



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



(11) Número de publicación: **2 365 732**

(51) Int. Cl.:

**C12Q 1/68** (2006.01)

**C12N 9/10** (2006.01)

(12)

### TRADUCCIÓN DE PATENTE EUROPEA

T3

(96) Número de solicitud europea: **04773526 .1**

(96) Fecha de presentación : **24.09.2004**

(97) Número de publicación de la solicitud: **1668156**

(97) Fecha de publicación de la solicitud: **14.06.2006**

(54) Título: **Método de diagnóstico del cáncer de mama.**

(30) Prioridad: **24.09.2003 US 505571 P**

(73) Titular/es: **ONCOTHERAPY SCIENCE, Inc.**  
**2-1, Sakado 3-chome, Takatsu-ku**  
**Kawasaki-shi, Kanagawa 213-0012, JP**

(45) Fecha de publicación de la mención BOPI:  
**10.10.2011**

(72) Inventor/es: **Nakamura, Yusuke;**  
**Katagiri, Toyomasa y**  
**Nakatsuru, Shuichi**

(45) Fecha de la publicación del folleto de la patente:  
**10.10.2011**

(74) Agente: **Elzaburu Márquez, Alberto**

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

## Método de diagnóstico del cáncer de mama

La presente invención se refiere a métodos para detectar y diagnosticar el cáncer de mama así como métodos para tratar y prevenir el cáncer de mama y la metástasis del cáncer de mama.

5 El cáncer de mama, una enfermedad genéticamente heterogénea, es la malignidad más habitual en mujeres. Se ha indicado un cálculo de aproximadamente de 800.000 nuevos casos cada año en todo el mundo (Parkin DM, Pisani, P, Ferlay J (1999). CA Cancer J Clin 49: 33-64). La mastectomía es la primera opción concurrente para el tratamiento de esta enfermedad. A pesar de la extracción quirúrgica de los tumores primarios, puede producirse la reincidencia en sitios locales o distantes debido a micrometástasis indetectable (Saphner T, Tommey DC, Gray R 10 (1996). J Clin Oncol, 14, 2738-2749) en el momento del diagnóstico. Normalmente se administran agentes citotóxicos como terapia adyuvante después de cirugía con objeto de destruir estas células residuales o premalignas.

15 El tratamiento con agentes quimioterapéuticos convencionales es con frecuencia empírico y se basa principalmente en parámetros tumorales histológicos y sin conocimiento del mecanismo específico. Por lo tanto los fármacos dirigidos a dianas se convierten en el tratamiento base para el cáncer de mama. Los inhibidores tamoxifen y aromatasa, dos representantes de este tipo, han demostrado tener enormes respuestas usados como adyuvantes o como quimioprevención en pacientes con cáncer de mama metastatizado (Fisher B, Costantino JP, Wickerham DL, Redmon CK, Kavanah M, Cronin WM, Vogel V, Robidoux A, Dimitrov N, Atkins J, Daly M, Wieand S, Tan-Chiu E, Ford L, Wolmark N (1998). J. Natl Cancer Inst, 90, 1371-1388; Cuzick J (2002). Lancet 360, 817-824). Sin embargo, el inconveniente es que solamente los pacientes que expresan receptores de estrógenos son sensibles a estos fármacos. Se plantearon incluso recientes cuestiones en cuanto a sus efectos secundarios particularmente sobre la posibilidad de producir cáncer endometrial durante el tratamiento prolongado de tamoxifen así como efectos perjudiciales de fracturas óseas en mujeres posmenopáusicas en pacientes a las que se prescribió aromatasa (Coleman RE (2004). Oncology 18 (5 Suppl. 3), 16-20). Debido a la aparición de efectos secundarios y resistencia a fármacos, obviamente es necesario buscar nuevas dianas moleculares para fármacos eficaces selectivos basándose en mecanismos de acción caracterizados.

20 El cáncer de mama es una enfermedad compleja asociada con numerosos cambios genéticos. Se sabe poco acerca de si estas anomalías son la causa de tumorigénesis de mama, aunque se ha indicado que se producen por un proceso multietapa que puede equipararse ampliamente a la transformación de células normales, mediante las etapas de hiperplasia ductal atípica, carcinoma ductal in situ (DCIS) y carcinoma ductal invasivo (IDC). Existen pruebas de que sólo una parte de las lesiones premalignas progresan obligatoriamente hacia cáncer invasivo mientras que las otras lesiones experimentan regresión espontánea. Esta explicación de participación molecular, que conduce al desarrollo de cáncer de mama primario, su progresión, y su transformación de metástasis, es el enfoque principal para nuevas estrategias dirigidas a la prevención y tratamiento.

25 35 Los perfiles de expresión de genes generados por análisis de micromatriz de ADNc pueden proporcionar bastantes más detalles acerca de la naturaleza de cánceres individuales que los métodos histopatológicos tradicionales son capaces de proporcionar. Lo prometedor de dicha información se basa en su potencial para mejorar estrategias clínicas para el tratamiento de enfermedades neoplásicas y desarrollo de nuevos fármacos (Petricoin, E. F., 3<sup>a</sup>, Hackett, J. L., Lesko, L. J., Puri, R. K., Gutman, S.L., Chumakov, K., Woodcock, J., Feigal, D. W., Jr., Zoon, K. C., y Sistare, F. D. Medical applications of microarray Technologies: a regulatory science perspective. Nat Genet, 32 Suplem: 474-479, 2002). Para esto, los autores de la presente invención han analizado los perfiles de expresión de tumores o tumores procedentes de diferentes tejidos por micromatrizes de ADNc (Okabe, H. et al., Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: identification of genes involved in viral carcinogenesis and tumor progression. Cancer Res, 61: 2129-2137, 2001; Hasegawa, S. et al., 40 Genome-wide analysis of gene expression in intestinal-type gastric cancers using a complementary DNA microarray representing 23.040 genes. Cancer Res, 62: 7012-7017, 2002.; Kaneta, Y. et al., y Ohno, R. Prediction of Sensitivity to ST1571 among Chronic Myeloid Leukemia Patients by Genome-wide cDNA Microarray Analysis. Jpn J Cancer Res, 93: 849-856, 2002.; Kaneta, Y. et al., Genome-wide analysis of gene-expression profiles in chronic myeloid leukemia cells using a cDNA microarray. Int J Oncol, 23: 681-691, 2003.; Kitahara, O. et al., Alterations of gene expression 45 during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissue and normal epithelia. Cancer Res, 61, 3544-3549, 2001.; Lin, Y. et al., Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. Oncogene, 21: 4120-4128, 2002; Nagayama, S. et al., Genome-wide analysis of gene expression in synovial sarcomas using a cDNA microarray. Cancer Res, 62: 5859-5866, 2002.; Okutsu, J. et al., Prediction of chemosensitivity for patients with acute myeloid leukemia, according to expression levels of 28 genes selected by genome-wide complementary DNA microarray analysis. Mol Cancer Ther, 1: 1035-1042, 2002; Kikuchi, T. et al., Expression profiles of non-small cell lung cancers on cDNA microarrays: identification of genes for prediction of lymph-node metastasis and sensitivity to anti-cancer drugs. Oncogen, 22: 2192-2205, 2003).

50 55 Exámenes recientes en los niveles de expresión de miles de genes, mediante el uso de micromatrizes de ADNc, han dado como resultado el descubrimiento de modelos diferentes en tipos diferentes de cáncer de mama (Sgroi, D. C. 60

et al., *In vivo* gene expression profile analysis of human breast cancer progresión. *Cancer Res*, 59: 5656-5661, 1999; Sorlie, T. et al., Gene expression patterns of breast carcinomas distinguish tumor subclases with clinical implications. *Proc Natl Acad Sci U S A*, 98: 10869-10874, 2001.; Kauraniemi, P. et al., New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. *Cancer Res*, 61: 8235-8240, 2001.; Gruvberger, S. et al., S. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res*, 61: 5979-5984, 2001.; Dressman, M. et al., Gene expression profiling detects gene amplification and differentiates tumor types in breast cancer. *Cancer Res*, 63: 2194-2199, 2003).

Estudios en perfiles de expresión de genes en cánceres de mama han dado como resultado la identificación de genes que pueden servir como candidatos para marcadores de diagnóstico o perfiles de pronóstico. Sin embargo, estos datos, derivados principalmente de masas tumorales, no pueden reflejar adecuadamente cambios de expresión durante la carcinogénesis de mama, porque las células del cáncer de mama existen como una masa sólida con una reacción altamente inflamatoria y contienen diversos componentes celulares. Por lo tanto, es probable que datos de micromatrices publicados anteriormente reflejen perfiles heterogéneos.

Estudio diseñados para revelar mecanismos de carcinogénesis ya han facilitado la identificación de dianas moleculares para determinados agentes antitumorales. Por ejemplo, inhibidores de farnesiltransferasa (FTI) que se han desarrollado originalmente para inhibir la ruta de señalización del crecimiento relacionada con Ras, cuya activación depende de la farnesilación post-traduccional, han mostrado ser eficaces en el tratamiento de tumores dependientes de Ras en modelos animales (He et al, *Cell* 99: 335-45 (1999)). De manera similar, ensayos clínicos en seres humanos usando una combinación de fármacos anticancerosos y el anticuerpo monoclonal anti-HER2, trastuzumab, con el objetivo de antagonizar el receptor del protooncogén HER2/neu, han conseguido respuesta clínica mejorada y supervivencia global de pacientes con cáncer de mama (Lin et al., *Cancer Res* 61: 6345-9 (2001)). Por último, se ha desarrollado un inhibidor de tirosina quinasa, ST1-571, que inactiva selectivamente proteínas de fusión bcr-abl, para tratar leucemias mielógenas crónicas en las que la activación constitutiva de la tirosina quinasa bcr-abl desempeña un papel crucial en la transformación de leucocitos. Agentes de estos tipos se diseñan para suprimir la actividad oncogénica de productos génicos específicos (Fujita et al., *Cancer res* 61: 7722-6 (2001)). Por consiguiente, es obvio que productos génicos normalmente regulados positivamente en células cancerosas pueden servir como dianas potenciales para el desarrollo de nuevos agentes anticancerosos.

Adicionalmente se ha demostrado que los linfocitos T citotóxicos (LTC) CD8+ reconocen péptidos epitópicos derivados de antígenos asociados a tumores (AAT) presentados en la molécula MHC de Clase I y células de lisis de tumores. Desde el descubrimiento de la familia MAGE como el primer ejemplo de AAT, se han descubierto muchos otros AAT usando estrategias inmunológicas (Boon, *Int J Cancer* 54: 177-80 (1993); Boon y van der Bruggen, *J Exp Med* 183: 725-9 (1996); van der Bruggen et al., *Science* 254: 1643-7 (1991); Brichard et al., *J Exp Med* 178: 489-95 (1993); Kawakami et al., *J. Exp med* 180: 347-52 (1994)). Algunos de los AAT recientemente descubiertos se someten actualmente a desarrollo clínico como dianas de inmunoterapia. Los AAT descubiertos hasta ahora incluyen MAGE (van der Bruggen et al., *Science* 254: 1643-7 (1991)), gp100 (Kawakami et al., *J Exp Med* 180: 347-52 (1994)), SART (Shichijo et al., *J Exp Med* 187: 277-88 (1998)) y NY-ESO-1 (Chen et al., *Proc Natl Acad Sci USA* 94: 1914-8 (1997)). Por otro lado, se ha demostrado que productos génicos que han mostrado estar específicamente sobreexpresados en células tumorales se reconocen como dianas que inducen respuestas inmunes celulares. Dichos productos génicos incluyen p53 (Umano et al., *Brit J Cancer* 84: 1052-7 (2001)), HER2/neu (Tanaka et al., *Brit J Cancer* 84: 94-9 (2001)), CEA (Nukaya et al., *Int J Cancer* 80: 92-7 (1999)), y similares.

A pesar del progreso significativo en la investigación básica y clínica con respecto a los AAT (Rosenberg et al., *Nature Med* 4: 321-7 (1998); Mukherji et al., *Proc Natl Acad Sci USA* 92: 8078-82 (1995); Hu et al., *Cancer res* 56: 2479-83 (1996)), actualmente solo se dispone de un número limitado de AAT candidatos para el tratamiento de adenocarcinomas, incluyendo el cáncer colorrectal. Los AAT se expresan abundantemente en células cancerosas aunque esta expresión se limita a células cancerosas serían candidatos prometedores como dianas inmunoterapéuticas. Además, se espera que la identificación de nuevos AAT que inducen respuestas inmunes antitumorales fuertes y específicas impulsen el uso clínico de estrategias de vacunación de péptidos contra diversos tipos de cáncer (Boon y can der Bruggen, *J Exp Med* 183: 725-9 (1996); van der Bruggen et al., *Science* 254: 1643-7 (1991); Brichard et al., *J Exp Med* 178: 489-95 (1993); Kawakami et al., *J Exp Med* 180: 347-52 (1994); Shichijo et al., *J Exp Med* 187: 277-88 (1998); Chen et al., *Proc Natl Acad Sci USA* 94: 1914-8 (1997); Harris, H *Natl Cancer Inst* 88: 1442-5 (1996); Butterfield et al., *Cancer res* 59: 3134-42 (1999); Vissers et al., *Cancer Res* 59: 5554-9 (1999); van der Burg et al., *J Immunol* 156: 3308-14 (1996); Tanaka et al., *Cancer Res* 57: 4465-8 (1997); Fujie et al., *Int J Cancer* 80: 169-72 (1999); Kikuchi et al., *Int J Cancer* 81: 459-66 (1999); Oiso et al., *Int J Cancer* 81: 387-94 (1999)).

Reiteradamente se ha informado que las células mononucleares de sangre periférica (CMSp) estimuladas por péptidos procedentes de determinados donantes sanos producen niveles significativos de IFN- $\gamma$  en respuesta al péptido, pero raras veces ejercen citotoxicidad contra células tumorales de manera restrictiva en HLA-A24 o -A0201 en ensayos de liberación de  $^{51}\text{Cr}$  (Kawano et al., *Cancer Res* 60: 3550-8 (2000); Nishizaka et al., *Cancer res* 60: 4830-7 (2000); Tamura et al., *Jpn J Cancer Res* 92: 762-7 (2001)). Sin embargo, tanto HLA-A24 como HLA-A0201 son alelos de HLA conocidos en la población japonesa así como en la caucasiana (Date et al., *Tissue Antigens* 47: 93-101 (1996); Kondo et al., *J. Immunol* 155: 4307-12 (1995); Kubo et al., *J Immunol* 152: 3913-24 (1994); Imanishi et al., *Proceeding of the eleventh International Histocompatibility Workshop and Conference Oxford University Press*, Oxford, 1065 (1992); Williams et al., *Tissue Antigen* 49: 129 (1997)). Por lo tanto, péptidos antigenicos de

carcinomas presentados por estos HLA pueden ser especialmente útiles para el tratamiento de carcinomas entre japoneses y caucasianos. Además, se sabe que la inducción de LTC de baja afinidad *in vitro* normalmente resulta del uso de péptidos a una concentración elevada, generando un nivel elevado de complejos péptido/MHC en células presentadoras de antígenos (CPA), que activaran eficazmente estos LTC (Alexander-Miller et al., Proc Natl Acad Sci USA 93: 4102-7 (1996)).

Anteriormente, se ha descrito la sobreexpresión de PBK en diversos tipos de tejidos como tejido pulmonar, tejido de próstata, tejido de colon o tejido de mama y se refieren a muestras de tumor de mama en términos generales que consideran que PBK puede presentar una mayor expresión en dichas muestras (documento WO02/29104).

Por consiguiente, en un esfuerzo por comprender los mecanismos carcinogénicos asociados con cáncer y para identificar posibles dianas para el desarrollo de nuevos agentes anticancerosos, los autores de la presente invención realizaron análisis pangenómicos a gran escala de los perfiles de expresión de genes encontrados en poblaciones purificadas de células de cáncer de mama, incluyendo 12 carcinomas ductales *in situ* (DCIS) y 69 carcinomas ductales invasivos (IDC), usando una micromatriz de ADNc que representa 23.040 genes.

La presente descripción se basa en el descubrimiento de un patrón de expresión génica que se correlaciona con cáncer de mama (BRC, siglas en inglés). Los genes que se expresan diferencialmente en cáncer de mama se denominan, en la presente memoria "ácidos nucleicos BRC" o "polinucleótidos BRC" y los correspondientes polipéptidos codificados se conocen como "polipéptidos BRC" o "proteínas BRC".

La presente invención se refiere a las realizaciones caracterizadas en las reivindicaciones. Por lo tanto, se refiere a los siguientes apartados:

20 1. Un método para el diagnóstico *in vitro* del carcinoma ductal invasivo (IDC) o una predisposición para el desarrollo de IDC en un sujeto, que comprende determinar un nivel de expresión de un gen asociado al cáncer de mama en una muestra biológica procedente de un paciente, en el que un aumento del nivel de expresión en dicha muestra en comparación con un nivel de control normal de dicho gen indica que el sujeto padece o corre el riesgo de desarrollar IDC, en el que dicho gen asociado al cáncer de mama es BRC No 456.

25 2. El método de 1, en el que dicho nivel de expresión en la muestra es al menos 10% superior que dicho nivel de control normal.

30 3. El método de 1, en el que el nivel de expresión del gen se determina por un método seleccionado del grupo que consiste en:

(a) detectar el ARNm de BRC No. 456;

35 (b) detectar la proteína codificada por BRC No. 456, y

(c) detectar una actividad quinasa de la proteína codificada por BRC No. 456.

4. El método de 3, en el que dicha detección se realiza en una matriz de ADN.

5. El método de 1, en el que dicha muestra biológica procedente de un paciente comprende

(a) una célula epitelial;

35 (b) una célula IDC; o

(c) una célula epitelial de tejido que se sabe o se sospecha que es IDC.

6. Un método para explorar un compuesto para tratar o prevenir IDC, comprendiendo dicho método las etapas de:

(a) poner en contacto un compuesto de ensayo con el polipéptido codificado por BRC No. 456; y

40 (b) detectar la actividad de unión entre dicho polipéptido y dicho compuesto de ensayo.

7. Un método para explorar un compuesto para tratar o prevenir IDC, comprendiendo dicho método la etapa de:

poner en contacto un compuesto candidato con una célula que expresa BRC No 456; y determinar el nivel de expresión de BRC No. 456.

8. El método de 7, en el que dicha célula es una célula IDC.

45 9. Un método para explorar un compuesto para tratar o prevenir IDC, comprendiendo dicho método las etapas de:

(a) poner en contacto un compuesto de ensayo con el polipéptido codificado por BRC No. 456; y

(b) detectar la actividad quinasa del polipéptido de la etapa (a).

10. El uso de una secuencia de nucleótidos complementaria a la secuencia codificante de BRC No. 456 para la preparación de una composición farmacéutica para el tratamiento o prevención de IDC en un sujeto.

5 11. El uso de una composición de ARNip, que reduce la expresión de BRC No. 456 para la preparación de una composición farmacéutica para el tratamiento o prevención de IDC en un sujeto

12. El uso de 11, en el que el ARNip comprende la cadena en sentido que comprende una secuencia de nucleótidos seleccionada del grupo que consiste en las secuencias de nucleótidos de laS SEC ID Nº: 25, 28 y 31 como la secuencia diana.

10 13.- Una composición farmacéutica que comprende ARNip contra BRC No 456, en la que dicho ARNip comprende la cadena en sentido que comprende una secuencia de nucleótidos seleccionada del grupo que consiste en las secuencias de nucleótidos de las SEC ID Nº: 25, 28 y 31 como la secuencia diana para usar en el tratamiento o prevención de IDC.

15 14. Una molécula bicatenaria que comprende una cadena en sentido y una cadena antisentido, en la que la cadena en sentido comprende una secuencia de ribonucleótidos que corresponde a una secuencia diana que consiste en las SEC ID Nº: 25, 28 y 31, y en la que la cadena antisentido comprende una secuencia de ribonucleótidos que es complementaria a dicha cadena en sentido, en la que dicha cadena en sentido y dicha cadena antisentido hibridan entre sí para formar dicha molécula bicatenaria para usar en el tratamiento o prevención de IDC.

15. La molécula bicatenaria de 14, que tiene la fórmula general seleccionada entre grupo que consiste en:

20 (a) gaacgauauaaagccagcc-[B]-ggcuggccuuauaucguuc,

(b) cuggaugaaucauaccaga-[B]-ucugguaugauucauccag, y

(c) guguggcuugcguaauaa-[B]-uuauuuacgcaagccacac,

en la que [B] es una secuencia en bucle de ribonucleótidos que consiste en de 3 a 23 nucleótidos.

25 Por consiguiente, se describe un método de diagnóstico o de determinación de una predisposición al cáncer de mama en un sujeto determinando un nivel de expresión de un gen asociado al BRC en una muestra biológica procedente de un paciente, tal como una muestra de tejido. La expresión "gen asociado al BRC" se refiere a un gen que se caracteriza por un nivel de expresión que es diferente en una célula de BRC en comparación con el de una célula normal. Una célula normal es una célula obtenida de tejido de mama. En el contexto de la presente invención, el gen asociado al BRC es el gen que aparece en la lista de las tablas 3-8 (es decir, el gen de BRC No. 456). Un cambio, por ejemplo, un aumento o una disminución en cuanto al nivel de expresión de un gen en comparación con un nivel de control normal del gen, indica que el sujeto padece o está en riesgo de desarrollar BRC.

30 En el contexto de la presente invención, la frase "nivel de control" se refiere a un nivel de expresión de proteína detectado en una muestra de control e incluye tanto un nivel de control normal como un nivel de control del cáncer de mama. Un nivel de control puede ser un solo patrón de expresión derivado de una sola población de referencia o de una pluralidad de patrones de expresión. Por ejemplo, el nivel de control puede ser una base de datos de patrones de nivel de expresión procedente de células previamente ensayadas. Un "nivel de control normal" se refiere a un nivel de expresión génica detectado en un individuo normal, sano o en una población de individuos que se sabe que no padecen cáncer de mama. Un individuo normal es aquel que no presenta síntomas clínicos de cáncer de mama. Por otro lado, un "nivel de control de BRC" se refiere a un perfil de expresión de genes asociados al BRC encontrados en una población que padece BRC.

35 Un aumento en el nivel de expresión de uno o más genes asociados al BRC mostrados los listados de las tablas 3, 5, y 7 (es decir, los genes de BRC Nos. 123-175, 374-398, y 448-471) detectados en una muestra de ensayo en comparación con un nivel de control normal indica que el sujeto (del que se obtuvo la muestra) padece o corre el riesgo de desarrollar BRC. En cambio, una disminución en el nivel de expresión de uno o más genes asociados al BRC mostrados los listados de las tablas 4, 6, y 8 (es decir, los genes de BRC Nos. 176-373, 399-447, y 472-512) detectados en una muestra de ensayo en comparación con un nivel de control normal indica que dicho sujeto padece o corre el riesgo de desarrollar BRC.

40 Como alternativa, la expresión de un panel de genes asociados al BRC en una muestra puede compararse con un nivel de control de BRC del mismo panel de genes. Una similitud entre una expresión en una muestra y una expresión del control de BRC indica que el sujeto (del cual se obtuvo la muestra) padece o corre el riesgo de desarrollar BRC.

45 De acuerdo con la presente invención, el nivel de expresión génica se considera "modificado" cuando la expresión génica aumenta o disminuye un 10%, 25%, 50% en comparación con el nivel de control. Como alternativa, un nivel de expresión se considera "aumentado" o "disminuido" cuando la expresión génica está aumentada o disminuida al

menos 0,1, al menos 0,2, al menos 1, al menos 2, al menos 5, o al menos 10 veces o más en comparación con un nivel de control. La expresión se determina detectando hibridación, por ejemplo, en una matriz de una sonda de genes asociados al BRC con un transcripto génico de la muestra derivada del paciente.

5 En el contexto de la presente invención, la muestra derivada del paciente es cualquier tejido obtenido de un sujeto de ensayo, *por ejemplo*, un paciente que se sabe o que se sospecha que tiene BRC. Por ejemplo, el tejido puede contener una célula epitelial. Más particularmente, el tejido puede ser una célula epitelial de un carcinoma ductal de mama.

10 También se describe un perfil de expresión de referencia BRC que comprende un nivel de expresión génica de dos o más genes asociados al BRC mostrados los listados de las tablas 3-8. Como alternativa, el perfil de expresión de referencia BRC puede comprender los niveles de expresión de dos o más genes asociados al BRC mostrados los listados de las tablas 3, 5 y 7, o genes asociados al BRC mostrados los listados de las tablas 4, 6, y 8.

15 La presente invención también proporciona métodos para identificar un agente que inhiba o mejore la expresión o la actividad de un gen asociado al BRC, es decir, el gen BCR No 456 asociado al BRC, poniendo en contacto una célula de ensayo que exprese un gen asociado al BRC con un compuesto de ensayo y determinando el nivel de expresión del gen asociado al BRC o la actividad de su producto génico. La célula de ensayo puede ser una célula epitelial, tal como una célula epitelial obtenida de un carcinoma de mama. Una disminución en el nivel de expresión de un gen asociado al BRC regulado positivamente o de la actividad de su producto génico en comparación con un nivel de control normal o con la actividad del gen o producto génico indica que el agente de ensayo es un inhibidor del gen asociado al BRC y puede usarse para reducir un síntoma de BRC, por ejemplo, la expresión de uno o más genes asociados al BRC mostrados los listados de las tablas 3, 5 y 7. Como alternativa, un aumento en el nivel de expresión de un gen asociado al BRC regulado negativamente o de la actividad de su producto génico en comparación con un nivel de control normal o con la actividad del gen o producto génico indica que el agente de ensayo es un amplificador de la expresión o función del gen asociado al BRC y puede usarse para reducir un síntoma de BRC, por ejemplo, la infraexpresión de uno o más genes asociados al BRC mostrados los listados de las tablas 4, 6 y 8.

20 También se describe un kit que comprende un reactivo de detección que se une a uno o más ácidos nucleicos de BRC o polipéptidos de BRC. También se proporcionan una matriz de ácidos nucleicos que se une a uno o más ácidos nucleicos de BRC.

