### 15 5.4.3 MEJORA DE LA SENSIBILIDAD A LAS DIFERENCIAS DE NIVEL DE EXPRESIÓN

20 **[0076]** Al usar los marcadores que aquí se dan a conocer y, de hecho, al usar cualquier conjunto de marcadores para diferenciar a un individuo con un fenotipo de otro individuo con un segundo fenotipo, podemos comparar la expresión absoluta de cada uno de los marcadores de una muestra con un control; por ejemplo, el control puede ser el del nivel medio de expresión de cada uno de los marcadores, respectivamente, en una reserva de individuos. Para aumentar la sensibilidad de la comparación, no obstante, los valores del nivel de expresión deben transformarse preferentemente de una serie de maneras.

25 **[0077]** Por ejemplo, el nivel de expresión de cada uno de los marcadores puede normalizarse por el nivel medio de expresión de todos los marcadores cuyo nivel de expresión se determina, o por el nivel medio de expresión de un conjunto de genes de control. Así, en una materialización, los marcadores se representan con sondas en un biochip y el nivel de expresión de cada uno de los marcadores se normaliza mediante la media o la mediana del nivel de expresión en todos los genes representados en el biochip, incluido cualquier gen no marcador. En una materialización específica, la normalización se lleva a cabo dividiendo la mediana o media del nivel de expresión de todos los genes del biochip. En otra materialización, el nivel de expresión de los marcadores se normaliza mediante la media o la mediana del nivel de expresión de un conjunto de marcadores de control. En una materialización específica, los marcadores de control comprenden un conjunto de genes constitutivos. En otra materialización específica, la normalización se lleva a cabo dividiendo por la mediana o media del nivel de expresión de los genes de control.

35 **[0078]** La sensibilidad de un ensayo basado en marcadores aumentará también si los niveles de expresión de marcadores individuales se comparan con la expresión de esos mismos marcadores en una reserva de muestras. Preferiblemente, la comparación se hará con la media o la mediana del nivel de expresión de cada uno de los genes marcadores en la reserva de muestras. Dicha comparación puede realizarse, por ejemplo, dividiendo por la media o la mediana del nivel de expresión de la reserva para cada uno de los marcadores del nivel de expresión cada uno de los marcadores de la muestra. Esto produce el efecto de acentuar las diferencias relativas de expresión entre los marcadores de la muestra y los marcadores de la reserva en conjunto, haciendo las comparaciones más sensibles y con más probabilidades de que produzcan resultados significativos que si solo se usan niveles de expresión absolutos. Los datos del nivel de expresión pueden transformarse de la forma que más convenga; preferiblemente, los datos del nivel de expresión para todos se transforman a logaritmo antes de tomar las medias o las medianas.

45 **[0079]** Para realizar comparaciones con una reserva pueden usarse dos métodos. Primero, los niveles de expresión de los marcadores de la muestra pueden compararse con el nivel de expresión de aquellos marcadores de la reserva donde el ácido nucleico procedente de la muestra y el ácido nucleico procedente de la reserva se hibridan en el curso de un único experimento. Dicho método exige que se genere nuevo ácido nucleico en la reserva para cada comparación o números limitados de comparaciones, por lo que se halla limitado por la cantidad de ácido nucleico disponible. Como alternativa, y preferiblemente, los niveles de expresión en una reserva, ya estén normalizados y/o transformados o no, se almacenarán en un ordenador, o en soportes informáticos, para usarlos en comparaciones con los datos individuales de nivel de expresión de la muestra (esto es, datos de un solo canal).

55 **[0080]** Así pues, la presente invención suministra el siguiente método para clasificar una primera célula u organismo como poseedor de uno de al menos dos diferentes fenotipos, donde los diferentes fenotipos comprenden un primer fenotipo y un segundo fenotipo. El nivel de expresión de cada gen de una pluralidad de ellos en una primera muestra procedente de la primera célula u organismo se compara con el nivel de expresión de cada uno de dichos genes, respectivamente, en una muestra colectiva de una pluralidad de células u

organismos, pluralidad de células u organismos que comprende diferentes células u organismos que muestran al menos dos diferentes fenotipos, respectivamente, como se dice más arriba, para producir un primer valor comparado. El primer valor comparado se compara a continuación con un segundo valor comparado, en donde dicho segundo valor comparado es el producto de un método que consiste en comparar el nivel de expresión de cada uno de dichos genes en una muestra de una célula u organismo que se caracteriza por tener el segundo fenotipo al nivel de expresión de cada uno de dichos genes, respectivamente, en la muestra colectiva. Opcionalmente, se puede comparar el primer valor comparado con valores comparados adicionales, respectivamente, donde cada valor comparado adicional es el producto de un método que consiste en comparar el nivel de expresión de cada uno de dichos genes en una muestra de una célula u organismo que se caracteriza por tener un fenotipo diferente de dichos primer y segundos fenotipos, pero incluido entre al menos dos diferentes fenotipos, al nivel de expresión de cada uno de dichos genes, respectivamente, en dicha muestra colectiva. Por último, se determina a cuál de dichos segundo, tercer y, si está presente, uno o más valores comparados adicionales se parece más dicho primer valor comparado, en donde se clasifica la primera célula u organismo como poseedor del fenotipo de la célula u organismo usado para producir dicho valor comparado más parecido a dicho primer valor comparado.

**[0081]** En una materialización específica de este método, cada uno de los valores comparados es un cociente de los niveles de expresión de cada uno de dichos genes. En otra materialización específica, se normaliza cada uno de los niveles de expresión de cada uno de los genes de la muestra colectiva antes de proceder a cualquier paso de la comparación. En una materialización más específica, la normalización de los niveles de expresión se lleva a cabo dividiendo por la mediana o media del nivel de expresión de cada uno de los genes o dividiendo por la media o mediana del nivel de expresión de uno o más genes constitutivos de la muestra colectiva de dicha célula u organismo. En otra materialización específica, los niveles de expresión normalizados se someten a una transformación logarítmica y los pasos de la comparación consisten en restar la transformación logarítmica del logaritmo de los niveles de expresión de cada uno de los genes de la muestra. En otra materialización específica, los dos o más diferentes fenotipos son diferentes etapas de una enfermedad o trastorno. En otra materialización específica, los dos o más diferentes fenotipos son diferentes pronósticos de una enfermedad o trastorno. En otra materialización específica más, los niveles de expresión de cada uno de los genes, respectivamente, de la muestra colectiva o de dichos niveles de expresión de cada uno de dichos genes de una muestra de la célula u organismo que se caracteriza por tener el primer fenotipo, segundo fenotipo o dicho fenotipo diferente de dichos primer y segundo fenotipos, respectivamente, se almacenan en un ordenador o en un soporte informático.

**[0082]** En otra materialización específica, los dos fenotipos tienen categoría de ER(+) o ER(-). En otra materialización específica, los dos fenotipos tienen categoría de *BRCA1* o de tumor esporádico. En otra materialización específica más, los dos fenotipos son de buen pronóstico y de pronóstico deficiente.

**[0083]** Se pueden usar también, por supuesto, datos de un solo canal sin comparación específica con una reserva de muestras matemática. Por ejemplo, se puede clasificar una muestra como poseedora de un primer o un segundo fenotipo, en donde el primer y el segundo fenotipos están relacionados, calculando la semejanza de la expresión de al menos 5 marcadores en la muestra, donde los marcadores están en correlación con el primer o el segundo fenotipo, con la expresión de los mismos marcadores en una primera plantilla de fenotipos y una segunda plantilla de fenotipos, (a) etiquetando ácidos nucleicos procedentes de una muestra con un fluoróforo para obtener una reserva de ácidos nucleicos etiquetados con fluoróforo; (b) poniendo en contacto dicho ácido nucleico etiquetado con fluoróforo con un biochip en condiciones en las que pueda darse la hibridación, detectando en cada uno de una pluralidad de diferentes loci del biochip una señal de emisión fluorescente procedente de dicho ácido nucleico etiquetado con fluoróforo que está ligado a dicho biochip en dichas condiciones, y (c) determinando la semejanza de expresión de genes marcadores de la muestra individual con la primera y segunda plantillas, en donde si dicha expresión es más parecida a la primera plantilla, se clasifica la muestra como poseedora del primer fenotipo, mientras que si dicha expresión es más parecida a la segunda plantilla, se clasifica la muestra como poseedora del segundo fenotipo.

## 5.5 DETERMINACIÓN DE LOS NIVELES DE EXPRESIÓN DE GENES MARCADORES

### 5.5.1 MÉTODOS

**[0084]** Los niveles de expresión de los genes marcadores de una muestra pueden determinarse por cualquier medio conocido en la técnica. El nivel de expresión puede determinarse aislando y determinando el nivel (esto es, la cantidad) de ácido nucleico transcrito de cada gen marcador. Como alternativa, o adicionalmente, se puede determinar el nivel de proteínas específicas convertidas desde el ARNm transcrito de un gen marcador.

**[0085]** El nivel de expresión de genes marcadores específicos puede hallarse determinando la cantidad de ARNm, o de polinucleótidos derivados del mismo, presente en una muestra. Puede usarse cualquier método para determinar los niveles de ARN. Por ejemplo, se aísla el ARN de una muestra y se lo separa en un gel de agarosa. A continuación el ARN separado se transfiere a un soporte sólido, como puede ser un filtro. Luego se hibridan con el filtro mediante hibridación northern sondas de ácido nucleico que representan uno o más marcadores y se determina la cantidad de ARN procedente del marcador. Dicha determinación puede ser visual



o asistida por ordenador, por ejemplo, mediante un densitómetro. Otro método de determinar los niveles de ARN consiste en el uso de un dot-blot (borrón de puntos) o un slot-blot (borrón de ranura). En este método se etiqueta el ARN de una muestra, o el ácido nucleico derivado del mismo. A continuación el ARN o el ácido nucleico derivado del mismo se hibrida con un filtro que contiene oligonucleótidos procedentes de uno o más genes marcadores, en donde los oligonucleótidos se colocan sobre el filtro en posiciones diferenciadas y fácilmente identificables. La hibridación, o la ausencia de la misma, del ARN etiquetado con los oligonucleótidos ligados al filtro se determina visualmente o mediante densitómetro. Los polinucleótidos se pueden etiquetar mediante radiomarcaje o con una etiqueta fluorescente (esto es, visible).

**[0086]** Estos ejemplos no pretenden ser restrictivos; en la técnica se conocen otros métodos para determinar la abundancia de ARN.

**[0087]** El nivel de expresión de genes marcadores concretos puede calcularse también determinando el nivel de la proteína específica expresado a partir de los genes marcadores. Esto puede realizarse, por ejemplo, mediante la separación de proteínas de una muestra en un gel de poliacrilamida, seguida de la identificación de proteínas específicas derivadas de marcador mediante anticuerpos en un borrón de western. Como alternativa, las proteínas se pueden separar mediante sistemas de electroforesis bidimensional en gel. La electroforesis bidimensional en gel es bien conocida en la técnica y normalmente consiste en un enfoque isoeléctrico a lo largo de una primera dimensión seguido de electroforesis de SDS-PAGE (sistema de dilución simple-electroforesis en gel de poliacrilamida) a lo largo de una segunda dimensión. Véase, por ej., Hames et al, 1990, GEL ELECTROPHORESIS OF PROTEINS: A PRACTICAL APPROACH, IRL Press, New York; Shevchenko et al., Proc. Nat'l Acad. Sci. USA 93: 1440-1445 (1996); Sagliocco et al., Yeast 12:1519-1533 (1996); Lander, Science 274:536-539 (1996). Los electroferogramas resultantes pueden analizarse con numerosas técnicas, incluidas las técnicas de espectrometría de masas, los borrones de western y el análisis de inmunotransferencia mediante anticuerpos policlonales y monoclonales.

**[0088]** Como alternativa, los niveles de proteínas derivadas de marcadores pueden determinarse construyendo un biochip de anticuerpos en el que los centros de unión comprenden anticuerpos específicos, inmovilizados y preferiblemente monoclonales, a una pluralidad de especies proteínicas codificadas por el genoma de la célula. Preferiblemente, los anticuerpos estarán presentes en una fracción sustancial de las proteínas derivadas de marcadores que interesen. Los métodos para hacer anticuerpos monoclonales son bien conocidos (véase, por ej., Harlow and Lane, 1988, ANTIBODIES: A LABORATORY MANUAL, Cold Spring Harbor, New York). En una materialización, se crían anticuerpos monoclonales contra fragmentos de péptido sintético diseñados sobre la base de la secuencia genómica de la célula. Con dicho despliegue de anticuerpos, las proteínas procedentes de la célula se ponen en contacto con el despliegue y su unión se somete a prueba con ensayos conocidos en la técnica.

De modo general, la expresión, y el nivel de expresión, de proteínas de diagnóstico o pronóstico de interés pueden detectarse por tinción inmunohistoquímica de rodajas o secciones de tejido.

**[0089]** Por último, la expresión de genes marcadores en una serie de especímenes de tejido puede caracterizarse mediante un "despliegue de tejidos" (Kononen et al., Nat. Med 4 (7): 844-7 (1998)). En un despliegue de tejidos se calculan múltiples muestras de tejidos en el mismo biochip. Los despliegues permiten la detección in situ de los niveles de ARN y proteínas; las secciones consecutivas permiten el análisis de múltiples muestras simultáneamente.

#### 5.5.2 BIOCHIPS

**[0090]** En materializaciones preferidas, se usan biochips para medir la expresión a fin de que la categoría de expresión de cada uno de los marcadores de más arriba se calcule simultáneamente. Se describen despliegues de oligonucleótidos y ADNc que comprenden sondas hibridables con los genes correspondientes a cada uno de los conjuntos de marcadores descritos más arriba (esto es, marcadores para distinguir la categoría de ER; marcadores para distinguir tumores de *BRCA1* de tumores esporádicos; marcadores para distinguir pacientes con buen pronóstico de pacientes con pronóstico deficiente; marcadores para distinguir tanto ER (+) de ER (-) como tumores de *BRCA1* de tumores esporádicos; marcadores para distinguir ER (+) de ER (-) y pacientes con buen pronóstico de pacientes con pronóstico deficiente; marcadores para distinguir tumores de *BRCA1* de tumores esporádicos y pacientes con buen pronóstico de pacientes con pronóstico deficiente; marcadores capaces de distinguir ER (+) de ER (-), tumores de *BRCA1* de tumores esporádicos y pacientes con buen pronóstico de pacientes con pronóstico deficiente, y marcadores únicos para cada categoría).

**[0091]** Los biochips pueden comprender sondas hibridables con los genes con marcadores capaces de distinguir la categoría de una, dos o hasta las tres categorías clínicas indicados más arriba. Se suministran despliegues polinucleótidos que comprenden sondas a un subconjunto o subconjuntos de al menos 50, 100, 200, 300, 400, 500, 750, 1.000, 1.250, 1.500, 1.750, 2.000 o 2.250 marcadores genéticos, hasta el conjunto total de 2.460 marcadores, que distinguen entre ER (+) y ER (-) pacientes o tumores. También se suministran sondas a subconjuntos de al menos 20, 30, 40, 50, 75, 100, 150, 200, 250, 300, 350 o 400 marcadores, hasta el conjunto

total de 430 marcadores, que distinguen entre tumores que contienen una mutación de *BRCA1* y tumores esporádicos dentro de un grupo de tumores ER (-). Además, se suministran sondas a subconjuntos de al menos 20, 30, 40, 50, 75, 100, 150 o 200 marcadores, hasta el conjunto total de 231 marcadores, que distinguen entre pacientes con buen pronóstico y pronóstico deficiente dentro de los tumores esporádicos. En una materialización específica, el despliegue comprende sondas a conjuntos o subconjuntos de marcadores dirigidos a las tres categorías clínicas.

**[0092]** En otra materialización específica, los biochips que se usan en los métodos que aquí se muestran poseen marcadores adicionales a al menos algunos de los marcadores enumerados en las Tablas 1-6. Por ejemplo, en una materialización específica, el biochip es un despliegue de cribado o escaneo como el que se describe en Altschuler *et al.*, International Pub. WO 02/18646, de 7 Marzo de 2002 y Scherer *et al.*, International Pub. WO 02/16650, del 28 de febrero de 2002. Los despliegues de cribado y escaneo comprenden sondas de posición direccionable situadas a intervalos regulares y derivadas de la secuencia del ácido nucleico genómico, tanto expresa como no. También pueden comprender sondas en un subconjunto, o la totalidad de marcadores enumerados en las Tablas 1-6, o un subconjunto de los mismos, como se explica más arriba, y se puede usar para monitorizar la expresión de marcadores de la misma forma que un biochip que contenga sólo los marcadores enumerados en las Tablas 1-6.

**[0093]** En otra materialización específica más, el biochip es un biochip de ADNc disponible en el mercado que comprende al menos cinco de los marcadores enumerados en las Tablas 1-6. Preferiblemente, dicho biochip de ADNc disponible en el mercado comprenderá todos los marcadores enumerados en las Tablas 1-6. No obstante, dicho biochip puede comprender 5, 10, 15, 25, 50, 100, 150, 250, 500, 1.000 o más marcadores en cualquiera de las Tablas 1-6, hasta el número máximo de marcadores de una Tabla, y puede comprender todos los marcadores de cualquier Tabla 1 a 6 y un subconjunto de otra de las Tablas 1-6, o subconjuntos de cada una, como se explica más arriba. En una materialización específica de los métodos en este documento, los marcadores que son todas o parte de Tablas 1-6 constituyen al menos el 50%, 60%, 70%, 80%, 90%, 95% o 98% de las sondas del biochip.

**[0094]** En las siguientes secciones se describen métodos generales concernientes a la construcción de biochips y que comprenden los conjuntos y/o subconjuntos de más arriba.

#### 5.5.2.1 CONSTRUCCIÓN DE BIOCHIPS

**[0095]** Los biochips se preparan seleccionando sondas que comprenden una secuencia de polinucleótidos, para a continuación inmovilizar dichas sondas en un soporte o una superficie sólidos. Por ejemplo, las sondas pueden comprender secuencias de ADN, secuencias de ARN o secuencias de copolímeros de ADN y ARN. Las secuencias de polinucleótidos de las sondas pueden comprender también análogos de ADN y/o ARN, o combinaciones de los mismos. Por ejemplo, las secuencias de polinucleótidos de las sondas pueden estar llenas de fragmentos parciales de ADN genómico. Las secuencias de polinucleótidos de las sondas pueden ser también secuencias sintetizadas de nucleótidos, como las secuencias sintéticas de oligonucleótidos. Las secuencias de sondas pueden sintetizarse bien de manera enzimática, *in vivo*, enzimática *in vitro* (por ej., por PCR) o no enzimática *in vitro*.

**[0096]** La sonda o sondas usadas en los métodos de la invención estarán preferiblemente inmovilizadas en un soporte sólido que puede ser tanto poroso como no poroso. Por ejemplo, las sondas de la invención pueden ser secuencias de polinucleótidos que están sujetas a una membrana o filtro de nitrocelulosa o nailon por unión covalente por cualquiera de los dos extremos, 3' o 5', del polinucleótido. Tales sondas de hibridación son bien conocidas en la técnica (véase, por ej., Sambrook *et al.*, MOLECULAR CLONING - A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989). Como alternativa, el soporte o superficie sólidos pueden ser una superficie de cristal o plástico. En una materialización especialmente preferida, los niveles de hibridación se miden con respecto a biochips de sondas que constan de una fase sólida en cuya superficie se halla inmovilizada una población de polinucleótidos, como una población de ADN o mímicos de ADN o, como alternativa, una población de ARN o mímicos de ARN. La fase sólida puede ser un material no poroso u opcionalmente poroso, como un gel

**[0097]** En la materializaciones preferidas, el biochip comprende un soporte o superficie con un despliegue ordenado de centros de unión (por ej., por hibridación) o "sondas", cada una de las cuales representa uno de los marcadores descritos en este documento. Preferiblemente, los biochips serán despliegues direccionables y, más preferiblemente, despliegues de posición direccionable. Más concretamente, cada sonda del despliegue estará situada preferiblemente en una ubicación conocida y predeterminada en el soporte sólido de modo que la identidad (esto es, la secuencia) de cada sonda pueda determinarse a partir de su posición en el despliegue (esto es, en el soporte o superficie). En las materializaciones preferidas, cada prueba está fijada al soporte sólido por unión covalente por un solo punto.

**[0098]** Los biochips pueden hacerse de varias maneras, algunas de las cuales se describen a continuación. Independientemente de la forma en que se produzcan, los biochips comparten ciertas características. Los biochips son reproducibles, por lo que se puede producir múltiples copias de un biochip concreto y compararlas una con otra. Preferiblemente, los biochips estarán hechos de materiales que sean estables en condiciones de unión (por ej., hibridación del ácido nucleico). Los biochips serán preferiblemente pequeños, por ej., entre 1 cm<sup>2</sup> y 25 cm<sup>2</sup>, entre 12 cm<sup>2</sup> y 13 cm<sup>2</sup>, o 3 cm<sup>2</sup>. Sin embargo, también se contemplan biochips de mayor tamaño, que hasta pueden ser preferibles, por ej. para usarlos en despliegues de cribado. Preferiblemente, un centro de unión concreto o un único conjunto de centros de unión en el biochip se unirá (por ej., hibridará) específicamente al producto de un único gen en una célula (por ej., a un ARNm específico, o a un ADN específico derivado del mismo). No obstante, en general, otras secuencias relacionadas o parecidas se vincularán a un determinado centro de unión por hibridación cruzada.

**[0099]** Los biochips de la presente invención incluyen una o más sondas de ensayo, cada una de las cuales tiene una secuencia polinucleótida que es complementaria con una subsecuencia de ARN o ADN que hay que detectar. Preferiblemente, se conocerá la ubicación de cada sonda en la superficie sólida. De hecho, los biochips serán preferiblemente despliegues de posición direccionable. En concreto, cada sonda del despliegue estará situada preferiblemente en una ubicación conocida y predeterminada del soporte sólido, de modo que la identidad (esto es, la secuencia) de cada sonda pueda determinarse a partir de su posición en el despliegue (esto es, en el soporte o superficie).

**[0100]** Según la invención, el biochip es un despliegue (esto es, una matriz) en el que cada posición representa uno de los marcadores descritos en este documento. Por ejemplo, cada posición puede contener un ADN o un análogo de ADN basado en el ADN genómico con el que puede hibridar específicamente un ARN o ADNc concretos transcritos del marcador genético. El ADN o análogo de ADN puede ser, por ej., un oligómero sintético o un fragmento de gen. En una materialización, las sondas que representan cada uno de los marcadores están presentes en el despliegue. En una materialización preferida, el despliegue comprende los 550 de los 2.460 marcadores de categoría ER-, 70 de los marcadores BRCA1/espórádicos y el total de los 231 marcadores de pronóstico.

#### 5.5.2.2 PREPARACIÓN DE SONDAS PARA BIOCHIPS

**[0101]** Como se indica más arriba, la "sonda" con la que según la invención se hibrida específicamente una molécula polinucleótida concreta contiene una secuencia polinucleótida genómica complementaria. Las sondas del biochip consistirán preferiblemente en secuencias nucleótidas de no más de 1.000 nucleótidos. En algunas materializaciones, las sondas del despliegue consisten en secuencias nucleótidas de 10 a 1.000 nucleótidos. En una materialización preferida, las secuencias nucleótidas de las sondas son del orden de 10-200 nucleótidos de longitud y son secuencias genómicas de una especie de organismo, de modo que se halla presente una pluralidad de diferentes sondas, con secuencias complementarias y por lo tanto capaces de hibridar con el genoma de dicha especie u organismo, superpuestas en secuencia por todo o parte de dicho genoma. En otras materializaciones específicas, las sondas son del orden de 10-30 nucleótidos de longitud, del orden de 10-40 nucleótidos de longitud, del orden de 20-50 nucleótidos de longitud, del orden de 40-80 nucleótidos de longitud, del orden de 50-150 nucleótidos de longitud, del orden de 80-120 nucleótidos de longitud y, más preferiblemente, de 60 nucleótidos de longitud.

**[0102]** Las pruebas pueden comprender ADN o "mímicos" de ADN (por ej., derivados y análogos) correspondientes a una porción del genoma de un organismo. En otra materialización, las sondas del biochip son complementarios de ARN o mímicos de ARN. Los mímicos de ADN son polímeros compuestos de subunidades capaces de hibridación específica de tipo Watson-Crick con ADN, o de hibridación específica con ARN. Los ácidos nucleicos pueden modificarse en la fracción de base, en la fracción de azúcar o en el esqueleto del fosfato. Como ejemplo de mímicos de ADN se puede citar, por ej. los fosforotioatos.

**[0103]** El ADN se puede obtener, por ej., por reacción en cadena de polimerasa (PCR), por amplificación del ADN genómico o por secuencias clonadas. Los cebadores de PCR se elegirán preferiblemente sobre la base de una secuencia conocida del genoma, que resultará en la amplificación de fragmentos específicos de ADN genómico. Hay programas informáticos bien conocidos en la técnica que sirven para diseñar cebadores con la especificidad requerida y las propiedades óptimas de amplificación, como la versión 5.0 de *Oligo* (National Biosciences). Típicamente, cada sonda del despliegue estará entre 10 bases y 50.000 bases, normalmente entre 300 bases y 1.000 bases de longitud. Los métodos de PCR son bien conocidos en la técnica y se describen, por ejemplo, en Innis et al., eds., PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS, Academic Press Inc., San Diego, CA (1990). Para los experimentados en la técnica resultará evidente que los sistemas robóticos de control son útiles para aislar y amplificar los ácidos nucleicos.

**[0104]** Un medio preferido alternativo para generar las sondas polinucleótidas del biochip lo constituye la síntesis de polinucleótidos u oligonucleótidos sintéticos, por ej., mediante químicas de N-fosfonato o fosforamidita (Froehler et al., Nucleic Acid Res. 14:5399-5407 (1986); McBride et al., Tetrahedron Lett. 24:246-248 (1983)).

Las secuencias sintéticas tienen típicamente de unas 10 a unas 500 bases de longitud, y más preferiblemente de unas 40 a unas 70 bases de longitud. En algunas materializaciones, los ácidos nucleicos sintéticos incluyen bases no naturales, como por ejemplo, pero en modo alguno limitadas a ella, la inosina. Como se indica más arriba, se pueden usar análogos de ácido nucleico como centros de unión para hibridación. Un ejemplo de ácido nucleico adecuado es el ácido nucleico péptido (véase por ej., Egholm et al., *Nature* 363:566-568 (1993); U.S. Patent No. 5,539,083). Las sondas se seleccionarán preferiblemente mediante un algoritmo que tiene en cuenta las energías de unión, la composición de la base, la complejidad de la secuencia, las energías de unión de la hibridación cruzada y la estructura secundaria (véase Friend *et al.*, International Patent Publication WO 01/05935, publicado el 25 de enero de 2001; Hughes et al., *Nat. Biotech.* 19:342-7 (2001)).

**[0105]** Un operario experimentado advertirá también que las sondas de control positivo, por ej., las sondas de las que se sabe que son complementarias e hibridables con secuencias en las moléculas del polinucleótido destino y las sondas de control negativo, por ej. las sondas de las que se sabe que no son complementarias e hibridables con secuencias en las moléculas del polinucleótido destino, deben incluirse en el despliegue. En una materialización, los controles positivos se sintetizan a lo largo del perímetro del despliegue. En otra materialización, los controles positivos se sintetizan en franjas diagonales a través del despliegue. En otra materialización más, el complemento inverso para cada sonda se sintetiza junto a la posición de la sonda para que sirva de control negativo. Y en otra materialización más, se usan secuencias de otras especies de organismo como controles negativos y como controles "enclavados".

#### 5.5.2.3 SUJECCIÓN DE SONDAS A LA SUPERFICIE SÓLIDA

**[0106]** Las sondas se sujetan a un soporte o superficie sólidos, que pueden estar hechos por ej. de vidrio, de plástico (por ej. polipropileno, nailon), poliacrilamida, nitrocelulosa, gel u otro material poroso o no poroso. Un método preferido para sujetar los ácidos nucleicos a una superficie es imprimiendo en placas de vidrio, como se explica de modo general en Schena et al, *Science* 270:467-470 (1995). Este método es especialmente útil para preparar biochips de ADNc (véase también DeRisi et al, *Nature Genetics* 14:457-460 (1996); Shalon et al., *Genome Res.* 6:639-645 (1996); and Schena et al., *Proc. Natl. Acad. Sci. U.S.A.* 93:10539-11286 (1995)).

**[0107]** Un segundo método preferido para hacer biochips es hacer despliegues de alta densidad. Se conocen técnicas para producir despliegues que contienen miles de oligonucleótidos complementarios para secuencias definidas en ubicaciones definidas en una superficie mediante técnicas fotolitográficas para síntesis *in situ* (véase Fodor et al., 1991, *Science* 251:767-773; Pease et al., 1994, *Proc. Natl. Acad. Sci. U.S.A.* 91:5022-5026; Lockhart et al., 1996, *Nature Biotechnology* 14:1675; Patentes U.S. Nos. 5,578,832, 5,556,752 y 5,510,270) u otros métodos para síntesis y deposición rápidas de oligonucleótidos definidos (Blanchard et al., *Biosensors & Bioelectronics* 11:687-690). Cuando se usan estos métodos, los oligonucleótidos (por ej., 60-meros) de secuencia conocida se sintetizan directamente sobre una superficie tal como un portaobjeto de cristal derivatizado. Por lo general, el despliegue producido es redundante, con varias moléculas oligonucleótidas por cada ARN.

**[0108]** Se pueden usar otros métodos para hacer biochips, por ej. por enmascaramiento (Maskos and Southern, 1992, *Nuc. Acids. Res.* 20: 1679-1684). En principio, y como se indica más arriba, se puede usar cualquier tipo de despliegue, por ejemplo, borrones de puntos (dot blots) en una membrana de hibridación de nailon (véase Sambrook et al., *MOLECULAR CLONING - A LABORATORY MANUAL* (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989)). No obstante, como advertirán los experimentados en la técnica, muchas veces son preferibles despliegues muy pequeños, ya que los volúmenes de hibridación serán menores.

**[0109]** En una materialización, los despliegues de la presente invención se preparan sintetizando sondas polinucleótidas en un soporte. En dicha materialización, las sondas polinucleótidas van sujetas al soporte por unión covalente por cualquiera de los dos extremos, 3' o 5', del polinucleótido.

**[0110]** En una materialización especialmente preferida, se fabrican biochips de la invención mediante un dispositivo de impresión por chorro de tinta para síntesis de oligonucleótidos, por ej. con los métodos y sistemas descritos por Blanchard en U.S. Pat. No. 6,028,189; Blanchard et al., 1996, *Biosensors and Bioelectronics* 11:687-690; Blanchard, 1998, in *SYNTHETIC DNA ARRAY IN GENETIC ENGINEERING*, Vol. 20, J.K. Setlow, Ed., Plenum Press, New York, páginas 111-123. Concretamente, las sondas polinucleótidas de dichos biochips se sintetizarán preferiblemente en despliegues, por ej. en un portaobjeto de cristal, depositando en serie bases nucleótidas individuales en "microgotitas" de un disolvente de alta tensión superficial, como el carbonato de propileno. Las microgotitas tienen un volumen muy pequeño (por ej. 100 pL o menos, más preferiblemente 50 pL o menos) y están separadas unas de otras en el biochip (por ej. por dominios hidrófobos) para formar pozos de tensión de superficie circular que definen las posiciones de los elementos del despliegue (esto es, las diferentes sondas). Los biochips fabricados con este método de chorro de tinta suelen ser de alta densidad, preferiblemente con una densidad de al menos 2,500 diferentes sondas por cm<sup>2</sup>. Las sondas polinucleótidas están sujetas al soporte por unión covalente por cualquiera de los dos extremos, 3' o 5', del polinucleótido.

5.5.2.4 MOLECULAS POLINUCLEÓTIDAS DESTINO

**[0111]** Las moléculas polinucleótidas que pueden analizarse mediante la presente invención (las "moléculas polinucleótidas destino") pueden ser de cualquier origen clínicamente relevante, pero se expresan como ARN o como un ácido nucleico derivado del mismo (por ej. ADNc o ARN amplificado derivado de ADNc que lleva incorporado un promotor de polimerasa de ARN), incluidas moléculas de ácido nucleico de generación natural, así como moléculas de ácido nucleico sintético. En una materialización, las moléculas polinucleótidas destino comprenden ARN, incluyendo, pero sin limitarse en absoluto a ellos, ARN celular total, ARN mensajero poli(A)+ (ARNm) o fracción del mismo, ARNm citoplásmico o ARN transcrito de ADNc (esto es, ARNc, véase, por ej., Linsley & Schelter, U.S. Patent Application No. 09/411,074, registrada el 4 de octubre de 1999, o las Patentes U.S. Nos. 5,545,522, 5,891,636, o 5,716,785). Los métodos para preparar ARN total y poli(A)+ son bien conocidos en la técnica y se describen de modo general en, por ej., Sambrook et al., MOLECULAR CLONING - A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989). En una materialización, se extrae ARN de células de los diversos tipos de interés de esta invención por lisis de tiocianato de guanidina seguida de centrifugación CsCl (Chirgwin et al., 1979, Biochemistry 18: 5294-5299). En otra materialización, se extrae ARN total mediante una columna de silicio con base de gel, comercializada por RNeasy (Qiagen, Valencia, California) y StrataPrep (Stratagene, La Jolla, California). En una materialización alternativa, que es preferida para *S. cerevisiae*, se extrae ARN de células mediante fenol y cloroformo, como se explica en Ausubel et al., eds., 1989, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Vol III, Green Publishing Associates, Inc., John Wiley & Sons, Inc., New York, págs. 13.12.1-13.12.5). El ARN Poli(A)+ puede seleccionarse, por ej., por selección con celulosa oligo-dT o, como alternativa, por transcripción inversa cebada de oligo-dT del ARN celular total. En una materialización, se puede fragmentar el ARN con métodos conocidos en la técnica, por ej., por incubación con ZnCl<sub>2</sub>, para generar fragmentos de ARN. En otra materialización, las moléculas polinucleótidas analizadas por la invención comprenden ADNc, o productos PCR de ARN o ADNc amplificados.

**[0112]** En una materialización, el ARN total, el ARNm o los ácidos nucleicos derivados de los mismos son aislados de una muestra tomada de una persona aquejada de cáncer de mama. Las moléculas polinucleótidas destino que están pobremente expresadas en células concretas pueden enriquecerse mediante técnicas de normalización (Bonaldi et al., 1996, Genome Res. 6:791-806).

**[0113]** Como se explica más arriba, los polinucleótidos destino se etiquetan de forma detectable en uno o más nucleótidos. Cualquier método conocido en la técnica sirve para etiquetar de forma detectable los polinucleótidos destino. Preferiblemente, este etiquetado llevará incorporada uniformemente la etiqueta todo a lo largo del ARN y, más preferiblemente, el etiquetado se llevará a cabo con el mayor grado posible de eficacia. Una materialización usa para este etiquetado transcripción inversa cebada de oligo-dT para incorporar la etiqueta; sin embargo, los métodos convencionales de este método tienden a generar fragmentos con extremo de 3'. Por consiguiente, en una materialización preferida, se usan cebadores aleatorios (por ej. 9-meros) en transcripción inversa para incorporar uniformemente nucleótidos etiquetados todo a lo largo de los polinucleótidos destino. Como alternativa, los cebadores aleatorios pueden usarse en conjunción con métodos de PCR o métodos de transcripción *in vitro* basados en el promotor T7 para amplificar los polinucleótidos destino.

**[0114]** En una materialización preferida, la etiqueta detectable es una etiqueta luminiscente. En la presente invención se pueden usar, por ejemplo, etiquetas fluorescentes, etiquetas bioluminiscentes, etiquetas quimioluminiscentes y etiquetas colorimétricas. En una materialización particularmente preferida, la etiqueta es una etiqueta fluorescente, como una fluoresceína, un fósforo, una rodamina o un derivado de tintura de polimetina. Como ejemplos de etiquetas fluorescentes disponibles en el mercado están, entre otras, las fosforamiditas fluorescentes como FluorePrime (Amersham Pharmacia, Piscataway, N.J.), Fluoredit (Millipore, Bedford, Mass.), FAM (ABI, Foster City, Calif.), y Cy3 o Cy5 (Amersham Pharmacia, Piscataway, N.J.). En otra materialización, la etiqueta detectable es un nucleótido radiomarcado.

**[0115]** En otra materialización preferida, las moléculas polinucleótidas destino de la muestra de una paciente se etiquetan de modo diferente a las moléculas polinucleótidas destino de un patrón. El patrón puede comprender moléculas polinucleótidas destino de individuos normales (esto es, no aquejados de cáncer de mama). En una materialización especialmente preferida, el patrón comprende moléculas polinucleótidas destino reunidas a partir de muestras de individuos normales o muestras de tumores de individuos con tumores cancerosos de tipo esporádico. En otra materialización, las moléculas polinucleótidas destino proceden del mismo individuo, pero se toman en diferentes puntos temporales y de este modo indican la eficacia de un tratamiento por el cambio de expresión en los marcadores, o la ausencia del mismo, durante y después del curso del tratamiento (esto es, quimioterapia, radioterapia o crioterapia), en donde el paso en la expresión de los marcadores de un patrón de pronóstico deficiente a un patrón de buen pronóstico indica que el tratamiento es eficaz. En esta materialización, puntos temporales diferentes se etiquetan de forma diferente.

#### 5.5.2.5 HIBRIDACIÓN CON BIOCHIPS

**[0116]** La hibridación del ácido nucleico y las condiciones de lavado se eligen de forma que las moléculas polinucleótidas destino se unan o hibriden específicamente con las secuencias polinucleótidas complementarias del despliegue, preferiblemente con un lugar de despliegue concreto donde esté situado su ADN complementario.

**[0117]** Los despliegues que contienen sondas de ADN de doble cadena situadas en los mismos se someterán preferiblemente a condiciones de desnaturalización para hacer el ADN de cadena simple antes de que entre en contacto con las moléculas polinucleótidas destino, por ej., eliminar estructuras en horquilla o dímeros que forman cuota con las secuencias autocomplementarias.

**[0118]** Las condiciones de hibridación óptimas dependerán de la longitud (por ej., oligómeros frente a polinucleótidos mayores de 200 bases) y del tipo (por ej., ARN o ADN) de los ácidos nucleicos destino y de sonda. Los experimentados en la materia advertirán que a medida que los oligonucleótidos se van haciendo más cortos, puede que sea necesario ajustar su longitud para conseguir una temperatura de fusión relativamente uniforme y por tanto unos resultados de hibridación satisfactorios. Los parámetros generales para las condiciones de hibridación específicas (esto es, rigurosas) de los ácidos nucleicos se describen en Sambrook et al., *MOLECULAR CLONING - A LABORATORY MANUAL* (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989), y en Ausubel et al., *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY*, vol. 2, Current Protocols Publishing, New York (1994). Las condiciones de hibridación típicas para los biochips de ADNc de Schena *et al.* son la hibridación en 5 X SSC más 0,2% SDS a 65°C durante cuatro horas, seguida de lavados a 25°C en tampón de lavado de baja dureza (1 X SSC más 0,2% SDS), seguidos de 10 minutos a 25°C en tampón de lavado de mayor dureza (0,1 X SSC más 0,2% SDS) (Schena et al., *Proc. Natl. Acad. Sci. U.S.A.* 93:10614 (1993)). También son de gran utilidad las condiciones de hibridación suministradas en, por ej., Tijessen, 1993, *HYBRIDIZATION WITH NUCLEIC ACID PROBES*, Elsevier Science Publishers B.V.; y Kricka, 1992, *NONISOTOPIC DNA PROBE TECHNIQUES*, Academic Press, San Diego, CA.

**[0119]** Condiciones de hibridación especialmente preferidas son la hibridación de las sondas a la temperatura media de fusión, o cercana a ella (por ej., a no más de 5 °C, más preferiblemente a no más de 2 °C) en 1 M NaCl, 50 mM de tampón MES (pH 6.5), 0.5% de sarcosina de sodio y 30% de formamida.

#### 5.5.2.6 SEÑAL DE DETECCIÓN Y ANÁLISIS DE DATOS

**[0120]** Cuando se usan sondas con etiqueta fluorescente, las emisiones de fluorescencia en cada lugar de un biochip podrán detectarse, preferiblemente, mediante microscopia confocal de barrido láser. En una materialización se efectúa un barrido por separado, mediante la adecuada línea de excitación, para cada uno de los dos fluoróforos usados. Como alternativa, se puede usar un láser que permita iluminación de espécimen simultánea a longitudes de onda específicas a los dos fluoróforos y que las emisiones de los dos fluoróforos se puedan analizar simultáneamente (véase Shalon et al., 1996, "A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization", *Genome Research* 6:639-645). En una materialización preferida, los despliegues se escanean con un escáner de láser fluorescente con una fase X-Y controlada por ordenador y con objetivo microscópico. La excitación secuencial de los dos fluoróforos se lleva a cabo con un láser de gas mixto multilínea y la luz emitida se fracciona por longitud de onda y se detecta mediante dos tubos fotomultiplicadores. En Schena et al., *Genome Res.* 6:639-645 (1996), y en otras referencias citadas en el presente documento, se describen dispositivos de escaneo por láser fluorescente. Como alternativa, se puede usar el haz de fibra óptica descrito por Ferguson et al., *Nature Biotech.* 14:1681-1684 (1996), para monitorizar la abundancia de los niveles de mRNA en un gran número de lugares a la vez.

**[0121]** Las señales se graban y, en una materialización preferida, se analizan por ordenador, por ej. mediante una placa análoga a digital de 12 o 16 bits. En una materialización, se eliminan las intermitencias de la imagen escaneada mediante un programa de gráficos (por ej., Hijaak Graphics Suite) y a continuación se analiza mediante un programa de ploteado de imágenes que crea una hoja de cálculo de la hibridación media por cada longitud de onda en cada lugar. Si es necesario, se puede hacer una determinada corrección experimental para "producir diafonía" (superponer) entre los canales para los dos fluores. Para cualquier lugar concreto de hibridación del despliegue de transcripción se puede calcular un cociente de la emisión de los dos fluoróforos. El cociente es independiente de la expresión absoluta del gen afín, pero es útil para los genes cuya expresión se modula significativamente en asociación con los diferentes estados relacionados con el cáncer de mama.

#### 5.6 ANÁLISIS ASISTIDO POR ORDENADOR

**[0122]** A continuación se describen kits que comprenden los conjuntos de marcadores de más arriba. En una materialización preferida, el kit contiene un biochip listo para hibridarse con moléculas polinucleótidas destino, más el software para los análisis de datos explicados más arriba.

**[0123]** Los métodos analíticos explicados en las anteriores secciones pueden ponerse en práctica mediante los siguientes sistemas informáticos y según los siguientes programas y métodos. El sistema informático comprende componentes internos conectados a componentes externos. Los componentes internos de un sistema informático típico incluyen un elemento procesador interconectado a una memoria principal. Por ejemplo, el sistema informático puede ser un Intel 8086-, 80386-, 80486-, Pentium™ o procesador basado en Pentium™ preferiblemente con 32 MB o más de memoria principal.

**[0124]** Los componentes externos pueden incluir almacenamiento masivo. Este almacenamiento masivo puede ser uno o más discos duros (que suele ir incluido en el mismo paquete que el procesador y la memoria). Dichos discos duros tendrán preferiblemente una capacidad de 1 GB o más de memoria. Los demás componentes externos son un dispositivo de interfaz de usuario, que puede ser un monitor, junto con un dispositivo de entrada, que puede ser un "ratón", u otros dispositivos de entrada gráfica, y/o un teclado. También se le puede conectar al ordenador un dispositivo de impresión.

**[0125]** Típicamente, el sistema informático está conectado también a un enlace de red, que puede ser parte de un enlace de Ethernet con otros sistemas informáticos locales, sistemas informáticos remotos o redes de comunicación de gran amplitud, como Internet. Este enlace de red permite que el sistema informático comparta datos y tareas de procesado con otros sistemas informáticos.

**[0126]** Durante el funcionamiento de este sistema se cargan en la memoria varios componentes de software, que son a la vez estándares en la técnica y especiales en la presente invención. Estos componentes de software hacen colectivamente que el sistema informático funcione según los métodos de esta invención. Estos componentes de software están almacenados típicamente en el dispositivo de almacenamiento masivo. Un componente de software comprende el sistema operativo, que es responsable de la dirección del sistema informático y sus interconexiones. Este sistema operativo puede ser, por ejemplo, de la familia de Microsoft Windows®, como Windows 3.1, Windows 95, Windows 98, Windows 2000 o Windows NT. El componente de software representa lenguajes y funciones comunes convenientemente presentes en este sistema para ayudar a los programas a poner en práctica los métodos específicos de esta invención. Se pueden usar muchos lenguajes informáticos de nivel alto o bajo para programar los métodos analíticos de esta invención. Las instrucciones pueden interpretarse durante el periodo de funcionamiento o resumirse. Entre los lenguajes preferidos están C/C++, FORTRAN y JAVA. Lo más preferible es que los métodos de esta invención se programen en paquetes de software matemático que permiten la entrada simbólica de ecuaciones y especificación de proceso de alto nivel, incluyendo algunos o todos los algoritmos que van a usarse, liberando de este modo al usuario de la necesidad de programar procesalmente ecuaciones individuales o algoritmos. Dentro de dichos paquetes están Mathlab de Mathworks (Natick, MA), Mathematica® de Wolfram Research (Champaign, IL), o S-Plus® de Math Soft (Cambridge, MA). Concretamente, el componente de software incluye los métodos analíticos de la invención tal como se programan en un lenguaje procesal o paquete simbólico.

**[0127]** El software que debe incluirse con el kit comprende los métodos de análisis de datos de la invención tal como se dan a conocer en el presente documento. Concretamente, el software puede incluir rutinas matemáticas para descubrimiento de marcadores, incluido el cálculo de coeficientes de correlación entre categorías clínicas (esto es, la categoría de ER) y la expresión de marcadores. El software también puede incluir rutinas matemáticas para calcular la correlación entre expresión de marcadores de muestra y expresión de marcadores de control, mediante datos de la fluorescencia generada por el despliegue, para determinar la clasificación clínica de una muestra.

**[0128]** En una materialización modelo, para poner en práctica los métodos de la presente invención, primero el usuario carga datos experimentales en el sistema informático. Estos datos puede introducirlos directamente el usuario desde un monitor, un teclado u cualquier otro sistema informático conectado por conexión de red, o en un medio de almacenamiento amovible como un CD-ROM, un disquete (que no se muestra en las ilustraciones), una unidad de cinta magnética (que tampoco se muestra), una unidad de ZIP® drive (que tampoco se muestra) o a través de la red. A continuación el usuario procede a la ejecución del software de análisis del perfil de expresión que lleva a cabo los pasos de la presente invención.

**[0129]** En otra materialización modelo, primero el usuario carga datos experimentales y/o bases de datos en el sistema informático. Estos datos se cargan en la memoria desde los medios de almacenamiento o desde un ordenador remoto, preferiblemente desde un sistema dinámico de bases de datos de conjuntos de genes, a través de la red. A continuación el usuario procede a la ejecución del software que lleva a cabo los pasos de la presente invención.

**[0130]** A los experimentados en la técnica les resultarán evidentes sistemas informáticos y software alternativos para llevar a cabo los métodos analíticos de esta invención, que deben entenderse incluidos en las reivindicaciones anexas. Concretamente, es propósito de las reivindicaciones incluir las estructuras de programa alternativo para llevar a cabo los métodos de esta invención, cuya evidencia advertirán sin dificultad los experimentados en la técnica.

## 6. EJEMPLOS

## Materiales y Métodos

**[0131]** Se recogieron 117 muestras de tumores de pacientes de cáncer de mama. A continuación se prepararon muestras de ARN y se trazó el perfil de cada una de dichas muestras mediante biochips impresos por chorro de tinta. Después se identificaron genes marcadores basándose en patrones de expresión; estos genes se usaron luego para las prácticas con clasificadores, que usaron estos genes marcadores para clasificar tumores en categorías de diagnóstico y pronóstico. Por último, estos genes marcadores se usaron para predecir el resultado del diagnóstico y del pronóstico en un grupo de individuos.

1. Recogida de muestras

**[0132]** Se seleccionó a 117 pacientes de cáncer de mama en tratamiento en el Instituto de Cáncer de los Países Bajos / Antoni van Leeuwenhoek Hospital, Amsterdam, Países Bajos, ateniéndose a los siguientes métodos clínicos (datos extraídos de las historias clínicas del Registro de Tumores NKI/AvL, Departamento de Biométrica).

**[0133]** El grupo 1 (n=97, 78 para prácticas, 19 para ensayos independientes) se clasificó según el método de: (1) carcinoma de mama invasor primario <5 cm (T1 o T2); (2) ausencia de metástasis axilar (N0); (3) edad de diagnóstico <55 años; (4) periodo del diagnóstico 1983-1996, y (5) sin tumores malignos con anterioridad (excluyendo tumor cervical maligno in situ o carcinoma basocelular de la piel). A todas las pacientes se las trató con mastectomía radical modificada (n=34) o con tratamiento conservador de la mama (n=64) incluida la disección del ganglio linfático axilar. El tratamiento conservador de la mama consistió en la extirpación del tumor, seguida de la radiación de toda la mama en una dosis de 50 Gy, seguida de un refuerzo de 15 a 25 Gy. Cinco pacientes recibieron terapia sistémica coadyuvante consistente en quimioterapia (n=3) o terapia hormonal (n=2), ninguna de las demás pacientes recibió tratamiento adicional. Todas las pacientes tuvieron como mínimo un seguimiento anual durante un período de al menos 5 años. La investigación de las pacientes se extrajo del Registro de Tumores del Departamento de Biométrica.

**[0134]** Las integrantes del grupo 2 (n=20) se clasificaron en: (1) portadoras de una mutación germinal en *BRCA1* o *BRCA2*; y (2) aquejadas de un carcinoma de mama invasor primario. No se hizo selección o exclusión alguna basándose en el tamaño del tumor, la categoría del ganglio linfático, la edad de diagnóstico, el año de la realización del diagnóstico ni otros tumores malignos. La clase de mutación germinal ya se conocía antes de realizar este protocolo de investigación.