25 Los métodos terapéuticos, como se describe en la presente memoria, incluyen un método para tratar o prevenir el BRC en un sujeto que incluye la etapa de administrar al sujeto una composición antisentido. En el contexto de la presente invención, la composición antisentido reduce la expresión del gen diana específico. Por ejemplo, la composición antisentido puede contener un nucleótido que es complementario a una secuencia génica asociada al BRC seleccionada del grupo que consiste en los genes asociados al BRC mostrados los listados de las tablas 3, 5, y 7. Como alternativa, el presente método puede incluir las etapas de administrar a un sujeto una composición de ARN de interferencia pequeño (ARNip) .En el contexto de la presente descripción, la composición de ARNip reduce la expresión de un ácido nucleico de BRC seleccionado del grupo que consiste en los genes asociados al BRC mostrados los listados de las tablas 3, 5 y 7. En otro método más, el tratamiento o prevención del BRC en un sujeto puede realizarse administrando a un sujeto una composición de ribozima. En el contexto de la presente descripción, la composición de ribozima específica de ácido nucleico reduce la expresión de un ácido nucleico de BRC seleccionado del grupo que consiste en los genes asociados al BRC mostrados los listados de las tablas 3, 5 y 7. Realmente, se confirmó el efecto de inhibición del ARNip para genes asociados al BRC mostrados en las listas de las tablas. Por ejemplo, en la sección de ejemplos, se demostró claramente que el ARNip para el BRC-456 de la tabla 7 (Acceso a GenBank No. AF237709, TOPK; proteína quinasa originada en células T-LAK) inhibe la proliferación celular de células de cancer de mama. Por lo tanto, en la presente invención, los genes asociados al BRC mostrados los listados de las tablas 3, 5 y 7, especialmente el BRC-456, es la diana terapéutica preferible del cancer de mama. Otros métodos terapéuticos incluyen aquellos en los que a un sujeto se le administra un compuesto que aumenta la expresión de uno o más de los genes asociados al BRC mostrados los listados de las tablas 4, 6 y 8 o la actividad de un polipéptido codificado por uno o más de los genes asociados al BRC mostrados los listados de las tablas 4 6 y 8.

30 También se describen vacunas y métodos de vacunación. Por ejemplo, un método de tratamiento o prevención del BRC en un sujeto puede implicar administrar al sujeto una vacuna que contenga un polipéptido codificado por un ácido nucleico seleccionado del grupo que consiste en genes asociados al BRC mostrados los listados de las tablas 3, 5 y 7 o un fragmento inmunológicamente activo de dicho polipéptido. En el contexto de la presente descripción, un fragmento inmunológicamente activo es un polipéptido de menor longitud que la proteína de origen natural de longitud completa que aun induce una respuesta inmunológica análoga a la inducida por la proteína de longitud completa. Por ejemplo, un fragmento inmunológicamente activo debe tener al menos una longitud de 8 restos y deber poder estimular una célula inmunológica tal como una célula T ó una célula B. La estimulación de células inmunológicas puede medirse detectando la proliferación celular, la elaboración de citocinas (por ejemplo, IL-2), o la producción de un anticuerpo.

35 50 55 60 Adicionalmente, se describen moléculas diana para tratar o prevenir metástasis de cáncer de mama. De acuerdo con

la presente invención, los genes mostrados en el listado de las tabla 11 (es decir, genes de BRC Nos. 719-752) se identificaron como genes que tenían patrones de expresión modificados únicos en células de cáncer de mama con metástasis en nódulos linfáticos. Por lo tanto, la metástasis del cáncer de mama puede tratarse o prevenirse mediante la supresión de la expresión o actividad de genes regulados positivamente o sus productos génicos seleccionados del grupo que consiste en VAMP3, MGC11257, GSPT1, DNM2, CFL1, CLNS1A, SENP2, NDUFS3, NOP5/NOP58, PSMD13, SUOX, HRB2, LOC154467, THTPA, ZRF1, LOC51255, DEAF1, NEUL, UGCGL1, BRAF, TUFM, FLJ10726, DNAJBL AP4S1 y MRPL40. Como alternativa, la metastasis del cáncer de mama puede tratarse o prevenirse potenciando la expresión o actividad de UBA52, GenBank nº de acceso AA634090, CEACAM3, C21orf97, KIAA1040, EEF1D, FUS, GenBank nº de acceso AW965200 y KIAA0475 en células cancerosas.

5 También se describen métodos para pronosticar metástasis de cáncer de mama. Específicamente, el presente método comprende la etapa de medir el nivel de expresión de genes marcadores seleccionados del grupo que consiste en los genes mostrados en el listado de la tabla 11. En la presente memoria, estos genes marcadores se identifican como genes que tienen patrones de expresión modificados únicos en células de cáncer de mama de pacientes con metástasis en nódulos linfáticos. Por lo tanto, la metastasis del cáncer de mama en un sujeto puede pronosticarse determinando si el nivel de expresión detectado en una muestra derivada del sujeto está más cerca del nivel de expresión medio de casos positivos o casos negativos de metastasis en nódulos linfáticos en muestras de referencia. Entre los genes regulados positivamente, se identificó el A7870, denominado proteína quinasa originado en células T-LAK (*TOPK*), que se sobreexpresaba tres veces mas en 30 de 39 (77%) casos de cáncer de mama que podían obtener datos de expresión, especialmente en 29 de 36 (81%) casos con muestras de ensayo de carcinoma ductal invasivo. Posterior RT-PCR semi-cuantitativa también confirmó que el A7870 se regulaba positivamente en 7 de 12 muestras clínicas de cáncer de mama y en 17 de 20 líneas celulares de cáncer de mama, en comparación con órganos humanos normales que incluyen células ductales de mama o de mama normales.

10 Análisis de transferencia de Northern revelaron que el transcripto de A7870 se expresaba solo en líneas de células de cáncer de mama y en testículos y timo de seres humanos normales. La tinción inmunocitoquímica con anticuerpo *TOPK* mostró que la localización subcelular de A7870 endógeno se observaba en el citoplasma y alrededor de la membrana nuclear en líneas de células de cáncer de mama, T47D, BT20 y HBC5. El tratamiento de células de cáncer de mama con los ARN de interferencia pequeños (ARNip) inhibió eficazmente la expresión de A7870 y suprimió el crecimiento de células/tumores de líneas de células de cáncer de mama, T47D y BT-20, lo que sugería que este gen desempeña una función principal en la proliferación del crecimiento celular. Estos hallazgos sugieren

15 20 25 30

que la sobreexpresión de A7870 podría estar implicada en la tumorigénesis de mama y en estrategias prometedoras para el tratamiento específico de pacientes con cáncer de mama.

A menos que se defina otra cosa, todos los términos técnicos y científicos usados en la presente memoria tienen el mismo significado que el comúnmente entendido por un experto habitual en la materia a la cual pertenece esta invención. Aunque pueden usarse métodos y materiales similares o equivalentes a los descritos en la presente memoria, en la realización práctica o ensayo de la presente invención, a continuación se describen métodos y materiales adecuados.

En caso de conflicto, se controlará la presente memoria descriptiva, incluyendo las definiciones. Adicionalmente, los materiales, métodos, y ejemplos son solo ilustrativos y no pretenden ser limitantes.

Otra ventaja de los métodos descritos en la presente memoria es que la enfermedad se identifica antes de detectar síntomas clínicos palpables. Otras características y ventajas de la invención serán apreciables a partir de la siguiente descripción detallada y de las reivindicaciones.

#### BREVE DESCRIPCIÓN DE LOS DIBUJOS

La Figura 1 representa imágenes de células premicrodisecionadas (carril A), postmicrodisecionadas (carril B) y microdisecionadas (carril C). La microdiseción de células DCIS, IDC y de células epiteliales ductales de mama normales se realizó usando microdissección con micro-haz láser (LMM). Se realizó la microdiseción de células DCIS (caso de 10326T), células IDC (10502T), y de células epiteliales ductales de mama normales (10341N) de cada muestra de ensayo partir de secciones teñidas con hematoxilina y eosina.

La figura 2 representa los resultados del análisis no supervisado de agrupación jerárquica bidimensional de 710 genes en 102 muestras. En la Figura 2(A), cada fila horizontal representa un paciente con cáncer de mama y cada columna vertical muestra un solo gen. El color de cada pocillo representado en rojo y verde indica niveles de transcripto por encima y por debajo de la media para esos genes en todas las muestras, respectivamente. Un asterisco indica el tipo histológico principal y una punta indica el tipo histológico inferior en el mismo caso. Un cuadrado indica un caso duplicado (10149a1 y 10149a1T). Un cuadrado en negrita indica expresión sin cambio. ER se refiere al estado de ER medido por EIA, LN a metástasis en nódulos linfáticos y ESRI a perfiles de expresión de ESR1 en esta micromatriz. La Figura 2(B) representa el análisis de agrupación jerárquica bidimensional de 89 genes en 16 muestras con 2 lesiones diferenciadas microdisecionadas de 8 pacientes con cáncer de mama. La Figura 2(C) representa el análisis de agrupación usando 25 genes que muestra expresión diferencial entre células de cáncer ductal invasivo bien diferenciadas y poco diferenciadas.

La Figura 3 representa el análisis supervisado de agrupación jerárquica de genes usando 97 genes seleccionados

por un ensayo de permutación al azar. En la fila horizontal, se muestran 41 muestras ER-positivas y 28 muestras ER-negativas (seleccionadas de pacientes premenopáusicas). En la columna vertical, se agruparon 97 genes en diferentes ramas de acuerdo con similitud en proporciones de expresión relativas. Los genes en la rama principal inferior, se expresaron preferencialmente de manera similar al nivel de expresión de ESR1 así como en la Figura 2(A). Los de la rama superior fueron en proporción inversa de ESR1.

La Figura 4 representa genes con expresión modificada en DCIS en relación con ductales normales y en IDC en relación con DCIS. La Figura 4(A) representa una agrupación de 251 genes normalmente regulados positiva ó negativamente en DCIS e IDC. La Figura 4(B) representa una agrupación de 74 genes que tienen expresión elevada o disminuida en transición desde DCIS a IDC. La Figura 4(C) representa una agrupación de 65 genes específicamente regulados positiva o negativamente en IDC.

La Figura 5 representa los resultados de validación por RT-PCR semi-cuantitativa de genes altamente expresados. Específicamente, se examinó la expresión de 5 genes (AI261804, AA205444 y AA167194 en 12 casos bien diferenciados y AA676987 y H22566 en 12 casos mal diferenciados) y GAPDH (control interno) por RT-PCR semi-cuantitativa. Las señales de la micromatriz correspondieron a los resultados de experimentos RT-PCR semi-cuantitativa. Se prepararon células ductales de mama procedentes de células epiteliales ductales normales, en 15 pacientes premenopáusicas usadas en esta micromatriz. MG se refiere a toda la glándula mamaria humana.

La Figura 6 representa los resultados de la RT-PCR semi-cuantitativa. Se muestran niveles de expresión de A7870 en células tumorales de (a) 12 pacientes con cáncer de mama, (b) líneas de células de cáncer de mama (HBC4, HBC5, HBL100, HCC1937, MCF7, MDA-MB-231, SKBR3, T47D, YMB1, BT-20, BT-474, BT-549, HCC1143, HCC1500, HCC1599, MDA-MB-157, MDA-MB-435S, MDA-MB-453, OCUB-F y ZR-75-1), y tejidos humanos normales.

La Figura 7 representa los resultados del análisis de transferencia de Northern de transcriptos de A7870 en (a) diversos tejidos humanos y (b) líneas de células de cáncer de mama y órganos vitales humanos normales.

La Figura 8 representa la localización subcelular de (a) A7870 exógeno en células COS7 transfectadas y (b) A7870 exógeno en células T47D, BT-20 y HBC5.

La Figura 9 representa el análisis supervisado de agrupación jerárquica de genes usando 206 genes seleccionados de un ensayo de permutación al azar. En la fila horizontal, se representan 69 muestras (seleccionadas de pacientes con IDC). En la columna vertical, se agruparon 97 genes en diferentes ramificaciones de acuerdo con similitud en proporciones de expresión relativa. Los genes en las ramificaciones 1 y 2, se expresaron preferencialmente de manera similar al nivel de expresión del tipo mal diferenciado y del tipo bien-diferenciado.

La Figura 10(A) representa los resultados de un análisis de agrupación jerárquica bidimensional usando 34 genes seleccionados por evaluación de ensayo de clasificación y exclusión después de un ensayo de permutación al azar para establecer un sistema de puntuación predictivo. Los genes, en la rama principal superior, se expresaron preferencialmente en casos que implicaban metástasis en nódulos linfáticos; los de la rama inferior se expresaron más altamente en casos negativos en nódulos linfáticos. La Figura 10 (B) representa la fuerza de genes que aparecen en 7 (A) para separar tumores no metastásicos (negativos en nódulos linfáticos) de tumores metastásicos (positivos en nódulos linfáticos). Los cuadrados representan casos positivos en nódulos, los triángulos indican casos negativos. Los 17 cuadrados en blanco representan casos de ensayos positivos en nódulos linfáticos y los 20 triángulos en blanco representan casos de ensayos negativos en nódulos linfáticos que no se usaron para establecer puntuaciones predictivas. La Figura 10 (C) representa la correlación entre la puntuación predictiva para la información de metastasis y clínica después de la operación.

Las palabras "un", "una" y "el", "la" como se usa en la presente memoria significa "al menos un (una)" a menos que se indique específicamente otra cosa.

Generalmente, las células de cáncer de mama se presentan como una masa sólida que tiene una reacción altamente inflamatoria y que contiene diversos componentes celulares. Por lo tanto, es probable que datos previamente publicados de micromatrizes reflejen perfiles heterogéneos.

Teniendo en cuenta estas cuestiones, los autores de la presente invención prepararon poblaciones purificadas de células de cáncer de mama y de células ductales epiteliales de mama normales mediante un método de microdissección con micro-haz láser (LMM) y analizaron perfiles de expresión de genes de todo el genoma de los 81 BRC, incluyendo 12 carcinomas ductales situ (DCIS) y 69 carcinomas ductales invasivos (IDC), usando una de micromatriz de ADNc que representa 23.040 genes. Estos datos no solo deben proporcionar información importante sobre carcinogenesis mamaria sino que deben facilitar la identificación de genes candidatos cuyos productos puedan servir como marcadores de diagnóstico y/o como dianas moleculares para el tratamiento de pacientes con cáncer de mama y proporcionar información clínicamente relevante.

La presente invención se basa, en parte, en el descubrimiento de cambios de patrones de expresión de ácidos nucleicos múltiples entre células epiteliales y carcinomas de pacientes con BRC. Las diferencias en la expresión de genes se identificaron usando un sistema de micromatriz de ADNc completo.

Los perfiles de expresión de genes de células de cáncer de los 81 BRC, incluyendo los 12 DCIS y los 69 IDC, se analizaron usando una micromatriz de ADNc que representaba 23,040 genes junto con microdissección por láser. Comparando patrones de expresión entre células cancerosas de pacientes a las que se había diagnosticado BRC y células epiteliales ductales normales totalmente seleccionadas con microdissección por láser, se identificaron 102

genes (mostrados en las tablas 3, 5 y 7) como normalmente regulados positivamente en células BRC y entre ellos se seleccionaron 100 genes como genes asociados al BRC de la presente descripción. De manera similar, también se identificaron 288 genes (mostrados en las tablas 4, 6 y 8) como regulados negativamente en células BRC. Además, se realizó la selección de marcadores moleculares candidatos que tenían la posibilidad de detectar proteínas relacionadas con cáncer en suero o esputo de pacientes, y se descubrieron algunas posibles dianas para el desarrollo de estrategias de supresión de señales en BRC humano. Entre estas, las tablas 3 y 4 proporcionan un listado de genes cuya expresión está modificada entre BRC, incluyendo DCIS e IDC y tejido normal. En la tabla 3 y 4, respectivamente, se muestran los genes normalmente regulados positiva o negativamente en DCIS y en IDC. Los genes que presentan expresión elevada o disminuida en transición desde DCIS a IDC se muestran en los listados de las tablas 5 y 6, respectivamente. Además, en los listados de las tablas 7 y 8, respectivamente, se muestran genes normalmente regulados positiva o negativamente, en IDC en comparación con tejido normal.

Los genes expresados diferencialmente en la presente memoria encuentran utilidad para el diagnóstico como marcadores de BRC y como dianas de genes BRC, cuya expresión puede modificarse para tratar o aliviar un síntoma de BRC. Como alternativa, los genes expresados diferencialmente entre DCIS e IDC identificados en la presente memoria encuentran utilidad para el diagnóstico como marcadores para diferenciar el IDC del DCIS y como dianas de genes BRC, cuya expresión puede modificarse para tratar o aliviar un síntoma de cáncer ductal invasivo, IDC.

En las tablas 3-8, se resumen los genes cuyo nivel de expresión se modula (es decir, se aumenta o disminuye) en pacientes con BRC y en la presente memoria se denominan, en su conjunto, "genes asociados a BRC", "ácidos nucleicos BRC" o "polinucleótidos BRC" y los polipéptidos codificados correspondientes se denominan "polipéptidos BRC" o "proteínas BRC." A menos que se indique otra cosa, "BRC" se refiere a cualquiera de las secuencias descritas en la presente memoria (por ejemplo, genes asociados a BRC que se muestran en los listados de las tablas 3-8). Los genes que se han descrito anteriormente de presentan junto con un número de acceso de una base de datos.

El cáncer de mama, BRC, puede diagnosticarse midiendo la expresión de los diversos genes en una muestra de células. De manera similar, la medición de la expresión de estos genes en respuesta a diversos agentes puede identificar agentes para el tratamiento del BRC.

La presente invención implica determinar (por ejemplo, medir) la expresión de al menos uno y hasta todos los genes asociados al BRC, mostrados en los listados de las tablas 3-8. Usando la información de secuencias proporcionada por las entradas de la base de datos del GenBank™ para secuencias conocidas, los genes asociados a BRC pueden detectarse y medirse usando técnicas bien conocidas por un experto habitual en la materia. Por ejemplo, para construir sondas para detectar secuencias de ARN que corresponden a genes asociados a BRC, pueden usarse secuencias dentro de las entradas de la base de datos de secuencias correspondientes a genes asociados a BRC, por ejemplo, en análisis de hibridación por transferencia Nortthern. Las sondas incluyen típicamente al menos 10, al menos 20, al menos 50, al menos 100, o al menos 200 nucleótidos de una secuencia de referencia. Como otro ejemplo, las secuencias pueden usarse para construir cebadores para amplificar específicamente el ácido nucleico BRC, por ejemplo, en métodos de detección basados en amplificación, tales como, reacción en cadena de polimerasa basada en transcripción inversa.

Después, el nivel de expresión de uno o más de los genes asociados a BRC en una población de células de ensayo, por ejemplo, una muestra de tejido derivada de un paciente, se compara con el nivel (o niveles) de expresión del mismo gen (o genes) en una población de referencia. La población de células de referencia incluye una o más células en las que se conoce el parámetro comparado, es decir, células de carcinoma ductal de mama (por ejemplo, células BRC) o células epiteliales ductales de mama normales (por ejemplo, células que no son de BRC).

El que un modelo de expresión de genes en una población de células de ensayo, en comparación con una población de células de referencia, indique BRC o una predisposición al mismo depende de la composición de la población de células de referencia. Por ejemplo, si la población de células de referencia está compuesta por células que no son de BRC, una similitud en el modelo de expresión de genes entre la población de células de ensayo y la población de células de referencia indica que la población de células de ensayo no es BCR. Por el contrario, si la población de células de referencia está constituida por células de BRC una similitud en el perfil de expresión de genes entre la población de células de ensayo y la población de células de referencia indica que la población de células de ensayo incluye células de BRC.

Se considera que un nivel de expresión de un gen marcador de BRC, en una población de células de ensayo, está "modificado" si este varía del nivel de expresión del gen marcador de BRC correspondiente en una población de células de referencia en más de 1,1, más de 1,5, más de 2,0, más de 5,0, más de 10,0 o más veces.

La expresión diferencial de genes entre una población de células de ensayo y una población de células de referencia

puede normalizarse con respecto a un ácido nucleico, por ejemplo, un gen constitutivo. Por ejemplo, un ácido nucleico de control es uno que se sabe que no difiere dependiendo del estado canceroso o no canceroso de la célula. El nivel de expresión de un ácido nucleico de control puede usarse para normalizar niveles de señal en las poblaciones de ensayo y de referencia. Los genes de control ejemplares incluyen, pero sin limitación, por ejemplo,  $\beta$ -actina, gliceraldehído 3- fosfato deshidrogenasa y proteína ribosomal P1.

La población de células de ensayo puede compararse con poblaciones de células de referencia múltiples. Cada una de las poblaciones de referencia múltiples puede diferir en el parámetro conocido. Por tanto, a población de células de ensayo puede compararse con una primera población de células de referencia que se sabe que contiene, por ejemplo, células de BRC, así como con una segunda población de referencia que se sabe que contiene, por ejemplo, células que no son de BRC (células normales). La célula de ensayo puede estar incluida en un tipo de tejido o en una muestra de células de un sujeto que se sabe, o se sospecha, que contiene células de BRC.

La célula de ensayo se obtiene de un tejido o fluido corporal, por ejemplo, un fluido biológico (como sangre o esputo, por ejemplo). Por ejemplo, la célula de ensayo puede purificarse de tejido de mama. Preferentemente, la población de las células de ensayo comprende una célula epitelial. Preferentemente, la célula epitelial es de un tejido que se sabe, o se sospecha, que es un carcinoma ductal de mama.

En la población de células de referencia, las células deben derivar de un tipo de tejido similar al de las células de ensayo. Opcionalmente, la población de células de referencia es una línea celular, por ejemplo, una línea celular de BRC (es decir, un control positivo) o una línea celular no BRC normal (es decir, un control negativo). Como alternativa, la población de células de control puede derivar de una base de datos de información molecular derivada de células en las que se conoce el parámetro o estado ensayado.

El sujeto es preferentemente un mamífero. Los mamíferos ejemplares incluyen, pero sin limitación, por ejemplo, un ser humano, un primate no humano, ratón, rata, perro, gato, caballo o vaca.

La expresión de los genes descritos en la presente memoria puede determinarse a nivel de la proteína o del ácido nucleico, usando métodos conocidos en la técnica. Por ejemplo, para determinar la expresión de genes puede usarse el análisis de hibridación de Northern, usando sondas que reconocen específicamente una o más de estas secuencias de ácido nucleico. Como alternativa, la expresión de genes puede medirse usando ensayos de PCR basados en transcripción inversa, usando, por ejemplo, cebadores específicos para las secuencias de genes expresadas diferencialmente. La expresión también puede determinarse a nivel de la proteína, es decir, midiendo el nivel de un polipéptido codificado por un gen descrito en la presente memoria, o la actividad biológica del mismo. Estos métodos se conocen bien en la técnica e incluyen, pero sin limitación, por ejemplo, inmunoensayos que usan anticuerpos contra proteínas codificadas por los genes. Generalmente, las actividades biológicas de las proteínas codificadas por los genes se conocen bien.

#### Diagnóstico del cáncer de mama:

En el contexto de la presente descripción, el BRC se diagnostica midiendo el nivel de expresión de uno o más ácidos nucleicos de BRC de una población de células de ensayo (es decir, una muestra biológica derivada de un paciente). Preferentemente, la población de células de ensayo contiene una célula epitelial, por ejemplo, una célula obtenida de tejido mamario. La expresión de genes también puede medirse a partir de sangre o de otros fluidos corporales tales como orina. Para medir niveles de proteína pueden usarse otras muestras biológicas. Por ejemplo, el nivel de proteína en sangre o suero derivado de un sujeto a diagnosticar puede medirse por inmunoensayo u otro ensayo biológico convencional.

La expresión de uno o más genes asociados a BRC, por ejemplo, *de los genes mostrados en los listados de las tablas 3-8*, se determina en la célula de ensayo o muestra biológica y se compara con el nivel de expresión de control normal asociado con uno o más genes asociados a BRC ensayados. Un nivel de control normal es un perfil de expresión de un gen asociado a BRC típicamente encontrado en una población que se sabe que no padece BRC. Una modificación (por ejemplo, un aumento o disminución) en el nivel de expresión en la muestra de tejido derivada del paciente de uno o más genes asociados a BRC indica que el sujeto padece o corre el riesgo de desarrollar un BRC. Por ejemplo, un aumento en la expresión de uno o más genes asociados a BRC regulados positivamente mostrados en los listados de las tablas 3, 5, y 7 en la población de ensayo, en comparación con el nivel de control normal, indica que el sujeto padece o corre el riesgo de desarrollar BRC. Por otro lado, una disminución en la expresión de uno o más genes asociados a BRC regulados negativamente mostrados en los listados de las tablas 4, 6 y 8 en la población de ensayo en comparación con el nivel de control normal indica que el sujeto padece, o corre el riesgo de desarrollar, BRC.

La modificación de uno o más de los genes asociados a BRC en la población de ensayo en comparación con el nivel de control indica que el sujeto padece o corre el riesgo de desarrollar BRC. Por ejemplo, la modificación de al menos 1%, al menos 5%, al menos 25%, al menos 50%, al menos 60%, al menos 80%, al menos 90% o más del panel de genes asociados a BRC (genes mostrados en los listados de las tablas 3-8) indica que el sujeto padece o corre el riesgo de desarrollar BRC.

#### Identificación de la diferenciación histopatológica del BRC:

También se describe un método para identificar la diferenciación histopatológica del BRC en un sujeto, comprendiendo el método las etapas de:

- (a) detectar un nivel de expresión de uno o más genes marcadores en una muestra de tejido extraída del sujeto a ensayar, en el que se seleccionan uno o más genes marcadores del grupo que consiste en los genes mostrados en los listados de las tablas .1 y 10; y
- (b) comparar el nivel de expresión detectado de uno o más genes marcadores con un nivel de expresión asociado con un caso bien diferenciado y con un caso mal diferenciado;
- (c) de manera que cuando el nivel de expresión detectado de uno o más genes marcadores es similar al del caso bien diferenciado, la muestra de tejido se determina como bien diferenciada y cuando el nivel de expresión detectado de uno o más genes marcadores es similar al del caso mal diferenciado, la muestra de tejido se determina como mal diferenciada.