**[0135]** La información sobre el individuo de quien se recogió muestras tumorales comprende: año de nacimiento; sexo; si el individuo es pre-menopáusico o pos-menopáusico; año del diagnóstico; número de ganglios linfáticos positivos y número total de ganglios; si hubo cirugía y, si fue así, si la cirugía fue conservadora de la mama o radical; si hubo radioterapia, quimioterapia o terapia hormonal. El tumor se graduó según la fórmula P=TNM, donde T es el tamaño del tumor (en una escala de 0-5); N es el número de ganglios que son positivos (en una escala de 0-4) y M es metástasis (0 = ausente, 1 = presente). El tumor también se clasificó según la fase, el tipo de tumor (*in situ* o invasor; lobular o ductal; el grado) y la presencia o ausencia de los receptores de estrógeno y progesterona. La progresión del cáncer se describió (donde correspondía) por: metástasis distante, año de la metástasis distante, año de fallecimiento, año del último seguimiento y genotipo de *BRCA1*.

2. Tumores

**[0136]** El ensayo de mutación germinal de *BRCA1* y *BRCA2* en el ADN aislado de linfocitos sanguíneos periféricos comprende un cribado de mutación mediante un Ensayo de Truncamiento de Proteínas (Protein Truncation Test, PTT) del exón 11 de *BRCA1* y exon 10 y 11 de *BRCA2*, delección PCR de *BRCA1*, delección genómica de los exones 13 y 22, así como Electroforesis de Gel por Gradiente de Desnaturalización (Denaturing Gradient Gel Electrophoresis, DGGE) de los exones restantes. Todas las bandas aberrantes se confirmaron mediante secuenciación genómica analizada en un secuenciador automático ABI3700 y se confirmaron en una muestra independiente. De todo ello, el material tumoral material se sometió a congelación instantánea en nitrógeno líquido en la hora siguiente a la cirugía. Del material tumoral congelado se preparó una sección coloreada con H&E (hematoxilina-eosina) antes y después de cortar rodajas para el aislamiento del ARN. Se calcularon estas secciones de H&E congeladas para hallar el porcentaje de células tumorales; sólo se seleccionaron las muestras con >50% de células tumorales para su posterior estudio.

**[0137]** Para todos los tumores especímenes quirúrgicos fijados en formaldehído y recibidos con parafina se evaluaron con procedimientos histopatológicos estándar. Las secciones de parafina coloreadas con H&E se examinaron para calcular el tipo de tumor (por ej., ductal o lobular, según la clasificación WHO); para calcular el grado histológico según el método descrito por Elston y Ellis (grado 1-3); y para calcular la presencia de crecimiento linfango-invasivo y la presencia de una infiltración linfocítica extensiva. Todos los factores



histológicos fueron calculados independientemente por dos histólogos (MV y JL); se alcanzó un consenso sobre las diferencias examinando conjuntamente los portaobjetos. Se usó un portaobjeto representativo de cada tumor para coloreado inmunohistoquímico con anticuerpos dirigidos contra el receptor de estrógenos y progesterona con procedimientos estándar. El resultado del coloreado se registró como porcentaje de núcleos de coloración positiva (0%, 10%, 20%, etc., y así hasta 100%).

### 3. Amplificación, etiquetado e hibridación

**[0138]** El perfil de la producción de ácidos nucleicos derivados de marcadores y de la hibridación de los ácidos nucleicos con un biochip se muestra en la FIG. 2. Se usó un total de 30 secciones congeladas de 30 µM de grosor para el aislamiento total del ARN de cada espécimen tumoral sometido a congelación instantánea. El ARN total se aisló con RNAzol™ B (Campco Scientific, Veenendaal, Países Bajos) según el protocolo del fabricante, que incluye la homogeneización del tejido mediante un Polytron PT-MR2100 (Merck, Amsterdam, Países Bajos) y por último se disolvió en H<sub>2</sub>O sin RNasa. Se calculó la calidad del ARN total mediante el cociente A260/A280 que tenía que estar entre 1,7 y 2,1, así como inspección visual del ARN en gel de agarosa, que debería indicar una banda de ARN de ribosomal 28S más fuerte en comparación con la banda de ARN de ribosomal 18S, posteriormente se trataron 25 µg de ARN total con DNasa mediante el kit Qiagen de DNasa sin RNasa y columnas de centrifugación (Qiagen Inc, GmbH, Germany) según el protocolo del fabricante. El ARN total tratado con DNasa se disolvió en H<sub>2</sub>O sin RNasa hasta llegar a una concentración final de 0,2 µg/µl.

**[0139]** Se usaron 5 µg de ARN total como entrada para síntesis de ARNc. Se usó un cebador oligo-dT que contenía una secuencia promotora de polimerasa de ARN T7 para cebar una síntesis de ADN de primera cadena y se usaron cebadores aleatorios (pdN6) para cebar una síntesis de ADNc mediante transcriptasa inversa de MMLV. Esta reacción produjo un ADN de doble cadena que contenía el promotor de polimerasa de ARN T7 (T7RNAP). A continuación el ADNc de doble cadena se transcribió a ARNc mediante el T7RNAP.

**[0140]** Se etiquetó ARNc con colorantes Cy3 o Cy5 mediante un proceso de dos fases. primero, se incorporaron enzimáticamente nucleótidos derivativos de alilamina a productos de ARNc. Para el etiquetado del ARNc, se sustituyó el UTP por una mezcla de 3:1 de 5-(3-aminoali)uridina 5'- trifosfato (Sigma) y UTP en la reacción de la transcripción in vitro (IVT). A continuación se hizo reaccionar los productos de ARNc derivados de alilamina con ésteres de N-hidroxí succinimida de Cy3 o Cy5 (CyDye, Amersham Pharmacia Biotech). Se mezcló 5mg de ARNc etiquetado con Cy5 de una paciente con cáncer de mama con la misma cantidad de producto etiquetado con Cy3 de una reserva de igual cantidad de ARNc de cada paciente esporádico individual.

**[0141]** Se hicieron por duplicado hibridaciones de biochip con inversiones de flúor. Antes de la hibridación, se fragmentaron los ARNc a un tamaño medio de -50-100nt calentándolos a 60 °C en presencia de 10 mM de ZnCl<sub>2</sub>. Se añadieron los ARNc fragmentados a un tampón de hibridación que contenía 1 M de NaCl, 0.5% de sarcosian de sodio y 50 mM de MES, pH 6.5, cuya dureza se reguló con la adición de formamida hasta una concentración final de 30%. Se llevaron a cabo hibridaciones en un volumen final de mls a 40 °C en la plataforma giratoria de un horno de hibridación (Robbins Scientific) durante 48 h. Tras la hibridación, se lavaron los portaobjetos y se escanearon mediante un barrido confocal de láser (Agilent Technologies). Las intensidades de fluorescencia de las imágenes escaneadas se cuantificaron, normalizaron y corrigieron.

### 4. Combinación de las muestras

**[0142]** Se formó una reserva de referencia de ARNc combinando igual cantidad de ARNc de cada paciente esporádica individual.

### 5.25K Biochip humano

**[0143]** Se sintetizaron oligonucleótidos unidos a la superficie siguiendo esencialmente lo propuesto por Blanchard et al., Biosens. Bioelectron. 6(7):687-690 (1996); véase también Hughes et al., Nature Biotech. 19(4):342-347 (2000). Como sustratos para síntesis de nucleótidos se usaron superficies hidrófobas de cristal (3 pulgadas x 3 pulgadas) que contenían grupos de hidroxilo expuestos. En las superficies de cristal se introdujeron monómeros de fosforamida en posiciones definidas por ordenador mediante impresión por chorro de tinta. A continuación se eliminaron por lavado los monómeros no reaccionados y se les quitó la protección a los extremos de los oligonucleótidos extendidos. Para cada síntesis de nucleótidos que se deseaba se repitió este ciclo de acoplamiento, lavado y desprotección de los monómeros. Las secuencias de oligonucleótidos que debían imprimirse se especificaron mediante archivos informáticos.

**[0144]** Para este estudio se usaron biochips que contenían aproximadamente 25.000 secuencias de genes humanos (biochips Hu25K). Se seleccionaron secuencias para biochips de RefSeq (una colección de secuencias de ARNm no redundantes situadas en el sitio de Internet [nlm.nih.gov/LocusLink/refseq.html](http://nlm.nih.gov/LocusLink/refseq.html)) y Phil Green EST contigs, que es una colección de secuencias contiguas de EST reunida por el Dr. Phil Green et al en la Universidad de Washington (Ewing and Green, Nat. Genet. 25(2):232-4 (2000)), disponible en el sitio de Internet [phrap.org/est\\_assembly/index.html](http://phrap.org/est_assembly/index.html). Cada secuencia contigua de ARN o EST se representó en biochip Hu25K

mediante un solo oligonucleótido 60-mero siguiendo esencialmente las explicaciones de Hughes et al., Nature Biotech. 19(4):342-347 y de la Publicación Internacional WO01/06013, publicada el 25 de enero de 2001, y en la Publicación Internacional WO01/05935, publicada el 25 de enero de 2001, con la diferencia de que se modificaron las reglas de oligocribado para eliminar los oligonucleótidos con más del 30% de C o con 6 o más residuos contiguos de C.

Ejemplo 1: Conjuntos de genes regulados diferenciadamente y patrones generales de expresión de tumores de cáncer de mama

**[0145]** De las aproximadamente 25.000 secuencias representadas en el biochip se seleccionó un grupo de aproximadamente 5.000 genes que estaban significativamente regulados por todo el grupo de muestras. Se determinó que un gen estaba regulado de manera significativamente diferenciada con el cáncer de mama si mostraba más del doble de cambios de transcripción en comparación con una reserva de tumores esporádicos y si el valor-p para la regulación (Hughes et al., Cell 102:109-126 (2000)) era inferior a 0,01, tanto hacia arriba como hacia abajo, en al menos cinco sobre 98 muestras tumorales.

**[0146]** Un logaritmo de agrupamiento no supervisado nos permitió agrupar pacientes sobre la base de sus semejanzas medidas en este conjunto de 5.000 genes significativos. La medida de semejanza entre dos pacientes  $x$  e  $y$  se define como

$$S = 1 - \left[ \frac{\sum_{i=1}^{N_p} \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sigma_{x_i} \sigma_{y_i}}}{\sqrt{\sum_{i=1}^{N_p} \left( \frac{x_i - \bar{x}}{\sigma_{x_i}} \right)^2 \sum_{i=1}^{N_p} \left( \frac{y_i - \bar{y}}{\sigma_{y_i}} \right)^2}} \right]$$

**Ecuación (5)**

En la Ecuación (5),  $x$  e  $y$  son dos pacientes con componentes de cociente logarítmico  $x_i, y_i, i = 1, \dots, N = 5,100$ . A cada valor  $x_i$  va asociado el error  $\sigma_{x_i}$ . Cuanto menor sea el valor  $\sigma_{x_i}$  más fiable será la medición.

$$\bar{x} = \frac{\sum_{i=1}^{N_p} \frac{x_i}{\sigma_{x_i}^2}}{\sum_{i=1}^{N_p} \frac{1}{\sigma_{x_i}^2}}$$

es la media aritmética ponderada de error. El uso de la correlación como métrica de semejanza pone de relieve la importancia de la coregulación en el agrupamiento, más que la amplitud de las regulaciones.

**[0147]** El conjunto de aproximadamente 5.000 genes puede agruparse según sus semejanzas medidas en la totalidad del grupo de 98 muestras tumorales. La medida de semejanza entre dos genes se definió de la misma forma que en la Ecuación (1), con la diferencia de que ahora por cada gen hay 98 componentes de mediciones de cociente logarítmico.

**[0148]** El resultado de dicho agrupamiento bidimensional aparece representado en la FIG.3. Del agrupamiento emergen dos patrones característicos. El primer patrón consiste en un grupo de pacientes en la parte inferior del diagrama cuyas regulaciones son muy diferentes de la reserva esporádica. El otro patrón lo constituye un grupo de pacientes en la parte superior del diagrama cuyas expresiones están solo moderadamente reguladas en comparación con la reserva esporádica. Estos patrones dominantes sugieren que los tumores pueden dividirse de modo inequívoco en dos grupos tipos diferenciados basándose en dicho conjunto de 5.000 genes significativos.

**[0149]** Para facilitar la identificación de estos patrones, se asociaron los mismos a receptor de estrógeno (ER), receptor de proestrógeno (PR), grado de tumor, presencia de infiltrado linfocítico y angioinvasión (FIG. 3). El grupo inferior de la FIG. 3, que constituye el patrón dominante, consta de 36 pacientes. De las 30 pacientes con ER negativo, 34 pacientes se concentran en este grupo. A partir de la FIG. 4 se observó que la expresión del gen alfa receptor de estrógeno *ESR1* y un grupo numeroso de genes coregulados se ajustan a este patrón de expresión.

**[0150]** De la FIG. 3 y la FIG. 4 se sacó en conclusión que se pueden usar patrones de expresión para clasificar muestras tumorales en subgrupos de interés diagnóstico. Así pues, genes coregulados en un total de 98 muestras tumorales contienen información sobre la base molecular de cánceres de mama. La combinación de datos clínicos y abundancia de genes, medida en biochip, de *ESR1* demuestra que los distintos tipos están relacionados con, o al menos indicados por, la categoría de ER.

Ejemplo 2: Identificación de marcadores genéticos que distinguen entre pacientes receptoras de estrógeno (+) y pacientes receptoras de estrógeno (-)

**[0151]** Los resultados descritos en este Ejemplo permiten la identificación de genes marcadores de expresión con los que se diferencian dos grupos principales de células tumorales: grupo “ER-negativo” y grupo “ER-positivo”. La diferenciación de muestras por categoría de ER(+) se realizó en tres etapas: (1) identificación de un conjunto de genes marcadores candidatos que se hallan en correlación con el nivel de ER; (2) ordenación jerárquica de estos genes candidatos por fuerza de correlación; (3) optimización del número de genes marcadores y (4) clasificación de muestras basándose en dichos genes marcadores.

#### 1. Selección de genes discriminantes candidatos

**[0152]** En la primera etapa se identificó un conjunto de genes discriminantes candidatos basándose en los datos de expresión de genes de muestras para las prácticas. En concreto, se calcularon los coeficientes de correlación  $\rho$  entre los números de categoría o relación de nivel de ER y expresión logarítmica  $\bar{r}$  en la totalidad de las muestras para cada gen por separado.

$$\rho = (\bar{c} \bullet \bar{r}) / (\|\bar{c}\| \cdot \|\bar{r}\|)$$

**Ecuación (2)**

El histograma de los coeficientes de correlación resultantes se representa en la FIG. 5A con línea gris. Aunque para la mayoría de los genes la amplitud de correlación o anticorrelación es pequeña, para algunos llega a 0,5. Los genes cuyos cocientes de expresión tengan buena correlación o buena anticorrelación con la categoría de diagnóstico de interés se usan como genes indicadores para la categoría.

**[0153]** Los genes con un coeficiente de correlación superior a 0,3 (“genes correlacionados”) o inferior a -0,3 (“genes anti-correlacionados”) se seleccionaron como genes indicadores. Se seleccionó el umbral de 0,3 basándose en la distribución de correlación para los casos en los que no se da una correlación real (se pueden usar las permutaciones para determinar esta distribución). Estadísticamente, esta anchura de distribución depende del número de muestras usadas en el cálculo de correlación. La anchura de distribución para casos de

control (sin correlación real) es aproximadamente  $1/\sqrt{n-3}$ , donde  $n$  = el número de muestras. En nuestro caso,  $n = 98$ . Por lo tanto, un umbral de 0,3 corresponde más o menos a  $3 - \sigma$  en la distribución  $(3 \times 1/\sqrt{n-3})$ .

**[0154]** Se descubrió que 2.460 de dichos genes reunían dichas condiciones. A fin de evaluar la relevancia del coeficiente de correlación de cada gen con el nivel de ER, se usó una técnica de remuestreo para generar datos Monte-Carlo que aleatorizaran la asociación entre los datos de expresión de genes de las muestras y sus categorías. La distribución de los coeficientes de correlación obtenidos de un ensayo Monte-Carlo aparece con línea discontinua en la FIG 5A. Para estimar la relevancia de los 2.460 genes marcadores como grupo, se generaron 10.000 ensayos Monte-Carlo. La colección de estos 10.000 ensayos Monte-Carlo forma la hipótesis nula. El número de genes que reúnen las mismas condiciones para los datos Monte-Carlo varía de un ensayo a otro. La distribución de frecuencia a partir de los 10.000 ensayos Monte-Carlo del número de genes con coeficientes de  $>0,3$  o  $<-0,3$  se muestra en la FIG. 5B. Tanto el valor máximo como el mínimo son mucho menores de 2.460. Por lo tanto, se estima que la relevancia de este grupo de genes como conjunto de genes discriminantes entre muestras de ER(+) y ER(-) es superior al 99,99%.

#### 2. Ordenación jerárquica de los genes discriminantes candidatos

**[0155]** En la segunda etapa, se ordenaron jerárquicamente los genes de la lista de candidatos basándose en la relevancia de cada gen como gen discriminante. Los marcadores se ordenaron jerárquicamente bien por amplitud de correlación o bien usando una métrica similar a una estadística Fisher:

$$t = \frac{(\langle x_1 \rangle - \langle x_2 \rangle)}{\sqrt{[\sigma_1^2(n_1 - 1) + \sigma_2^2(n_2 - 1)] / (n_1 + n_2 - 1) / (1/n_1 + 1/n_2)}}$$

**Ecuación (3)**

En la Ecuación (3),  $\langle x_1 \rangle$  es la media ponderada de error del cociente logarítmico dentro del grupo de ER (-) y  $\langle x_2 \rangle$  es la media ponderada de error del cociente logarítmico dentro del grupo de ER (+).  $\sigma_1$  es la varianza de cociente logarítmico dentro del grupo ER(-) y  $n_1$  es el número de muestras que tuvieron mediciones válidas de cocientes logarítmicos.  $\sigma_2$  es la varianza de cociente logarítmico dentro del grupo ER(+) y  $n_2$  es el número de muestras que tuvieron mediciones válidas de cocientes logarítmicos. El valor- $t$  en la Ecuación (3) representa la diferencia de varianza compensada entre dos medios. El nivel de fiabilidad de cada gen en la lista de candidatos se estipuló con respecto a una hipótesis nula derivada del conjunto de datos reales mediante una técnica de remuestreo; esto es, se generaron muchos conjuntos de datos artificiales aleatorizando la asociación entre los datos clínicos y los datos de expresión de genes.

### 3. Optimización del número de genes marcadores

**[0156]** Para la validación cruzada se usó el método de “dejar uno fuera” a fin de optimizar los genes discriminantes. Un clasificador practicó con 97 ejemplos de un conjunto de genes marcadores extraído de la lista de candidatos por orden jerárquico y luego predijo la categoría de la muestra restante. Ese mismo procedimiento se repitió para cada uno de los ejemplos de la reserva y se contó el número de casos en que la predicción del que se había dejado fuera resultó errónea.

**[0157]** La evaluación del rendimiento de la validación cruzada de dejar uno fuera que se acaba de mencionar se repitió añadiendo sucesivamente más genes marcadores de la lista de candidatos. El rendimiento como función del número de genes marcadores se muestra en la FIG. 6. Las tasas de error para el tipo 1 y el tipo 2 variaron con el número de genes marcadores usados, pero las dos fueron mínimas mientras el número de genes marcadores se mantuvo en el orden de 550. Por lo tanto, estimamos que este conjunto de 550 genes debe considerarse el conjunto óptimo de genes marcadores que pueden usarse para clasificar los tumores de cáncer de mama en grupo “ER-negativo” y grupo “ER-positivo”. La FIG. 7 muestra la clasificación de pacientes en ER(+) o ER(-) basándose en este conjunto de 550 marcadores. La FIG. 8 muestra la correlación de cada tumor con cada plantilla ER-negativa frente a la correlación de cada tumor con la plantilla ER-positiva.

### 4. Clasificación basada en genes marcadores

**[0158]** En la tercera etapa se calculó un conjunto de parámetros clasificadores para cada tipo de conjunto de datos de prácticas basándose en cualquiera de los métodos de ordenación de más arriba. Se generó una plantilla para el grupo ER(-) ( $\bar{z}_1$ ) mediante la media ponderada de error del cociente logarítmico del grupo de genes seleccionado. Igualmente, se generó una plantilla para el grupo ER(+) (llamada  $\bar{z}_2$ ) mediante la media ponderada de error del cociente logarítmico del grupo de genes seleccionado. Se definieron dos parámetros clasificadores ( $P_1$  y  $P_{12}$ ) basándose en la correlación o en la distancia.  $P_1$  mide la semejanza entre una muestra y la plantilla ER(-)  $\bar{z}_1$  en ese grupo de genes seleccionado.  $P_2$  mide la semejanza entre una muestra y la plantilla ER(+)  $\bar{z}_2$  en ese grupo de genes seleccionado. La correlación  $P_1$  se define como:

$$P_i = (\bar{z}_i \cdot \bar{y}) / (\|\bar{z}_i\| \cdot \|\bar{y}\|)$$

Ecuación (1)

**[0159]** Se usó un método de “dejar uno fuera” para la validación cruzada basándose en los genes marcadores. En este método se reservó una muestra para la validación cruzada cada vez que actuaba un clasificador de prácticas. De un conjunto de 550 genes marcadores óptimos, el clasificador practicó con 97 de las 98 muestras y luego predijo la categoría de la muestra restante. Este procedimiento se repitió con cada una de las 98 pacientes. Se contó el número de casos en los que la predicción resultó errónea o acertada. Se determinó además que con subconjuntos de sólo -50 del total de 2.460 genes se pueden clasificar tumores en ER(+) o ER(-) casi igual de bien que usando el conjunto total.

**[0160]** En un pequeño número de casos hubo discordancia entre la clasificación mediante el conjunto de los 550 marcadores y la clasificación clínica. Al comparar el cociente logarítmico de expresión de ESR1 medida por biochip con la decisión binaria clínica (negativa o positiva) de la categoría de ER para cada paciente, se vio que la expresión medida concordaba con la categoría cualitativa de mediciones clínicas (mezcla de dos métodos) para la mayoría de tumores. Por ejemplo, dos pacientes a las que se diagnosticó clínicamente como ER(+) mostraron en realidad una baja expresión de ESR1 a partir de las mediciones con biochip y se las clasificó como ER negativas según los 550 genes marcadores. Asimismo, 3 pacientes a las que se diagnosticó clínicamente como ER(-) mostraron una alta expresión de ESR1 a partir de las mediciones con biochip y se las clasificó como ER(+) según los mismos 550 genes marcadores. Estadísticamente, sin embargo, la expresión de genes de ESR1 medida con biochip se ajusta más a los patrones dominantes que a la categoría de ER determinada clínicamente.

Ejemplo 3: Identificación de marcadores genéticos que distinguen entre tumores de *BRCA1* y tumores esporádicos en pacientes de receptor de estrógeno (-)

**[0161]** La mutación del *BRCA1* es una de las principales categorías clínicas en los tumores del cáncer de mama. Se determinó que, de los tumores de 38 pacientes del grupo ER(-), 17 mostraban la mutación del *BRCA1*, mientras que 21 eran tumores esporádicos. Por consiguiente, se desarrolló un método que hizo posible la diferenciación de los tumores de la mutación del *BRCA1* a partir de los 21 tumores esporádicos del grupo ER(-).

### 1. Selección de genes discriminantes candidatos

**[0162]** En el primer paso, se identificó una serie de genes candidatos basándose en los patrones de expresión de genes de los dichos 38 ejemplos. En primer lugar calculamos la correlación entre el número de categoría de mutación de *BRCA1* y el cociente de expresión en la totalidad de los 38 ejemplos para cada gen por separado

mediante la Ecuación (2). La distribución de los coeficientes de correlación se representa como un histograma definido por línea continua en la FIG. 9A. Se observó que, aunque la mayoría de los genes no se hallaban en correlación con la categoría de la mutación de *BRCA1*, un pequeño grupo de los mismos sí se correspondían a niveles significativos. Es probable que los genes con mayores coeficientes de correlación sirvan como indicadores para distinguir a las portadoras de tumores de la mutación de *BRCA1* de las de tumores esporádicos dentro del grupo ER(-).

**[0163]** A fin de evaluar la relevancia de cada coeficiente de correlación frente a una hipótesis nula de que dicho coeficiente de correlación pudiera hallarse de manera aleatoria, se usó una técnica de remuestreo para generar datos Monte-Carlo que aleatorizaran la asociación entre los datos de expresión de genes de las muestras y sus categorías. Se generaron 10.000 de dichos ensayos Monte-Carlo como control con el fin de estimar la relevancia de los genes marcadores como grupo. Se aplicó un umbral de 0,35 en la amplitud absoluta de coeficientes de correlación (o correlación o anticorrelación) tanto a los datos reales como a los datos Monte-Carlo. Siguiendo este método, se hallaron 430 genes que reunían estas condiciones para los datos experimentales. El valor-p de la relevancia, medido contra los 10.000 ensayos Monte-Carlo, es aproximadamente 0,0048 (FIG. 9B). Esto es, la probabilidad de que este conjunto de 430 genes contenga información útil sobre tumores de tipo *BRCA1* frente a tumores esporádicos es de más del 99%.

## 2. Ordenación jerárquica de genes discriminantes candidatos

**[0164]** En el segundo paso, se ordenaron jerárquicamente los genes de la lista de candidatos basándose en la relevancia de cada gen como gen discriminador. Aquí se usó la amplitud absoluta de coeficientes de correlación para ordenar jerárquicamente los genes marcadores.