Como se describe en la presente memoria, el gen (o genes) marcador para identificar la diferenciación histopatológica del BRC puede ser al menos un gen seleccionado del grupo que consiste en 231 genes mostrados en las Tablas 1 y 10. En la técnica se conocen las secuencias de nucleótidos de los genes y por tanto las secuencias de aminoácidos codificadas. Para los Números de Acceso de los genes, véanse las Tablas 1 y 10.

Identificación de agentes que inhiben o mejoran la expresión de genes asociados a BRC:

Un agente que inhibe la expresión de un gen asociado a BRC o la actividad de su producto génico puede identificarse poniendo en contacto una población de células de ensayo, que expresan un gen regulado positivamente asociado al BRC, con un agente de ensayo y después determinar el nivel de expresión del gen asociado a BRC o la actividad de su producto génico. Una disminución del nivel de expresión del gen asociado a BRC o del nivel de actividad de su producto génico en presencia del agente, en comparación con el nivel de expresión o actividad en ausencia del agente de ensayo indica que el agente es un inhibidor de un gen regulado positivamente asociado a BRC y útil inhibiendo el BRC.

Como alternativa, un agente que mejora la expresión de un gen regulado negativamente asociado a BRC o la actividad de su producto génico puede identificarse poniendo en contacto una población de células de ensayo que expresa un gen asociado a BRC con un agente de ensayo y después determinar el nivel de expresión o actividad del gen regulado negativamente asociado a BRC. Un aumento del nivel de expresión del gen asociado a BRC o del nivel de actividad de su producto génico en comparación con el nivel de expresión o actividad en ausencia del agente de ensayo indica que el agente de ensayo aumenta la expresión del gen regulado negativamente asociado a BRC o la actividad de su producto génico.

La población de células de ensayo puede ser cualquier célula que exprese los genes asociados a BRC. Por ejemplo, la población de células de ensayo puede contener una célula epitelial, tal como una célula derivada de tejido mamario. Además, la célula de ensayo puede ser una línea de células inmortalizada derivada de una célula de carcinoma. Como alternativa, la célula de ensayo puede ser una célula que se ha transfectado con un gen asociado a BRC o que se ha transfectado con una secuencia reguladora (por ejemplo, una secuencia promotora) de un gen asociado a BRC unido operativamente a un gen indicador.

Valoración de la eficacia del tratamiento del BRC en un sujeto:

Los genes asociados a BRC expresados diferencialmente identificados en la presente memoria también permiten controlar el ciclo de tratamiento del BRC. En este método, se proporciona una población de células de ensayo de un sujeto sometido a tratamiento para BRC. Si se desea, las poblaciones de células de ensayo se obtienen del sujeto a diversos puntos de tiempo, antes, durante y/o después del tratamiento. Despues, se determina la expresión de uno o más genes asociados a BRC en la población de células y se compara con una población de células de referencia que incluye células cuyo estado BRC se conoce. En el contexto de la presente descripción, las células de referencia no deben haberse expuesto al tratamiento de interés.

Si la población de células de referencia contiene células que no son BRC, una similitud en la expresión de un gen asociado a BRC en la población de células de ensayo y en la población de células de referencia indica que el tratamiento de interés es eficaz. Sin embargo, una diferencia en la expresión de un gen asociado a BRC en la población de ensayo y en la población de células de referencia de control normal indica un resultado o pronóstico clínico favorable. De manera similar, si la población de células de referencia contiene células BRC, una diferencia entre la expresión de un gen asociado a BRC en la población de células de ensayo y en la población de células de referencia indica que el tratamiento de interés es eficaz, mientras que una similitud en la expresión de un gen asociado a BRC en la población de ensayo y en una población de células de referencia de control del cancer indica un resultado o pronóstico clínico menos favorable.

Adicionalmente, el nivel de expresión de uno o más genes asociados a BRC determinado en una muestra biológica derivada de un sujeto obtenida después del tratamiento (es decir, niveles posteriores al tratamiento) puede compararse con el nivel de expresión de uno o más genes asociados a BRC determinado en una muestra biológica

derivada del sujeto obtenida antes de comenzar el tratamiento (es decir, niveles previos al tratamiento). Si el gen asociado a BRC es un gen regulado positivamente, una disminución del nivel de expresión en una muestra posterior al tratamiento indica que el tratamiento de interés es eficaz mientras que un aumento o mantenimiento del nivel de expresión en la muestra posterior al tratamiento indica un resultado o pronóstico clínico menos favorable. A la inversa, si el gen asociado a BRC es un gen regulado negativamente, un aumento del nivel de expresión en una muestra posterior al tratamiento puede indicar que el tratamiento de interés es eficaz mientras que una disminución o mantenimiento del nivel de expresión en la muestra posterior al tratamiento indica un resultado o pronóstico clínico menos favorable.

Como se usa en la presente memoria, el término "eficaz" indica que el tratamiento conduce a una reducción de la expresión de un gen regulado positivamente de manera patológica, a un aumento de la expresión de un gen regulado negativamente de manera patológica o a una disminución del tamaño, frecuencia, o posible metástasis de carcinoma ductal de mama en un sujeto. Cuando un tratamiento de interés se aplica profilácticamente, el término "eficaz" significa que el tratamiento retrasa o previene la formación de un tumor de mama o retrasa, previene o alivia un síntoma de BRC clínico. La valoración de tumores de mama puede realizarse usando protocolos clínicos convencionales.

Adicionalmente, la eficacia puede determinarse en asociación con cualquier método conocido para diagnosticar o tratar BRC. El BRC puede diagnosticarse, por ejemplo, identificando anomalías sintomáticas, por ejemplo, pérdida de peso, dolor abdominal, dorsalgia, anorexia, náuseas, vómitos y malestar, debilidad e ictericia generalizados.

Selección de un agente terapéutico, que sea apropiado para un individuo particular, para el tratamiento de BRC:

Las diferencias en la composición genética de individuos pueden dar como resultado diferencias en sus capacidades relativas para metabolizar diversos fármacos. Un agente que se metaboliza en un sujeto para actuar como un agente anti-BRC puede manifestarse por sí mismo induciendo un cambio en un patrón de expresión de genes en las células del sujeto a partir de esta característica de un estado canceroso con respecto a un patrón de expresión de genes característico de un estado no canceroso. Por consiguiente, los genes asociados a BRC expresados diferencialmente descritos en la presente memoria permiten ensayar un supuesto inhibidor terapéutico o profiláctico de BCR en una población de células de ensayo de un sujeto seleccionado para determinar si el agente es un inhibidor adecuado de BCR en el sujeto.

Para identificar un inhibidor de BRC que sea apropiado para un sujeto específico, se expone una población células de ensayo del sujeto a un agente terapéutico y se determina la expresión de uno o más de los genes asociados a BRC mostrados en los listados de las tablas 3 y en 8.

En el contexto del método de la presente descripción, la población de células de ensayo contiene una célula de BCR que expresa un gen asociado a BRC. Preferentemente, la célula de ensayo es una célula epitelial. Por ejemplo, una población de células de ensayo puede incubarse en presencia de un agente candidato y el modelo de expresión génica de la población de células de ensayo puede medirse y compararse con uno o más perfiles de referencia, por ejemplo, un perfil de expresión de referencia BRC o un perfil de expresión de referencia no BRC.

Una disminución en la expresión de uno o más de los genes asociados a BRC mostrados en los listados de las tablas 3, 5 y 7 o un aumento en la expresión de uno o más de los genes asociados a BRC mostrados en los listados de las tablas 4, 6 y 8 en una población de células de ensayo con respecto a una población de células de referencia que contiene BRC indica que el agente tiene potencial terapéutico.

En el contexto de la presente descripción, el agente de ensayo puede ser cualquier compuesto o composición. Los agentes de ensayo ejemplares incluyen, pero sin limitación, agentes inmunomoduladores.

Ensayos de exploración para identificar agentes terapéuticos

Los genes asociados a BRC expresados diferencialmente descritos en la presente memoria también pueden usarse para identificar agentes terapéuticos candidatos para el tratamiento del BRC. El método como se describe en la presente memoria, implica explorar un agente terapéutico candidato para determinar si este puede convertir un perfil de expresión de uno o más genes asociados a BRC mostrados en los listados de las tablas 3-8 característico de un estado BRC en un patrón de expresión génico característico de un estado no BRC.

En el presente método, se expone una célula a un agente de ensayo o a una pluralidad de agentes de ensayo (secuencialmente o en combinación) y se mide la expresión de uno o más de los genes asociados a BRC, mostrados en los listados de las tablas 3-8, en la célula. El perfil de expresión del gen (o genes) asociado a BRC ensayado en la población de ensayo se compara con el nivel de expresión del mismo gen (o genes) asociado a BRC en una población de células de referencia que no se expone al agente de ensayo.

Un agente que puede estimular la expresión de un gen infraexpresado o suprimir la expresión de un gen sobreexpresado tiene posible beneficio clínico. Además, dichos agentes pueden ensayarse para determinar su capacidad para prevenir el crecimiento de carcinoma ductal de mama en animales o sujetos de ensayo.

Adicionalmente, se describen métodos para explorar agentes candidatos que actúan sobre posibles dianas en el tratamiento de BRC. Como se describe con detalle anteriormente, controlando los niveles de expresión de genes marcadores o las actividades de sus productos génicos, se puede controlar la aparición y progresión de BRC. Por lo tanto, pueden identificarse agentes candidatos, que actúen sobre posibles dianas en el tratamiento de BRC,

5 mediante métodos de exploración que usan dichos niveles de expresión y actividades como índices del estado canceroso o no canceroso. En el compuesto de la presente invención, dicha exploración puede comprender, por ejemplo, las siguientes etapas:

- a) poner en contacto un compuesto de ensayo con un polipéptido codificado por un polinucleótido seleccionado del grupo que consiste en los genes mostrados en las listas de las tablas 3, 4, 5, 6, 7 u 8;
- 10 b) detectar la actividad de unión entre el polipéptido y el compuesto de ensayo; y
- c) seleccionar el compuesto de ensayo que se une al polipéptido.

Como alternativa, el método de exploración de la presente invención puede comprender las siguientes etapas:

- a) poner en contacto un compuesto candidato con una célula que exprese uno o más genes marcadores, en el que el gen (o genes) marcador se selecciona del grupo que consisten en los genes mostrados en las listas de las tablas 3, 4, 5, 6, 7 u 8; y
- 15 b) seleccionar el compuesto candidato que reduce el nivel de expresión de uno o más genes marcadores seleccionados del grupo que consiste en los genes mostrados en las listas de las tablas 3, 5 y 7 o elevar el nivel de expresión de uno o más genes marcadores seleccionados del grupo que consiste en los genes mostrados en los listados de las tablas 4, 6 y 8.

20 Las células que expresan un gen marcador incluyen, por ejemplo, líneas celulares establecidas para BRC, dichas células pueden usarse para la exploración anterior de la presente invención.

Como alternativa, el método de exploración de la presente invención puede comprender las siguientes etapas:

- a) poner en contacto un compuesto de ensayo con un polipéptido codificado por un polinucleótido seleccionado del grupo que consiste en los genes mostrados en los listados de las tablas 3, 4, 5, 6, 7 u 8;
- 25 b) detectar la actividad biológica del polipéptido de la etapa (a); y
- c) seleccionar un compuesto que suprima la actividad biológica del polipéptido codificado por el polinucleótido seleccionado del grupo que consiste en los genes mostrados en los listados de las tablas 3, 4 y 5 en comparación con la actividad biológica detectada en ausencia del compuesto de ensayo o potenciar la actividad biológica del polipéptido codificado por el polinucleótido seleccionado del grupo que consiste en los genes mostrados en los listados de las tablas 4, 6 y 8 en comparación con la actividad biológica detectada en ausencia del compuesto del ensayo.

30 Usando la secuencia de nucleótidos del gen marcador puede obtenerse una proteína como una proteína recombinante para su uso en el método de exploración de la presente invención. Basándose en la información relativa al gen marcador y a su proteína codificada, un experto en la materia puede seleccionar cualquier actividad biológica de la proteína como una clasificación para explorar y cualquier método de medición adecuado para ensayar para determinar la actividad biológica seleccionada.

Como alternativa, el método de exploración de la presente invención puede comprender las siguientes etapas:

- a) poner en contacto un compuesto candidato con una célula, en la que se ha introducido un vector, que comprende la región reguladora transcripcional de uno o más genes marcadores y un gen indicador que se expresa bajo el control de la región reguladora transcripcional, en el uno o más genes marcadores se seleccionan entre el grupo que consiste en los genes mostrados en los listados de las tablas 3, 4, 5, 6, 7 u 8;
- 40 b) medir la expresión o actividad de dicho gen indicador; y
- c) seleccionar el compuesto candidato que reduzca la expresión o actividad de dicho gen indicador en el que dicho gen marcador es un gen marcador regulado positivamente seleccionado entre grupo que consiste en los genes mostrados en los listados de las tablas 3, 5 y 7, o que mejore el nivel de expresión de dicho gen indicador en el que dicho gen marcador es un gen marcador regulado negativamente seleccionado entre el grupo que consiste en los genes mostrados en los listados de las tablas 4, 6 y 8, en comparación con un control.

45 En la técnica se conocen bien genes indicadores y células hospedadoras adecuados. Usando la región reguladora transcripcional de un gen marcador puede prepararse una construcción indicadora adecuada para el método de exploración de la presente invención. Cuando un experto en la materia conoce la región reguladora transcripcional del gen marcador, puede prepararse una construcción indicadora usando la información de la secuencia previa.

Cuando se desconoce la región reguladora transcripcional del gen marcador, puede aislarse un segmento de nucleótidos que contenga la región reguladora transcripcional de una genoteca basándose en la información de la secuencia de nucleótidos del gen marcador.

5 Un compuesto aislado por la exploración sirve como un candidato para el desarrollo de fármacos que inhiban la expresión del gen marcador o la actividad de la proteína codificada por el gen marcador y puede aplicarse al tratamiento o prevención del cáncer de mama.

Además, también se incluyen compuestos en los que una parte de la estructura del compuesto que inhibe la actividad de proteínas codificadas por genes marcadores se convierte por adición, delección y/o sustitución, como los compuestos que pueden obtenerse mediante el método de exploración de la presente invención.

10 Cuando un compuesto aislado por el método como se describe en la presente memoria se administra como un compuesto farmacéutico en seres humanos y en otros animales, tales como ratones, ratas, cobayas, conejos, gatos, perros, ovejas, cerdos, vacas, monos, babuinos y chimpancés, el compuesto aislado puede administrarse directamente o puede formularse en una forma de dosificación usando métodos de preparación farmacéuticos conocidos. Por ejemplo, de acuerdo con la necesidad, los fármacos pueden administrarse por vía oral, como comprimidos recubiertos con azúcares, cápsulas, elixires y microcápsulas, o por vía no oral, en forma de inyecciones de soluciones o suspensiones estériles con agua o con cualquier otro líquido farmacéuticamente aceptable. Por ejemplo, los compuestos pueden mezclarse con vehículos o medios farmacéuticamente aceptables, específicamente, agua esterilizada, solución salina fisiológica, aceites vegetales, emulsionantes, agentes de suspensión, tensioactivos, estabilizantes, saporíferos, excipientes, vehículos, conservantes, aglutinantes, y similares, 15 en una forma de dosificación unitaria necesaria para la implementación farmacológica generalmente aceptada. La cantidad de ingrediente activo contenido en una preparación de este tipo hace una dosificación adecuada dentro del intervalo indicado obtenido.

20 Los ejemplos de aditivos que pueden mezclarse en los comprimidos y las capsulas incluyen, pero sin imitación aglutinantes, tales como gelatina, almidón de maíz, goma de tragacanto y goma arábica; excipientes, tales como celulosa cristalina; agentes expansores, tales como almidón de maíz, gelatina y ácido algínico; lubricantes, tales como estearato de magnesio; edulcorantes, tal como sacarosa lactosa o sacarina; saporíferos tales como menta, aceite de *Gaultheria adenothrix* y cereza.

25 Cuando la forma de dosificación unitaria es una cápsula, en los ingredientes anteriores puede incluirse además un vehículo líquido, tal como un aceite. Las composiciones estériles para inyección pueden formularse siguiendo 30 implementaciones farmacológicas normales usando vehículos, tales como agua destilada, adecuada para inyección.

Como soluciones acuosas para inyección pueden usarse solución salina fisiológica, glucosa y otros líquidos isotónicos, incluyendo adyuvantes, tales como D-sorbitol, D-mannosa, D-mannitol y cloruro de sodio. Estas pueden usarse junto con solubilizantes adecuados, tales como alcohol, por ejemplo, etanol; poli alcoholes, tales como propilenglicol y polietilenglicol; y tensioactivos no iónicos, tales como Polisorbato 80 (TM) y HCO-50.

35 El aceite de sésamo o de semilla de soja que puede usarse como un líquido oleaginoso, puede usarse junto con benzoato de bencilo o alcohol bencílico como un solubilizante y puede formularse con un tampón, tal como tampón fosfato y tampón de acetato de sodio; con un analgésico, tal como clorhidrato de procaína; con un estabilizante, tal como alcohol bencílico y fenol; y/o con un antioxidante. Una inyección preparada puede cargarse en una ampolla adecuada.

40 Para administrar a pacientes la composición farmacéutica de la presente invención pueden usarse métodos bien conocidos por los expertos en la materia, por ejemplo, como una inyección intraarterial, intravenosa o percutánea o como una administración intranasal transbronquial, intramuscular u oral. La dosificación y el método de administración varían de acuerdo con el peso corporal, con la edad de un paciente y con el método de administración; sin embargo, un experto en la materia puede seleccionar rutinariamente un método de administración adecuado. Si un ADN puede codificar dicho compuesto, el ADN puede insertarse en un vector para terapia génica y el vector puede administrarse a un paciente para realizar la terapia. La dosificación y el método de administración varía de acuerdo con el peso corporal, la edad y los síntomas del paciente; sin embargo, un experto en la materia puede seleccionarlos adecuadamente.

45 Por ejemplo, aunque la dosis de un compuesto que se une a una proteína de la presente invención y regule su actividad dependa de los síntomas, la dosis es generalmente de aproximadamente 0,1 mg a aproximadamente 100 mg por día, preferentemente de aproximadamente 1,0 mg a aproximadamente 50 mg por día y más preferentemente de aproximadamente 1,0 mg a aproximadamente 20 mg por día, cuando se administra por vía oral a un ser humano adulto normal (60 Kg de peso).

50 Cuando el compuesto se administra por vía parenteral, en forma de una inyección a un ser humano adulto normal (60 Kg de peso), aunque existan algunas diferencias de acuerdo con el paciente, órgano diana, síntomas y método de administración, es conveniente inyectar por vía intravenosa una dosis de aproximadamente 0,01 mg a aproximadamente 30 mg por día, preferentemente de aproximadamente 0,1 a aproximadamente 20 mg por día y más preferentemente de aproximadamente 0,1 a aproximadamente 10 mg por día. En el caso de otros animales,

la cantidad de dosificación apropiada puede calcularse rutinariamente por conversión a 60 Kg de peso corporal.

Ensayos de exploración para identificar agentes terapéuticos para la metástasis del cáncer de mama:

Se describen moléculas diana para el tratamiento o prevención de la metástasis del cáncer de mama. Los ensayos de exploración para la metástasis del BRC, como se describe en la presente memoria, pueden realizarse de acuerdo con el método para BRC descrito anteriormente, usando genes marcadores asociados con la metástasis del BRC.

En la presente descripción, los genes marcadores seleccionados entre el grupo que consiste en los genes mostrados en el listado de la tabla 11 son útiles para la exploración. Los 34 genes mostrados en la Tabla están asociados con metástasis en nódulos linfáticos. Entre estos genes, 25 genes (+) se regularon más o menos positivamente y 9 genes (-) se regularon negativamente en tumores positivos nodales (Tabla 11 y Figura 10). Un agente que suprime la expresión de uno o más genes regulados positivamente o la actividad de sus productos génicos, obtenidos por la presente descripción, es útil para el tratamiento o prevención de BRC con metástasis en los nódulos linfáticos. Como alternativa, un agente que potencia la expresión de uno o más genes regulados negativamente o la actividad de sus productos génicos, obtenidos por la presente descripción, también es útil para el tratamiento o prevención de BRC con metástasis en los nódulos linfáticos.

En la presente descripción, el agente que regula un nivel de expresión de los genes mostrados en los listados de la tabla 11, puede identificarse de la misma manera para identificar agentes que inhiban o potencien la expresión de genes asociados a BRC. Como alternativa, el agente que regula la actividad de sus productos génicos también puede identificarse de la misma manera para identificar agentes que inhiban o potencien productos génicos asociados a BRC.

20 Valoración del pronóstico de un sujeto con cáncer de mama:

También se describe un método para valorar el pronóstico de un sujeto con BRC que incluye la etapa de comparar la expresión de uno o más genes asociados a BRC en una población de células de ensayo con respecto a la expresión de los mismos genes asociados a BRC en una población de células de referencia derivada de pacientes sobre un espectro de fases de enfermedad. El pronóstico del sujeto puede valorarse comparando la expresión génica de uno o más genes asociados a BRC en la población de células de ensayo y la población (o poblaciones) de células de referencia o comprando el modelo de expresión génica con el tiempo en poblaciones de células de ensayo derivadas del sujeto.

Por ejemplo, un aumento en la expresión de uno o más genes asociados a BRC regulados positivamente, tales como los mostrados en los listados de las tablas 3, 5, 7, en comparación con un control normal o una disminución en la expresión de uno o más genes asociados a BRC regulados negativamente, tales como los mostrados en los listados de las tablas 4, 6 u 8, en comparación con un control normal, indica pronóstico menos favorable. A la inversa, una similitud en la expresión de uno o más genes asociados a BRC mostrados en los listados de las tablas 3-8 en comparación con un control normal indica un pronóstico más favorable para el sujeto. Preferentemente, el pronóstico de un sujeto puede valorarse comparando el perfil de expresión del gen seleccionado del grupo que consiste en los genes mostrados en los listados de las tablas 3, 4, 5, 6, 7 y 8. Para comparar el perfil de expresión puede usarse una puntuación de clasificación (CS).

Kits:

También se describe un reactivo de detección de BRC, por ejemplo, un ácido nucleico que se une específicamente a o identifica uno o más ácidos nucleicos BRC, tales como secuencias de oligonucleótidos que son complementarias a una parte de un ácido nucleico BRC, o un anticuerpo que se une a una o más proteínas codificadas por un ácido nucleico BRC. Los reactivos de detección pueden envasarse juntos en forma de un kit. Por ejemplo, los reactivos de detección pueden envasarse en envases individuales, por ejemplo, un ácido nucleico o anticuerpo (unido a una matriz sólida o envasado por separado con reactivos para unirlos a la matriz), un reactivo de control (positivo y/o negativo), y/o un marcador detectable. En el kit también pueden incluirse instrucciones (por ejemplo, escritas, grabadas, en VCR, CD-ROM, etc.) para realizar el ensayo. El formato del kit de ensayo puede ser una hibridación de Northern o un ELISA de tipo sándwich, ambos conocidos en la técnica.

Por ejemplo, un reactivo de detección de BRC puede inmovilizarse en una matriz sólida, tal como una tira porosa, para formar al menos un sitio de detección de BRC. La región de medición o de detección de la tira porosa puede incluir una pluralidad de sitios, conteniendo cada uno un ácido nucleico. Una tira de ensayo también puede contener sitios para controles negativos y/o positivos. Como alternativa, los sitios de control pueden localizarse en una tira separada de la tira de ensayo. Opcionalmente, los diferentes sitios de detección pueden contener cantidades diferentes de ácidos nucleicos inmovilizados, es decir, una cantidad mayor en el primer sitio de detección y menores cantidades en sitios posteriores. Después de la adicción de la muestra de ensayo, el número de sitios que muestran una señal detectable proporciona una indicación cuantitativa de la cantidad de BRC presente en la muestra. Los sitios de detección pueden configurarse de cualquier forma adecuadamente detectable y tienen típicamente forma de una barra o punto que se extiende a lo ancho de una tira de ensayo. Como alternativa el kit puede contener una matriz de substrato de ácido nucleico que comprenda uno o más ácidos nucleicos. Los ácidos nucleicos en la matriz identifican específicamente una o más secuencias de ácidos nucleicos representadas por los genes asociados a

5 BRC mostrados en las listas de las tablas 3-8. La expresión de 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 40 ó 50 o más de los ácidos nucleicos representados por los genes asociados a BRC mostrados en los listados de las tablas 3-8 pueden identificarse en virtud del nivel de unión a una tira o chip de ensayo micromatriz. La matriz de substrato puede ser, por ejemplo, sobre un substrato sólido tal, como un "chip" descrito en la patente de Estados Unidos Nº 5.744.305, cuyo contenido se incorpora por referencia en la presente memoria en su totalidad.

#### Matrices y pluralidades:

10 También se describe una matriz de substrato de ácido nucleico que comprende uno o más ácidos nucleicos. Los ácidos nucleicos en la matriz corresponden específicamente a una o más secuencias de ácidos nucleicos representadas por los genes asociados a BRC mostrados en los listados de las tablas 3-8. El nivel de expresión de 2 ,3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 40 ó 50 ó más de los ácidos nucleicos representados por los genes asociados a BRC mostrados en los listados de las tablas 3-8 pueden identificarse detectando la unión del ácido nucleico a la matriz.

15 También se describe una pluralidad de ácidos nucleicos (es decir, una mezcla de dos o más ácidos nucleicos) aislados. Los ácidos nucleicos pueden estar en una fase líquida o en una fase sólida, por ejemplo, inmovilizados en un soporte sólido tal como una membrana de nitrocelulosa. La pluralidad incluye uno o más de los ácidos nucleicos representados por los genes asociados a BRC mostrados en los listados de las tablas 3-8. En diversas realizaciones, la pluralidad incluye 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 40 ó 50 o más de los ácidos nucleicos representados por los genes asociados a BRC mostrados en los listados de las tablas 3-8.