**[0165]** En el tercer paso, se usó para la clasificación un subconjunto de genes de los puestos más altos de la lista jerarquizada. Se definió una plantilla del grupo *BRCA1* (llamada  $\bar{z}_1$ ) mediante la media ponderada de error del cociente logarítmico del grupo de genes seleccionado. Igualmente, se definió una plantilla del grupo *no-BRCA1* (llamada  $\bar{z}_2$ ) mediante la media ponderada de error del cociente logarítmico del grupo de genes seleccionado. Se definieron dos parámetros clasificadores (P1 y P2) basándose o en la correlación o en la distancia. P1 mide la semejanza entre una muestra y y la plantilla de *BRCA1*  $\bar{z}_1$  en la totalidad del grupo de genes seleccionado. P2 mide la semejanza entre una muestra y y la plantilla de *no-BRCA1*  $\bar{z}_2$  en la totalidad del grupo de genes seleccionado. Para la correlación, P1 y P2 se definieron de la misma forma que en la Ecuación (4).

**[0166]** Se usó el método de dejar uno fuera para validación cruzada a fin de optimizar los genes discriminantes como se explica en el Ejemplo 2. Para un conjunto de genes marcadores de la lista jerárquica de candidatos, el clasificador practicó con 37 ejemplos y predijo el restante. El procedimiento se repitió para todas las muestras de la reserva y se contó el número de casos en los que la predicción resultó errónea o acertada.

**[0167]** Para determinar el número de marcadores que constituyen un subconjunto viable, se repitió la ejecución de la anterior validación cruzada añadiendo acumulativamente más genes marcadores de la lista de candidatos. La ejecución como función del número de genes marcadores se muestra en la FIG. 10. Las tasas de error para el tipo 1 (falso negativo) y el tipo 2 (falso positivo) (Bendat & Piersol, RANDOM DATA ANALYSIS AND MEASUREMENT PROCEDURES, 2D ED., Wiley Interscience, p. 89) alcanzaron niveles óptimos cuando el número de los genes marcadores fue de aproximadamente 100. Por consiguiente, se considera que un conjunto de unos 100 genes es el conjunto óptimo de genes marcadores que pueden usarse para clasificar tumores en el grupo ER(-) tanto en tumores relacionados con *BRCA1* como en tumores esporádicos.

**[0168]** Los resultados de clasificación usando los 100 genes óptimos se muestran en las FIGS. 11A y 11B. Como se ve en la FIG 11A, los patrones de coregulación de las pacientes esporádicas difieren de los de las pacientes *BRCA1* sobre todo en la amplitud de regulación. Sólo se clasificó un tumor esporádico en el grupo de *BRCA1*. Las pacientes del grupo esporádico no son necesariamente negativas a la mutación de *BRCA1*; sin embargo, se estima que sólo un 5% aproximadamente de tumores esporádicos son portadores reales de mutación de *BRCA1*.

Ejemplo 4: Identificación de marcadores genéticos que distinguen las pacientes con tumor esporádico con >5 años de supervivencia frente a las de <5 años de supervivencia

**[0169]** Se usaron 78 tumores de pacientes de cáncer de mama para explorar predictores de pronóstico de datos de expresión de genes. De los 78 ejemplos de este grupo de cáncer de mama esporádico, se tenía constancia clínica de que 44 muestras no habían tenido metástasis distante dentro de los 5 años siguientes al diagnóstico inicial ("grupo sin metástasis distante") y de que 34 muestras habían tenido metástasis distante dentro de los 5 años siguientes al diagnóstico inicial ("grupo con metástasis distante"). Se identificó un grupo de 231 marcadores, y dentro de él un grupo óptimo de 70 marcadores, que permitía la diferenciación entre los dos grupos de pacientes mencionados.

#### 1. Selección de genes discriminantes candidatos

**[0170]** En el primer paso, se identificó un grupo de genes discriminantes candidatos basándose en los datos de expresión de genes de las dichas 78 muestras. Se calculó la correlación entre el número de categoría de pronóstico (metástasis distante frente a metástasis no distante) y el cociente de expresión logarítmica en todas las muestras para cada gen por separado mediante la Ecuación (2). La distribución de los coeficientes de correlación se representa con línea continua en la FIG. 12A. La FIG. 12A muestra también el resultado de un ensayo Monte-Carlo en línea discontinua. Se observó que aunque la mayoría de genes no están en correlación con las categorías de pronóstico, un pequeño grupo de genes sí lo está. Es probable que genes con mayores coeficientes de correlación sean más útiles como indicadores del pronóstico de interés – grupo de metástasis distante y grupo de metástasis no distante.

**[0171]** A fin de evaluar la relevancia de cada coeficiente de correlación frente a una hipótesis nula de que dicha correlación pueda hallarse de manera aleatoria, se usó una técnica de remuestreo para generar datos a partir de 10.000 ensayos Monte-Carlo como control (FIG. 12B). Luego se seleccionaron genes que o tenían un coeficiente de correlación mayor de 0,3 ("genes correlacionados") o menor de -0.3 ("genes anti-correlacionados"). Ese mismo criterio de selección se aplicó tanto a los datos reales como a los datos Monte-Carlo. Por medio de esta comparación se identificaron 231 marcadores que cumplieran con ese requisito. La probabilidad de que dicho conjunto de genes para diferenciar pacientes del grupo de metástasis distante del grupo de metástasis no distante sea elegido por fluctuación aleatoria es aproximadamente de 0,003.

#### 2. Ordenación jerárquica de genes discriminantes candidatos

**[0172]** En el segundo paso, se ordenaron los genes de la lista de candidatos basándose en la relevancia de cada gen como gen discriminante. Concretamente, se usó una métrica similar a una estadística "Fisher", definida en la Ecuación (3), para la ordenación jerárquica. El nivel de fiabilidad de cada gen en la lista de candidatos se estimó frente a una hipótesis nula derivada del conjunto de datos reales mediante la técnica de remuestreo. Los genes de la lista de candidatos también pueden ordenarse por la amplitud de coeficientes de correlación.

#### 3. Optimización de genes discriminantes

**[0173]** En el tercer paso, se seleccionó un subconjunto de 5 genes de los puestos más altos de esta lista jerarquizada para usarlos como genes discriminantes en la clasificación de 78 tumores como "grupo de metástasis distante" o "grupo de metástasis no distante". Para la validación cruzada se usó el método de dejar uno fuera. Concretamente, 77 muestras definieron un clasificador basado en el conjunto de genes discriminantes seleccionados, los cuales se usaron para predecir la muestra restante. Se repitió este procedimiento hasta que se predijo cada una de las 78 muestras. Se contó el número de casos en los que las predicciones resultaron correctas o incorrectas. El rendimiento del clasificador se midió por las tasas de error del tipo 1 y del tipo 2 para este conjunto de genes seleccionado.

**[0174]** Se repitió el anterior procedimiento de evaluación del rendimiento añadiendo cada vez 5 genes marcadores más de los puestos más altos de la lista jerarquizada, hasta que se hubo usado el total de los 231 genes. Como se ve en la FIG. 13, el número de predicciones erróneas de errores de tipo 1 y tipo 2 cambia radicalmente según el número de genes marcadores empleados. La tasa combinada de error bajó al mínimo cuando no se usaron nunca los 70 genes de los puestos más altos de nuestra lista de candidatos. Por lo tanto, este conjunto de 70 genes es el conjunto de genes marcadores óptimo y preferido, que sirve para la clasificación de pacientes con tumores esporádicos en grupo de metástasis distante y grupo de metástasis no distante. Un número mayor o menor de marcadores también actúa como predictor, pero es menos eficiente, bien por las tasas de error más altas o bien por la introducción de ruido estadístico.

#### 4. Curvas de probabilidad de recaída

**[0175]** Se predijo la clasificación de pronóstico de 78 pacientes con tumores esporádicos de cáncer de mama en dos subgrupos diferenciados basándose en su expresión de 70 genes marcadores óptimos (FIGS. 14 y 15).

**[0176]** Para evaluar la clasificación de pronóstico de pacientes esporádicas, se predijo el resultado de cada paciente por medio de un clasificador que practicó con los 77 pacientes restantes basándose en los 70 genes marcadores óptimos. En la FIG. 16 se muestra en diagrama la probabilidad de las metástasis distantes como función de tiempo desde el diagnóstico inicial para los dos grupos previstos. Hay una diferencia significativa entre estas dos curvas de recaída. Mediante la prueba  $\chi^2$  (S-PLUS 2000 Guide to Statistics, vol. 2, MathSoft, p. 44), se estimó el valor-p en  $\sim 10^{-9}$ . También se comparó la probabilidad de metástasis distante como función de tiempo desde el diagnóstico inicial entre individuos con ER(+) y ER(-) (FIG. 17), individuos con PR(+) y PR(-) (FIG. 18) y entre individuos con diferentes grados de tumor (FIGS. 19A, 19B). En comparación, los valores-p para las diferencias entre dos grupos de pronóstico basadas en datos clínicos son mucho menos significativas que las que se basan en datos de expresión de genes, que van de  $10^{-3}$  a 1.

**[0177]** Para fijar los parámetros de la probabilidad de recaída como función de tiempo desde el diagnóstico inicial, se ajustó la curva a un tipo de modelo de supervivencia “normal”:

$$P = \alpha \times \exp\left(-t^2/\tau^2\right) \quad (4)$$

Para un  $\alpha = 1$  fijado, hallamos que  $\tau = 125$  meses para las pacientes del grupo de metástasis no distante y  $\tau = 36$  meses para las pacientes del grupo de metástasis distante. Usando grados de tumor, hallamos que  $\tau = 100$  meses para las pacientes con grados de tumor 1 y 2 y  $\tau = 60$  para las pacientes con grado de tumor 3. En la práctica clínica se admite que los grados de tumor son el mejor predictor de pronóstico de que se dispone. No obstante, la diferencia entre los dos grupos de pronóstico clasificados sobre la base de los 70 genes marcadores es mucho más significativa que la de los grupos clasificados mediante la mejor información clínica disponible.

#### 10 5. Predicción de pronóstico para 19 tumores esporádicos independientes

**[0178]** Para confirmar el método de clasificación de pronóstico propuesto y para asegurar la reproducibilidad, robustez y potencia de predicción de los 70 genes marcadores de pronóstico, se aplicó el mismo clasificador a 19 muestras tumorales independientes tomadas de pacientes con cáncer de mama esporádico, preparadas por separado en el Instituto de Cáncer de los Países Bajos (NKI). Se usó la misma reserve de referencia.

15 **[0179]** Los resultados de la clasificación de 19 tumores esporádicos independientes se muestran en la Figura 20. La FIG. 20A muestra el cociente logarítmico de regulación de expresión de esos mismos 70 genes marcadores óptimos. Basándonos en nuestro modelo clasificador, esperábamos la clasificación errónea de  $19 \times (6+7)/78 = 3,2$  tumores. En consecuencia,  $(1+3) = 4$  de 19 tumores se clasificaron erróneamente.

#### 6. Parámetros clínicos como grupo frente a datos de biochip - Resultados de la regresión logística

20 **[0180]** En la sección anterior, la potencia de predicción de cada parámetro clínico por separado se comparó con la de los datos de expresión. Sin embargo, tiene más sentido combinar todos los parámetros clínicos como grupo y luego compararlos con los datos de expresión. Para ello se requiere una modelización multivariante; el método elegido fue la regresión logística. Este enfoque demuestra también la mejora que aporta el método de biochip a los resultados de los datos clínicos.

25 **[0181]** Los parámetros clínicos usados para la modelización multivariante fueron: (1) grado de tumor; (2) categoría de ER; (3) presencia o ausencia del receptor progestativo (PR); (4) tamaño del tumor; (5) edad de la paciente, y (6) presencia o ausencia de angiovasión. Para los datos de biochip se usaron dos coeficientes de correlación. Uno es la correlación con la media del grupo de buen pronóstico (C1) y el otro es la correlación con la media del grupo de pronóstico deficiente (C2). Cuando se calculan los coeficientes de correlación para una  
30 paciente dada, a dicha paciente se la excluye de cualquiera de las dos medias.

**[0182]** La regresión logística optimiza el coeficiente de cada parámetro de entrada para predecir óptimamente el resultado de cada paciente. Una manera de juzgar la potencia de predicción de cada parámetro de entrada es cuánta desviación (parecida a ji-cuadrado en la regresión lineal, véase, por ejemplo, Hasomer & Lemeshow, APPLIED LOGISTIC REGRESSION, John Wiley & Sons, (2000)) es causante el parámetro. El mejor predictor  
35 debería ser el causante de la mayoría de la desviación. Para calcular con precisión la potencia de predicción, se modeló cada parámetro por separado. Los parámetros del biochip explican la mayoría de la desviación y por consiguiente son predictores potentes.

**[0183]** A continuación los parámetros clínicos, y los dos parámetros del biochip, se monitorizaron como grupo. La desviación total explicada por los seis parámetros clínicos fue de

40 31,5 y la desviación total explicada por los parámetros del biochip fue de 39,4. No obstante, cuando se modelizaron primero los datos clínicos y luego se añadieron los dos parámetros del biochip, la desviación final fue causante del 57,0.

**[0184]** La regresión logística computa la probabilidad de que una paciente pertenezca al grupo de buen pronóstico o al del deficiente. Las FIGS. 21A y 21B muestran la sensibilidad frente a la (1-especificidad). Se generaron diagramas variando el umbral en la probabilidad prevista del modelo. La curva que atraviesa la esquina superior izquierda es la mejor sensibilidad (alta sensibilidad con alta especificidad). El biochip superó los  
45 datos clínicos por un amplio margen. Por ejemplo, a una sensibilidad establecida de alrededor del 80%, la especificidad fue de ~80% de los datos del biochip y de ~65% de los datos clínicos para el grupo de buen pronóstico. Para el grupo de pronóstico deficiente, las especificidades correspondientes fueron ~80% y ~70%,  
50 también a una sensibilidad establecida del 80%. La combinación de los datos del biochip con los datos clínicos mejoró los resultados. El resultado se puede mostrar también como la tasa total de error como función del umbral

en la FIG. 21C. En todos los umbrales posibles, la tasa de error del biochip fue siempre más pequeña que la de los datos clínicos. Al añadir los datos del biochip a los datos clínicos, la tasa de error se reduce más aún, como se puede ver en la Figura 21C.

- 5 **[0185]** Se pueden crear tablas de cocientes impares a partir de la predicción de la regresión logística. La probabilidad de que una paciente esté en el grupo de buen pronóstico se calcula mediante la regresión logística basándose en diferentes combinaciones de parámetros de entrada (clínicos y/o de biochip). Las pacientes se dividen en los cuatro grupos siguientes según la predicción el resultado real: (1) predicción buena y en realidad buena; (2) predicción buena pero en realidad deficiente; (3) predicción deficiente pero en realidad buena y (4) predicción deficiente y en realidad deficiente. Los grupos (1) & (4) representan predicciones correctas, mientras que los grupos (2) & (3) representan predicciones erróneas. La división para la predicción se establece en una probabilidad del 50%, aunque se pueden usar otros umbrales. Los resultados se enumeran en la Tabla 7. En la
- 10 tabla 7 se ve claramente que el perfil de biochip (Tabla 7.3 & 7.10) supera a cualquier dato clínico simple (Tabla 7.4-7.9) y la combinación de los datos clínicos (Tabla 7.2). La añadidura del perfil de biochip, junto con los datos clínicos, da los mejores resultados (Tabla 7.1).
- 15 **[0186]** Para el perfil de biochip se puede hacer también una tabla parecida (Tabla 7.11) sin usar la regresión logística. En este caso, la predicción se basó simplemente en C1-C2 (mayor que 0 significa buen pronóstico, menor que 0 significa deficiente

Tabla 7.1 Predicción por clínico+biochip		
	Predicción buena	Predicción deficiente
en realidad buena	39	5
en realidad deficiente	4	30
Tabla 7.2 Predicción solo por clínico		
	Predicción buena	Predicción deficiente
en realidad buena	34	10
en realidad deficiente	12	22
Tabla 7.3 Predicción por biochip		
	Predicción buena	Predicción deficiente
en realidad buena	39	5
en realidad deficiente	10	24
Tabla 7.4 Predicción por grado		
	Predicción buena	Predicción deficiente
en realidad buena	23	21
en realidad deficiente	5	29
Tabla 7.5 Predicción por ER		
	Predicción buena	Predicción deficiente
en realidad buena	35	9
en realidad deficiente	21	13
Tabla 7.6 Predicción por PR		
	Predicción buena	Predicción deficiente
en realidad buena	35	9



Tabla 7.1 Predicción por clínico+biochip		
	Predicción buena	Predicción buena
en realidad deficiente	18	16
Tabla 7.7 Predicción por tamaño		
	Predicción buena	Predicción deficiente
en realidad buena	35	9
en realidad deficiente	13	21
Tabla 7.8 Predicción por edad		
	Predicción buena	Predicción deficiente
en realidad buena	33	11
en realidad deficiente	15	19
Tabla 7.9 Predicción por angioinvasión		
	Predicción buena	Predicción deficiente
en realidad buena	37	7
en realidad deficiente	19	15
Tabla 7.10 Predicción por dC (C1-C2)		
	Predicción buena	Predicción deficiente
en realidad buena	36	8
en realidad deficiente	6	28
Tabla 7.11 Sin regresión logística, simplemente juzgada por dC (C1-C2)		
	Predicción buena	Predicción deficiente
en realidad buena	37	7
en realidad deficiente	6	28

Ejemplo 5. Concepto de biochip a efectos diagnósticos.

- 5 **[0187]** Todos los genes de la lista de genes marcadores a efectos de diagnóstico y pronóstico pueden sintetizarse en un biochip a pequeña escala mediante tecnología por chorro de tinta. Se puede hacer un biochip con genes para diagnóstico y pronóstico respectiva o colectivamente. Cada gen de la lista está representado por sondas oligonucleótidas sencillas o múltiples, dependiendo de la unicidad de su secuencia en todo el genoma. Este biochip diseñado a medida, en combinación con el protocolo de preparación de muestras, puede usarse en clínicas como kit de diagnóstico/pronóstico.

Ejemplo 6. Relevancia biológica de los genes marcadores de diagnóstico

- 10 **[0188]** Se registró el dominio en busca de anotaciones funcionales disponibles para los 430 genes marcadores para la diagnosis de BRCA1 de la Tabla 3. Los 430 genes de la Tabla 3 pueden dividirse en dos grupos: (1) 196 genes cuyas expresiones están claramente expresadas en el grupo de tipo BRCA1, y (2) 234 genes cuyas expresiones están altamente expresadas como grupo esporádico. De los 196 genes del grupo BRCA1, 94 están anotados. De los 234 genes del grupo esporádico, 100 están anotados. Los términos "célula-T", "célula-B" o "inmunoglobulina" están implicados en 13 de los 94 genes anotados y en 1 de los 100 genes anotados, respectivamente. De 24.479 genes representados en los biochips, hay hasta la fecha 7.586 genes con anotaciones. "Célula-T", "célula-B" e "inmunoglobulina" están presentes en 207 de estos 7.586 genes. En vista de ello, el valor-p de los 13 genes de "célula-T", "célula-B" o "inmunoglobulina" del grupo BRCA1 es muy
- 15

relevante (valor- $p = 1,1 \times 10^{-6}$ ). En comparación, la observación de 1 gen relacionado con "célula-T", "célula-B" o "inmunoglobulina" en el grupo esporádico no es relevante (valor- $p = 0,18$ ).

**[0189]** La observación de que las pacientes con *BRCA1* tienen genes linfocitos altamente expresados (célula-T y célula-B) concuerda con lo que se ha visto en la patología, es decir, que el tumor de mama de *BRCA1* está asociado a la alta infiltración linfocítica más a menudo que los casos esporádicos (Chappuis et al., 2000, Semin Surg Oncol 18:287-295).

Ejemplo 7. Relevancia biológica de los genes marcadores de pronóstico

**[0190]** Se realizó una búsqueda de anotaciones funcionales disponibles para los 231 genes marcadores de pronóstico (Tabla 5). Los marcadores se dividen en dos grupos: (1) 156 marcadores cuyas expresiones están altamente expresadas en el grupo de pronóstico deficiente, y (2) 75 genes cuyas expresiones están altamente expresadas en el grupo de buen pronóstico. De los 156 marcadores, 72 genes están anotados; de los 75 genes, 28 genes están anotados.

**[0191]** Doce de los 72 marcadores, pero ninguno de los 28 marcadores, son, o están asociados a, quinasas. Por contraste, de los 7.586 genes del biochip que hasta la fecha tienen anotaciones, sólo 471 tienen quinasas. Partiendo de esta base, es significativo el valor- $p$  de que doce marcadores relacionados con la quinasa se hallen en el grupo de pronóstico (valor- $p = 0,001$ ). Las quinasas son importantes reguladores de los procedimientos de transducción por señal intracelular que intervienen en la proliferación, diferenciación y apoptosis de células. Normalmente, su actividad se controla y vigila estrechamente. La sobreexpresión de ciertas quinasas es bien conocida y está implicada en la oncogénesis, como receptor1 del factor de crecimiento endotelial vascular (VEGFR1 o FLT1), una quinasa de tirosina del grupo de pronóstico deficiente, que juega un papel muy importante en la angiogénesis tumoral. Curiosamente, el factor de crecimiento endotelial vascular (VEGF), ligando de VEGFR, se halla también en el grupo de pronóstico, lo que significa que tanto el ligando como el receptor son sobreexpresados en individuos con pronóstico deficiente por un mecanismo desconocido.

**[0192]** Igualmente, 16 de los 72 marcadores, y solo 2 de los 28 marcadores, son, o están asociados a, proteínas de enlace de ATP o de enlace de GTP. Por contraste, de los 7.586 genes del biochip que hasta la fecha tienen anotaciones, solo 714 y 153 implican enlace de ATP y enlace de GTP, respectivamente. Partiendo de esta base, es significativo el valor- $p$  de que 16 marcadores relacionados con el enlace de GTP o de ATP se hallen en el grupo de pronóstico deficiente (valor- $p$  0,001 y 0,0038). Así pues, los marcadores relacionados con el enlace de GTP o de ATP dentro de los 72 marcadores pueden usarse como indicadores de pronóstico.

**[0193]** El cáncer se caracteriza por la proliferación desregulada de células. Al nivel más simple, esto requiere la división de la célula o mitosis. Buscando por palabra clave, hallamos "división de células" o "mitosis" incluida en las anotaciones de 7 genes respectivamente en los 72 marcadores anotados del total de los 156 marcadores de pronóstico deficiente, pero en ninguno de los 28 genes anotados del total de los 75 marcadores de buen pronóstico. De los 7.586 marcadores de biochip con anotaciones, hallamos "división de célula" en 62 anotaciones y "mitosis" en 37 anotaciones. Basándose en estos hallazgos, se considera altamente significativo el valor- $p$  de que siete marcadores relacionados con la división de célula o la mitosis se hallen en el grupo de buen pronóstico (valor- $p = 3,5 \times 10^{-5}$ ). En comparación, la ausencia de marcadores relacionados con la división de célula o la mitosis en el grupo de buen pronóstico no es significativa (valor- $p = 0,69$ ). Así pues, los siete marcadores relacionados con la división de célula o la mitosis pueden usarse como marcadores para el diagnóstico deficiente.

Ejemplo 8: Elaboración de una reserva de referencia artificial

**[0194]** La reserva de referencia para el perfil de expresión en los Ejemplos de más arriba se hizo usando igual cantidad de ARNs de cada paciente individual del grupo esporádico. Para tener una gran cantidad de reserva de referencia fiable y fácil de hacer, se puede elaborar una reserva de referencia para el diagnóstico y el pronóstico de cáncer de mama mediante ácido nucleico sintético que represente, o que se derive de, cada gen marcador. La expresión de genes marcadores para muestra de paciente individual se monitoriza solo contra la reserva de referencia, no contra una reserva derivada de otras pacientes.

**[0195]** Para hacer la reserva de referencia, se sintetizan oligonucleótidos 60-meros según la secuencia 60-mera de sonda de biochip por chorro de tinta para cada uno de los genes indicadores de diagnóstico/pronóstico, luego se bicatenan y clonan al vector pBluescript SK (Stratagene, La Jolla, CA), adyacente a la secuencia del promotor T7. Se aíslan los clones por separado y se verifican las secuencias de sus insertos mediante secuenciación de ADN. Para generar ARNs sintéticos, se linealizan con *EcoRI* y se lleva a cabo una reacción de transcripción *in vitro* (IVT) de T7 según el kit MegaScript kit (Ambion, Austin, TX). La IVT va seguida del tratamiento del producto con ADNasa. Se purifican ARNs sintéticos en columnas RNeasy (Qiagen, Valencia, CA). Dichos ARNs sintéticos se transcriben, amplifican, etiquetan y mezclan unos con otros para elaborar la reserva de referencia. La abundancia de esos ARNs sintéticos se ajusta a aproximadamente la abundancia de las correspondientes transcripciones derivadas de marcadores en la reserva de tumores reales.

Ejemplo 9: Uso de datos de un solo canal y una reserva de muestras representada por valores almacenados

#### 1. Creación de una reserva de referencia de valores almacenados ("reserva matemática de muestras")

**[0196]** El uso de datos basados en cocientes en los Ejemplos 1-7 de más arriba requiere una muestra física de referencia. En dichos Ejemplos se usó una muestra de tumores esporádicos como referencia. El uso de dicha referencia, aunque permite predicciones de pronóstico y diagnóstico sólidas, puede causar problemas, ya que la reserva es típicamente un recurso limitado. Por lo tanto, se desarrolló un método clasificador que no requiere una reserva física de muestras, haciendo así la aplicación de esta técnica predictiva y de diagnóstico mucho más sencilla para las aplicaciones clínicas.

**[0197]** Para comprobar si se podían usar los datos de un solo canal, se desarrolló el siguiente procedimiento. Primero, se seleccionaron los datos de intensidad del canal único para los 70 genes óptimos, descrito en el Ejemplo 4, de las 78 muestras esporádicas para prácticas, descritas en los Materiales y Métodos, de los datos de muestra esporádica frente a los datos de hibridación de reserva de tumores. Las 78 muestras consistían en 44 muestras de pacientes con buen pronóstico y 34 muestras de pacientes con pronóstico deficiente. A continuación se normalizaron las intensidades de hibridación para estas muestras dividiendo por la mediana de intensidad de todas las manchas biológicas del mismo biochip. Allí donde se usaron múltiples biochips por muestra, se halló la media en todos los biochips. Se realizó una transformación logarítmica en los datos de intensidad para cada uno de los 70 genes, o para la intensidad media para cada uno de los 70 genes donde se hibrida más de un biochip, y se calculó una intensidad logarítmica media para cada gen en todas las 78 muestras esporádicas. Para cada muestra, las intensidades logarítmicas medias calculadas de este modo se restaron de la intensidad logarítmica de la muestra individual. Esta cifra, la media de intensidad logarítmica sustraída, se trató a continuación como el cociente logarítmico de dos colores para el clasificador por sustitución por la Ecuación (5). Para nuevas muestras, la intensidad logarítmica media se resta de la misma forma que se indica más arriba y se calcula una intensidad logarítmica sustraída media.

**[0198]** La creación de un conjunto de intensidades logarítmicas medias para cada gen hibridado crea a su vez una "reserva matemática de muestras" que sustituye la "reserva material de muestras". Esta reserva matemática de muestras puede aplicarse entonces a cualquier muestra, incluidas las muestras actualmente disponibles y las que se obtengan en el futuro. Esta "reserva matemática de muestras" puede actualizarse a medida que se va disponiendo de más muestras.

#### 2. Resultados

**[0199]** Para demostrar que la reserva matemática de muestras desempeña una función equivalente a la reserva de muestras de referencia, se plotó la intensidad logarítmica sustraída media (datos de un solo canal relativos a la reserva matemática) frente al cociente logarítmico (hibridaciones relativas a la reserva de muestras) para los 70 genes marcadores óptimos en la totalidad de las 78 muestras esporádicas, como se ve en la FIG. 22. El cociente y las cantidades de un solo canal están en estrecha correlación, lo que indica que ambos tienen capacidad para informar de cambios relativos en la expresión de genes. A continuación se construyó un clasificador mediante la intensidad logarítmica sustraída media, siguiendo exactamente el mismo procedimiento que se usó para los datos de cociente, como en el Ejemplo 4.

**[0200]** Como se ve en las FIGS. 23A y 23B, con los datos de un solo canal se consiguió clasificar muestras basadas en patrones de expresión de genes. En la FIG. 23A se ven muestras agrupadas según el pronóstico mediante datos de hibridación de un solo canal. La línea blanca separa muestras de pacientes clasificados como de pronóstico deficiente (abajo) y de buen pronóstico (arriba). En la FIG. 23B se hace un diagrama de cada muestra según sus datos de expresión se hallen en correlación con el parámetro clasificador de pronóstico bueno (círculos abiertos) o deficiente (cuadrados rellenos). Mediante el método de validación cruzada de dejar uno fuera, el clasificador predijo 10 falsas positivas de un total de 44 muestras de pacientes con buen pronóstico y 6 falsas negativas de un total de 34 muestras de pacientes con pronóstico deficiente, donde un pronóstico deficiente se considera un "positivo". Este resultado es comparable al uso del clasificador basado en cociente, que predijo 7 de 44 y 6 de 34, respectivamente.

**[0201]** En aplicaciones clínicas, es muy preferible tener pocos falsos positivos, lo que resulta en menos pacientes con infratratamiento. Para adecuar los resultados a esta preferencia, se construyó un clasificador ordenando jerárquicamente la muestra del paciente según su coeficiente de correlación con la plantilla de "buen pronóstico" y se eligió un umbral para este coeficiente de correlación que permitiera aproximadamente un 10% de falsos negativos, esto es, la clasificación de una muestra de un paciente con pronóstico deficiente como la de un paciente con buen pronóstico. De las 34 muestras de pronóstico deficiente usadas aquí, ello representa una tolerancia de 3 sobre 34 pacientes con pronóstico deficiente clasificadas incorrectamente. Este límite de tolerancia corresponde a un umbral de coeficiente umbral de 0,2727 de correlación con la plantilla de "buen pronóstico". Los resultados que usan este umbral se muestran en las FIGS. 24A y 24B. La FIG. 24A muestra datos de hibridación de un solo canal para muestras ordenadas jerárquicamente según los coeficientes de correlación con el clasificador de buen pronóstico; las mezclas clasificadas como de "buen pronóstico" se hallan

por encima de la línea blanca, mientras que las clasificadas como de “mal pronóstico” se hallan por debajo de la misma. La FIG. 24B muestra un diagrama de dispersión de coeficientes de correlación, con tres muestras clasificadas incorrectamente situadas a la derecha del valor del coeficiente de correlación umbral. Usando este umbral, el clasificador tuvo una tasa de falso positivo de 15 sobre 44 muestras de buen pronóstico. Este resultado no es muy diferente si se compara con la tasa de error de 12 sobre 44 para el clasificador basado en el cociente.

**[0202]** En resumen, los 70 genes indicadores transmiten información consistente sobre el pronóstico; los datos de un solo canal pueden predecir el resultado del tumor casi tan bien como los datos basados en el cociente y a la vez son más convenientes en un entorno clínico.

## REIVINDICACIONES

1. Método para clasificar a un individuo aquejado de cáncer de mama como poseedor de buen pronóstico o pronóstico deficiente, en donde dicho individuo es un humano, en donde dicho pronóstico indica que se espera que dicho individuo no tenga metástasis distante dentro de los cinco años siguientes al diagnóstico inicial de cáncer de mama y en donde dicho pronóstico deficiente indica que se espera que dicho individuo tenga metástasis distante dentro de los cinco años siguientes al diagnóstico inicial de cáncer de mama, que comprende
  - (ia) calcular un primer parámetro clasificador entre un primer perfil de expresión y una plantilla de buen pronóstico, o
  - (ib) calcular un segundo parámetro clasificador entre dicho primer perfil de expresión y dicha plantilla de buen pronóstico y un tercer parámetro clasificador entre dicho primer perfil de expresión y una plantilla de pronóstico deficiente;
 comprendiendo dicho primer perfil de expresión los niveles de expresión de una primera pluralidad de genes en una muestra de células tomada del individuo,
  - comprendiendo dicha plantilla de buen pronóstico, para cada gen en dicha primera pluralidad de genes, el nivel medio de expresión de dicho gen en una pluralidad de pacientes que no tengan metástasis distante dentro de los cinco años siguientes al diagnóstico inicial de cáncer de mama;
  - y comprendiendo dicha plantilla de pronóstico deficiente, para cada gen en dicha primera pluralidad, el nivel medio de expresión de dicho gen en una pluralidad de pacientes que tengan metástasis distante dentro de los cinco años siguientes al diagnóstico inicial de cáncer de mama;
  - consistiendo dicha primera pluralidad de genes en al menos 5 de los genes cuyos marcadores se enumeran en la Tabla 5, y
  - (iia) clasificar a dicho individuo como poseedor de buen pronóstico si dicho primer parámetro clasificador está por encima de un umbral elegido o si dicho primer perfil de expresión es más parecido a dicha plantilla de buen pronóstico que a dicha plantilla de pronóstico deficiente, o
  - (iib) clasificar a dicho individuo como poseedor de dicho pronóstico deficiente si dicho primer parámetro clasificador está por debajo de dicho umbral elegido o si dicho primer perfil de expresión es más parecido a dicha plantilla de pronóstico deficiente que a dicha plantilla de buen pronóstico.
2. El método de la reivindicación 1, en donde dicha primera pluralidad consiste en al menos 20 de los genes cuyos marcadores se enumeran en la lista de la Tabla 5.
3. El método de la reivindicación 1, en donde dicha primera pluralidad consiste en al menos 100 de los genes cuyos marcadores se enumeran en la lista de la Tabla 5.
4. El método de la reivindicación 1, en donde dicha primera pluralidad consiste en al menos 150 de los genes cuyos marcadores se enumeran en la lista de la Tabla 5.
5. El método de la reivindicación 1, en donde dicha primera pluralidad consiste en cada uno de los genes cuyos marcadores se enumeran en la lista de la Tabla 5.
6. El método de la reivindicación 1, en donde dicha primera pluralidad consiste en al menos 70 de los genes cuyos marcadores se enumeran en la lista de la Tabla 6.
7. El método de la reivindicación 1, que además comprende los pasos de:
  - (a) generar dicha plantilla de buen pronóstico por hibridación de ácidos nucleicos derivados de dicha pluralidad de pacientes que no tengan metástasis distante dentro de los cinco años siguientes al diagnóstico inicial de cáncer de mama contra ácidos nucleicos derivados de una reserva de tumores de una pluralidad de pacientes que tengan cáncer de mama;
  - (b) generar dicha plantilla de buen pronóstico por hibridación de ácidos nucleicos derivados de dicha pluralidad de pacientes que tengan metástasis distante dentro de los cinco años siguientes al diagnóstico inicial de cáncer de mama contra ácidos nucleicos derivados de una reserva de tumores de dicha pluralidad de pacientes;

- (c) generar dicho primer perfil de expresión hibridando ácidos nucleicos derivados de dicha muestra de células tomada dicho individuo contra dicha reserva, y
- (d) calcular (d1) dicho segundo parámetro clasificador entre dicho primer perfil de expresión y la plantilla de buen pronóstico y (d2) dicho tercer parámetro clasificador entre dicho primer perfil de expresión y la plantilla de pronóstico deficiente, en donde si dicho primer perfil de expresión es más parecido a la plantilla de buen pronóstico que a la plantilla de pronóstico deficiente, se clasifica al individuo como poseedor de buen pronóstico, y si dicho perfil de expresión es más parecido a la plantilla de pronóstico deficiente que a la plantilla de buen pronóstico, se clasifica al individuo como poseedor de pronóstico deficiente.
- 5
8. El método de la reivindicación 1, que comprende además:
- 10
- iv) clasificar a dicho individuo como ER(+) (receptor positivo de estrógeno) o ER(-) (receptor negativo de estrógeno) basándose en un segundo perfil de expresión que comprende los niveles de expresión de una segunda pluralidad de genes en una muestra de células tomada del individuo, consistiendo dicha segunda pluralidad de genes en al menos 5 de los genes cuyos marcadores se enumeran en la Tabla 1, en donde dicha clasificación de dicho individuo como ER(+) o como ER(-) se lleva a cabo con un método que comprende:
- 15
- a) calcular una primera medida de semejanza entre dicho segundo perfil de expresión y una plantilla de ER(+) y una segunda medida de semejanza entre dicho segundo perfil de expresión y una plantilla de ER (-);
- 20
- (b) comprendiendo dicha plantilla de ER(+), para cada gen en dicha segunda pluralidad de genes, el nivel medio de expresión de dicho gen en una pluralidad de pacientes ER(+);
- comprendiendo dicha plantilla de ER(-), para cada gen en dicha segunda pluralidad de genes, el nivel medio de expresión de dicho gen en una pluralidad de pacientes ER(-), y
- 25
- (b) clasificar (b1) a dicho individuo como ER(+) si dicho segundo perfil de expresión tiene una mayor semejanza con dicha plantilla de ER(+) que con dicha plantilla de ER(-), o (b2) como ER(-) si dicho segundo perfil de expresión tiene una menor semejanza con dicha plantilla de ER(+) que con dicha plantilla de ER(-).
9. El método de la reivindicación 1, que comprende además:
- 30
- (iv) clasificar a dicho individuo como *BRCA1* o esporádico basándose en un segundo perfil de expresión que comprende los niveles de expresión de una segunda pluralidad de genes en una muestra de células tomada del individuo, consistiendo dicha segunda pluralidad de genes en al menos 5 de los genes cuyos marcadores se enumeran en la Tabla 3, en donde dicha clasificación de dicho individuo como *BRCA1* o esporádico se lleva a cabo con un método que comprende:
- 35
- (a) calcular una primera medida de semejanza entre dicho segundo perfil de expresión y una plantilla de *BRCA1* y una segunda medida de semejanza entre dicho segundo perfil de expresión y una plantilla de *no-BRCA1*;
- comprendiendo dicha plantilla de *BRCA1*, para cada gen en dicha segunda pluralidad de genes, el nivel medio de expresión de dicho gen en una pluralidad de pacientes *BRCA1*;
- 40
- comprendiendo dicha plantilla de *no-BRCA1*, para cada gen en dicha segunda pluralidad de genes, el nivel medio de expresión de dicho gen en una pluralidad de pacientes *no-BRCA1*, y
- 45
- (b) clasificar (b1) a dicho individuo como *BRCA1* si dicho segundo perfil de expresión tiene una mayor semejanza con dicha plantilla de *BRCA1* que con dicha plantilla de *no-BRCA1*, o (b2) como esporádico si dicho segundo perfil de expresión tiene una menor semejanza con dicha plantilla de *BRCA1* que con dicha plantilla de *no-BRCA1*.
10. El método de la reivindicación 1, en donde dicho nivel de expresión de cada gen de dicho primer perfil de expresión es un nivel relativo de expresión de dicho gen en dicha muestra de células frente al nivel de expresión de dicho gen en una reserva de referencia.
- 50
11. El método de la reivindicación 10, en donde dicha reserva de referencia procede de una línea celular de cáncer de mama.

12. El método de la reivindicación 10, en donde dicha reserva de referencia procede de una línea celular de una mama normal.
13. El método de la reivindicación 10, en donde dicho nivel relativo de expresión se representa como cociente logarítmico.
- 5 14. El método de la reivindicación 1, en donde dicho paso (i) comprende calcular dicho primer parámetro clasificador entre dicho primer perfil de expresión y dicha plantilla de buen pronóstico, y dicho paso (ii) comprende clasificar a dicho individuo como poseedor de buen pronóstico si dicho primer parámetro clasificador se halla por encima de un umbral elegido.
- 10 15. El método de la reivindicación 14, en donde dicha media es una media ponderada de error del cociente logarítmico.
- 15 16. El método de la reivindicación 14, en donde dicho nivel de expresión de cada gen en dicho primer perfil de expresión es un nivel relativo de expresión de dicho gen en dicha muestra de células frente al nivel de expresión de dicho gen en una reserva de referencia, representado como cociente logarítmico, y en donde el nivel medio de expresión de cada gen en dicha primera pluralidad de genes en dicha plantilla de buen pronóstico es una media de niveles relativos de expresión, siendo cada nivel relativo de expresión el nivel de expresión de dicho gen en una de dicha pluralidad de pacientes que no tienen metástasis distante en los cinco años siguientes al diagnóstico inicial de cáncer de mama frente al nivel de expresión de dicho gen en una reserva de referencia, representado como cociente logarítmico que es una media de los cocientes logarítmicos para dicho gen en dicha pluralidad de pacientes que no tienen metástasis distante en los cinco años siguientes al diagnóstico inicial de cáncer de mama.
- 20 17. El método de la reivindicación 14, en donde dicho primer parámetro clasificador es un coeficiente de correlación entre dicho primer perfil de expresión y dicha plantilla de buen pronóstico.
- 25 18. El método de la reivindicación 1, que comprende además determinar dicho primer perfil de expresión midiendo los niveles de expresión de dicha primera pluralidad de genes en dicha muestra de células de dicho individuo.
19. El método de la reivindicación 8, en donde dicho nivel de expresión de cada gen en dicho segundo perfil de expresión es un nivel relativo de expresión de dicho gen frente al nivel de expresión de dicho gen en una reserva de referencia.
- 30 20. El método de la reivindicación 19, en donde dicha reserva de referencia procede de una línea celular de una mama normal.
21. El método de la reivindicación 19, en donde dicha reserva de referencia procede de una línea celular de un cáncer de mama.
22. El método de la reivindicación 19, en donde dicho nivel relativo de expresión se representa como cociente logarítmico.
- 35 23. El método de la reivindicación 8, en donde dicha media es una media ponderada de error del cociente logarítmico.
- 40 24. El método de la reivindicación 8, en donde dicho nivel de expresión de cada gen en dicho segundo perfil de expresión es un nivel relativo de expresión de dicho gen en dicha muestra de células frente al nivel de expresión de dicho gen en una reserva de referencia, representado como cociente logarítmico; en donde el nivel medio de expresión de cada gen en dicha segunda pluralidad de genes en dicha plantilla de ER (+) es una media de niveles relativos de expresión, siendo cada nivel relativo de expresión el nivel de expresión de dicho gen en una paciente de dicha pluralidad de pacientes ER(+) frente al nivel de expresión de dicho gen en una reserva de referencia, representada como cociente logarítmico que es una media de los cocientes logarítmicos para dicho gen en dicha pluralidad de pacientes ER(+), y en donde el nivel medio de expresión de cada gen en dicha segunda pluralidad de genes en dicha plantilla de ER(-) es una media de los niveles relativos de expresión, siendo cada nivel relativo de expresión el nivel de expresión de dicho gen en una de dicha pluralidad de pacientes ER(-) con respecto al nivel de expresión de dicho gen en una reserva de referencia, representado como cociente logarítmico que es una media de los cocientes logarítmicos para dicho gen en dicha pluralidad de pacientes ER(-).
- 45 25. El método de la reivindicación 8, en donde dicha primera medida de semejanza entre dicho segundo perfil de expresión y dicha plantilla de ER(+) es un coeficiente de correlación entre dicho segundo perfil de expresión y dicha plantilla de ER(+), en donde dicha segunda medida de semejanza entre dicho segundo perfil de expresión y dicha plantilla de ER(-) es un coeficiente de correlación entre dicho segundo perfil de expresión y dicha plantilla de ER(-) y en donde de dicho segundo perfil de expresión se dice que tiene una semejanza mayor con dicha plantilla de ER(+) que con dicha plantilla de ER(-) si
- 50
- 55

- dicho coeficiente de correlación entre dicho segundo perfil de expresión y dicha plantilla de ER(+) es mayor que dicho coeficiente de correlación entre dicho segundo perfil de expresión y dicha plantilla de ER(-), o del que se dice que tiene una semejanza menor con dicha plantilla de ER(+) que con dicha plantilla de ER(-) si dicho coeficiente de correlación entre dicho segundo perfil de expresión y dicha plantilla de ER(+) es menor que dicho coeficiente de correlación entre dicho segundo perfil de expresión y dicha plantilla de ER(-).
26. El método de la reivindicación 8, que comprende además determinar dicho segundo perfil de expresión midiendo los niveles de expresión de dicha segunda pluralidad de genes en dicha muestra de células de dicho individuo.
27. El método de la reivindicación 9, en donde dicho nivel de expresión de cada gen en dicho segundo perfil de expresión es un nivel relativo de expresión de dicho gen en dicha muestra de células frente al nivel de expresión de dicho gen en una reserva de referencia.
28. El método de la reivindicación 27, en donde dicha reserva de referencia procede de una línea celular de una mama normal.
29. El método de la reivindicación 27, en donde dicha reserva de referencia procede de una línea celular de cáncer de mama.
30. El método de la reivindicación 27, en donde dicho nivel relativo de expresión se representa como cociente logarítmico.
31. El método de la reivindicación 9, en donde cada una de dichas medias es una media ponderada de error del cociente logarítmico.
32. El método de la reivindicación 9, en donde dicho nivel de expresión de cada gen en dicho segundo perfil de expresión es un nivel relativo de expresión de dicho gen en dicha muestra de células frente al nivel de expresión de dicho gen en una reserva de referencia, representado como cociente logarítmico; en donde el nivel medio de expresión de cada gen en dicha segunda pluralidad de genes en dicha plantilla de *BRCA1* es una media de niveles relativos de expresión, siendo cada nivel relativo de expresión el nivel de expresión de dicho gen en una paciente de dicha pluralidad de pacientes *BRCA1* frente al nivel de expresión de dicho gen en una reserva de referencia, representado como cociente logarítmico que es una media de cocientes logarítmicos de dicho gen en dicha pluralidad de pacientes *BRCA1* y en donde el nivel medio de expresión de cada gen en dicha segunda pluralidad de genes en dicha plantilla de *no BRCA1* es una media de niveles relativos de expresión, siendo cada nivel relativo de expresión el nivel de expresión de dicho gen en una paciente de dicha pluralidad de pacientes *no-BRCA1* frente al nivel de expresión de dicho gen en una reserva de referencia, representado como un cociente logarítmico que es una media de cocientes logarítmicos de dicho gen en dicha pluralidad de pacientes *no-BRCA1*.
33. El método de la reivindicación 32, en donde dicha primera medida de semejanza entre dicho segundo perfil de expresión y dicha plantilla de *BRCA1* es un coeficiente de correlación entre dicho segundo perfil de expresión y dicha plantilla de *BRCA1*, en donde dicha segunda medida de semejanza entre segundo perfil de expresión y dicha plantilla de *no-BRCA1* es un coeficiente de correlación entre dicho segundo perfil de expresión y dicha plantilla de *no-BRCA1* y en donde de dicho segundo perfil de expresión se dice que tiene una semejanza mayor con dicha plantilla de *BRCA1* que con dicha plantilla de *no-BRCA1* si dicho coeficiente de correlación entre dicho segundo perfil de expresión y dicha plantilla de *BRCA1* es mayor que dicho coeficiente de correlación entre dicho segundo perfil de expresión y dicha plantilla de *no-BRCA1*, o del que se dice que tiene una semejanza menor con dicha plantilla de *no-BRCA1* que con dicha plantilla de *BRCA1* si dicho coeficiente de correlación entre dicho segundo perfil de expresión y dicha plantilla de *BRCA1* es menor que el coeficiente de correlación entre dicho segundo perfil de expresión y dicha plantilla de *no-BRCA1*.
34. El método de la reivindicación 9, que comprende además determinar dicho segundo perfil de expresión midiendo los niveles de expresión de dicha segunda pluralidad de genes en dicha muestra de células de dicho individuo.
35. El método de cualquiera de las reivindicaciones 10-14, 15 y 16, en donde dicha pluralidad consiste en al menos 20 de los genes cuyos marcadores se enumeran en la Tabla 5.
36. El método de cualquiera de las reivindicaciones 10-14, 15 y 16, en donde dicha pluralidad consiste en al menos 50 de los genes cuyos marcadores se enumeran en la Tabla 5.
37. El método de cualquiera de las reivindicaciones 10-14, 15 y 16, en donde dicha pluralidad consiste en al menos 75 de los genes cuyos marcadores se enumeran en la Tabla 5.



- 5 38. El método de cualquiera de las reivindicaciones 1-6, en donde dicho paso (i) comprende calcular dicho segundo parámetro clasificador entre dicho segundo perfil de expresión y dicha plantilla de buen pronóstico y dicho tercer parámetro clasificador entre dicho primer perfil de expresión y dicha plantilla de pronóstico deficiente, consistente dicho paso (ii) en clasificar a dicho individuo como poseedor de buen pronóstico si dicho primer perfil de expresión es más parecido a dicha plantilla de buen pronóstico que a dicha plantilla de pronóstico deficiente, o clasificar a dicho individuo como poseedor de pronóstico deficiente si dicho primer perfil de expresión es más parecido a dicha plantilla de pronóstico deficiente que a dicha plantilla de buen pronóstico.
- 10 39. El método de la reivindicación 38, en donde dicho segundo parámetro clasificador es un coeficiente de correlación entre dicho primer perfil de expresión y dicha plantilla de buen pronóstico y en donde dicho tercer parámetro clasificador es un coeficiente de correlación entre dicho primer perfil de expresión y dicha plantilla de pronóstico deficiente.
40. El método de cualquiera de las reivindicaciones 1-6 y 14, en donde dicho cáncer de mama es cáncer de mama esporádico.
- 15 41. El método de la reivindicación 38, en donde dicho cáncer de mama es cáncer de mama esporádico.