#### Métodos de inhibición del cáncer de mama:

20 También se describe un método para tratar o aliviar un síntoma de BRC en un sujeto disminuyendo la expresión de uno o más de los genes asociados a BRC mostrados en los listados de las tablas 3, 5 y 7 (o la actividad de sus productos génicos) o aumentando la expresión de uno o más de los genes asociados a BRC mostrados en los listados de las tablas 4, 6 y 8 (o la actividad de sus productos génicos). Los compuestos terapéuticos adecuados pueden administrarse profiláctica o terapéuticamente a un sujeto que padece o que está en riesgo (o es susceptible a) de desarrollar BRC. Dichos sujetos pueden identificarse usando métodos clínicos convencionales o detectando un nivel de expresión anómalo de uno o más de los genes asociados a BRC mostrados en los listados de las tablas 3-8 o una actividad anómala de sus productos génicos. En el contexto como se describe en la presente memoria, los agentes terapéuticos adecuados incluyen, por ejemplo, inhibidores de la regulación del ciclo celular, de la proliferación celular y de la actividad proteína quinasa.

25 30 El método terapéutico como se describe en la presente memoria incluye la etapa de aumentar la expresión, la función o ambas cosas, de uno o más productos génicos de genes cuya expresión está disminuida (genes "regulados negativamente" o "infraexpresados") en una célula BRC con respecto células normales del mismo tipo de tejido a partir del cual derivan las células BRC. En estos métodos, el sujeto se trata con una cantidad eficaz de un compuesto que aumenta la cantidad de uno o más de los genes infraexpresados (regulados negativamente) en el sujeto. La administración puede ser sistémica u oral. Los compuestos terapéuticos adecuados incluyen un producto polipeptídico de un gen infraexpresado, un fragmento biológicamente activo del mismo, un ácido nucleico que codifica un gen infraexpresado y que tiene elementos de control de expresión que permiten la expresión en las células BRC; por ejemplo, un agente que aumenta el nivel de expresión de dicho gen endógeno con respecto a las células BRC (es decir, que regulan positivamente la expresión del gen o genes infraexpresados). La administración de dichos compuestos contrarresta los efectos de un gen o genes anómalamente infraexpresados en las células mamarias de sujetos y puede mejorar el estado clínico del sujeto.

35 40 Como alternativa, el método terapéutico como se describe en la presente memoria puede incluir la etapa de disminuir la expresión, la función, o ambas cosas, de uno o más productos génicos de genes cuya expresión está anómalamente aumentada (genes "regulados positivamente" o "sobreexpresados") en células mamarias. La expresión puede inhibirse de cualquiera de las diversas maneras conocidas en la técnica. Por ejemplo, la expresión puede inhibirse administrando al sujeto un ácido nucleico que inhibía o antagonizó la expresión del gen o genes sobreexpresados, por ejemplo, un oligonucleótido antisentido o ARN de interferencia pequeño que interrumpa la expresión del gen o genes sobreexpresados.

#### Ácidos nucleicos antisentido:

45 50 Como se ha indicado anteriormente, los ácidos nucleicos antisentido, que corresponden a la secuencia de nucleótidos de los genes asociados a BRC mostrados en los listados de las tablas 3, 5 y 7 pueden usarse para reducir el nivel de expresión de los genes. Los ácidos nucleicos antisentido correspondientes a los genes asociados a BRC mostrados en los listados de las tablas 3, 5 y 7 que están regulados positivamente en cáncer de mama son útiles para el tratamiento del cáncer de mama. Específicamente, los ácidos nucleicos antisentido como se describe en la presente memoria pueden actuar uniéndose a los genes asociados a BRC mostrados en los listados de tablas 3, 5 y 7 o a los ARNm correspondientes de los mismos, por lo tanto inhibiendo la transcripción o traducción de los genes, promoviendo la degradación de los ARNm y/o inhibiendo la expresión de proteínas codificadas por los genes asociados a BRC mostrados en los listados de las tablas 3, 5 y 7, inhibiendo de esta manera la función de las

proteínas. La expresión "ácidos nucleicos antisentido" como se usa en la presente memoria incluye nucleótidos que son totalmente complementarios con la secuencia diana y los que tienen un emparejamiento erróneo de uno o más nucleótidos, siempre y cuando los ácidos nucleicos antisentido puedan hibridarse específicamente con las secuencias diana. Por ejemplo, los ácidos nucleicos antisentido de la presente invención incluyen polinucleótidos que tienen una homología de al menos 70% o mayor, preferentemente al menos 80% o mayor, más preferentemente al menos 90% o mayor, incluso más preferentemente al menos 95% o mayor sobre un tramo de al menos 15 nucleótidos consecutivos. Para determinar la homología pueden usarse algoritmos conocidos en la técnica.

El acido nucleico antisentido como se describe en la presente memoria actúa en células que producen las proteínas codificadas por los genes marcadores asociados a BRC uniéndose a los ADN o a los ARNm que codifican las proteínas, inhibiendo su transcripción o traducción, promoviendo la degradación de los ARNm e inhibiendo la expresión de las proteínas, dando de esta manera como resultado la inhibición de la función de las proteínas.

Un acido nucleico antisentido, como se describe en la presente memoria, puede prepararse en una preparación externa, tal como un linimento o una cataplasma, mezclándolo con un material base adecuado que es inactivo contra el ácido nucleico.

Además, según se necesite, los ácidos nucleicos antisentido, como se describe en la presente memoria, pueden formularse en comprimidos, polvos, gránulos, cápsulas, cápsulas de liposomas, inyecciones, soluciones, gotas nasales y agentes liofilizados mediante la adición de excipientes, agentes isotónicos, solubilizantes, estabilizantes, conservantes, analgésicos y similares. Estos pueden prepararse siguiendo métodos conocidos.

Los ácidos nucleicos antisentido, como se describen en la presente memoria, pueden administrarse al paciente por aplicaron directa en el sitio enfermo o por inyección en un vaso sanguíneo de manera que alcance el sitio de la enfermedad. También puede usarse un medio de soporte antisentido que aumente la durabilidad y la permeabilidad de membrana. Los ejemplos incluyen, pero sin limitación, liposomas, poli-L-lisina, lípidos, colesterol, lipofectina o derivados de los estos.

La dosificación del derivado del ácido nucleico antisentido, como se describe en la presente memoria, puede ajustarse adecuadamente de acuerdo con la afección del paciente y usarse en cantidades deseadas. Por ejemplo, puede administrarse un intervalo de dosificación de 0,1 a 100 mg/kg, preferentemente de 0,1 a 50 mg/kg. Los ácidos nucleicos antisentido, como se describe en la presente memoria, inhiben la expresión de una proteína como se describe en la presente memoria, y por lo tanto son útiles para suprimir la actividad biológica de la proteína como se describe en la presente memoria. Además, los inhibidores de expresión que comprenden ácido nucleicos antisentido de la presente invención, son útiles ya que pueden inhibir la actividad biológica de una proteína de la presente invención.

El método como se describe en la presente memoria, puede usarse para modificar la expresión en una célula de un gen asociado a BRC regulado positivamente, por ejemplo, una regulación positiva resultante de la transformación maligna de las células. La unión del ARN ip a un transcripto correspondiente a uno de los genes asociados a BRC mostrados en las listas de las tablas 3, 5 y 7 en la célula diana da como resultado la reducción de la producción de proteína por la célula. La longitud del oligonucleótido es de al menos 10 nucleótidos y puede ser tan larga como el transcripto de origen natural. Preferentemente, el oligonucleótido tiene una longitud de 19-25 nucleótidos. Más preferentemente, el oligonucleótido tiene una longitud menor de 75, 50, 25 nucleótidos.

Los ácidos nucleicos antisentido, como se describe en la presente memoria, incluyen oligonucleótidos modificados. Por ejemplo, pueden usarse oligonucleótidos tioados para conferir existencia nucleasa a un oligonucleótido.

Además, para reducir el nivel de expresión del gen marcador puede usarse un ARNip contra un gen marcador. En la presente memoria, el término "ARNip" se refiere a una molécula de ARN bicatenaria que impide la traducción de un ARNm diana. Para introducir ARNip en la célula pueden usarse técnicas convencionales, incluyendo aquellas en las que el ADN es un molde a partir del cual se transcribe el ARN. En el contexto de la presente invención, el ARNip comprende una secuencia de ácido nucleico con sentido y una secuencia de ácido nucleico antisentido contra un gen marcador regulado positivamente, tal como un gen asociado a BRC mostrado en los listados de las tablas 3, 5, y 7. El ARNip se construye de manera que un solo transcripto tiene las dos secuencias antisentido y sentido complementarias del gen diana, por ejemplo, una horquilla.

Un ARNip de un gen asociado a BRC, como se muestra en los listados de las tablas 3, 5 y 7, se hibrida con ARNm diana y por lo tanto diminuye o inhibe la producción de los polipéptidos codificados por el gen asociado a BRC mostrado en los listados de las tablas 3, 5 y 7 por asociación con el transcripto de ARNm normalmente monocatenario, interfiriendo por lo tanto con la traducción y de esta manera, la expresión de la proteína. En el contexto de la presente invención, un ARNip tiene preferentemente una longitud menor de 500, 200, 100, 50 ó 25 nucleótidos. Más preferentemente un ARNip tiene una longitud de 19-25 nucleótidos. La secuencia de ácidos nucleicos ejemplar para la producción de ARNip TOPK incluye las secuencias de nucleótidos de la SEC Nos: 25, 28 y 31 como la secuencia diana. Para potenciar la actividad de inhibición del ARNip, en el extremo 3' de la cadena antisentido de la secuencia diana pueden añadirse "u" nucleótidos. La cantidad de "u" añadidos es de al menos 2, generalmente de 2 a 10, preferentemente de 2 a 5. Los "u" añadidos forman una sola cadena en el extremo 3' de la

cadena antisentido del ARNip.

Un ARNip de un gen asociado a BRC, como se muestra en los listados de las tablas 3, 5 y 7 puede introducirse directamente en las células de forma que sea capaz de unirse a los transcritos de ARNm. Como alternativa, un vector puede incluir un ADN que codifica el ARNip.

- 5 Los vectores pueden producirse, por ejemplo, clonando una secuencia diana de genes asociados a BRC en un vector de expresión que tiene secuencias reguladoras unidas operativamente flanqueando la secuencia de manera que se permita la expresión (por transcripción de la molécula de ADN) de ambas cadenas (Lee, N.S., Dohjima, T., Bauer, G., Li, H., Li, M.-J., Ehsani, A., Salvaterra, P., y Rossi, J (2002)). Una molécula de ARN que es antisentido con respecto al ARNm de un gen asociado a BRC se transcribe por un primer promotor (por ejemplo, una secuencia promocional 3' del ADN clonado) y una molécula de ARN que es la cadena con sentido para el ARNm de un gen asociado a BRC se transcribe por un segundo promotor (por ejemplo, una secuencia promocional 5' del ADN clonado). Las cadenas con sentido y antisentido se hibridan *in vivo* para generar construcciones de ARNip para el silenciamiento de los genes asociados a BRC. Como alternativa, las dos construcciones pueden utilizarse para crear las cadenas con sentido y antisentido de una construcción de ARNip. Los genes asociados a BRC clonados pueden codificar una construcción que tenga una estructura secundaria, por ejemplo, horquillas, en las que un solo transcríto tiene tanto la secuencia antisentido complementaria como la secuencia con sentido del gen diana.

Entre la secuencia con sentido y antisentido puede localizarse una secuencia en bucle que consiste en una secuencia de nucleótidos arbitraria para formar la estructura en bucle de horquilla. Por tanto, la presente invención también proporciona ARNip que tiene la fórmula general 5'-[A]-[B]-[A']-3', en la que [A] es una secuencia de ribonucleótidos correspondiente a una secuencia de genes seleccionados de las tablas 3, 5 ó 7,

[B] es una secuencia de ribonucleótidos que consiste en de 3 a 23 nucleótidos y [A'] es una secuencia de ribonucleótidos que consiste en la secuencia complementaria de [A]. La región [A] se hibrida con [A'], y después se forma un bucle que consiste en la región [B]. Preferentemente, la secuencia en bucle puede tener una longitud de 3 a 23 nucleótidos. La secuencia en bucle, por ejemplo, puede seleccionarse del grupo que consiste en las siguientes secuencias ([http://www.ambion.com/techlib/tb/tb\\_506.html](http://www.ambion.com/techlib/tb/tb_506.html)). Adicionalmente, la secuencia en bucle que consiste en 23 nucleótidos también proporciona ARNpi activo (Jacque, J.-M., Triques, K., y Stevenson, M. (2002) Modulation of HIV-1 replication by RNA interferente. Nature 418: 435-438.

CCC, CCACC or CCACACC: Jacque, J. M, Triques, K., and Stevenson, M (2002) Modulation of HIV-1 replication by RNA interferente. Nature, Vol. 418: 435-438.

30 UUCG: Lee, N. S., Dohjima, T., Bauer, G., Li, H., Li, M.-J., Ehsani, A., Salvaterra, P., and Rossi, J. (2002) Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. Nature Biotechnology 30 : 500-505. Fruscoloni, P., Zamboni, M., and Tocchini-Valentini, G. P. (2003) Exonucleolytic degradation of double-stranded RNA by an activity in Xenopus laevis germinal vesicles. Proc. Natl. Acad. Sci. USA 100(4): 1639-1644.

35 UUCAAGAGA: Dykxhoorn, D. M., Novina, C. D., and Sharp, P. A. (2002) Killing the Messenger: Short RNAs that silence gene expresión. Nature Reviews Molecular Cell Biology 4: 457-467.

Por consiguiente, la secuencia en bucle también puede seleccionarse del grupo que consiste en, CCC, UUCG, CCACC, CCACACC, Y UUCAAGAGA. Preferentemente la secuencia en bucle es UUCAAGAGA ("ttcaagaga" en ADN). El ARNip en horquilla ejemplar adecuado para su uso en el contexto de la presente invención incluye:

Para ARNip- TOPK

- 40 gaacgauuaaggcagcc- [b]- ggcuggcuuuauaucguuc (para la secuencia diana de la SEC ID Nº: 25);  
cuggaugaaucauaccaga- [b]- ucugguaugauucauccag (para la secuencia diana de la SEC ID Nº: 28);  
guguggcuugcguaaaaa- [b]- uuauuuacgcaagccacac (para la secuencia diana de la SEC ID Nº: 31)

45 La secuencia de nucleótidos de los ARNip adecuados puede diseñarse usando un programa informático de diseño de ARNip disponible en la página web Ambion ([http://www.ambion.com/techlib/misc/siRNA\\_finder.html](http://www.ambion.com/techlib/misc/siRNA_finder.html)). El programa informático selecciona secuencias de nucleótidos para la síntesis de ARNip basándose en el siguiente protocolo.

#### Selección de Sitios Diana del ARNip:

- 50 1. Comenzar con el codón de inicio AUG del transcríto objeto, explorar aguas abajo secuencias de dinucleótidos AA. Registrar la aparición de cada AA y los 19 nucleótidos adyacentes a 3' como posibles sitios diana de ARNip. Tuschl, et al. no recomiendan el diseño contra ARNip en las regiones 5' y 3' no traducidas (UTR) ni en regiones cercanas al codón de inicio (en 75 bases) ya que estas pueden ser más ricas en sitios de unión a proteínas reguladoras. Las proteínas de unión a UTR y/o los complejos de inicio de traducción pueden interferir con la unión del complejo endonucleasa ARNip.
2. Comparar los posibles sitios diana con la base de datos del genoma humano y eliminar considerar cualquiera

de las secuencias diana con homología significativa con respecto a otras secuencias codificantes. La búsqueda de homología puede realizarse usando BLAST, que puede encontrarse en el servidor del NCBI en: [www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/).

3. Seleccionar secuencias diana que cumplan los requisitos para la síntesis. Preferentemente, en Ambion, pueden seleccionarse diversas secuencias diana a lo largo de la longitud del gen a evaluar.

Las secuencias reguladoras que flanquean las secuencias de genes asociados a BRC pueden ser idénticas o diferentes, de manera que su expresión pueda modularse independientemente o de una manera temporal o espacial. Los ARNip se transcriben intracelularmente clonando los moldes de genes asociados a BRC, respectivamente, en un vector que contiene, por ejemplo, una unidad de transcripción de ARN pol III a partir del promotor de ARN nuclear pequeño (ARNnp) U6 o de ARN H1 humano. Para introducir el vector en la célula, puede usarse un agente potenciador de la transfección. Como agentes potenciadores de la transfección son útiles FuGENE (Rochediagnostics), Lipofectamina 2000 (Invitrogen), Oligofectamina (Invitrogen) y Nucleofactor (Wako pure Chemical).

El oligonucleótido antisentido o ARNip de la presente invención inhiben la expresión de un polipéptido de la presente invención y es por lo tanto útil para suprimir la actividad biológica de un polipéptido de la invención. Del mismo modo, inhibidores de expresión, que comprenden el oligonucleótido antisentido o ARNip de la invención, son útiles en el punto en el que pueden inhibir la actividad biológica del polipéptido de la invención. Por lo tanto, para el tratamiento del cáncer de mama es útil una composición que comprenda un oligonucleótido antisentido o ARNip de la presente invención.

20 Anticuerpos:

Como alternativa, la función de uno o más productos génicos de los genes sobreexpresados en BCR puede inhibirse administrando un compuesto que se une a, o de otra manera, inhiba la función de los productos génicos. Por ejemplo, el compuesto es un anticuerpo que se une al producto (o a los productos) genico sobreexpresado.

Se describe el uso de anticuerpos, particularmente anticuerpos contra una proteína codificada por un gen marcador regulado positivamente o un fragmento de dicho anticuerpo. Como se usa en la presente memoria, el término "anticuerpo" se refiere a una molécula de inmunoglobulina que tiene una estructura específica, que solo interacciona (es decir, se une) al antígeno cuando se usa para sintetizar el anticuerpo (es decir, el producto genico de un marcador regulado positivamente) o con un antígeno estrechamente relacionado con el mismo. Adicionalmente, un anticuerpo puede ser un fragmento puede ser un fragmento de un anticuerpo o de un anticuerpo modificado, siempre y cuando se una a una o más de las proteínas codificadas por los genes marcadores. Por ejemplo, el fragmento de anticuerpo puede ser Fab, F(ab') 2, Fv o Fv monocatenario (scFv) en el cual los fragmentos Fv de las cadenas H y L están ligados mediante un conector apropiado (Huston J. S. et al. Proc. Natl. Acad. Sci. U.S.A. 85:5879-5883 (1988)). Más específicamente, un fragmento de anticuerpo puede generarse tratando un anticuerpo con una enzima, tal como papaína o pepsina. Como alternativa, puede construirse un gen que codifique en fragmento de anticuerpo, insertarse en un vector de expresión y expresarse en una célula hospedadora apropiada (véase, por ejemplo Co M. S. et al. J. Immunol. 152:2968-2976 (1994); Better M. y Horwitz A. H. Methods Enzymol. 178:476-496 (1989); Pluckthun A. y Skerra A. Methods Enzymol. 178:497-515 (1989); Larnoyi E. Methods Enzymol. 121:652-663 (1986); Rousseaux J. et al. Methods Enzymol. 121:663-669 (1986); Bird R. E. y Walker B. W. Trends Biotechnol. 9:132-137 (1991)).

40 Un anticuerpo puede modificarse por conjugación con una diversidad de moléculas, tales como polietilenglicol (PEG). Se describen dichos anticuerpos modificados. El anticuerpo modificado puede obtenerse por modificación química de un anticuerpo. Dichos métodos de modificación son convencionales en el campo.

Como alternativa, un anticuerpo puede comprender un anticuerpo químérico que tenga una región variable derivada de un anticuerpo no humano y una región constante derivada de un anticuerpo humano o un anticuerpo humanizado, que comprenda una región determinante de la complementariedad (CRD) derivada de un anticuerpo no humano, una región marco conservada (FR) y una región constante derivada de un anticuerpo humano. Dichos anticuerpos pueden prepararse usando tecnologías conocidas.

50 Las terapias contra el cáncer dirigidas a modificaciones moleculares específicas que se producen en células cancerosas se han validado mediante desarrollo clínico y aprobación reguladora de fármacos anticancerosos tales como trastuzumab (Herceptina) para el tratamiento del cáncer de mama avanzado, metilato de imatinib (Gleevec) para leucemia mieloide crónica, gefitinib (Iressa) para cáncer de pulmón no microcítico (NSCLC) y rituximab (mAb anti-CD20) para linfoma de células C y linfoma de células del manto (Ciardiello F, Tortora G. A novel approach in the treatment of cancer: targeting the epidermal growth factor receptor. Clin Cancer Res. Octubre 2001; 7 (10): 2958-70. Review.; Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajarnonde A, Fleming T, Eiermann W, Wolter J, Pegram M, Baselga J, Norton L. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med. 15 de marzo 2001; 344 (11): 783-92.; Rehwald U, Schluz H, Reiser M, Sieber M, Staak JO, Morshhauser F, Driessens C, Rudiger T, Muller-Hermelink K, Diehl V, Engert A. Treatment of relapsed CD20+ Hodgkin lymphoma with the monoclonal antibody rituximab is effective and well

- tolerated: results of a phase 2 trial of the German Hodgkin Lymphoma Study Group. *Blood*. 15 enero 2002; 101 (2): 420-424.; Fang G, Kim CN, Perkins CL, Ramadevi N, Winton E, Wittmann S and Bhalla KN. (2000). *Blood*, 96, 2246-2253.). Estos fármacos son clínicamente eficaces y se toleran mejor que los agentes anticancerosos tradicionales porque solo se dirigen a células transformadas. Por lo tanto, dichos fármacos no solamente mejoran la supervivencia y la calidad de vida de los pacientes con cáncer, sino que también validan el concepto de terapia contra el cáncer dirigida molecularmente. Adicionalmente, los fármacos dirigidos pueden potenciar la eficacia de la quimioterapia convencional cuando se usan en combinación con esta (Gianni L. (2002). *Oncology*, 63 Suplem 1, 47-56.; Klejman A, Rushen L, Morrione A, Slupianek A and Skorski T. (2002). *Onogene*, 21, 5868-5876.). Por lo tanto, los futuros tratamientos contra el cáncer implicarán probablemente la combinación de fármacos convencionales con agentes específicos diana dirigidos a diferentes características de células tumorales tales como angiogénesis e invasividad.
- Estos métodos moduladores pueden realizarse *ex vivo* o *in vivo* (por ejemplo, cultivando la célula con el agente) o, como alternativa, *in vivo* (por ejemplo administrando el agente a un sujeto). Los métodos implican administrar una proteína o combinación de proteínas o una molécula de ácido nucleico o combinación de moléculas de ácido nucleico como terapia para contrarrestar la incorrecta expresión de los genes expresados diferencialmente o la incorrecta actividad de sus productos génicos.
- Las enfermedades y trastornos que se caracterizan por un aumento (con respecto a un sujeto que no padece la enfermedad o trastorno) de los niveles de expresión o actividades biológicas de los genes y productos génicos, respectivamente, pueden tratarse con agentes terapéuticos que antagonicen (es decir, reduzcan o inhiban) la actividad del gen o genes sobreexpresados. Los agentes terapéuticos que antagonizan la actividad pueden administrarse terapéutica o profilácticamente.
- Por consiguiente, los agentes terapéuticos que pueden usarse en el contexto como se describe en la presente memoria incluyen, por ejemplo, (i) un polipéptido del gen o genes sobreexpresados o infraexpresados, o análogos, derivados, fragmentos u homólogos de los mismos; (ii) anticuerpos del gen o de los productos génicos sobreexpresados; (iii) ácidos nucleicos que codifican el gen o genes sobreexpresados o infraexpresados; (iv) ácidos nucleicos antisentido o ácidos nucleicos que son "disfuncionales" (es decir, debido a una inserción heteróloga en los ácidos nucleicos de uno o más genes o genes sobreexpresados; (v) ARN de interferencia pequeño (ARNip); o (vi) moduladores (es decir, inhibidores, agonistas y antagonistas que modifican la interacción entre un polipéptido sobreexpresado o infraexpresado y su compañero de unión). Las moléculas antisentido disfuncionales se usan para inactivar ("knockout") la función endógena de un polipéptido por recombinación homóloga (véase, por ejemplo, Capecchi, *Science* 244:1288-1292 1989).
- Las enfermedades y trastornos que se caracterizan por una disminución de la actividad biológica (con respecto a un sujeto que no padece la enfermedad o trastorno) pueden tratarse con agentes terapéuticos que aumenten (es decir, sean agonistas) la actividad. Los agentes terapéuticos que regulan positivamente la actividad pueden administrarse de una manera terapéutica o profiláctica. Los agentes terapéuticos que pueden usarse incluyen, pero sin limitación, un polipéptido (o análogos, derivados, fragmentos u homólogos del mismo); o un agonista que aumente la biodisponibilidad.
- El aumento o la disminución de los niveles puede detectarse fácilmente cuantificando el péptido y/o el ARN, obteniendo una muestra del tejido de un paciente (por ejemplo, de tejido de biopsia) y sometiéndola a un ensayo *in vitro* con respecto a los niveles de ARN o péptido, estructura y/o actividad de los péptidos expresados (o los ARNm de un gen cuya expresión está modificada). Los métodos bien conocidos en la técnica incluyen, pero sin limitación, inmunoensayos (por ejemplo, por análisis de transferencia de Western, immunoprecipitación seguida de electroforesis en gel de poliacrilamida y dodecil sulfato sódico (SDS), inmunohistoquímica, etc.) y/o ensayos de hibridación para detectar la expresión de los ARNm (por ejemplo, ensayos de Northern, transferencias puntuales, hibridación *in situ*, etc.).
- La administración profiláctica se realiza antes de la manifestación de síntomas clínicos palpables de la enfermedad, para prevenir o, como alternativa, retrasar el avance del trastorno o enfermedad.
- Los métodos terapéuticos, como se describe en la presente memoria, pueden incluir la etapa de poner en contacto una célula con un agente que module una o más de las actividades de los productos génicos de los genes expresados diferencialmente. Los ejemplos de agentes que modulan actividades de proteínas incluyen, pero sin limitación, ácidos nucleicos, proteínas, ligandos afines de origen natural de dichas proteínas, péptidos, peptidomiméticos y otras moléculas pequeñas. Por ejemplo, un agente adecuado puede estimular una o más actividades de proteínas de uno o más genes infraexpresados diferencialmente.
- Vacunación contra el cáncer de mama:
- También se describe un método para tratar o prevenir el cáncer de mama en un sujeto que comprende la etapa de administrar, a dicho sujeto, una vacuna que comprende un polipéptido codificado por un ácido nucleico seleccionado entre el grupo que consiste en los genes asociados a BRC mostrados en los listados de las tablas 3, 5 y 7 (es decir, genes regulados positivamente), un fragmento inmunológicamente activo de dicho polipéptido, o un polinucleótido que codifica dicho polipéptido o fragmento del mismo. La administración del polipéptido induce una inmunidad

antitumoral en un sujeto. Para inducir la inmunidad antitumoral en un sujeto que lo necesita, se administra un polipéptido codificado por un ácido nucleico seleccionado del grupo que consiste en los genes asociados a BRC mostrados en los listados de las tablas 3, 5 y 7, un fragmento inmunológicamente activo de dicho polipéptido, o un polinucleótido que codifica dicho polipéptido o fragmento del mismo. Adicionalmente, el polipéptido codificado por un ácido nucleico seleccionado entre el grupo que consiste en los genes asociados a BRC mostrados en los listados de las tablas 5 y 7 puede inducir inmunidad antitumoral contra la invasión del cáncer de mama y carcinoma ductal invasivo (IDC), respectivamente. El polipéptido o los fragmentos inmunológicamente activos del mismo son útiles como vacunas contra BRC. En algunos casos, las proteínas o fragmentos de las mismas pueden administrarse de forma unida al receptor de células T (RCT) o presentada por una célula presentadora de抗igenos (CPA), tales como macrófagos, células dendríticas (CD), o células B. Entre las CPA, el uso de las CD es más preferente debido a la fuerte capacidad para presentar al antígeno.

Como se describe en la presente memoria, una vacuna contra BRC se refiere a una sustancia que puede inducir una inmunidad antitumoral tras la inoculación en animales. De acuerdo con la presente descripción, se sugirió que los polipéptidos codificados por los genes asociados a BRC mostrados en los listados de las tablas 3, 5 y 7, o fragmentos de los mismos, eran péptidos epítópicos restringidos por HLA-A24 o HLA-A\*0201 que podían inducir una respuesta inmunitaria potente y específica contra células BRC que expresan genes asociados a BRC mostrados en los listados de las tablas 3, 5 y 7. Por tanto, también se describe un método para inducir inmunidad antitumoral usando los polipéptidos. En general, la inmunidad antitumoral incluye respuestas inmunitarias tales como:

- inducción de linfocitos citotóxicos contra tumores,
- inducción de anticuerpos que reconocen tumores, e
- inducción de producción de citocinas antitumorales.

Por lo tanto, cuando una proteína determinada induce cualquiera de estas respuestas inmunitarias tras la inoculación en un animal, se establece que la proteína tiene efecto inductor de inmunidad antitumoral. La inducción de la inmunidad antitumoral por una proteína puede detectarse observando *in vivo* o *in vitro* la respuesta del sistema inmunitario en el hospedador contra la proteína.

Por ejemplo, se conoce bien un método para detectar la inducción de linfocitos T citotóxicos. Específicamente, por la acción de las células presentadoras de抗igenos (CPA), una sustancia extraña que entra en el organismo se presenta a las células T y a las células B. Las células T que responden al antígeno, presentado por las CPA de una manera específica para el antígeno, se diferencian en células T citotóxicas (o linfocitos T citotóxicos, LTC) debido a la estimulación por el antígeno, y después proliferan (esto se denomina activación de células T). Por lo tanto, la inducción de LTC por un péptido determinado puede evaluarse presentando el péptido a una célula T mediante una CPA y detectando la inducción de los LTC. Adicionalmente, las CPA tienen el efecto de activación de células T CD4+, células T CD8+, macrófagos, eosinófilos y células NK (del inglés Natural Killer que significa linfocitos citolíticos naturales). Dado que las células T CD4+ y las células T CD8+ también son importantes en la inmunidad antitumoral, la acción inductora de inmunidad antitumoral del péptido puede evaluarse usando el efecto de activación de estas células como indicadores.

En la técnica se conoce bien un método para evaluar la acción inductora de los LTC usando células dendríticas (CD) como CPA. Entre las CPA, las CD son un representante de las CPA que tienen la acción inductora de LTC más fuerte. En este método, el polipéptido de ensayo se pone inicialmente en contacto con las CD y después las CD se ponen en contacto con las células T. La detección de las células T, que tienen efectos citotóxicos contra las células de interés, después del contacto con las CD, demuestra que el polipéptido de ensayo tiene una actividad de inducción de las células T citotóxicas. La actividad de los LTC contra tumores puede detectarse, por ejemplo, usando la lisis de células tumorales marcadas con <sup>51</sup>Cr como indicador. Como alternativa, también se conoce bien el método de evaluación del grado de lesión celular tumoral usando la actividad de captación de <sup>3</sup>H-timidina o liberación de LDH (lactosa deshidrogenasa) como indicador.

Como CPA, además de las CD, también pueden usarse las células mononucleares de sangre periférica (CMSP). Se ha indicado que la inducción de los LTC se potencia cultivando CMSP en presencia de Factor Estimulante de Colonias de Macrófagos y Granulocitos FEC-MG (GM-CSF, siglas en inglés) e IL-4. De manera similar, se ha demostrado que los LTC, se inducen cultivando CMSP en presencia de hemocianina de lapa californiana (HLC) e IL-7.

Los polipéptidos de ensayo que, a través de estos métodos, confirmaron poseer actividad inductora de LTC se consideraron que eran polipéptidos que tenían efecto de activación de CD y posterior actividad inductora de LTC. Por lo tanto, los polipéptidos que inducen LTC contra células tumorales son útiles como vacunas contra tumores. Adicionalmente, las CPA que han adquirido la capacidad de inducir LTC contra tumores a través del contacto con los polipéptidos también son útiles como vacunas contra tumores. Adicionalmente, los LTC que han adquirido citotoxicidad debido a la presentación de los抗igenos polipeptídicos por las CPA también pueden usarse como vacunas contra tumores. Dichos métodos terapéuticos para tumores, que usan inmunidad antitumoral debido a las CPA y a los LTC reciben el nombre de inmunoterapia celular.

Generalmente, cuando para inmunoterapia celular se usa un polipéptido, se sabe que la eficacia de la inducción de los LTC aumenta combinando una pluralidad de polipéptidos que tienen diferentes estructuras y poniéndolos en contacto con las CD. Por lo tanto, cuando las CD se estimulan con fragmentos de proteínas, es ventajoso usar una mezcla de tipos de fragmentos múltiples.

- 5 Como alternativa, la inducción de inmunidad antitumoral por un polipéptido puede confirmarse observando la inducción de la producción de anticuerpos contra tumores. Por ejemplo, cuando en un animal de laboratorio inmunizado con el polipéptido se inducen anticuerpos contra un polipéptido y cuando estos anticuerpos suprimen el crecimiento de células tumorales, se considera que el polipéptido puede inducir inmunidad antitumoral
- 10 La inmunidad antitumoral se induce administrando la vacuna, como se describe en la presente memoria, y la inducción de la inmunidad antitumoral permite el tratamiento y la prevención del BRC. La terapia contra el cáncer o la prevención de la aparición del cáncer incluye cualquiera de las siguientes etapas, tales como inhibición del crecimiento de células cancerosas, involución del cáncer, y supresión de la aparición de cáncer. En la terapia o prevención del cáncer también se incluye una disminución en la mortalidad y morbilidad de individuos que padecen cáncer, disminución de los niveles de marcadores tumorales en la sangre, alivio de síntomas detectables que acompañan al cáncer y similares. Dichos efectos terapéuticos y preventivos son, de manera preferente, estadísticamente significativos. Por ejemplo, en la observación, a un nivel de significación estadística del 5% o menor, en el que el efecto terapéutico o preventivo de una vacuna contra enfermedades proliferativas celulares se compara con un control sin administración de vacuna. Para análisis estadísticos, puede usarse, por ejemplo, el ensayo *t* de Student, el ensayo *U* de Mann-Whitney o ANOVA.
- 15 20 La proteína mencionada anteriormente que tiene actividad inmunológica o un vector que codifica la proteína puede combinarse con un adyuvante. Un adyuvante se refiere a un compuesto que potencia la respuesta inmunológica contra la proteína cuando se administra junto (o sucesivamente) con la proteína que tiene actividad inmunológica. Los adyuvantes ejemplares incluyen, pero sin limitación, la toxina del cólera, la toxina de salmonella, alumbre y similares. Adicionalmente, la vacuna como se describe en la presente memoria puede combinarse de manera apropiada con un vehículo farmacéuticamente aceptable. Los ejemplos de dichos vehículos incluyen agua esterilizada, solución salina fisiológica, tampón fosfato, líquido de cultivo y similares. Adicionalmente, la vacuna puede contener, si fuera necesario, estabilizantes, suspensiones, conservantes, tensioactivos y similares. La vacuna puede administrarse por vía sistémica o por vía local. La administración de la vacuna puede realizarse por administración sencilla o de refuerzo por administraciones múltiples.
- 25 30 Cuando se usa una CPA o un LTC como la vacuna, como se describe en la presente memoria, los tumores pueden tratarse o prevenirse, por ejemplo, por el método *ex vivo*. Más específicamente, se extraen las CMSP del sujeto que recibe el tratamiento o la prevención, se ponen en contacto con el polipéptido *ex vivo* y después de la inducción de las CPA o los LTC, las células pueden administrarse al sujeto. Las CPA también pueden inducirse introduciendo un vector que codifique el polipéptido en las CMSP *ex vivo*. Las CPA o los LTC inducidos *in vitro* pueden clonarse antes de la administración. La inmunoterapia celular puede realizarse más eficazmente clonando y cultivando células que tengan elevada actividad para producir daños a células diana. Adicionalmente, las CPA y los LTC aislados de esta manera pueden usarse para inmunoterapia celular no solo contra individuos de los que derivan las células, sino también contra tipos de tumores similares de otros individuos.
- 35 40 Adicionalmente, se describe una composición farmacéutica para el tratamiento o prevención de una enfermedad proliferativa celular, tal como cáncer, que comprende una cantidad terapéuticamente eficaz del polipéptido. La composición farmacéutica puede usarse para provocar inmunidad antitumoral.

#### Composiciones farmacéuticas para inhibir BRC o BRC maligno:

- 45 En el contexto de la presente descripción, las formulaciones farmacéuticas adecuadas incluyen aquellas que son adecuadas para la administración oral, rectal, nasal, tópica (incluyendo bucal y sublingual), vaginal o parenteral (incluyendo intramuscular, subcutánea e intravenosa), o para la administración por inhalación o insuflación. Preferentemente, la administración es intravenosa. Opcionalmente, las formulaciones se envasan en unidades de dosificación individuales.
- 50 Las formulaciones farmacéuticas adecuadas para la administración oral incluyen cápsulas, obleas o comprimidos, conteniendo cada uno de estos una cantidad predeterminada del principio activo. Las formulaciones adecuadas también incluyen polvos, gránulos, soluciones, suspensiones y emulsiones. El ingrediente activo se administra opcionalmente como un electuário o concentrado en bolo. Los comprimidos y cápsulas para la administración oral pueden contener excipientes convencionales, tales como agentes de unión, cargas, lubricantes, agentes desgregantes y/o humectantes. Un comprimido puede fabricarse por compresión o moldeo, opcionalmente con uno o más principios de la fórmula. Los comprimidos fabricados por compresión pueden prepararse en una compresora adecuada comprimiendo los principios activos en una forma fluida, tal como polvos o gránulos, opcionalmente mezclados con un agente aglutinante, lubricante, diluyente inerte, lubricante, tensioactivo y/o dispersante. Los comprimidos moldeados pueden prepararse en una moldeadora adecuada moldeando una mezcla del compuesto en polvo humedecido con un diluyente líquido inerte. Los comprimidos pueden recubrirse de acuerdo con métodos bien conocidos en la técnica. Las preparaciones fluidas orales pueden estar, por ejemplo, en forma de suspensiones,
- 55 55

soluciones, emulsiones acuosas u oleosas, jarabes o elixires o pueden estar presentes como un producto seco para la constitución con agua u otro vehículo adecuado antes del uso. Dichas preparaciones líquidas pueden contener aditivos convencionales, tales como agentes de suspensión, agentes emulsionantes, vehículos no acuosos (que pueden incluir aceites comestibles) y/o conservantes. Opcionalmente, los comprimidos pueden formularse para proporcionar una liberación lenta o controlada del principio activo en su interior. Un envase de comprimidos puede contener un comprimido para tomar mensualmente.

5 Las formulaciones adecuadas para la administración parenteral que incluyen soluciones acuosas y no acuosas para inyección estéril, opcionalmente contienen antioxidantes, tampones, bacteriostáticos y solutos que hacen que la formulación sea isotónica con la sangre del receptor deseado; así como suspensiones estériles acuosas y no acuosas que incluyen agentes de suspensión y/o agentes espesantes. Las formulaciones pueden presentarse en envases monodosis o multidosis, por ejemplo como ampollas y viales herméticamente cerrados y pueden conservarse en estado seco congelado (lioofilizado), que solo requiere la adición del vehículo líquido estéril, por ejemplo, solución salina, agua para inyección, inmediatamente antes del uso. Como alternativa, las formulaciones pueden presentarse para infusión continua. Pueden prepararse soluciones y suspensiones para inyección improvisada a partir de polvos, gránulos y comprimidos estériles del tipo previamente descrito.

10 Las formulaciones adecuadas para la administración rectal incluyen supositorios con vehículos convencionales tales como manteca de cacao o polietilenglicol. Las formulaciones adecuadas para la administración tópica en la boca, por ejemplo, por vía bucal o sublingual, incluyen pastillas para chupar que contienen el principio activo en una base con sabor tal como sacarosa y goma arábiga o tragacanto y pastillas que comprenden el principio activo en una base tal como gelatina y glicerina o sacarosa y goma arábiga. Para la administración intranasal, los compuestos de la invención pueden usarse como pulverización líquida, un polvo dispersable o en forma de gotas. Las gotas pueden formularse con una base acuosa o no acuosa que también comprende uno o más agentes dispersantes, agentes solubilizantes y/o agentes de suspensión.

15 Para la administración por inhalación los compuestos pueden administrarse convenientemente a partir de un insuflador, nebulizador, envases presurizados y otros medios convenientes de administración en pulverización por aerosol. Los envases presurizados pueden contener un propulsor adecuado tal como díclorodifluorometano, triclorofluorometano, díclorotetrafluoroetano, dióxido de carbono u otro gas adecuado. En el caso de un aerosol presurizado, la unidad de dosificación puede determinarse proporcionando una válvula para administrar una cantidad medida.

20 30 Como alternativa, para la administración por inhalación o insuflación, los compuestos pueden adoptar la forma de una composición en polvo seco, por ejemplo una mezcla en polvo del compuesto y una base en polvo adecuada, tal como lactosa o almidón. La composición en polvo puede presentarse en forma de dosificación unitaria, por ejemplo, como cápsulas, cartuchos, envases de gelatina o blisters, a partir de los cuales se administra el polvo usando un inhalador o insufladores.

25 35 Otras formulaciones incluyen dispositivos implantables y parches adhesivos que liberan el agente terapéutico.

Cuando se desee, las formulaciones descritas anteriormente, pueden emplearse adaptadas para proporcionar la liberación prolongada del principio activo. Las composiciones farmacéuticas también pueden contener otros ingredientes activos, tales como agentes antimicrobianos, inmunosupresores y/o conservantes.

40 45 Debe entenderse que, además de los principios mencionados en particular anteriormente, las formulaciones de la presente invención pueden incluir otros agentes convencionales en la técnica con respecto al tipo de formulación en cuestión. Por ejemplo, las formulaciones adecuadas para administración oral pueden incluir agentes saporíferos.

Las formulaciones de dosificación unitaria preferidas contienen una dosis eficaz, como se ha indicado anteriormente, o una fracción apropiada de las mismas del principio activo.

50 55 Para cada una de las afecciones indicadas anteriormente, las composiciones, es decir, polipéptidos y compuestos orgánicos, pueden administrarse por vía oral o por inyección a una dosis que varía de aproximadamente 0,1 a aproximadamente 250 mg/kg por día. El intervalo de la dosis para seres humanos adultos es generalmente de aproximadamente 5 mg a aproximadamente 17,5 g/día, preferentemente de aproximadamente 5 mg a aproximadamente 10 g/día y más preferentemente de aproximadamente 100 mg a aproximadamente 3 g/día. Los comprimidos u otras formas de presentación en dosificación unitaria, proporcionados en unidades individuales, pueden contener convenientemente una cantidad que sea eficaz a dicha dosificación o una cantidad múltiple de la misma, por ejemplo, unidades que contienen de aproximadamente 5 mg a aproximadamente 500 mg, normalmente de aproximadamente 100 mg a aproximadamente 500 mg.

La dosis empleada dependerá de numerosos factores, incluyendo la edad y sexo del sujeto, el trastorno exacto a tratar y su gravedad. También la vía de administración puede variar dependiendo de la afección y su gravedad. En cualquier caso, considerando los factores mencionados anteriormente, los expertos en la materia pueden calcular las dosificaciones apropiadas y óptimas.

En los siguientes ejemplos se describen aspectos de la presente invención, que no pretenden limitar el alcance de la

invención descrita en las reivindicaciones. Los siguientes ejemplos ilustran la identificación y caracterización de genes expresados diferencialmente en células BRC (de cáncer de mama).

#### EJEMPLOS

Para identificar genes que se expresan diferencialmente o una patología, por ejemplo BRC, se evaluaron tejidos obtenidos de tejido enfermo (por ejemplo, células epiteliales de BRC) y tejidos normales. Los ensayos se realizaron de la siguiente manera.

#### Pacientes y muestras de tejidos:

Los cánceres de mama primarios se obtuvieron de 81 pacientes, que habían dado su consentimiento con total conocimiento de causa, (12 de carcinoma ductal *in situ* y 69 de carcinoma ductal invasivo de 2 a 5 cm (T2), edad media de 45 años en un intervalo de 21 a 68 años de edad) tratados en el Departamento de Cirugía de Mama, Cancer Institute Hospital, Tokyo, Japón, en lo que respecta a todos los pacientes que habían dado su consentimiento (Tabla 12). La información clínica se obtuvo de registros médicos y cada tumor se diagnosticó por patólogos, de acuerdo con subtipos y grados histopatológicos. El tejido tumoral se usó para evaluar el tipo de tumor (de acuerdo con la clasificación de la Organización Mundial de la Salud y la clasificación de la sociedad japonesa contra el cáncer). El estado clínico se evaluó de acuerdo con la clasificación TNM de JBCS. No se observaron diferencias significativas entre los casos nódulo-positivos y nódulo-negativos. Los patólogos determinaron la presencia de crecimiento angioinvasivo e infiltrado linfocítico exhaustivo, la expresión del receptor de estrógenos (ER) y del receptor de progesterona (PgR) se determinó por EIA (ER negativo cuando es menor que 13 fmol/mg de proteína, BML). Se usó una mezcla de células ductales de mama normales de las 15 pacientes premenopáusicas con cáncer de mama o de las 12 pacientes postmenopáusicas como controles normales, respectivamente. Todas las muestras se congelaron inmediatamente y se conservaron a -80°C.

#### Muestras de Tejido y LMM:

La información clínica y patológica sobre el tumor se detalla en la Tabla 12. Las muestras se incluyeron en medio OCT TissueTek (Sakura) y después se conservaron a -80°C hasta su uso. Las muestras congeladas se seccionaron en serie en cortes de 8 µm con un criostato y se tiñeron con hematoxilina y eosina para definir las regiones analizadas. Para evitar la contaminación cruzada de células cancerosas y no cancerosas, se prepararon estas dos poblaciones mediante el sistema LMM EZ Cut (SL Microtest GmbH) siguiendo el protocolo del fabricante con algunas modificaciones. Para minimizar los efectos durante el proceso de conservación y extracción de tejido, los tejidos cancerosos se trataron cuidadosamente mediante el mismo procedimiento. Para controlar la calidad de los ARN, el ARN total extraído del tejido residual de cada caso se sometió a electroforesis con gel de agarosa degenerativo y se confirmó su calidad por una presencia de bandas de ARN ribosomal.

#### Extracción de ARN y Amplificación de ARN basada en T7:

Se extrajo ARN total de cada población de células capturadas con láser en 350 µl de tampón de lisis RLT (QIAGEN). El ARN extraído se trató durante 30 minutos a temperatura ambiente con 30 unidades de DNasa I (QIAGEN). Después de la inactivación a 70°C durante 10 minutos, los ARN se purificaron con un Kit RNeasy Mini (QIAGEN) de acuerdo con las recomendaciones del fabricante. Todo el ARN tratado con DNasa I se sometió a amplificación basada en T7 usando el Kit de Transcripción Ampliscribe T7 (Epicentre Technologies). Dos rondas de amplificación produjeron 28,8-329,4 µg de los ARN amplificados (ARNa) para cada muestra, mientras que cuando se amplificaron los ARN de muestras normales de 15 pacientes premenopáusicas o de 12 pacientes postmenopáusicas se produjeron un total de 2240,2 µg y 2023,8 µg, respectivamente. Se transcribieron, de manera inversa, alícuotas de 2,5 µg de ARNa de cada una de las células cancerosas y células ductales de mama no cancerosas, en presencia de Cy5-dCTP y Cy3-dCTP (Amersham Biosciences), respectivamente.

#### Micromatrizes de ADNc:

Se estableció un sistema de micromatriz de ADNc "pangenómico" que contenía 23,040 de los ADNc seleccionados de la base de datos UniGene (construcción nº 131) del National Center for Biotechnology Information (NCBI). La fabricación de portaobjetos para micromatriz de ADNc se describe en otro documento (Ono K, Tanaka T, Tsunoda T, Kitahara O, Kihara C, Okamoto A, Ochiai K, Katagiri T and Nakamura Y. Identification by cDNA Microarray of Genes Involved in Ovarian Carcinogenesis. *Cancer Res.*, 60, 5007-11,2000.). En resumen, los ADNc se amplificaron por PCR de transcripción inversa usando, como moldes, poli(A)+ARN aislado de diversos órganos humanos; las longitudes de los amplicones variaban de 200 a 1100 pb sin secuencias repetitivas o poli(A). Los productos de la PCR se aplicaron puntualmente por duplicado en portaobjetos de vidrio de tipo 7 (Amersham Bioscience) usando un aplicador puntual de matriz Lucidea (Amersham Biosciences); en un solo portaobjeto se aplicaron 4.608 o 9.216 genes por duplicado. Se prepararon tres conjuntos de portaobjetos diferentes (en total 23.040 genes), en los que se aplicaron puntualmente en cada uno de ellos los mismos 52 genes constitutivos y así como dos tipos de genes de control negativo.

#### Hibridación y recogida de datos:

La hibridación y el lavado se realizaron de acuerdo con los protocolos descritos anteriormente con la excepción de que todos los procesos se realizaron con un Procesador de Portaobjetos Automatizado (Amersham Biosciences) (Giuliani, N., et al., V. Human myeloma cells stimulate the receptor activator of nuclear factor-kappa B ligand (RANKL) in T lymphocytes: a potential role in 5 multiple myeloma bone disease. *Blood*, 100: 4615-4621, 2002.). La intensidad de cada señal de hibridación se calculó fotométricamente mediante el programa informático ArrayVision (Amersham Biosciences) y se restó la intensidad de fondo. Las intensidades de fluorescencia de Cy5 (tumor) y Cy3 (control) para cada aplicación diana se ajustaron de manera que la proporción Cy5/Cy3 promedio se realizó usando señales promedio de los 52 genes constitutivos. Como los datos derivados de intensidades de señales bajas son menos fiables, se determinó un valor de límite para intensidades de señal en cada portaobjetos y se excluyeron genes de análisis posteriores cuando los dos colorantes Cy3 y Cy5 proporcionaron intensidades de señal inferiores al límite. De acuerdo con la fluctuación de fondo, para cada nivel de expresión se calculó automáticamente un valor de límite. Cuando las dos intensidades de señal, Cy5 y Cy3, eran menores que los valores límite, la expresión del gen correspondiente en esa muestra se evaluó como ausente. La proporción Cy5/Cy3 se calculó como la proporción de expresión relativa. Para otros genes, la proporción Cy5/Cy3 se calculó usando los datos sin procesar para cada muestra.

Las intensidades de señal de Cy3 y Cy5 de las 23.040 aplicaciones puntuales se cuantificaron y se analizaron por sustitución de fondos, usando el programa informático ArrayVision (Imaging Research, Inc., St. Catharines, Ontario, Canadá). Posteriormente, se ajustaron las intensidades de fluorescencia de Cy5 (tumor) y Cy3 (control) para cada aplicación diana, de manera que la proporción Cy3/Cy5 media de 52 genes constitutivos en la matriz era igual a uno. Como los datos derivados de intensidades de señales bajas son menos fiables, se determinó un valor de límite en cada portaobjetos, como se ha descrito anteriormente (Ono, K., et al., Identification by cDNA microarray of genes involved in ovarian 25 carcinogenesis. *Cancer Res*, 60: 5007-5011, 2000.) y se excluyeron genes de análisis posteriores cuando los dos colorantes Cy3 y Cy5 proporcionaron intensidades de señal inferiores al límite (Saito-Hisaminato, A., Katagiri, T., Kakiuchi, S., Nakamura, T., Tsunoda, T., and Nakamura, Y. Genome-wide profiling of gene expression in 29 normal human tissues with a cDNA microarray. *DNA Res*, 9: 35-45, 2002.). Para otros genes, la proporción Cy5/Cy3 se calculó usando los datos sin procesar para cada muestra.