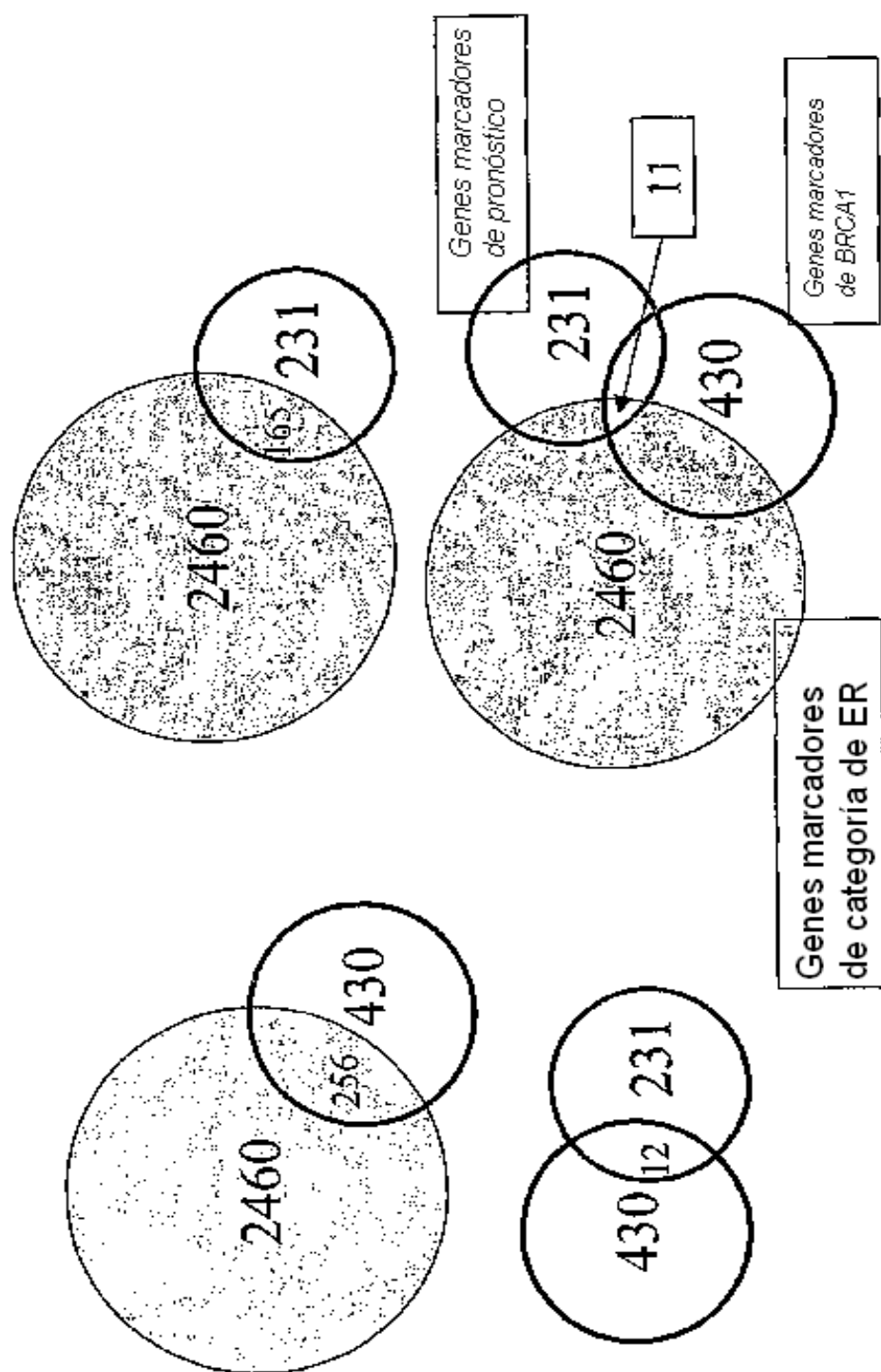


FIG. 1

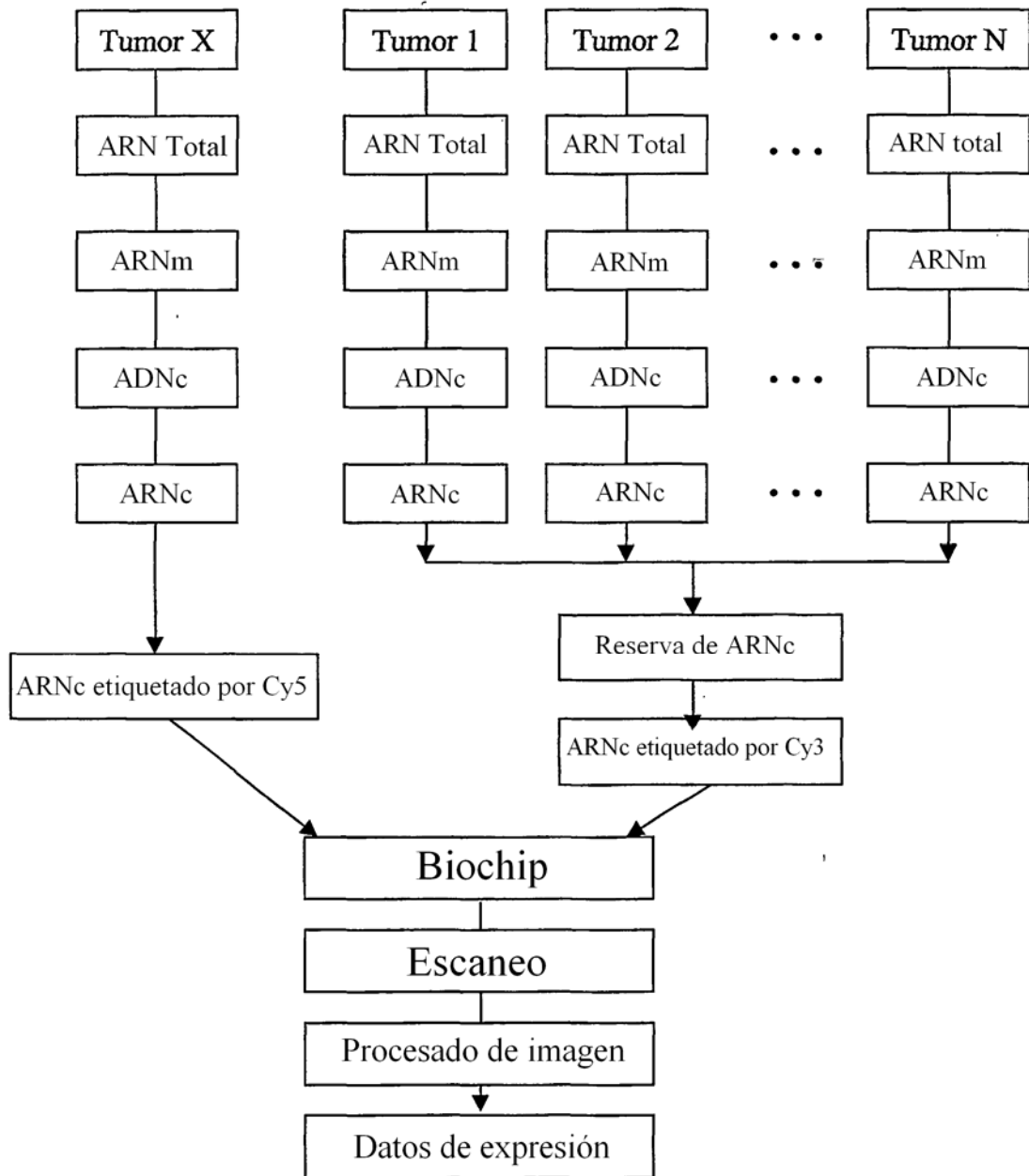


FIG. 2

Dendograma para muestras

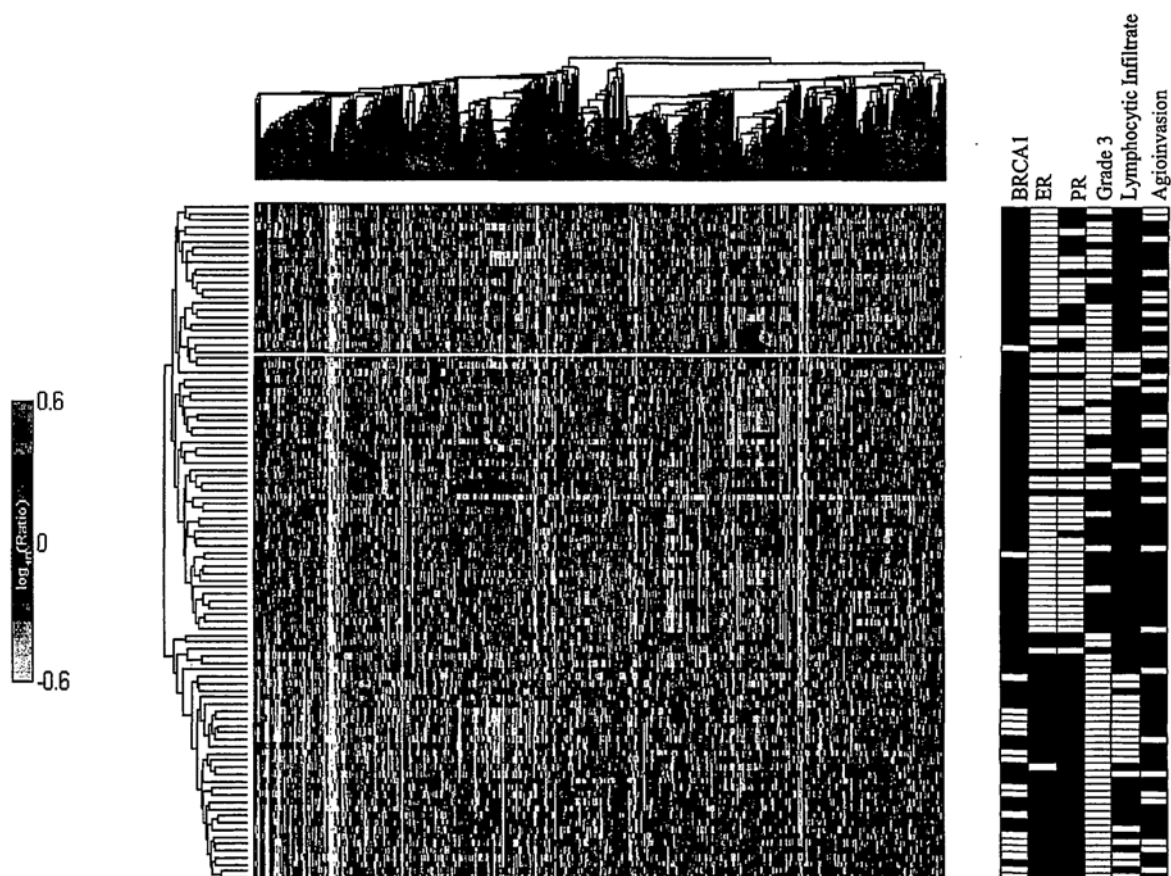


FIG. 3

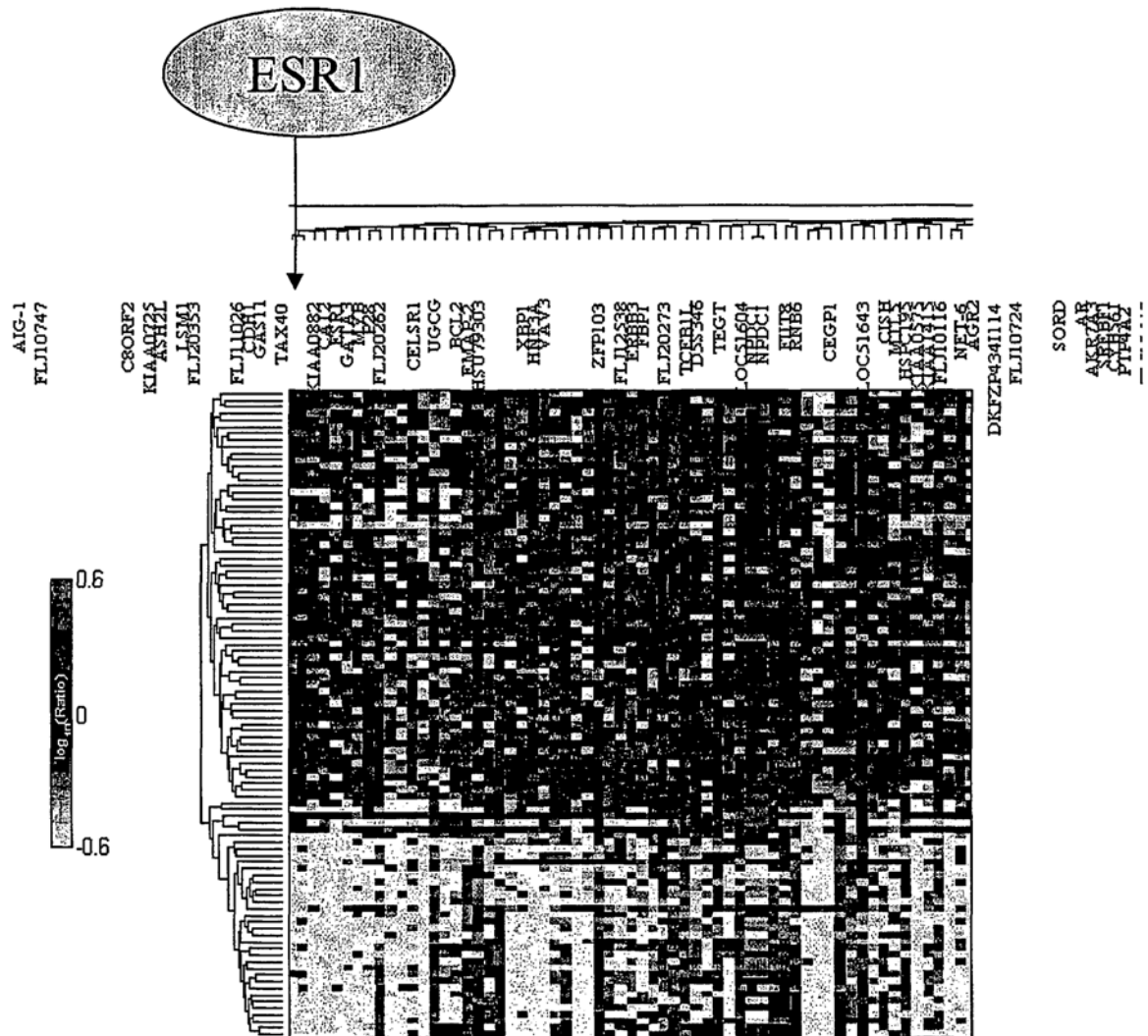


FIG. 4

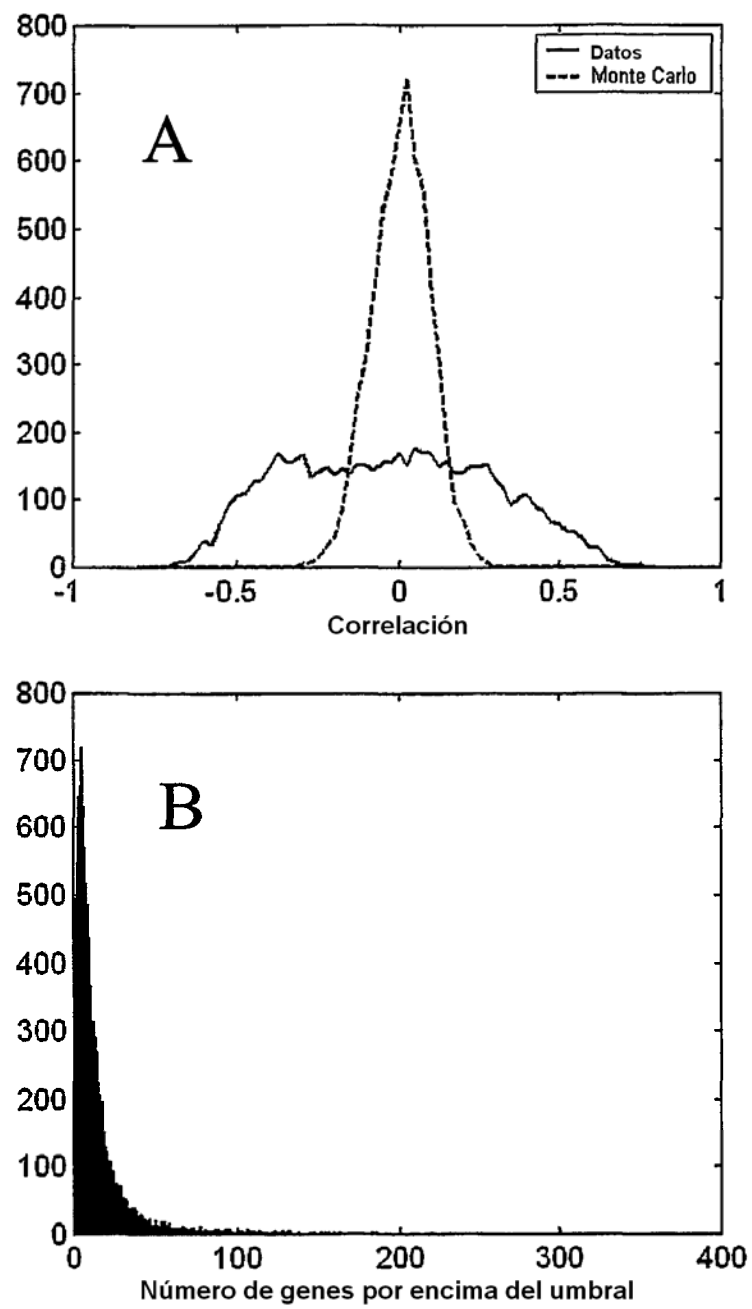


FIG. 5

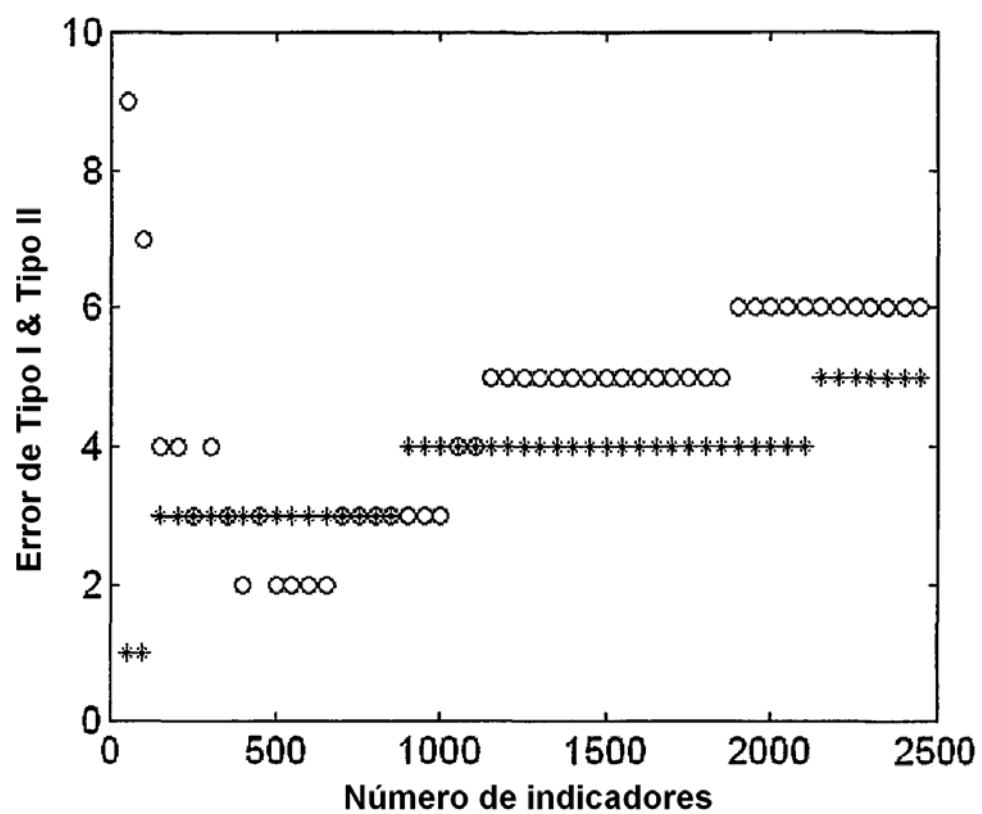


FIG. 6

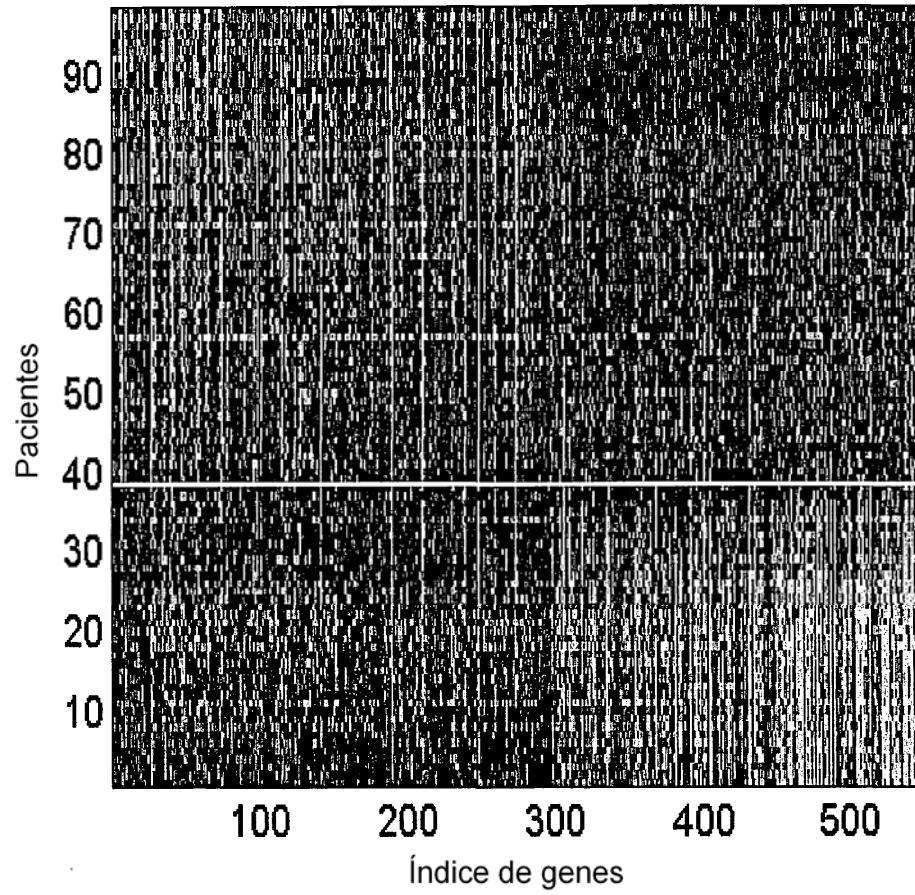


FIG. 7



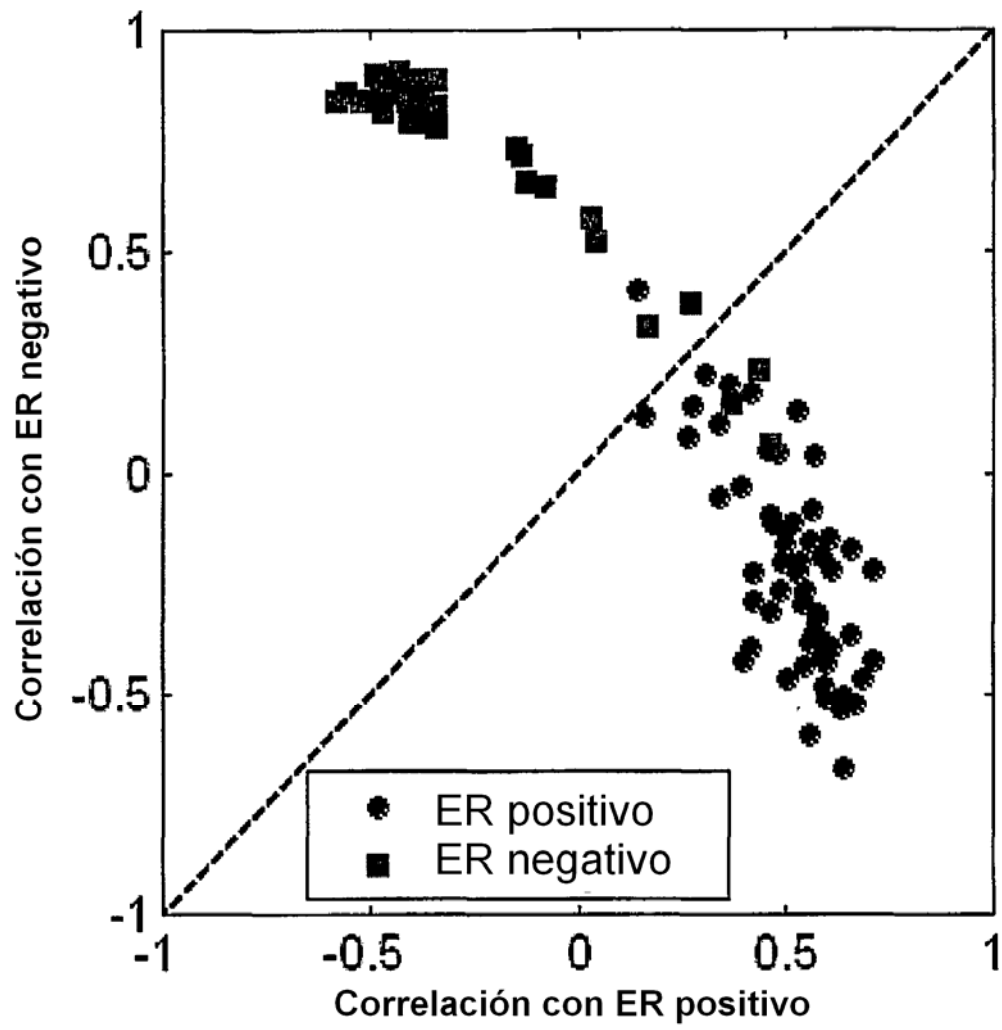


FIG. 8

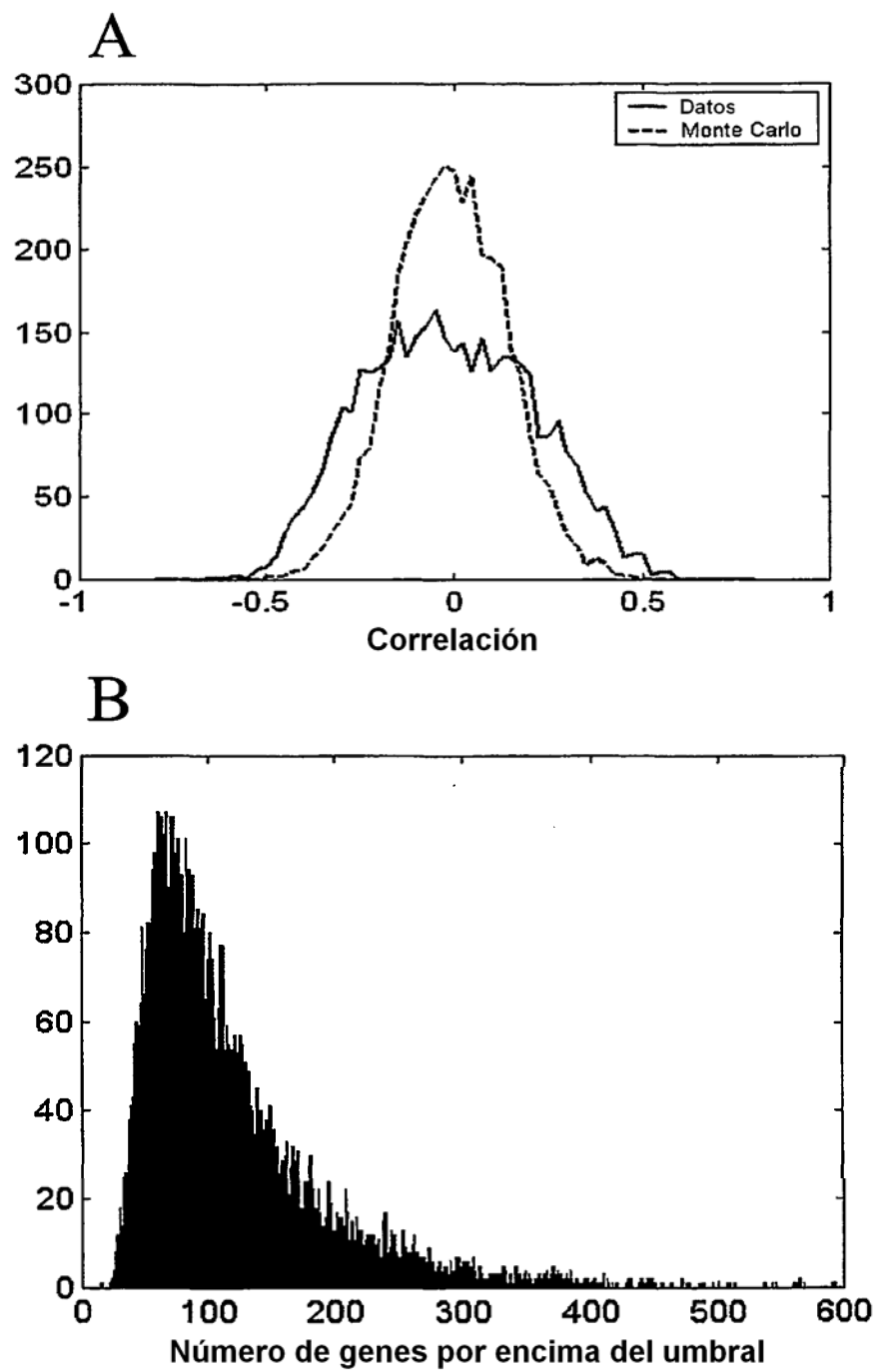


FIG. 9

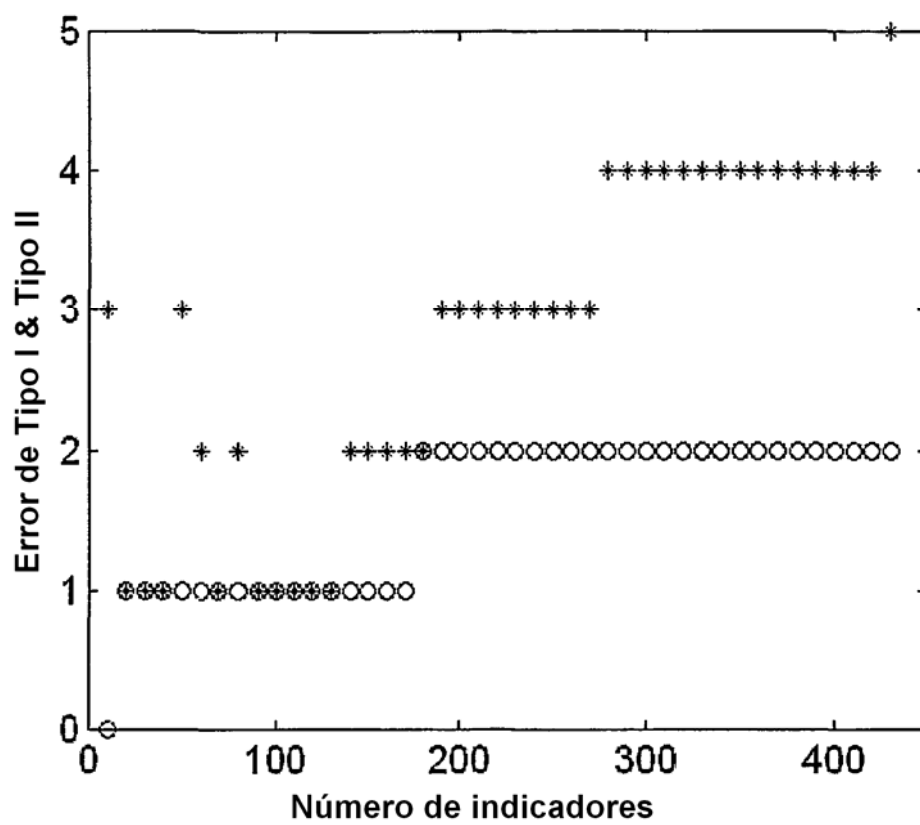


FIG. 10

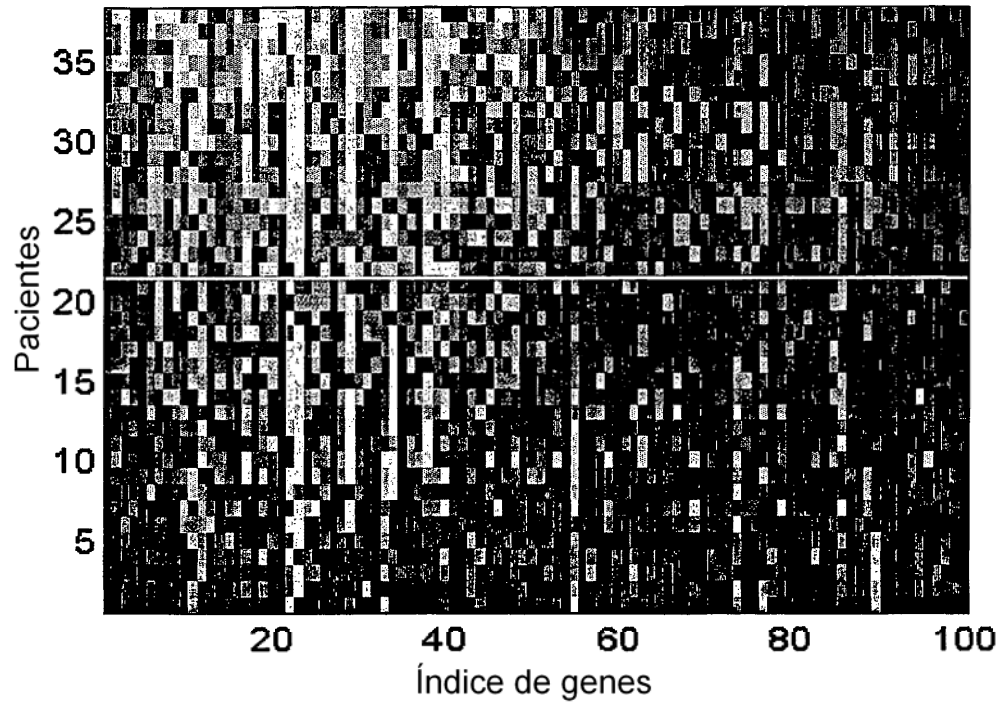