#### Cálculo del porcentaje de contaminación:

Se expresó perlipina (PLIN) y la proteína 4 de unión a ácidos grasos 4 (FABP4) exclusivamente en tejido adiposo y en tejido de glándula mamaria por perfiles de expresión de genes en 29 tejidos humanos normales con una micromatriz de ADNc (Saito-Hisaminato, A. et al., Genome-wide profiling of gene expression in 29 normal human tissues with a cDNA microarray. *DNA Res*, 9: 35-45, 2002.). Esto se usó para evaluar la proporción de adipocitos presentes en la población de células epiteliales ductales de mama normales microdisecionadas. Cada ARNa de poliA<sup>+</sup>ARN aislado de glándula mamaria completa normal (Clontech) y de células epiteliales ductales de mama normales microdisecionadas se transcribió de manera inversa en presencia de Cy5-dCTP y Cy3-dCTP, respectivamente. Después de la hibridación en portaobjetos de micromatriz, se calculó la proporción Cy5/Cy3. El promedio de cada proporción se decidió por el resultado de tejido de glándula mamaria y células ductales de mama normales microdisecionadas usados en pacientes premenopáusicas y en pacientes postmenopáusicas.

#### Análisis de grupo de 102 Muestras Con 81 Carcinomas de Mama de Acuerdo con Perfiles de Expresión de genes:

Se aplicó un método de agrupamiento jerárquico no supervisado tanto a genes como a tumores. Para obtener grupos reproducibles para la clasificación de las 102 muestras, se seleccionaron 710 genes para los que se obtuvieron datos válidos en el 80% de los experimentos y cuyas relaciones de expresión variaron por desviaciones convencionales en más de 1,1. El análisis se realizó usando el programa informático disponible en la red ("Cluster" y "TreeView") confeccionado por M. Eisen (<http://genome-www5.stanford.edu/MicroArray/SMD/restech.html>). Después de aplicar el algoritmo de agrupamiento, la relación de fluorescencia para cada mancha se transformó en logaritmo y después se centró la mediana de los datos para cada muestra para eliminar sesgos experimentales y se usó como unión promedio.

#### Identificación de Genes Regulados Positiva o Negativamente entre DCIS e IDC:

La relación de expresión relativa de cada gen (relación de intensidad Cy5/Cy3) se clasificó en una de cuatro categorías: (A) regulado positivamente (proporción de expresión>2,0); (B) regulado negativamente (proporción de expresión<0,5); (C) sin cambios (proporción de expresión entre 0,5 y 2,0); y (D) no expresado (o ligera expresión pero por debajo del nivel límite para detección). Estas categorías se usaron para detectar un conjunto de genes para los que los cambios en las relaciones de expresión eran normales entre las muestras. Para detectar genes candidatos que normalmente se regulaban positiva o negativamente en cada grupo, primero se exploraron los patrones de expresión en conjunto de 23.040 genes para seleccionar genes con relaciones de expresión>3,0 o <1/3 que estaban presentes en >50% de los grupos categorizados.

#### RT-PCR semicuantitativa:

Se seleccionaron cinco genes regulados positivamente y sus niveles de expresión se examinaron aplicando los experimentos RT-PCR semicuantitativos. Se transcribió de manera inversa 1 µg de alícuota de ARNa de cada

5 muestra para ADNc monocatenarios usando un cebador al azar (Taniguchi, K., et al., Mutational spectrum of beta-catenin, AXIN1, and AXTN2 in hepatocellular carcinomas and hepatoblastomas. *Oncogene*, 21: 4863-4871, 2002.) y Superscript II (Life Technologies, Inc.). Cada mezcla de ADNc se diluyó para amplificación PCR posterior con los conjuntos de cebadores que se muestran en la Tabla 9. La expresión de GAPDH sirvió como un control interno. Las reacciones PCR se optimizaron para la cantidad de ciclos para garantizar la intensidad del producto dentro de la fase de amplificación lineal.

Identificación de genes responsables del estado histopatológico, estado ER y metástasis en nódulos linfáticos en cáncer de mama:

10 Se seleccionaron genes discriminantes usando los dos criterios siguientes: (1) intensidades de señal superiores al nivel límite al menos en un 70% (estado ER) ó 50% (estado histopatológico y metástasis en nódulos linfáticos) de los casos; (2)  $|Med_r - Med_n| > 1$  (estado ER) o 0,5 (estado histopatológico y metástasis en nódulos linfáticos) de los casos, en la que Med indica la mediana derivada de relaciones de expresión relativas transformadas en logaritmo en casos nódulo-positivos o nódulo-negativos. Después, se aplicó un ensayo de permutación al azar para identificar genes que se expresaban diferencialmente entre un grupo (grupo A) y otro (grupo B). Se calculó la media ( $\mu$ ) y desviaciones típicas ( $\sigma$ ) a partir de relaciones de expresión relativas transformadas en logaritmo de cada gen en los casos del grupo A ( $r$ ) y grupo B ( $n$ ). Para cada gen se definió una puntuación de discriminación (DS) de la siguiente manera:

15

$$DS = (\mu_r - \mu_n) / (\sigma_r + \sigma_n)$$

20 Se realizaron ensayos de permutación para calcular la capacidad de que genes individuales distinguiesen entre el grupo A y el grupo B; las muestras se permutaron al azar entre las dos clases 10.000 veces. Dado que el conjunto de datos DS de cada gen mostró una distribución normal, se calculó el valor P para el agrupamiento definido por el usuario (Golub, T. et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 286: 531-537, 1999.).

Cálculo de la puntuación de predicción para metástasis en nódulos linfáticos:

25 Se calcularon las puntuaciones de predicción de acuerdo con los procedimientos descritos anteriormente (Golub, T. et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 286: 531-537, 1999.). Cada gen ( $g_i$ ) vota por nódulo linfático negativo o nódulo linfático positivo dependiendo si el nivel de expresión ( $x_i$ ) en la muestra está más próximo al nivel de expresión medio del nódulo negativo o positivo en muestras de referencia. La magnitud del voto ( $v_i$ ) refleja la desviación del nivel de expresión 30 en la muestra a partir del promedio de las dos clases:

$$Vi = | xi - (\mu_r + \mu_n) / 2 |$$

Los votos se sumaron para obtener el total de votos para el nódulo negativo ( $V_r$ ) y nódulo positivo ( $V_n$ ) y los valores PS se calcularon de la siguiente manera:

35 PS =  $(V_r - V_n) / (V_r + V_n) \times 100$ , reflejando el margen de triunfo en la dirección del nódulo positivo o nódulo negativo. Los valores PS varían de -100 a 100; un valor absoluto superior de PS refleja una predicción más fuerte.

Evaluación de clasificación y ensayo de una exclusión:

La puntuación de clasificación (CS) se calculó usando las puntuaciones de predicción de nódulos linfáticos negativos (PSr) y positivos (PSn) en cada conjunto de genes, de la siguiente manera:

40  $CS = (\mu_{PSr} - \mu_{PSn}) / (\sigma_{PSr} + \sigma_{PSn})$

Un valor mayor de CS indica mejor separación de los dos grupos mediante el sistema de puntuación predictiva. Para el ensayo de una exclusión, se descartó una muestra, el valor p de permutación y los niveles de expresión promedio se calcularon usando las muestras restantes y la clase de la muestra excluida se evaluó posteriormente calculando su puntuación de predicción. Este procedimiento se repitió para cada una de las 20 muestras.

45 Líneas celulares

Se adquirieron líneas celulares de cáncer de mama humano HBL-100, HCC1937, MCF-7, MDA-MB-435s, YMB1, SKBR3, T47D, BT-20, BT-474, BT-549, HCC1143, HCC1500, HCC1599, MDA-MB-157, MDA-MB453, OUCB-F, ZR-75-1, COS-7 de la Colección Americana de Cultivos Tipo (ATCC) y se cultivaron según recomendación de los respectivos depositantes. Las líneas celulares HBC4, HBC5 y MDA-MD-231 son donaciones generosas del Dr.

50 Yamori del Molecular Pharmacology, Cancer Chemotherapy Centre of the Japanese Foundation for Cancer Research. Todas las células se cultivaron en medios apropiados; es decir RPMI-1640 (Sigma, St. Louis, MO) para HBC4, HBC5, T47D, YMB1, OUCB-F, ZR-75-1, BT-549, HCC1143, HCC1500, HCC1599 y HCC1937 (con L-glutamina 2 mM); medio de Eagle modificado por Dulbecco (Invitrogen, Carlsbad, CA) para BT474, HBL100, COS7;

EMEM (Sigma) con aminoácido esencial 0,1 mM (Roche), piruvato de sodio 1 mM (Roche), 0,01 mg/ml de insulina (Sigma) para BT-20 y MCF-7; McCoy (Sigma) para SKBR3 (con L-glutamina 1,5 mM); L-15 (Roche) para MDA-MB-231, MDA-MB-157, MDA-MB453 y MDA-MB-435S. Cada medio se complementó con suero bovino fetal al 10% (Cansera) y solución de antibiótico/antimicótico al 1% (Sigma). Las células MDA-MB-231 y MDA-MB-435S se mantuvieron a 37°C a una atmósfera de aire húmedo sin CO<sub>2</sub>. Las otras líneas celulares se mantuvieron a 37°C a una atmósfera de aire húmedo con CO<sub>2</sub> al 5%. Las muestras clínicas (de cáncer de mama y ductales de mama normales) se obtuvieron de muestras quirúrgicas, respecto a las cuales todas las pacientes habían dado consentimiento con total conocimiento de causa.

#### Análisis de transferencia de Northern

Se extrajeron ARN totales de todas las líneas celulares de cáncer de mama usando el kit RNeasy (QIAGEN) de acuerdo con las instrucciones del fabricante. Después del tratamiento con DNase I (Nippon Gene, Osaka, Japón), se aisló ARNm con el kit de purificación de ARNm (Amersham Biosciences) siguiendo las instrucciones del fabricante. Una alícuota de 1 µg de cada ARNm, junto con los ARN poliA(+) aislados de mama humana adulta normal (Biochain), pulmón, corazón, hígado, riñón, médula ósea (BD, Clontech, Palo Alto, CA), se separaron sobre geles desnaturalizantes de agarosa al 1% y se transfirieron a membranas de nylon (transferencias de Northern para cáncer de mama). Las transferencias de Northern de tejido múltiple humano y de cáncer de mama (Clontech, Palo Alto, CA) se hibridaron con un producto PCR marcado con [ $\alpha^{32}$ P]-dCTP de A7870 preparado por RT-PCR (véase a continuación). La prehibridación, hibridación y lavado se realizaron de acuerdo con las recomendaciones del proveedor. Las transferencias se auto radiografiaron con exploraciones de intensificación a -80°C durante 14 días. Se prepararon sondas específicas para A7870 (320 pb) por RT-PCR usando el conjunto cebador siguiente: 5'-AGACCCTAACGATCGCTTCTG-3' (SEC ID Nº: 13) y 5'-GTGTTTTAACGTCAGCATGAGCAG-3' (SEC ID Nº: 14) y se marcó radiactivamente con un sistema de marcaje de ADN megaprime (Amersham bioscience).

#### Tinción inmunocitoquímica.

Para la construcción de vectores de expresión A7870, se amplificó toda la secuencia codificante del ADNc de A7870 por PCR usando KOD-Plus ADN polimerasa (Toyobo, Osaka, Japón). Los productos de la PCR se insertaron en los sitios EcoRI y Xho I del vector de expresión pCAGGSn3FH-HA. Esta construcción (pCAGGS-A7870-HA) se confirmó por secuenciación de ADN. A continuación, para examinar inicialmente la localización subcelular de A7870 exógeno, se sembraron células COS7 a 1x10<sup>5</sup> por pocillo para expresión exógena. Después de 24 horas, se transfectaron transitoriamente con 1 µg de pCAGGS-A7870-HA en células COS7 usando reactivo de transfección FuGENE 6 (Roche) de acuerdo con las instrucciones del fabricante, respectivamente. Después, las células se fijaron con PBS que contenía paraformaldehído al 4% durante 15 minutos y se hicieron permeables con PBS que contenía Triton X-100 a 0,1% durante 2,5 minutos a 4°C. Posteriormente las células se cubrieron con BSA al 3% en PBS durante 12 horas a 4°C para bloquear la hibridación no específica. Después, se incubaron células COS7 transfectadas con A7870-HA con un anticuerpo anti-HA de ratón (SANTA CRUZ) a una dilución de 1:1000 y anticuerpo anti-TOPK policlonal (señalización celular) a una dilución de 1:1000. Después de lavar con PBS, ambas células transfectadas se tiñeron con un anticuerpo secundario anti-ratón conjugado con Alexa594 (Molecular Probe) a una dilución de 1:5000.

Posteriormente se confirmó la localización subcelular de la proteína A7870 endógena en líneas celulares de cáncer de mama, T47D, BT-20 y HBC5 a 2x10<sup>5</sup> células por pocillo. Las células con un anticuerpo policlonal anti-TOPK de conejo se fabricaron con péptido sintético correspondiente a aminoácidos en el extremo c de PBK/TOPK humano a una dilución de 1:1000. Después de lavar con PBS, las células se tiñeron con un anticuerpo secundario anti-conejo conjugado con Alexa4888 (Molecular Probe) a una dilución de 1:3000. Los núcleos se contratiñeron con dihidrocloruro de 4',6'-diamidina-2'-fenilindol (DAPI). Se obtuvieron imágenes de fluorescencia al microscopio TCS SP2 AOBS (Leica, Tokyo, Japón).

#### Construcción del vector de expresión de ARNip específico de A7870 usando psiU6BX3.0

Se estableció un sistema de ARNip basado en vectores usando el vector de expresión de ARNip psiU6BX3.0 de acuerdo con la descripción anterior (Shimokawa T, Furukawa Y, Sakai M, Li M, MiwaN, Lin YM, NakamuraY (2003). Cancer Res, 63,6116-6120). Se preparó un vector de expresión de ARNip frente a A7870 (psiU6BX-A7870) clonando los oligonucleótidos bicatenarios mostrados en la Tabla 13 en el sitio BbsI en el vector psiH1BX3.0. Se prepararon plásmidos de control, psiU6BX-SC y psiU6BX-LUC clonando oligonucleótidos bicatenarios de 5'-TCCCGCGCGCTTGAGGATTGTTCAAGAGACGAATCCTACAAAGCGCGC-3' (SEC ID Nº: 15) y 5'-AAAAGCGCGCTTGAGGATTGTTCAAGAGACGAATCCTACAAAGCGCGC-3' (SEC ID Nº: 16) para SC (control codificado); 5'-TCCCCGTACGCCGAATACTTCGATTCAAGAGATCGAAGTATTCCGCGTACG-3'(SEC ID Nº: 17) y 5'-AAAACGTACGCCGAATACTTCGATCTCTGAATCGAAGTATTCCGCGTACG-3' (SEC ID Nº: 18) para LUC (control luciferasa) en el sitio BbsI y en el vector psiU6BX3.0 respectivamente.

#### Efecto del silenciamiento génico de A7870

Se sembraron líneas de células de cáncer de mama humano, T47D o BT-20, en placas de 15 cm (4x10<sup>6</sup> células/placa) y se transfectaron con 16 µg de cada psiU6BX-LUC (control luciferasa), psiU6BX-SC (control

codificado) como controles negativos y psiU6BX-A7870 usando reactivo FuGENE6 de acuerdo con las recomendaciones del proveedor (Roche). Después de 24 horas de transfección, las células se volvieron a sembrar de nuevo para ensayo de formación de colonias ( $2 \times 10^6$  cél/placa de 10 cm), RT-PCR ( $2 \times 10^6$  cél/placa 10 cm) y ensayo con MTT ( $2 \times 10^6$  cél/pocillo). Se seleccionaron las células que introducían A7870 con medio que contenía 0,7 mg/ml o 0,6 mg/ml de neomicina (Geneticin, Gibco) en 20 células T47D o BT-20, respectivamente. Después de esto, el medio se cambió cada dos días durante 3 semanas. Para evaluar el funcionamiento del ARNip, se extrajo ARN total de las células 11 días después de la selección con neomicina y después se confirmó el efecto de inactivación de los ARNip mediante una RT-PCR semicuantitativa usando conjuntos cebadores específicos para A7870 y GAPDH; 5'- ATGGAAATCCCATCACCATCT -3' (SEC ID Nº: 19) y 5'-GGTTGAGCACAGGGTACTTTATT -3' (SEC ID Nº: 20) para GAPDH como un control interno, y 5'- GCCTTCATCATCCAAACATT-3' (SEC ID Nº: 21) y 5'- GGCAAATATGTCTGCCTTGT-3' (SEC ID Nº: 22) para A7870.

Además, se cultivaron transfectantes que expresaban los ARNip usando líneas celulares T47D o BT-20 durante 23 días en medio selectivo que contenía neomicina, respectivamente. Después de la fijación con paraformaldehído al 4%, las células transfectadas se tiñeron con solución Giemsa para evaluar la formación de colonias. Los ensayos con MTT se realizaron para cuantificar la viabilidad celular. Después de 10 días de cultivo en el medio que contenía neomicina, se añadió una solución de MTT (3-(4,5-dimetiltiazol-2-il)-2,5-difenil tetrazolio bromuro) (Sigma) a una concentración de 0,5 mg/ml. Después de incubación a 37°C durante 2,5 horas, se añadió ácido-SDS (HCl 0,01 N/SDS al 10%); la suspensión se mezcló enérgicamente y después se incubó durante una noche a 37°C para disolver los cristales azul oscuro. La absorbancia se midió a 570 nm con un Lector Microplaca 550 (BioRad). Para evaluar el funcionamiento del ARNip, se extrajo ARN total de las células 7 días después de la selección, el ensayo con MTT se realizó 10 días después de la selección usando el Kit-8 de Recuento Celular (Dojindo) de acuerdo con el protocolo del fabricante. La absorbancia se midió a una longitud de onda de 570 nm con un Lector Microplaca 550 (BioRad). Para el ensayo de formación de colonias las células se fijaron con paraformaldehído al 4% durante 15 minutos antes de la tinción con solución de Giemsa (Merck). Cada experimento se realizó por triplicado.

## 25 RESULTADOS

### Análisis de clasificación basándose en perfiles de expresión de genes exactos de cáncer de mama:

Dado que el cáncer de mama contiene una población baja de células cancerosas en masas tumorales y se origina de células ductales epiteliales normales, la microdiseccción se realizó para impedir la contaminación de las células no cancerosas o células epiteliales ductales no normales circundantes. Como en el tejido mamario la gran mayoría de células son adipocitos, se consideró que no era adecuado usar todo el tejido mamario para analizar los perfiles de expresión específicos de cáncer en este órgano. Como se muestra en la Figura 1, los ejemplos representativos de DCIS (caso 10326T), IDC (10502T) y epitelio ductal normal (10341N) se microdisecionaron para cada muestra clínica. Esto permite obtener perfiles de expresión de genes posteriores de manera más precisa. La proporción de adipocitos que contaminaron la población microdisecionada de células epiteliales ductales de mama normales que sirvieron como un control universal se examinó midiendo las intensidades de señal de dos genes (es decir, *PLIN* y *FABP4*) que se expresan altamente en los tejidos adiposo y de glándula mamaria, como se ha descrito anteriormente (Saito-Husaminato, A., et al., Genomewide profiling of gene expression in 29 normal human tissues with a cDNA microarray. DNA 30 Res, 9: 35-45,2002.). Cuando las intensidades de señal de estos genes se investigaron en todo el tejido glandular mamario, que contenía un gran cantidad de adipocitos, el promedio de relación de intensidades de señal de estos genes era de aproximadamente 99,4%; la relación en células epiteliales ductales de mama normales microdisecionadas era de aproximadamente 0,6% (véase el apartado *Porcentaje de Contaminación en Materiales y Métodos*). Por lo tanto, se calculó que la proporción promedio de adipocitos contaminantes en las poblaciones de células de control era del 0,6% después de la microdiseción. En primer lugar, se aplicó un algoritmo de agrupamiento jerárquico bidimensional no supervisado al grupo de genes basándose en la similitud en su patrón de expresión sobre 102 muestras clínicas: 81 ejemplares de cáncer de mama clínico diferentes microdisecionados, 11 tipos histológicos diferentes microdisecionados en 10 individuos, 2 tejidos de cáncer de mama completo, 6 células ductales de mama normales microdisecionadas y dos tejidos glandulares mamarios completos. Los grupos reproducibles se obtuvieron con 710 genes (véase Material y Métodos); sus patrones de expresión a través de las 102 muestras se muestran en la Figura 2A. En el eje de muestras, las 102 muestras se agruparon en tres grupos principales (Grupo A, B y C) basándose en sus perfiles de expresión. Después, esta clasificación se asoció con parámetros clínicos, especialmente receptor de estrógeno (ER) como se determina con EIA. De los 55 tumores ER positivos, 45 casos se agruparon en la misma rama (Grupo B) del dendrograma tumoral, lo que sugiere una tendencia con el estado ER. Además, 7 de 10 casos con diferente tipo histológico (muestras nº 10864, 10149, 10818, 10138, 10005, 10646 y 10435) que se marcaron y se hibridaron en experimentos independientes se agruparon más estrechamente dentro del mismo grupo. En particular, entre ellos, el caso duplicado (10149a1 y 10159a1T) también se agrupó en la rama más corta, lo que respalda la reproducibilidad y fiabilidad de los datos de micromatriz. Notablemente, el Grupo C contenía células no cancerosas microdisecionadas y tejidos completos de cáncer de mama, a excepción de un caso de tumor microdisecionado, lo que sugiere que estos datos representan perfiles de expresión exactos específicos del cáncer de mama.

60 Además, se realizó un análisis de agrupamiento jerárquico bidimensional de 89 genes a través de 16 muestras con 2 lesiones diferenciadas microdisecionadas de 8 pacientes con cáncer de mama. Como resultado, las muestras de cáncer de mama con diferentes lesiones fenotípicas estaban estrechamente adyacentes (Figura 2B). Después, se

realizó una permutación al azar para identificar los genes que se expresaban diferencialmente en las lesiones emparejadas de pacientes fenotípicamente bien o mal diferenciadas de 8 ejemplares de cánceres microdisecionados. Como se muestra en la Figura 2C, el análisis de agrupación usando 25 genes que mostraban expresión diferencial puede separarse entre células de cáncer ductal invasivo bien o mal diferenciadas. Estos 25 genes (Tabla 1) incluyeron algunos factores clave cuyas posibles funciones en la invasión y crecimiento celular se habían descrito anteriormente: TNFSF11, ITGA5 y NFAT5 (Giuliani, N., et al., Human myeloma cells stimulate the receptor activator of nuclear factor-kappa B ligand (RANKL) in T lymphocytes: a potential role in multiple myeloma bone disease. *Blood*, 100: 4615-4621, 2002.; Sébastien J. et al., The role of NFAT transcription factors in integrin-mediated carcinoma invasion. *Nature cell biology*, 4: 540-544, 2002.; Klein, S. et al., Alpha 5 beta 1 integrin activates an NF-kappa B-dependent program of gene expression important for angiogenesis and inflammation. *Mol Cell Biol*, 22: 5912-5922, 2002.).

A continuación, se realizó un ensayo de permutación al azar para identificar los genes que se expresaban diferencialmente en 41 tumores ER positivos y en 28 tumores ER negativos en IDC. Todas estas muestras procedían de pacientes premenopáusicas. Se enumeraron 97 genes que pudieron diferenciarse entre ER positivo y negativo con valores P de permutación menores de 0,0001 (véase "Materiales y Métodos") (Figura 3 y Tabla 2). Entre estos se seleccionaron 96 genes como genes relacionados con BRC de la presente invención. Los niveles de expresión aumentaron para 92 de estos genes y disminuyeron para los otros cinco en el grupo ER positivo, en comparación con el grupo ER negativo. Entre estos genes, la proteína 3 de unión a GATA (GATA3), el factor trébol 3 (TFF3), la ciclina D1 (CCND1), el homólogo de MAPKK (MAP2K4) y el inhibidor de metaloproteasa 1 en tejidos (TIMP1), el sustrato del receptor de insulina 1 (IRS1), la proteína de unión a la caja X 1 (XBP1), el miembro GLI3 de la familia GLI-Kruppel (GLI3), se sobreexpresaron en los ER positivos (Tabla 2). Además, dado que el receptor de estrógeno (ESR1) se ordenó en la categoría en el sexto gen basándose en la magnitud del valor p (panel superior en la Figura 3), puede ser posible diferenciar cánceres de mama de acuerdo con perfiles de expresión de ER.

#### Identificación de Genes Normalmente Regulados Positiva o Negativamente en DCIS o IDC:

Para aclarar adicionalmente los mecanismos subyacentes en la carcinogénesis del cáncer de mama, se investigaron genes normalmente regulados positiva o negativamente en DCIS e IDC, respectivamente. En 77 tumores de mama (8 DCIS y 69 IDC en pacientes premenopáusicas) los perfiles de expresión de genes identificaron 325 genes con expresión normalmente modificada (Figura 4A, 4B); 78 genes que normalmente se regulaban positivamente más de tres veces sobre sus niveles en células ductales de mama normales (Figura 4A, 4B, Tabla 3, 5), mientras que 247 genes cuya expresión se redujo menos de 1/3 en células de cáncer de mama (Figura 4A, 4B, Tabla 4, 6). En particular, como se muestra en la Figura 4B, el nivel de expresión de 25 genes aumentó y en 49 genes disminuyó en transición de DCIS a IDC (Tabla 5 y 6). Entre los genes con expresión elevada, se incluyó la fibronectina (FN1) (Tabla 4) que ya se había descrito como sobreexpresado en cánceres de mama (Mackay, A. et al., cDNA microarray analysis of genes associated with ERBB2 (HER2/neu) overexpression in human mammary luminal epithelial cells. *Oncogene*, 22: 2680-2688, 2003.; Lalani, E. N. et al., Expression of the gene coding for a human mucin in mouse mammary tumor cells can affect their tumorigenicity. *J Biol Chem*, 266: 15420-15426, 1991.; 22. Martin-Lluesma, S., et al., A Role of Heel in spindle checkpoint signaling and kinetochore recruitment of 5 Mad1/Mad2. *Science*, 291: 2267-2270, 2002.). Por otro lado, entre los genes con expresión disminuida, también se incluyeron ST5 y SCHIP1 que se sabía que actuaban como supresores tumorales (Tabla 6).

A continuación, se investigaron genes con expresión específicamente modificada exclusivamente en IDC. Como resultado, se identificaron 24 genes regulados positivamente (Figura 4C, Tabla 7) y 41 genes regulados negativamente (Figura 4C, Tabla 8). De los genes regulados positivamente, ya se sabía que ERBB2, CCNB1, BUB1B estaban implicados en carcinogénesis de cánceres de mama (Latta, E. K., et al., The role of HER2/neu overexpression/amplification in the progression of ductal carcinoma in situ to invasive carcinoma of the breast. *Mod Pathol*, 15: 1318-1325, 2002.; Takeno, S., et al., Prognostic value of cyclin B1 in patients with esophageal squamous cell carcinoma. *Cancer*, 15: 94: 2874-2881, 2002.; Slamon, D. J., et al., Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science*, 235: 177-182, 1987.). De los genes regulados negativamente, se incluyó AXUD1, un gen inducido por AXIN, que con frecuencia se regulaba negativamente en cánceres de pulmón, hígado, colon y riñón (Ishiguro, H., et al., Identification of AXUD1, a novel human gene induced by AXIN1 and its reduced expression in human 20 carcinomas of the lung, liver, colon and kidney. *Oncogene*, 20: 5062-5066, 2001.) lo que sugería que AXUD1 también podía estar implicado en la carcinogénesis del cáncer de mama.

#### Verificación de genes seleccionados por RT-PCR semicuantitativa:

Para confirmar la fiabilidad de los datos de expresión obtenidos por análisis de micromatriz de ADNc, se realizaron experimentos RT-PCR semicuantitativos para 3 genes (nºs de acceso Al261804, AA205444, AA167194) que se regulaban altamente de manera positiva en casos informativos con tipo bien diferenciado y 2 genes (AA676987 y H22566) que también se regulaban altamente de manera positiva en casos informativos con tipo mal diferenciado. Los resultados de la RT-PCR eran muy concordantes con los del análisis de micromatriz en la gran mayoría de los casos ensayados (Figura 5, Tabla 9).

Identificación de A7870, denominado proteína quinasa originado en células T-LAK, como un gen regulado

positivamente en células de cáncer de mama

Se identificaron 24 genes que se regulaban positivamente en IDC (Tabla 7). Entre estos, los autores de la invención se centraron en A7870, denominado proteína quinasa originado en células T-LAK, TOPK (Acceso Genbank NM\_018492), que se localiza en el cromosoma 8p21.2 con un transcripto de ARNm de 1899 bases de longitud que consiste en 8 exones. La expresión de A7870 era elevada en 30 de 39 casos (77%) de cáncer de mama que pudieron obtener datos de expresión, especialmente en 29 de 36 casos (81%) con muestras de ensayo de carcinoma ductal invasivo. Para confirmar el modelo de expresión de este gen en cánceres de mama, se realizó análisis RT-PCR semicuantitativo usando líneas celulares de cáncer de mama y tejidos humanos normales incluyendo células de mama normales. Como resultado, se encontró que A7870 mostró expresión elevada en 7 de 12 muestras clínicas de cáncer de mama (tipo bien diferenciado) en comparación con células ductales de mama normales y otros tejidos normales (Figura 6a) y se sobreexpresó en 17 de 20 líneas celulares de cáncer de mama (Figura 6b). Para examinar adicionalmente el modelo de expresión de este gen, se realizaron análisis de transferencia de Northern con tejidos humanos múltiples y líneas celulares de cáncer de mama usando un fragmento de ADNc (320 pb) de A7870 como una sonda (Figura 7a). Como resultado, se observó que se expresaban dos transcritos (de aproximadamente 1,9 kb y 1,8 kb) exclusivamente en testículos y timo de ser humano normales. Cuando se examinó adicionalmente el modelo de expresión de estos transcritos con transferencia de Northern para cáncer de mama, se encontró que ambos transcritos se sobreexpresaban específicamente en líneas celulares de cáncer de mama, en comparación con tejidos normales humanos (Figura 7b).

Aislamiento del transcripto de A7870 expresado específicamente en cáncer de mama

Mediante el análisis de secuenciación de dos transcritos de A7870, ya que dos variantes de A7870 contienen la misma fase de lectura abierta (ORF), los autores de la invención se centraron en TOPK, (número de acceso Genbank NM\_018492), que codifica una proteína que es una serina/treonina quinasa relacionada con la familia quinasa de proteínas quinasas activadas por mitógeno (MAPKK) específicas duales. La predicción informatizada SMART muestra que TOPK contiene pfam, motivo pquinasa en 32 a 320 restos, lo que sugiere que esta proteína podría estar implicada en una ruta de transducción de señales que desempeña una función en la morfogénesis celular y el crecimiento celular.

Localización subcelular de A7870

Para examinar adicionalmente la caracterización de A7870, se examinó la localización subcelular de estos productos génicos en células de mamífero. En primer lugar, se transfecaron transitoriamente plásmidos que expresaban la proteína A7870 (pCAGGS-A7870-HA) en células COS7, un análisis inmunocitoquímico con anticuerpo anti-etiquetaHA y anticuerpo polyclonal TOPK reveló que la proteína A7870 exógena se localizaba en el citoplasma y especialmente, una fuerte señal alrededor de la membrana del núcleo en todas las células COS7 transfecadas (Figura 8a). Además, se examinó la localización subcelular de la proteína endógena con tinción inmunocitoquímica usando un anticuerpo polyclonal anti-TOPK. De manera similar, también se observó que la proteína A7870 estaba en el aparato citoplasmático y alrededor del núcleo en células T47D, BT-20 y HBC5 (Figura 8b).

Efecto inhibitorio sobre el crecimiento de ARN de interferencia pequeño (ARNip) diseñado para reducir la expresión de A7870

Para evaluar la función promotora del crecimiento de A7870, se inactivó la expresión de A7870 endógeno en la línea de cáncer de mama T47D y BT-20, que habían mostrado la sobreexpresión de A7870, mediante la técnica de ARN de interferencia (ARNi) basada en vectores de mamífero (véase Materiales y Métodos). Se examinaron los niveles de expresión de A7870 por experimentos RT-PCR semicuantitativos. Los ARNip específicos de A7870 (si1, si3 y si4) suprimieron significativamente la expresión, en comparación con construcciones de ARNip de control (psiU6BX-LUC o SC). Para confirmar la inhibición del crecimiento celular con los ARNip específicos de A7870, se realizaron ensayos de formación de colonias y de MTT, respectivamente. Como resultado, la introducción de construcciones de ARNip de A7870 suprimió el crecimiento de estas células de cáncer de mama, de acuerdo con el resultado de expresión reducida anterior de este gen. Cada resultado se verificó mediante tres experimentos independientes. Por tanto, el descubrimiento de los autores de la invención sugiere que A7870 tiene una función significativa en el crecimiento celular del cáncer de mama.

Identificación de genes expresados diferencialmente en tipos histopatológicos y diferencia fenotípica en pacientes individuales

Un objetivo de la presente invención era descubrir genes regulados positiva o negativamente de manera consistente a diferentes fenotipos en algunos pacientes. Sin embargo, dado que el cáncer de mama muestra fenotipos heterogéneos y diversos, la diferenciación histopatológica por microscopía no estaba muy clara usando clasificación no supervisada por modelos de expresión de genes como se muestra en la Figura 2. Para examinar esta observación más de cerca se realizó un ensayo de permutación al azar y se extrajeron 206 genes que pueden diferenciar entre casos bien diferenciados y mal diferenciados. Estos 206 genes discriminativos eran todos significativos en el nivel de P<0,01 entre 31 cánceres bien diferenciados y 24 mal diferenciados (Figura 9, Tabla 10). El análisis de agrupamiento jerárquico bidimensional usando estos 206 genes también pudo clasificar los grupos con

respecto a los distintos componentes de IDC (bien diferenciado, moderadamente diferenciado y mal diferenciado). La agrupación del grupo A contenía genes con expresión notablemente aumentada en muestras mal diferenciadas (rama 1 en la línea horizontal); estructura de matriz extracelular (*COL1A2*, *COL3A1* y *P4HA2*), adhesión celular (*LOXL2*, *THBS2* y *TAGLN2*), mientras que la agrupación del grupo B contenía los genes con expresión aumentada en principalmente en muestras bien diferenciadas y moderadamente diferenciadas (rama 2 en la línea horizontal); regulación de la transcripción (*BTF*, *WTAP*, *HTATSF1*), regulador del ciclo celular (*CDC5L*, *CCT7*). Sin embargo, dos muestras mal diferenciadas (muestra nº 10709 y 10781) en el grupo B, mostraron un patrón de expresión que era similar a la identificación bien diferenciada en lugar de a los tipos mal diferenciados. Algunas muestras bien diferenciadas demostraron coexpresión de algunos genes que eran característicos de la identificación mal diferenciada.

#### Desarrollo de puntuaciones predictivas para metástasis de nódulo linfático:

En cáncer de mama, la invasión en los nódulos linfáticos axilares es el factor de pronóstico más importante (Shek, L. L. and Godolphin, W. Model for breast cancer survival: relative prognostic roles of axillary nodal status, TNM stage, estrogen receptor concentration, and 20 tumor necrosis. *Cancer Res*, 48: 5565-5569, 1988.). Para desarrollar una ecuación para conseguir un parámetro de puntuación para la predicción de metástasis de nódulo linfático axilar usando perfiles de expresión de genes seleccionados, se compararon los perfiles de expresión de 20 casos positivos en nódulos y 20 casos negativos en nódulos. Siguiendo los criterios descritos anteriormente, primero se seleccionaron los 93 genes discriminantes que mostraron valores *p* de permutación de menos de 0,0001. Después, se obtuvieron los 34 genes superiores en la lista de candidatos que mostraron la mejor separación de casos positivos en nódulos de los negativos (Tabla 11). Como se muestra en la Figura 10A, un análisis de agrupación jerárquica usando estos 34 genes clasificó claramente los 40 casos de cáncer de mama en uno de los dos grupos de acuerdo con el estado de nódulo linfático.

Finalmente, se construyó un sistema de puntuación predictivo que pudiera distinguir claramente casos positivos en nódulos de casos negativos en nódulos usando los perfiles de expresión del conjunto de 34 genes. Para validar adicionalmente este sistema de puntuación, se calcularon puntuaciones para 20 casos positivos en nódulos y 20 casos negativos en nódulos linfáticos que no habían estado entre los usados para la construcción del sistema de puntuación (véase "Materiales y Métodos"). Cuando se separó claramente una puntuación límite de 15,8 para 40 pacientes pertenecientes al grupo de metástasis positiva y negativa (Figura 10B), las puntuaciones de más de 15,8 eran "positivas" y las de 15,8 o inferiores eran "negativas". Para aclarar el sistema adicionalmente, se calculó la puntuación de predicción de metástasis de tumores primarios, 17 casos positivos y 20 casos negativos en nódulos que no habían formado parte del procedimiento original para seleccionar genes discriminatorios. Como se muestra en la Figura 10B y 10C, entre los 17 casos con metástasis en nódulos linfáticos, todos los casos tenían puntuaciones positivas de acuerdo con la definición en la presente memoria, mientras que 18 (90%) de los 20 casos sin metástasis en nódulos linfáticos mostraron puntuaciones negativas. 75 (97%) casos de 77 se ubicaron correctamente de acuerdo con su estado de nódulo linfático pero dos casos negativos en nódulos se ubicaron mal o se ubicaron en el límite o en la región positiva.

#### DISCUSIÓN

El cáncer de mama es una enfermedad multifactor que se desarrolla como resultado de interacciones entre factores genéticos, ambientales y hormonales. Aunque se han descrito distintas fases patológicas del cáncer de mama, las diferencias moleculares entre estas fases se desconocen en gran medida (McGuire, W. L. Breast cancer prognostic factors: evaluation guidelines. *J Natl Cancer Inst*, 83: 154-155, 1991.; Eifel, P., et al., National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, November 1-3, 2000. *J Natl Cancer Inst*, 93: 979-989, 2001.; Fisher, B., et al., Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. *N Engl J Med*, 347: 1233-1241, 2002.).

El desarrollo de análisis pangenómicos de expresión de genes y aislamiento por microdissección con micro-haz láser (LMM) de poblaciones de células cancerosas puras de cáncer de mama permiten la búsqueda de genes diana moleculares con clasificación, tratamiento y predicción obtenida específica del cáncer en una diversidad de tipos tumorales especialmente en cáncer de mama.

Dado que los adipocitos representan más del 90% del tejido glandular mamario y que las células epiteliales en el órgano, a partir de las cuales se origina carcinoma, corresponden a un porcentaje muy pequeño, un análisis de los perfiles de expresión de genes usando tejidos de cáncer completo y glándula mamaria completa normal está significativamente influenciado por la mezcla particular de células en los tejidos examinados; diferencias proporcionales de adipocitos, fibroblastos y células antiinflamatorias pueden enmascarar significativamente la expresión significativa de genes implicados en la carcinogénesis de mama. Por lo tanto, se usó un sistema LMM para purificar tanto como fuera posible las poblaciones de células cancerosas y células epiteliales normales obtenidas a partir de muestras quirúrgicas (Hasegawa, S., et al. Genome wide analysis of gene expression in intestinal-type gastric cancers using a complementary DNA microarray representing 23,040 genes. *Cancer Res*, 62: 7012-7017, 2002.; Kitahara, et 5 al., and Tsunoda, T. Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. *Cancer Res*,

61: 3544-3549,2001.; Kikuchi, T., et al. Expression profiles of non- small cell lung cancers on cDNA microarrays: identification of genes for prediction of lymph- node metastasis and sensitivity to anti-cancer drugs. *Oncogene*, 22: 2192-2205,2003.; 10 Gjerdrum, L. M., et al., Laser-assisted microdissection of membrane-mounted paraffin sections for polymerase chain reaction analysis: identification of cell populations using immunohistochemistry and in situ hybridization. *J Mol Diagn*, 3: 105-110, 2001.), (Figura 1). Para evaluar la pureza de la poblaciones de células microdisecionadas, se analizó la expresión de *PLIN* y *FABP4*, que se expresaban altamente en tejido adiposo y glándula mamaria, por perfiles de expresión de genes en 29 tejidos normales humanos usando una micromatriz de ADNc (Saito-Hisamino, A., et al., Genome-wide profiling of gene expression in 29 normal human tissues with a cDNA microarray. *DNA Res*, 9: 35-45,2002.). Después del procedimiento de disección se calculó que la proporción de adipocitos contaminantes entre las células epiteliales ductales de mama normales era menor del 0,6%. En particular, cuando se examinaron los niveles de expresión de *PLIN* (Nishiu, 20 J., et al., Isolation and chromosomal mapping of the human homolog of perilipin (PLIN), a rat adipose tissue-specific gene, by differential display method. *Genomics*, 48: 254-257,1998.), la pureza de las poblaciones de células sometidas a la técnica LMM podría ser por lo tanto aproximadamente del 100%. Como se muestra en la Figura 2, el análisis de agrupación no supervisado representa que los tejidos completos de cáncer de mama se separaron de células de cáncer de mama microdisecionadas por LMM, mientras que células ductales de mama normales y glándulas mamarias se agruparon en la misma rama. Por lo tanto, para obtener de manera precisa los perfiles de expresión específicos del cáncer de mama en algunos estudios, es esencial microdisecionar células de cáncer de mama y células epiteliales ductales de mama a partir de las cuales se origina el cáncer de mama. El uso combinado de análisis LMM y micromatriz de ADNc proporciona una estrategia poderosa para aclarar sucesos moleculares exactos que rodean el desarrollo y progresión del cáncer de mama y conducen al entendimiento del mecanismo de carcinogénesis multietapa de las células cancerosas de mama y heterogeneidad tumoral.

Como se muestra en la Figura 2A, mediante un análisis de clasificación no supervisado basándose en perfiles de expresión, el cáncer de mama primario puede dividirse en dos grupos y demuestra asociarse con el estado ER mediante EIA. Se descubrió que tumores ER+ y ER- muestran fenotipos de expresión de genes muy diferentes. Este resultado sugiere que estas dos lesiones histológicamente distintas tienen diferentes naturalezas biológicas que pueden desempeñar una función importante en la carcinogénesis del cáncer de mama y además sugiere que el estado de ER puede usarse para establecer la necesidad de terapia hormonal en la composición adyuvante (Eifel, P., et al National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, November 1-3, 2000. *J Natl Cancer Inst*, 93: 979-989, 2001.; Hartge, P. Genes, hormones, and pathways to breast cancer. *N Engl J Med*, 348: 2352-2354,2003.). Además, mediante análisis estadístico supervisado, se seleccionó un subconjunto de genes que podía separar ER positivo de ER negativo para investigar la progresión dependiente de hormonas y se exploraron nuevas dianas moleculares para fármacos anticancerosos. Se identificaron 97 genes cuya expresión es significativamente diferente entre estos dos grupos que consisten en pacientes premenopáusicas mediante un ensayo de permutación al azar (Figura 3). Entre estos genes, se incluyó *MAP2K4*, que es un mediador de las rutas SAPK ubicado centralmente. En este y otros estudios también se incluyó la *ciclina D1*, un gen que está fuertemente asociado con la expresión de ER en el cáncer de mama (May, F. E. and Westley, B. R. Expression of human intestinal trefoil factor in malignant cells and its regulation by oestrogen in breast cancer cells. *J Pathol*, 182: 404-413, 1997.). Los estrógenos son reguladores importantes del crecimiento y diferenciación en la glándula mamaria normal y también son importantes en el desarrollo y progresión del carcinoma de mama (Shek, L. L. and Godolphin, W. Model for breast cancer survival: relative prognostic roles of axillary nodal status, TNM stage, estrogen receptor concentration, and tumor necrosis. *Cancer Res*, 48: 5565-5569,1988.). Los estrógenos regulan la expresión de genes mediante ER (receptores de estrógenos), sin embargo, los detalles del efecto de los estrógenos sobre dianas de genes aguas abajo, la función de cofactores y la interacción entre otras rutas de señalización no se conocen del todo. Dado que aproximadamente dos tercios de todos los cánceres de mama son ER + en el momento del diagnóstico, la expresión del receptor tiene implicaciones importantes para su biología y terapia. Dado que recientemente se han desarrollado nuevos moduladores selectivos de receptores de estrógenos (MSRE) como tratamiento hormonal contra pacientes con cancer de mama ER- positivos, estos genes asociados con el estado ER podrían ser nuevas posibles dianas moleculares para los MSRE (Smith, I. E. and Dowsett, M. Aromatase inhibitors in breast cancer. *N Engl J Med*, 348: 2431-2442, 2003.). Estos hallazgos sugieren que la comparación de los perfiles de expresión y el estado ER proporcionan información útil para aclarar la regulación hormonal de la proliferación celular y la progresión de células de cáncer de mama independientes de ER.

El desarrollo y uso de terapia basada en genética molecular contra el cáncer de mama y otros tumores humanos requiere un análisis genético molecular detallado de los tejidos de los pacientes. Las pruebas histológicas sugieren que existen diversos estados pre-neoplásicos anteriores a tumores de mama invasivos. Estas lesiones histológicas incluyen hiperplasia ductal atípica, hiperplasia lobular atípica, carcinoma ductal in situ (DCIS) y carcinoma lobular in situ (Lakhani, S. R. The transition from hyperplasia to invasive carcinoma of the breast. *J Pathol*, 187: 272-278, 1999.). Se piensa que estas lesiones se incluyen en una escala histológica entre epitelio normal de mama o las unidades lobulares ductales terminales a partir de las cuales surge el cáncer de mama y el cáncer de mama invasivo final. Se han propuesto diversos modelos para explicar las anomalías genéticas entre pre-neoplasia y neoplasia.

Se observaron diversos genes que normalmente mostraron expresión aumentada o disminuida entre los estados patológicamente separados, tales como la comparación entre DCIS e IDC, dando como resultado una identificación total de 325 genes. Estos genes pueden estar bajo la base molecular del grado patológico del cáncer de mama y

niveles de expresión de estos se correlacionan con un grado tumoral avanzado. También se identificaron 78 genes normalmente regulados positivamente (Tablas 3, 5) y 247 genes normalmente regulados negativamente (Tablas 4, 6) en DCIS e IDC. Dentro de los genes regulados positivamente, *NAT1*, *HEC*, *GATA3* y *RAI3*, que se habían descrito que se sobreexpresaban en cáncer de mama, se advirtieron como potencialmente expresados en estados preinvasivos (Geylan, Y. S., et al., Arylamine N-acetyltransferase activities in human breast cancer tissues. *Neoplasma*, 48: 108-111,2001.; Chen, Y., et al., HEC, a novel nuclear protein rich in leucine heptad repeats specifically involved in mitosis. *Mol Cell Biol*, 17: 6049-6056,1997.; Bertucci, F., et al., Gene expression profiling of primary breast carcinomas using arrays of candidate genes. *Hum Mol Genet*, 9: 2981-2991,2000.; Cheng, Y. and Lotan, R. Molecular cloning and characterization of a novel retinoic acid-inducible gene that encodes a putative G protein- coupled receptor. *J Biol Chem*, 273: 35008-35015,1998.). Por otro lado, *TGFBR2*, incluido como un gen regulado positivamente en la presente invención, se sabe que conduce a malignidad reducida (Sun, L., et al., Expression of transforming growth factor beta type II receptor 30 leads to reduced malignancy in human breast cancer MCF-7 cells. *J Biol Chem*, 269: 26449- 26455,1994.). Estos hallazgos sugieren que estos genes pueden estar implicados en la transición de DCIS a IDC.

En particular, se identificaron 25 genes regulados positivamente (Tabla 5) y 49 genes regulados negativamente (Tabla 6) con expresión elevada o disminuida de acuerdo con la transición de DCIS a IDC. La lista de elementos regulados positivamente incluyó genes que codifican factores transcripcionales y proteínas implicadas en la ruta de traducción de señales y en el ciclo celular y que desempeñan un importante papel en la tumorigénesis invasiva. La sobreexpresión de *FoxM1* y *ciclina B1* se ha descrito en diversos tipos de tumores. La sobreexpresión de *FoxM1* estimula la expresión de *ciclina B1* (Leung TW, 2001). CCNB1 es una proteína de control del ciclo celular que es necesaria para el paso a través de G2 y mitosis (Pines, J. and Hunter, T. Cyclins A and B1 in the human cell cycle. *Ciba Found Symp*, 170: 187-196; discussion 196-204,1992.). Los inhibidores *TOP2A* se usan ampliamente como agentes quimioterapéuticos en el tratamiento contra el cáncer de pulmón (Miettinen, H. E., et al., High topoisomerase II alpha expression associates with high proliferation rate and poor prognosis in oligodendroglomas. *Neuropathol Appl Neurobiol*, 26: 504-512, 2000.). *BUB1B* puede ser responsable de un fenotipo de inestabilidad cromosómica que contribuye a la progresión tumoral en el punto de control mitótico e inestabilidad genética (Bardelli, A., et al. Carcinogen specific induction of genetic instability. *Proc Natl Acad Sci USA*, 98: 5770-5775,2001.). La expresión de *MMP11*, mostró tener un efecto negativo directo sobre la supervivencia de pacientes (Boulay, A., et al. High cancer cell death in syngeneic tumors developed in host mice deficient for the stromelysin-3 matrix metalloproteinase. *Cancer Res*, 61:2189-2193,2001.). *ECM1* tiene propiedades angiogénicas y se expresa por células tumorales de mama (Han, Z., et al., 20 Extracellular matriz protein 1 (ECM1) has angiogenic properties and is expressed by breast tumor cells. *Faseb J*, 15: 988-994,2001.). Aunque la mayoría de estas funciones aún se desconocen, la evaluación del análisis funcional de estos genes puede indicar que estos desempeñan una función en la mediación de la actividad invasiva.

En este documento, a través de los perfiles de expresión exactos del cáncer de mama por medio de micromatriz de ADN del genoma completo, se aislaron nuevos genes, comparado con tejidos humanos normales A7870 se sobreexpresaba significativamente en células de cáncer de mama. Además, se demostró que el tratamiento de células de cáncer de mama con ARNip inhibía eficazmente la expresión del gen diana, A7870 y suprimía significativamente el crecimiento celular/tumoral del cáncer de mama. Estos hallazgos sugieren que A7870 podría desempeñar funciones clave en la proliferación del crecimiento celular tumoral y podría ser una diana prometedora para el desarrollo de fármacos anticancerosos.

A7870, denominado *TOPK*, un nuevo miembro de la familia MAPKK , se seleccionó para estudio debido a su expresión elevada significativa en cáncer de mama. Se identificaron los transcritos de aproximadamente 1,8 y 1.9 kb que mostraron expresión específica de cáncer. Estos transcritos tienen secuencias diferentes de UTR 5' pero la misma ORF. Se demostró que el tratamiento de células de cáncer de mama con ARNip inhibía eficazmente la expresión de A7870 y suprimía significativamente el crecimiento celular/tumoral del cáncer de mama. Estos hallazgos sugieren que A7870 podría desempeñar funciones clave en la proliferación del crecimiento celular tumoral y podrían ser una diana prometedora para el desarrollo de los fármacos anticancerosos.

Sin embargo, la capacidad de algunos criterios para predecir la progresión de enfermedades y el resultado clínico es imperfecta. Pacientes con enfermedades más agresivas pueden beneficiarse de quimioterapia con adyuvantes o terapia hormonal y normalmente se identifican de acuerdo con una combinación de criterios: edad, tamaño del tumor, estado del nódulo axilar, tipo histológico y grado patológico de cáncer, estado receptor hormonal. Los tumores histológicamente diferentes se clasificaron por subconjuntos de genes, un proceso que proporciona información patológicamente relevante. La mayoría de los investigadores han sugerido que los pacientes tienen un pronóstico más malo si el tumor muestra un porcentaje significativamente superior de histología mal diferenciada.