FIG. 11A

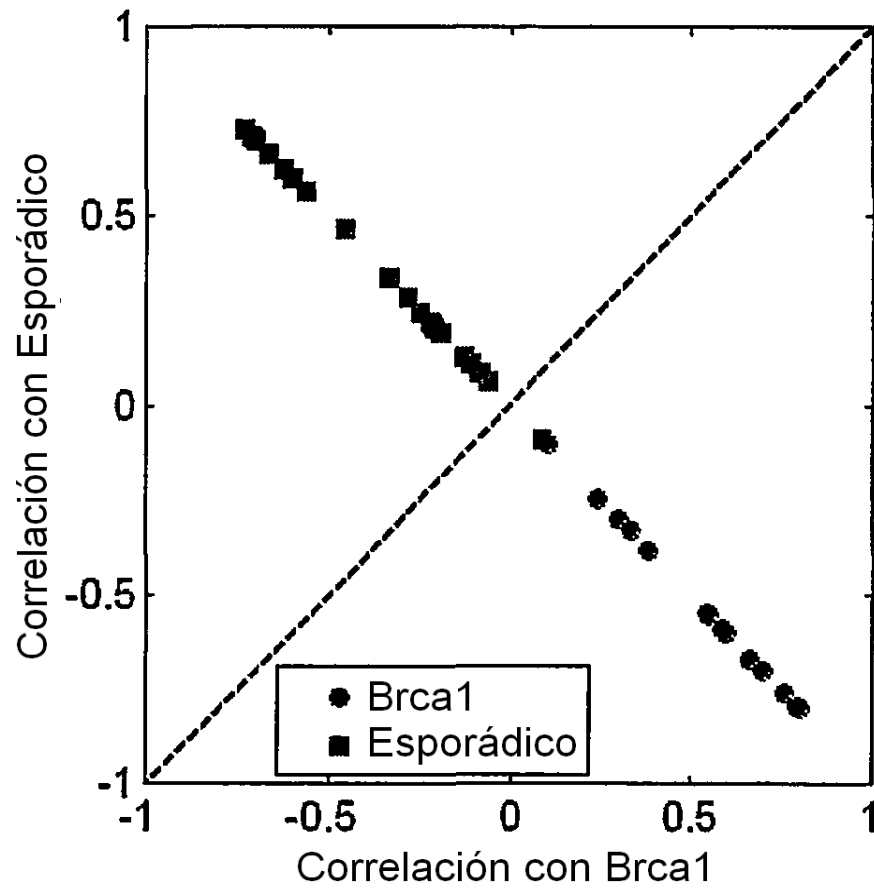


FIG. 11B

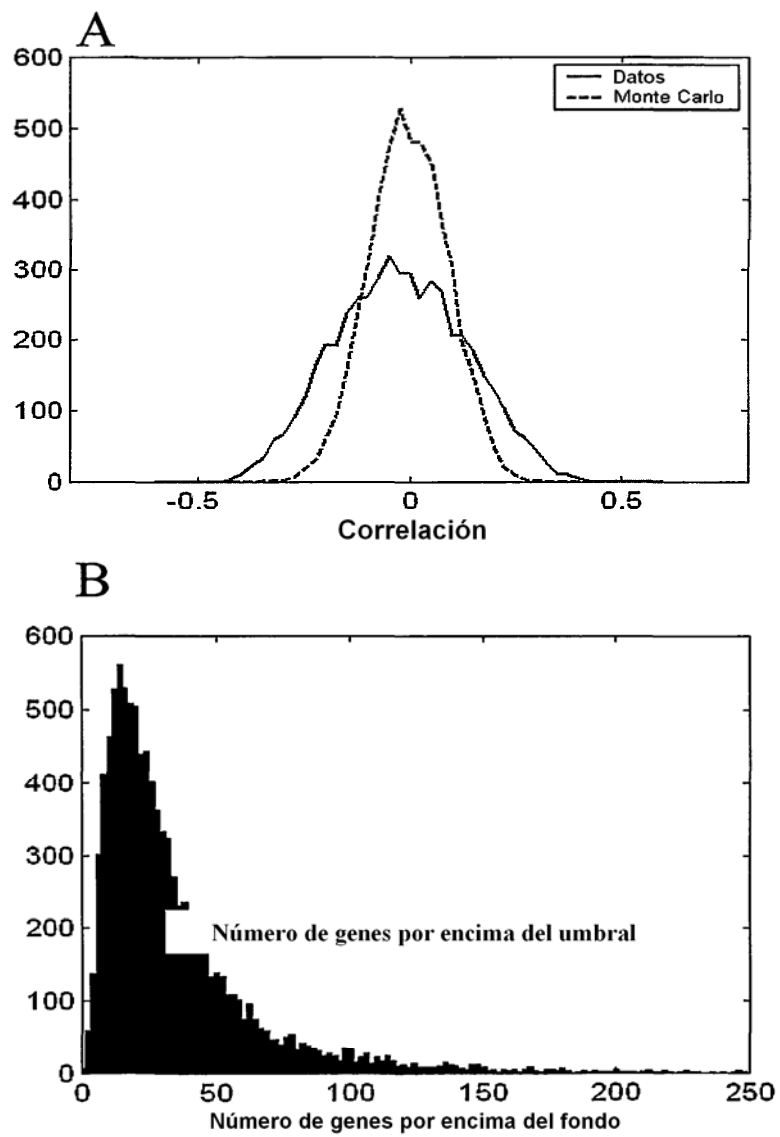


FIG. 12

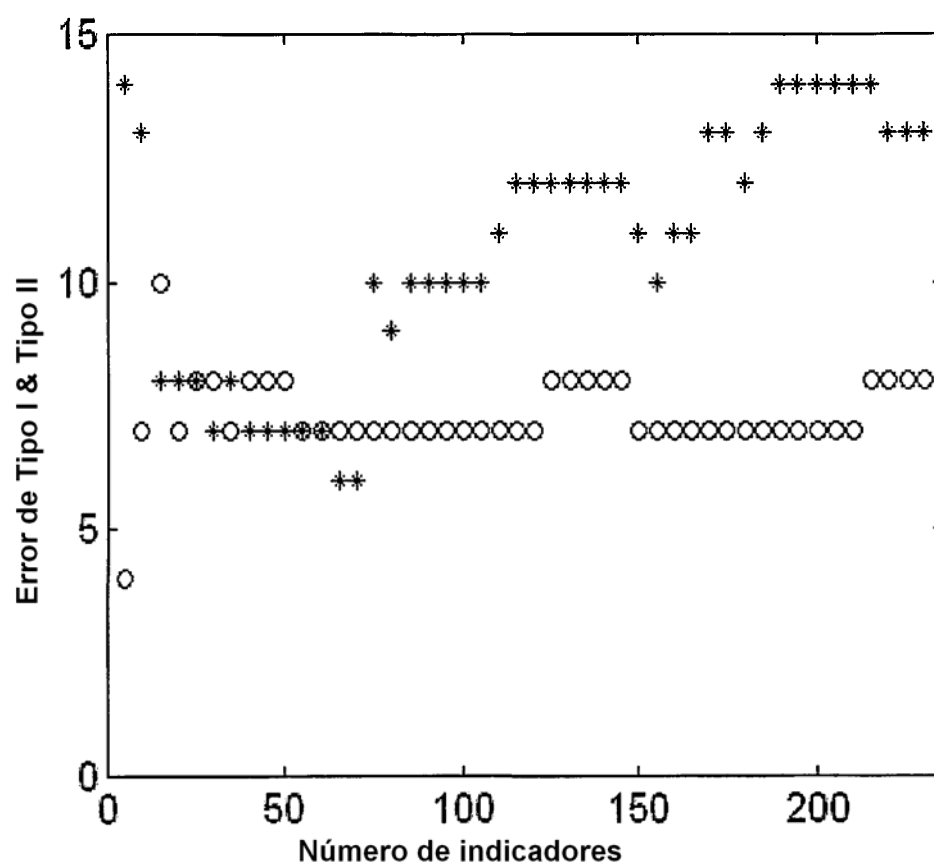


FIG. 13

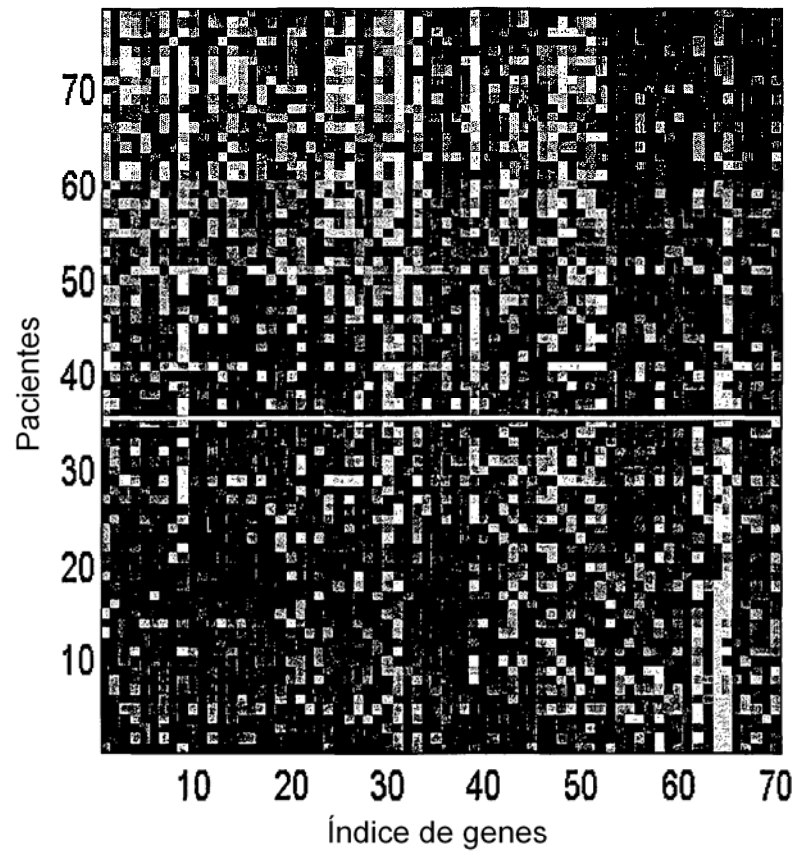


FIG. 14



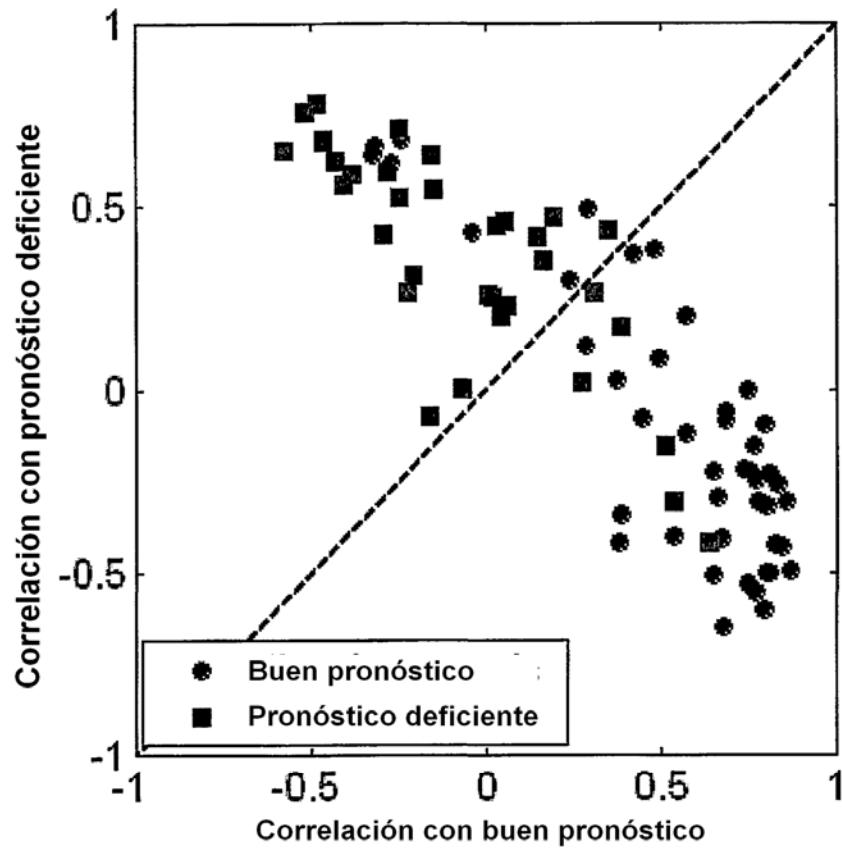


FIG. 15

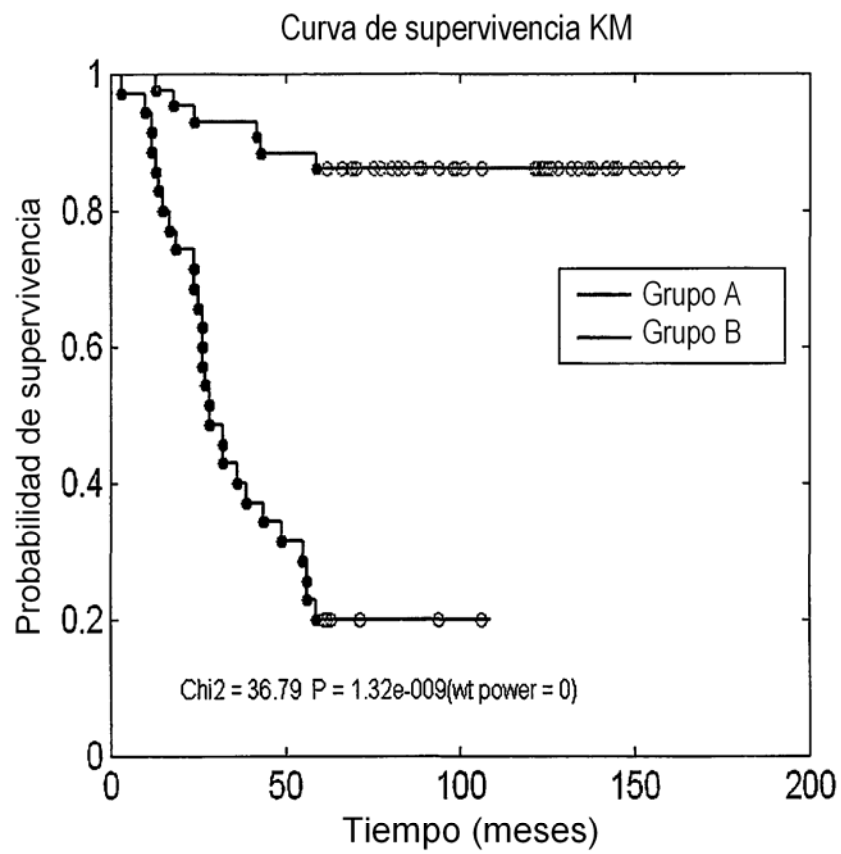


FIG. 16

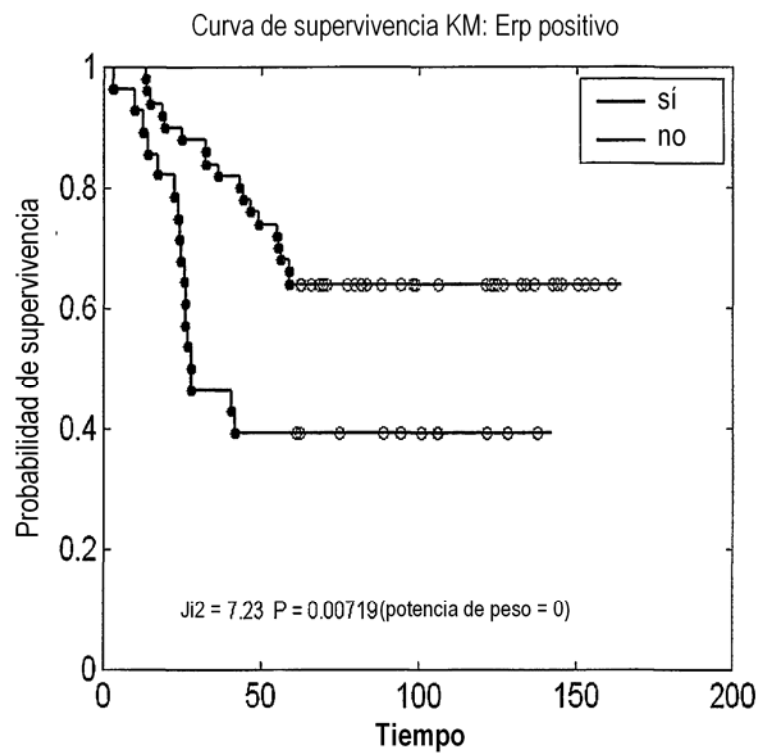


FIG. 17

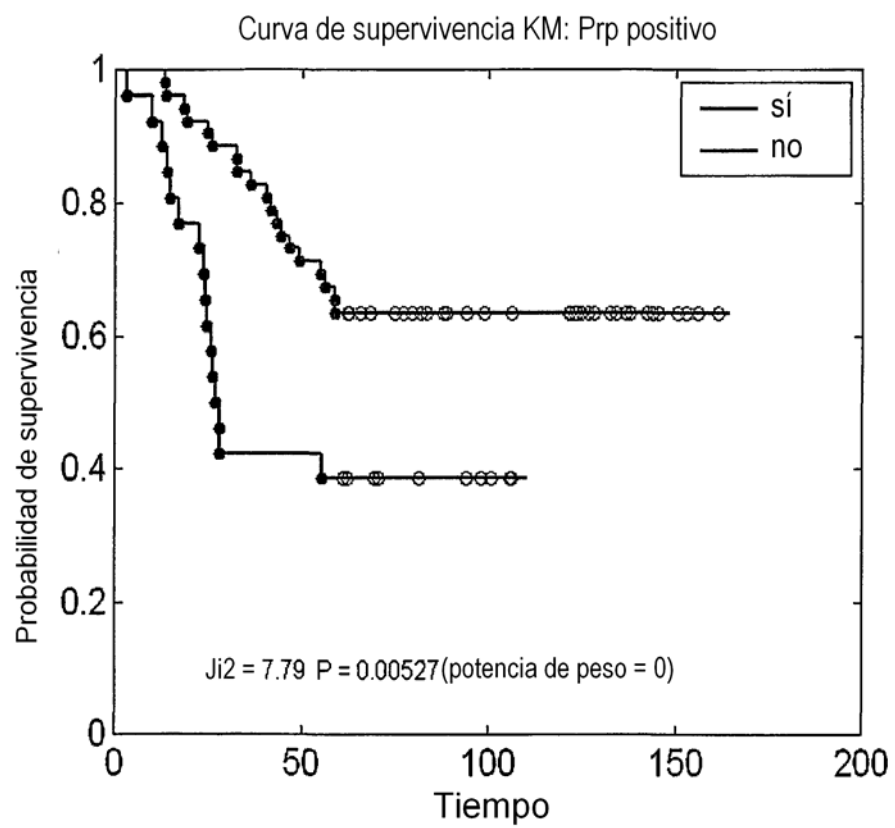


FIG. 18

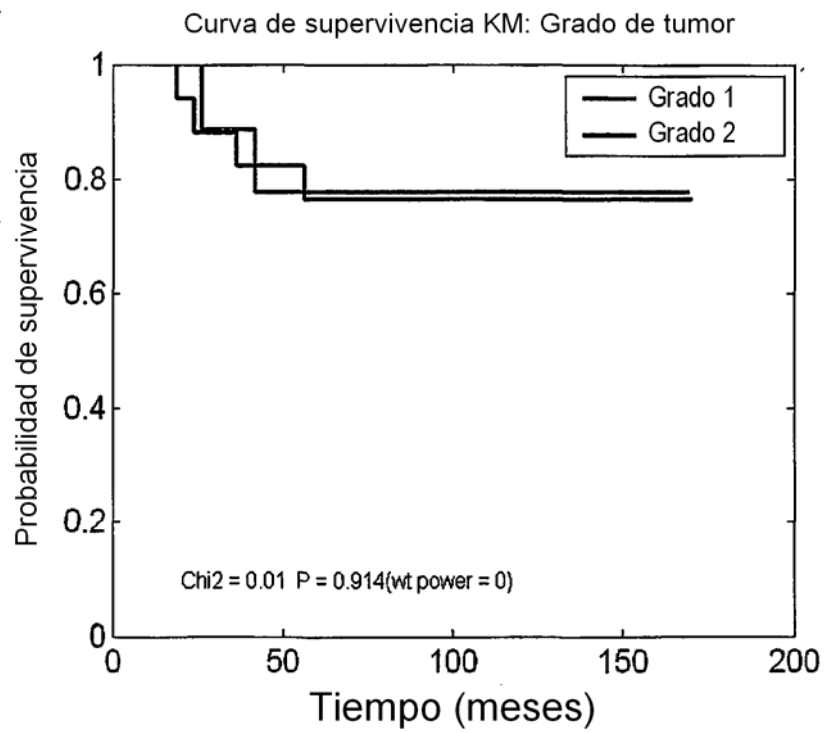


FIG. 19A

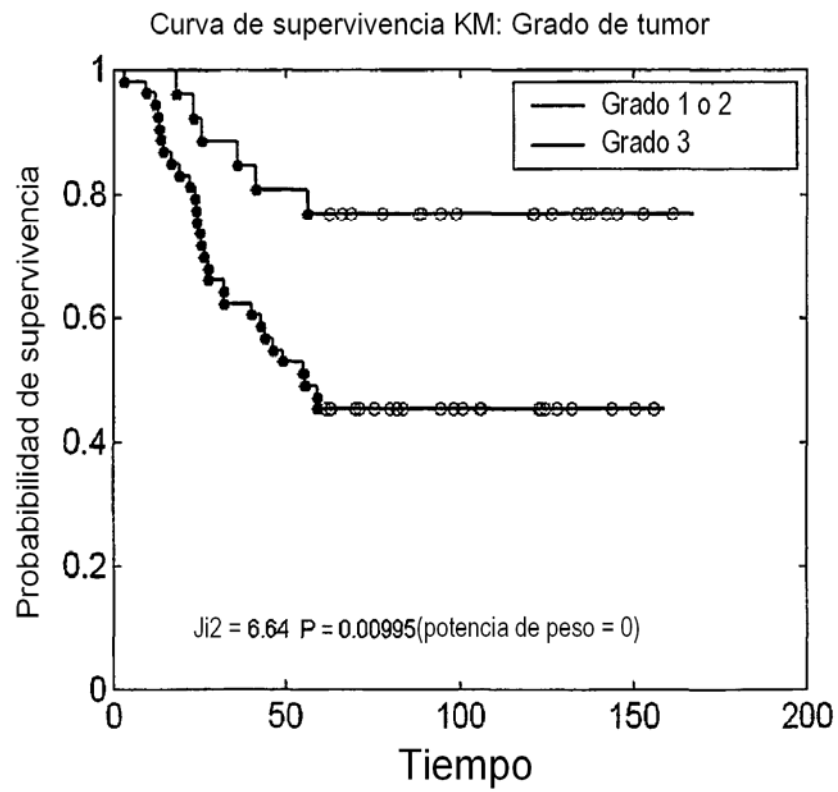


FIG. 19B

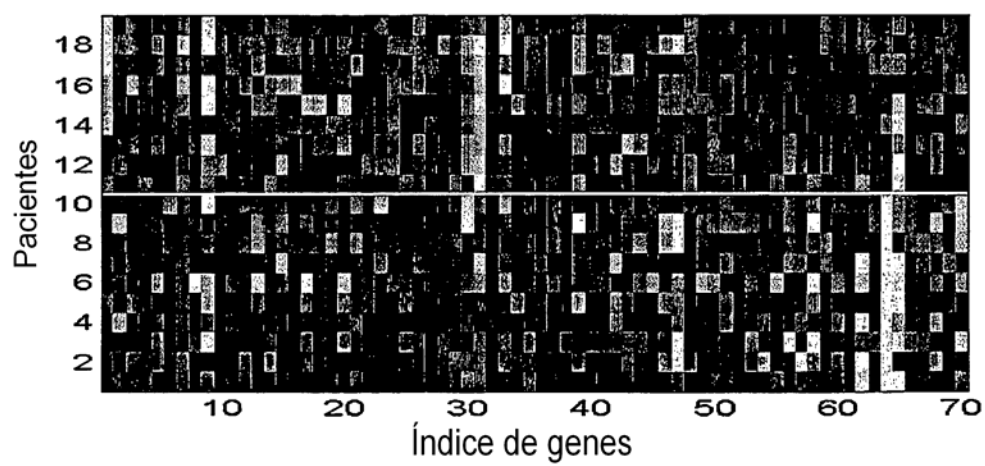


FIG. 20A

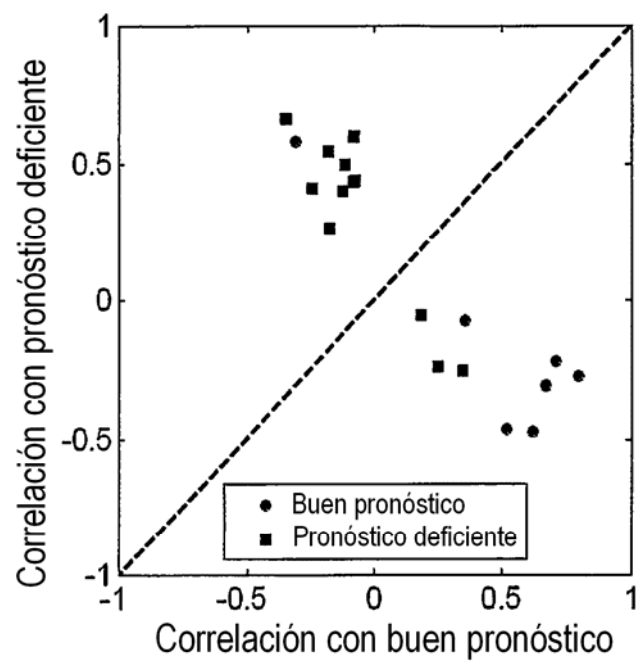


FIG. 20B



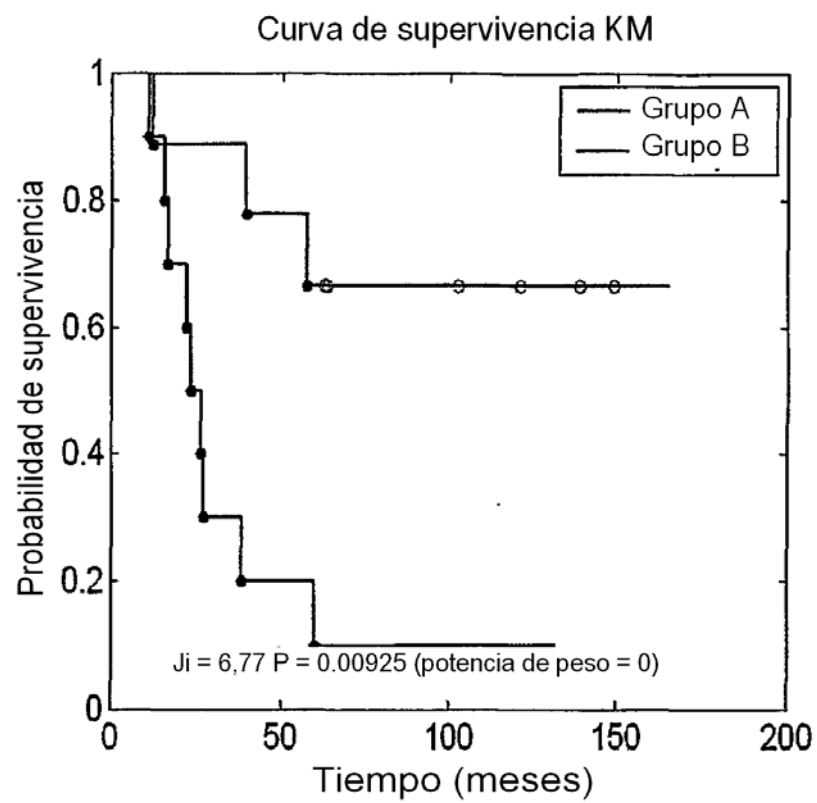


FIG. 20C

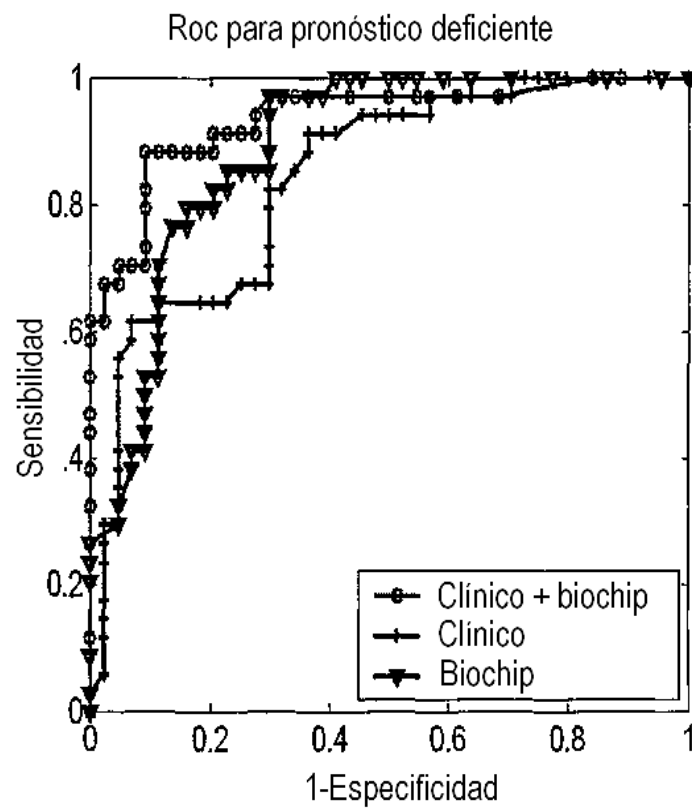


FIG. 21A

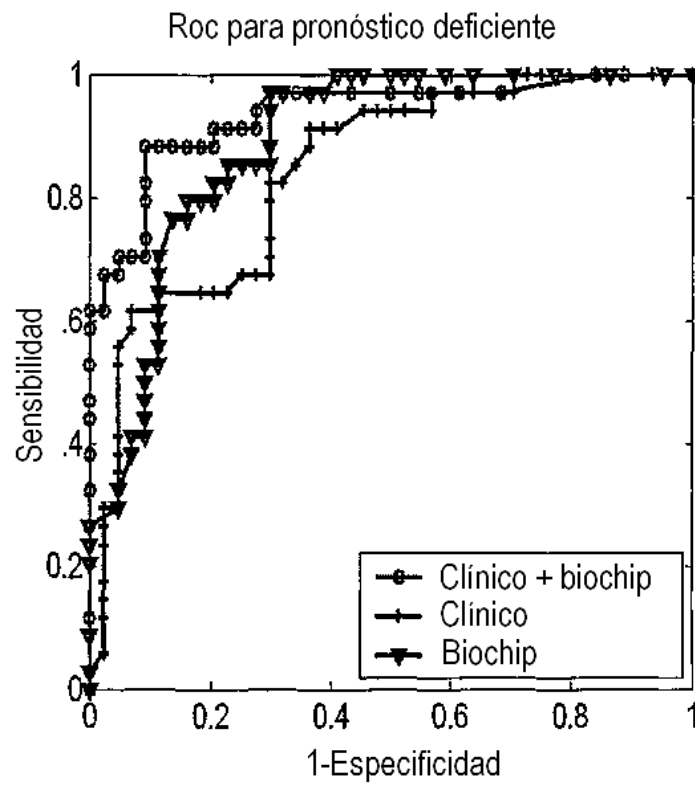


FIG. 21B

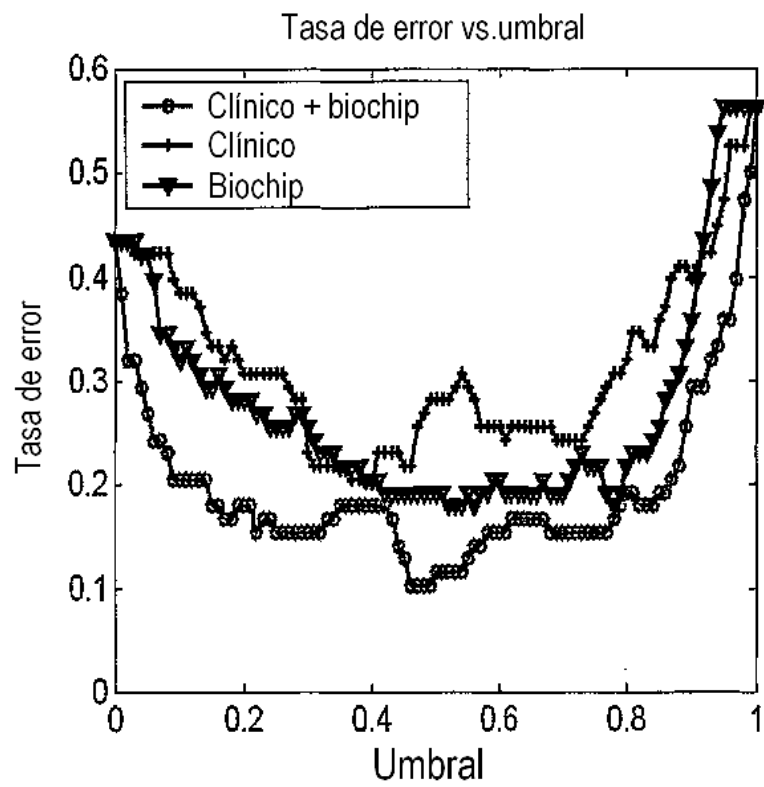
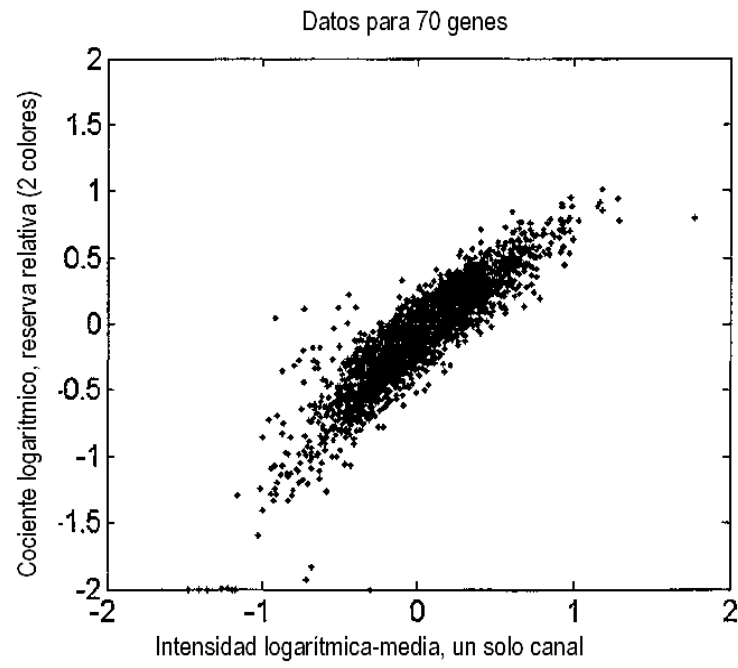


FIG. 21C



**FIG. 22**

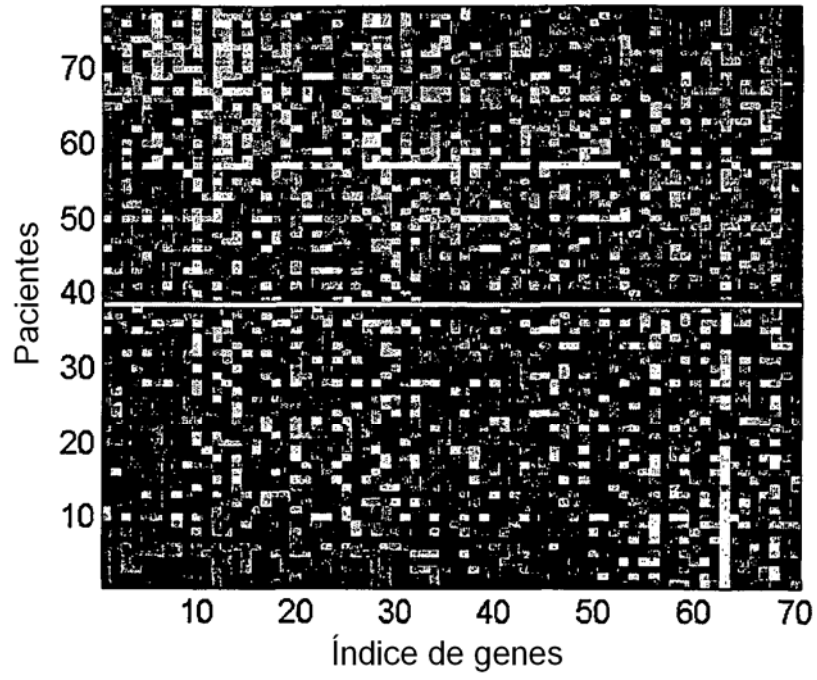


FIG. 23A

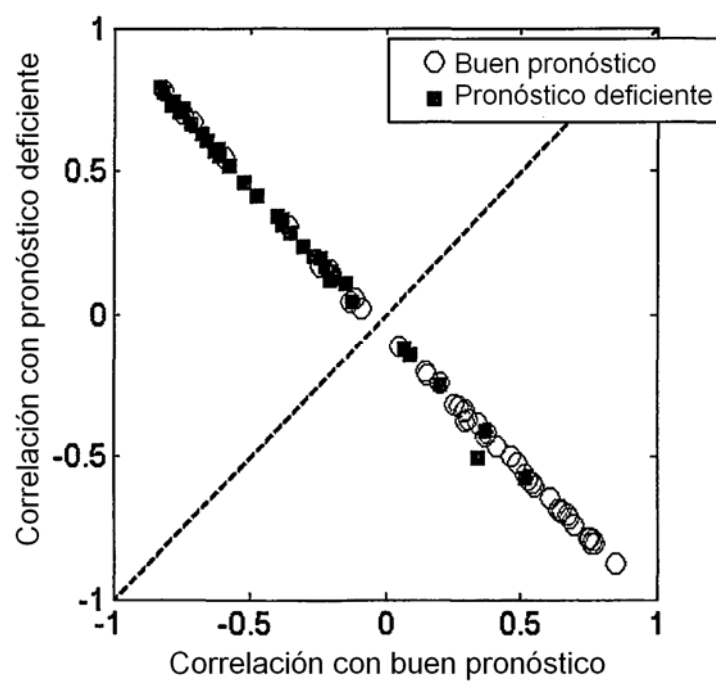


FIG. 23B

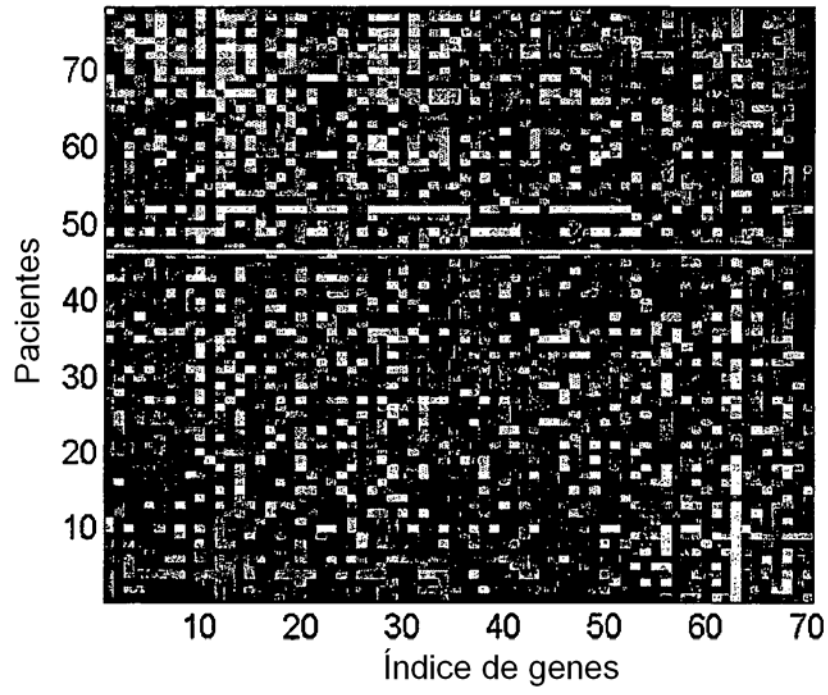


FIG. 24A



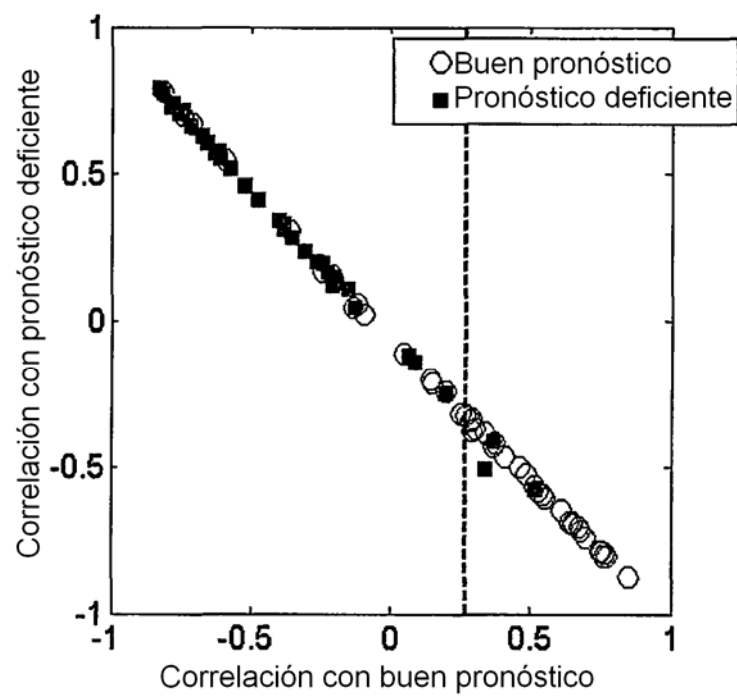


FIG. 24B