Un sorprendente resultado de este estudio fue la notable similitud en los perfiles de expresión de diferentes tipos histológicos en cada paciente. Mediante microdissección y análisis de expresión global de genes, se examinaron cambios en la expresión de genes asociados con invasión y pronóstico usando perfiles de expresión de ARNm procedentes de células de cáncer de mama de tipo bien diferenciado y de tipo mal diferenciado usando análisis supervisado. Mediante un análisis de clasificación no supervisado basándose en perfiles de expresión, el cáncer de mama puede dividirse en dos grupos y muestra estar asociado con lesiones patológicamente diferentes. Mediante un ensayo de permutación al azar (Figura 2C), se identificaron 25 genes cuya expresión es significativamente

diferente entre estos dos grupos constituidos en cada paciente. Entre estos genes, el factor nuclear 5 de células T activadas (*NFAT5*) se limita para promover la migración celular del carcinoma, lo que destaca la posibilidad de diferenciar genes que están inducidos por estos factores de transcripción (Sebastien J. et al., The role of NFAT transcription factors in integrin-mediated carcinoma invasion. *Nature cell biology*, 4: 540-544,2002.). La 5 trombospondina 2 (*THSB2*) es una proteína de la matriz extracelular que parece desempeñar una función en la adhesión celular y migración celular. Una ventaja importante de la estrategia basada en LMM es la capacidad de seleccionar células cancerosas de diferentes fenotipos de la muestra de ensayo. El análisis sistemático de modelos de expresión de genes proporciona una apertura en la biología y patogénesis de invasión.

Adicionalmente, la metástasis en nódulos linfáticos es una etapa crítica en la progresión de tumor y uno de los 10 principales componentes de pronóstico malo en los pacientes con cáncer de mama (Shek, L. L. and Godolphin, W. Model for breast cancer survival: relative prognostic roles of axillary nodal status, TNM stage, estrogen receptor concentration, and tumor necrosis. *Cancer Res*, 48: 5565-5569, 1988.), pero solamente una minoría de pacientes muestran metástasis clínicamente detectable en el diagnóstico. El estado del nódulo linfático en el diagnóstico es la medición más importante para futura reaparición y supervivencia global, es un sustituto de algo imperfecto.

Aproximadamente un tercio de pacientes sin afectación detectable en nódulos linfáticos, por ejemplo, desarrollará enfermedad recurrente a los 10 años (Saphner, T., et al., Annual hazard rates of recurrence for breast cancer after primary therapy. *J Clin Oncol*, 14: 2738-2746,1996.). Se demostró que la biopsia del nódulo linfático centinela era un procedimiento preciso en el estudio de nódulos linfáticos axilares, esto permitió disminuir notablemente la morbilidad relacionada con la cirugía del cáncer de mama y poder evitar la disección axilar. Otros parámetros, tales como 15 clasificación nuclear, edad del paciente, tamaño del tumor no pueden predecir el estado del nódulo linfático axilar y no es posible diagnosticar eficazmente el estado del nódulo linfático mediante biopsia de nódulo linfático centinela. Por lo tanto, la presente identificación de un subconjunto de genes expresados diferencialmente entre tumores 20 nódulo-positivos y nódulo-negativos pueden contribuir a mejorar el diagnóstico clínico y a comprender los sucesos biofísicos exactos. El análisis de grupo (Figura 10) sugirió separar casos con metástasis en nódulos linfáticos de aquellos que no presentaban metástasis. Los genes que contribuyen a la separación de los dos grupos de pacientes 25 de acuerdo con el estado de la metástasis del nódulo linfático pueden servir como marcadores moleculares para la metástasis (Ramaswamy, S., et al., A molecular signature of metastasis in primary solid tumors. *Nat Genet*, 33: 49-54, 2003.). Por ejemplo, entre estos 34 genes, FUS, conocido como TLS por translocarse en el liposarcoma, que 30 está disminuido en cánceres negativos en nódulos se transloca con el gen que codifica el factor de transcripción ERG-1 en leucemias mieloideas humanas. Una de las funciones importantes del FUS de tipo silvestre es el mantenimiento del genoma, particularmente el mantenimiento de la estabilidad genómica (Hicks, G. G., et al., Fus deficiency in mice results in defective B-lymphocyte development and activation, high levels of chromosomal instability and perinatal death. *Nat Genet*, 24: 175-179,2000.). Los niveles de expresión aumentaron en algunos de 35 estos genes en el grupo positivo a metástasis en comparación con el grupo negativo. Por ejemplo, en lo que respecta a *EEF1D*, la mayor expresión de EF-1 delta en los tumores sugirió que la transformación cancerosa *in vivo* requiere un aumento en el factor de traducción del ARNm y síntesis de proteína para la entrada en y transición a través del ciclo celular. La transducción de señal de la proteína Rho, CFL1, y GTPasas de la familia Rho regulan el citoesqueleto y la migración celular y se sobreexpresan frecuentemente en tumores (Yoshizaki, H., et al., Activity of Rho-family GTPases during cell division as visualized with FRET-based probes. *J Cell Biol*, 162: 223-232,2003.;

Arthur, W. T., et al., Regulation of Rho family GTPases by cell-cell and cell- matrix adhesion. *Biol Res*, 35: 239-246, 2002.). Se demostró que BRAF, la B-Raf quinasa era capaz de fosforilar y activar MEK como resultado de estimulación del factor de crecimiento. Aunque la función de alguno de estos genes aún se desconoce, el entendimiento de la función de estos genes puede aclarar sus funciones en la metástasis en cáncer de mama.

Las causas y el ciclo clínico de reaparición se desconocen actualmente. Además no es posible predecir resultados 45 fiables basándose en la disponibilidad clínica, patológica y marcadores genéticos. Aunque se piensa que el sistema de puntuación predictiva de la presente invención, usando los perfiles de expresión de estos 34 genes puede ser útil para mejorar el pronóstico, la verificación usando una gran cantidad de casos puede ser necesaria para la introducción en etapas clínicas. En cualquier caso, la presente invención parece proporcionar información exacta sobre la naturaleza biológica de las células cancerosas que se ha malinterpretado por diagnóstico histológico convencional.

Las terapias contra el cáncer dirigidas a modificaciones moleculares específicas que se producen en células 50 cancerosas se han validado mediante desarrollo clínico y aprobación reguladora de fármacos anticancerosos tales como trastuzumab (Herceptina) para el tratamiento de cáncer de mama avanzado (Coussens, L., et al. Tyrosine kinase receptor with extensive homology to EGF receptor shares 20 chromosomal location with neu oncogene. *Science*, 230: 1132-1139, 1985.). Este fármaco es clínicamente eficaz y se tolera mejor que los agentes anticancerosos tradicionales porque se dirige solamente a células transformadas. Por lo tanto este fármaco no solo mejora la supervivencia y la calidad de vida de los pacientes con cáncer, sino que también valida el concepto de terapia contra el cáncer dirigida molecularmente. Además, los fármacos dirigidos pueden potenciar la eficacia de 55 quimioterapia convencional cuando se usan en combinación con los mismos (Gianni, L. and Grasselli, G. Targeting the epidermal growth factor receptor a new strategy in cancer treatment. *Suppl Tumori*, 1: S60-61, 2002.; Klejman, A., et al., Phosphatidylinositol-3 kinase inhibitors enhance the anti-leukemia effect of ST1571. *Oncogene*, 21: 5868-5876,2002.). Por lo tanto, tratamientos futuros contra el cáncer probablemente implicarán la combinación de 60 fármacos convencionales con agentes específicos-diana dirigidos a diferentes características de células tumorales

tales como angiogénesis e invasividad. Adicionalmente, la presente invención demuestra que los nuevos marcadores tumorales, sustancias que pueden estar presentes en la sangre en cantidades anómalas, o aspirados del pezón de una mujer con cáncer de mama, pueden ser lo suficientemente fiables para usarse rutinariamente para detectar el cáncer de mama temprano.

- 5 Actualmente, no se dispone de tratamiento eficaz para pacientes con cáncer de mama avanzado. Por lo tanto, se requieren urgentemente nuevas estrategias terapéuticas y tratamiento a medida. Los perfiles de expresión específicos contra el cáncer de la presente invención, incluyendo genes regulados negativa y positivamente, deben proporcionar información útil para identificar dianas moleculares para el tratamiento de pacientes.

Tabla 1 Listado de genes con expresión modificada entre el tipo bien y mal diferenciado en fenotipo histológico.

| Nº de BRC | Nº de acceso | Símbolo   | Título  | Valor de p |
|-----------|--------------|-----------|---|------------|
| 1         | AF053712     | TNFSF11   | Superfamilia del factor de necrosis tumoral (ligando), miembro 11   | 1.2E-06    |
| 2         | BF973104     | LOC201725 | Proteína teórica LOC201725  | 3.2E-05    |
| 3         | AV752313     | KPNA6     | carioferina alfa 6 (importina alfa 7)   | 1.1E-04    |
| 4         | AK026898     | FOXP1     | Caja de cabeza de tenedor P1  | 7.4E-04    |
| 5         | AA148107     | ITGA5     | integrina, alfa 5 (receptor de fibronectina, polipéptido alfa)  | 7.9E-04    |
| 6         | AK001067     | NFAT5     | Factor nuclear de células T activadas 5, sensible a tonicidad   | 8.2E-04    |
| 7         | AB007919     | KIAA0450  | Producto génico KIAA0450  | 1.8E-03    |
| 8         | BG026429     | SFRS2     | Factor de corte y empalme, rico en arginina/serina 2  | 2.0E-03    |
| 9         | M87770       | FGFR2     | Receptor del factor de crecimiento de fibroblastos 2 (quinasa expresada en bacterias, receptor del factor de crecimiento de queratinocitos, diostosis craneofacial 1, síndrome de Crouzon, síndrome de Pfeiffer, síndrome de Jackson-Weiss) | 2.1E-03    |
| 10        | L02785       | SLC26A3   | Familia transportadora de solutos 26, miembro 3   | 2.7E-03    |
| 11        | BF037402     |           | Homo sapiens, clon MGC: 17296 IMAGEN:3460701, ARNm, cds completo  | 2.8E-03    |
| 12        | L12350       | THBS2     | trombospondina 2  | 2.8E-03    |
| 13        | N36875       |           | Homo sapiens, clon IMAGEN.-4994678, ARNm  | 3.8E-03    |
| 14        | AL135342     |           | EST (etiquetas de secuencia expresada), débilmente similar a la proteína de la cadena neural [Homo sapiens] [H.sapiens]   | 4.3E-03    |
| 15        | AL049426     | SDC3      | syndecan 3 (N-syndecan)   | 4.5E-03    |
| 16        | AW961424     | KIAA1870  | Proteína KIAA1870   | 5.2E-03    |
| 17        | AA523117     | DC-TM4F2  | tetraspanina similar a TM4SF9   | 5.5E-03    |
| 18        | Z11531       | EEF1G     | Factor 1 gamma de alargamiento de la traducción   | 6.1E-03    |
| 19        | AI423028     | SMARCD3   | Regulador de cromatina dependiente de actina, asociado a la matriz, relacionado con SWI/SNF, subfamilia d, miembro 3  | 6.8E-03    |

| Nº de BRC | Nº de acceso | Símbolo | Título  | Valor de p |
|-----------|--------------|---------|---|------------|
| 20        | AB002391     | MN7     | D15F3 7 (pseudogen)                           | 7.1E-03    |
| 21        | D32050       | AARS    | alanil-ARNt sintetasa                         | 7.2E-03    |
| 22        | BE876949     | RAB7    | RAB7, miembro de la familia de oncogenes RAS  | 7.9E-03    |
| 23        | AW291083     |         | EST (ETIQUETAS DE SECUENCIA EXPRESADAS)       | 8.0E-03    |
| 24        | AI568910     |         | EST (ETIQUETAS DE SECUENCIA EXPRESADAS)       | 8.2E-03    |
| 25        | AK023480     | SRP72   | Partícula de reconocimiento de señal de 72kDa | 8.7E-03    |

Tabla 2 Listado de genes con expresión modificada entre tumores ER-positivos ER-negativos

| Nº de BRC | Nº de acceso | Símbolo           | Título   | Valor de p |
|-----------|--------------|-------------------|--|------------|
| 26        | AW949747     | GATA3             | Proteína 3 de unión a GATA   | 3.2E-20    |
| 27        | BE868254     | ESTs              | EST  | 2.2E-14    |
| 28        | AF037335     | CA12              | Anhidrasa carbónica XII  | 1.6E-13    |
| 29        | BF724977     | ASB13             | Repetición de anquirina y que contiene caja SOCS 13  | 8.5E-13    |
| 30        | NM_004636    | SEMA3B            | dominio sema, dominio de immunoglobulina (Ig), dominio corto básico, secretado, (semaforina) 3B                    | 9.7E-13    |
| 31        | NM_000125    | ESR1              | receptor de estrógenos 1   | 1.2E-12    |
| 32        | M73554       | CCND1             | ciclina D1 (PRADI: adenomatosis paratiroidea 1)  | 3.9E-12    |
| 33        | NM_005544    | IRS1              | sustrato del receptor de insulina 1  | 4.4E-12    |
| 34        | M14745       | BCL2              | CLL célula -B/Linfoma 2  | 5.1E-12    |
| 35        | BE826171     | BCMP11            | proteína de membrana de cáncer de mama 11  | 2.8E-11    |
| 36        | AI087270     | SIAH2             | siete en ausencia de homólogo 2 ( <i>Drosophila</i> )  | 2.8E-11    |
| 37        | L07033       | HMGCL             | 3 -hidroximetil-3 -metilglutaril-Coenzima A liasa (hidroximetilglutaricaciduria)                                   | 2.8E-11    |
| 38        | AB014523     | ULK2              | quinasa 2 de tipo 51-unc ( <i>C. elegans</i> )   | 4.0E-11    |
| 39        | AL137588     | DKFZp43<br>4K1210 | proteína teórica DKFZp434K1210   | 5.2E-11    |
| 40        | AL137566     | EST               | ARNm de Homo sapiens; ADNc de DKFZp586G0321 (del clón DKFZp586G0321)   | 5.4E-11    |
| 41        | AF038421     | GFRA1             | receptor alfa 1 de la familia GDNF   | 8.4E-11    |
| 42        | AI194045     | FE65L2            | Proteína 2 de tipo FE65-   | 9.2E-11    |
| 43        | BG163478     | EST               | EST, ligeramente similar al precursor del inhibidor 1 de angiogénesis específico de cerebro BAI1_HUMAN [H.sapiens] | 1.1E-10    |
| 44        | M31627       | XBP1              | proteína 1 de unión a la caja X  | 1.1E-10    |
|           | AA156269     | EST               | Homo sapiens, clón IMAGE:4794107, ARNm   | 1.3E-10    |
| 46        | NM_006763    | BTG2              | familia BTG, miembro 2   | 1.9E-10    |

| Nº de BRC | Nº de acceso | Símbolo   | Título   | Valor de p |
|-----------|--------------|-----------|--|------------|
| 47        | AW504052     | SEC15L    | SEC 15 similar a ( <i>S. cerevisiae</i> )                                      | 2.1E-10    |
| 48        | NM 005400    | PRKCE     | proteína quinasa C, epsilon  | 2.3E-10    |
| 49        | AI628151     | XBP1      | proteína 1 de unión a la caja X  | 2.7E-10    |
| 50        | AF043045     | FLNB      | filamina B, beta (proteína de unión a actina 278)                              | 3.5E-10    |
| 51        | U31383       | GNG10     | proteína de unión al nucleótido guanina (proteína G), gamma 10                 | 4.6E-10    |
| 52        | L10333       | RTN1      | reticulón 1  | 5.6E-10    |
| 53        | AK025099     | SIGERR    | Molécula relacionada con IL-IR de Ig sencilla                                  | 6.2E-10    |
| 54        | AL039253     | LIV-1     | proteína LIV-1, regulada por estrógenos  | 7.4E-10    |
| 55        | AW949662     | KIAA023 9 | proteína KIAA0239  | 8.0E-10    |
| 56        | D13629       | KTN1      | quinectina 1 (receptor de quinesina)   | 1.5E-09    |
| 57        | NM_000165    | GJA1      | proteína de unión a hueco, alfa 1, 43 kDa (conexina 43)                        | 1.5E-09    |
| 58        | AA533079     | Clorf21   | Fase de lectura abierta 21 del cromosoma 1                                     | 1.8E-09    |
| 59        | AF251056     | CAPS2     | Calcifofina 2  | 1.9E-09    |
| 60        | AF061016     | UGDH      | UDP-glucosa deshidrogenasa   | 2.0E-09    |
| 61        | U92544       | MAGED2    | Antígeno de melanoma, familia D, 2   | 2.1E-09    |
| 62        | BE617536     | RPL13A    | Proteína ribosomal L13a  | 2.4E-09    |
| 63        | AK024102     | MYST1     | MYST histona acetiltransferasa 1   | 2.5E-09    |
| 64        | BF212902     | EST       | ARNm de homo sapiens; ADNc DKFZp564F053 (clon DKFZp564F053)                    | 2.8E-09    |
| 65        | AK025480     | FLJ21827  | proteína teórica FLJ21827  | 3.0E-09    |
| 66        | AI376713     | ESTs      | ESTs, muy similar a la proteína hipotética FLJ20378 [Homo sapiens] [H.sapiens] | 3.6E-09    |
| 67        | AI028483     | ESTs      | ESTs   | 3.8E-09    |
| 68        | AK022249     | EST       | ADNc FLJ12187 fis de homo sapiens cDNA, clon MAMMA1000831.                     | 4.2E-09    |
| 69        | AI568527     | EST       | ADNc FLJ34849 fis de homo sapiens, clon NT2NE2011687.                          | 5.0E-09    |
| 70        | AL133074     | TP53INP1  | proteína 1 nuclear inducible por p53 tumoral proteína tumoral                  | 5.3E-09    |
| 71        | AF022116     | PRKAB1    | Proteína quinase, activada por AMP, subunidad no catalítica beta 1             | 6.1E-09    |
| 72        | AF007170     | Clorf34   | Fase de lectura abierta 34 del cromosoma 1                                     | 9.7E-09    |
| 73        | AF042081     | SH3BGR L  | Dominio SH3 de unión a ácido glutámico rico de tipo proteína                   | 1.2E-08    |
| 74        | AK027813     | MGC1074 4 | Proteína teórica MGC10744  | 1.4E-08    |

| Nº de BRC | Nº de acceso | Símbolo   | Título   | Valor de p |
|-----------|--------------|-----------|--|------------|
| 75        | M57609       | GLI3      | Miembro GLI3 de la familia GLI-Kruppel (síndrome de cefalopolisintactilia de Greig)                        | 1.7E-08    |
| 76        | AL359600     | EST       | ARNm de homo sapiens; ADNc DKFZp547C136 (del clon DKFZp547C136)  | 1.9E-08    |
| 77        | BQ006049     | TTMP1     | Inhibidor tisular de metaloproteinasa 1 (actividad potenciadora eritroide, inhibidor de colagenaza)        | 2.1E-08    |
| 78        | AF111849     | HELO1     | Homólogo de enzima 2 de alargamiento de ácidos grasos poliinsaturados de cadena larga homólogo de levadura | 2.2E-08    |
| 79        | AL157499     | RAB5EP    | Rabaptina-5  | 2.2E-08    |
| 80        | AK023199     | EST       | ADNc FLJ13137 fis de homo sapiens, clon NT2RP3003150.  | 2.5E-08    |
| 81        | J05176       | SERPINA 3 | Inhibidor proteinasa de serina (o cisteína), clado A (alfa-1 antiproteinasa, antitripsina), miembro 3      | 3.2E-08    |
| 82        | AA028101     | KIAA030 3 | proteína KIAA0303  | 3.3E-08    |
| 83        | AI300588     | MAP2K4    | proteína quinasa quinasa 4 activada por mitógeno   | 4.1E-08    |
| 84        | AA682861     | ESTs      | EST, moderadamente similar a la proteína teórica FLJ20378 [Homo sapiens] [H. sapiens]                      | 4.6E-08    |
| 85        | M26393       | ACADS     | acil-Coenzima A deshidrogenasa, C-2 a C-3 de cadena corta  | 5.4E-08    |
| 86        | NM_001609    | ACADSB    | acil-Coenzima A deshidrogenasa, cadena corta/ramificada  | 5.5E-08    |
| 87        | U91543       | CHD3      | proteína 3 de unión a ADN helicasa dominio cromo   | 5.7E-08    |
| 88        | AK023813     | FLJ10081  | Proteína teórica FLJ10081  | 6.0E-08    |
| 89        | BF111711     | FLJ20727  | Proteína teórica FLJ20727  | 7.0E-08    |
| 90        | AL049987     | EST       | ARNm de homo sapiens; ADNc de DKFZp564F112 (del clon DKFZp564F112)   | 7.2E-08    |
| 91        | AW081894     | EST       | EST  | 8.2E-08    |
| 92        | AK000350     | FLJ20343  | Proteína teórica FLJ20343  | 1.1E-07    |
| 93        | AA418493     | DPP7      | Dipeptidilpeptidasa 7  | 1.1E-07    |
| 94        | BE674061     | PIN4      | proteína (peptidil-prolil cis/trans isomerasa) interactúa con NIMA, 4 (parvulina)                          | 1.2E-07    |
| 95        | AB011155     | DLG5      | Homólogo 5 del disco grande, (Drosophila)  | 1.2E-07    |
| 96        | L15203       | TFF3      | Factor 3 trébol (intestinal)   | 1.4E-07    |
| 97        | NM_001552    | IGFBP4    | Proteína 4 de unión al factor de crecimiento de tipo insulina  | 1.4E-07    |
| 98        | M57230       | IL6ST     | Transductor de señal de interleucina 6 (gp130, receptor M de oncostatina)                                  | 1.5E-07    |
| 99        | N92706       | EST       | ADNc de FLJ38461 fis de homo sapiens, clon FEBRA2020977.   | 1.5E-07    |

| Nº de BRC | Nº de acceso | Símbolo    | Título  | Valor de p |
|-----------|--------------|------------|---|------------|
| 100       | M30704       | AREG       | Anfiregulina (factor de crecimiento derivado de schwannoma)                               | 1.8E-07    |
| 101       | AB004066     | BHLHB2     | Dominio que contiene hélice-bucle-hélice básico, clase B, 2                               | 2.2E-07    |
| 102       | M15518       | PLAT       | Activador de plasminógeno, tejidos  | 2.3E-07    |
| 103       | BM697477     | ShrmL      | Proteína relacionada con Shroom   | 2.4E-07    |
| 104       | R45979       | CELSR1     | Cadherina, receptor 1 de tipo G EGF LAG de siete pasos (homólogo del flamenco, Drosofila) | 3.0E-07    |
| 105       | AL049365     | EST        | ARNm de homo sapiens; ADNc de DKFZp586A0618 (del clon DKFZp586A0618)                      | 6.5E-07    |
| 106       | NM_003225    | TFF1       | factor trébol 1 (cancer de mama, secuencia inducible por estrógenos expresada en)         | 7.1E-07    |
| 107       | AI733356     | EST        | ADNc de FLJ31746 fis de homo sapiens, clon NT2RI2007334.                                  | 7.8E-07    |
| 108       | AF078853     | KIAA124 3  | Proteína KIAA1243   | 8.2E-07    |
| 109       | N30179       | PLAB       | Factor de diferenciación de próstata  | 1.0E-06    |
| 110       | BG026429     | SFRS2      | Factor de corte y empalme, rico en arginina/serina 2                                      | 2.4E-06    |
| 111       | AU149272     | ESTs       | EST   | 2.5E-06    |
| 112       | J03827       | NSEP1      | Proteína 1 de unión al elemento sensible a nucleasa                                       | 3.0E-06    |
| 113       | AJ276469     | C20orf35   | Fase de lectura abierta 35 del cromosoma 20   | 3.4E-06    |
| 114       | AW295100     | LOC2015 62 | Proteína teórica LOC201562  | 3.9E-06    |
| 115       | J03817       | GSTM1      | Glutatión S-transferasa M1  | 4.8E-06    |
| 116       | AF288571     | LEF1       | Factor 1 de unión a potenciador linfoide  | 5.1E-06    |
| 117       | AF069301     | PECI       | Isomerasa peroxisomal D3,D2-enoil-CoA   | 5.3E-06    |
| 118       | AA621665     | EST        | EST   | 6.7E-06    |
| 119       | AI739486     | ESTs       | EST   | 8.0E-06    |
| 120       | X81438       | AMPH       | Anfifisina (síndrome de Stiff-Man con cáncer de mama autoantígeno 128 kDa)                | 8.7E-06    |
| 121       | U89606       | PDXK       | Piridoxal (piridoxina, vitamina B6) quinase   | 8.8E-06    |
| 122       | NM_017555    | EGLN2      | Homólogo 2 de egl nueve ( <i>C. elegans</i> )   | 9.2E-06    |

Tabla 3 Genes normalmente regulados positivamente en DCIS e IDC

| BRC N°. | Nº de acceso | Símbolo   | TÍTULO   |
|---------|--------------|-----------|--|
| 123     | D90041       | NAT1      | N-acetiltransferasa 1 (arilamina N-acetiltransferasa)  |
| 124     | M13755       | G1P2      | interferón, proteína alfa inducible (clon IFI-15K)   |
| 125     | D88308       | SLC27A2   | familia 27 transportadora de solutos (transportador de ácidos grasos), miembro 2   |
| 126     | AW23506 1    | NM_004170 | SLC1A1 familia 1 transportadora de solutos (transportadora de glutamato de alta afinidad neuronal/epitelial, sistema Xag), miembro 1 |
| 127     | K02215       | AGT       | inhibidor proteinasa angiotensinógeno (serina (o cisteína), clado A (alfa-1 antiproteinasa, antitripsina), miembro 8)                |
| 128     | AB032261     | SCD       | estearoil-CoA desaturasa (delta-9-desaturasa)  |
| 129     | NM 0009 09   | NPY1R     | receptor Y1 del neuropéptido Y   |
| 130     | AF017790     | HEC       | altamente expresado en cáncer, septeto de repeticiones rico en leucina   |
| 131     | NM 0070 19   | UBE2C     | enzima E2C conjugante de ubiquitina  |
| 132     | AF065388     | TSPAN-1   | tetraspanina 1   |
| 133     | N70334       | DUSP10    | fosfatasa 10 de especificidad dual   |
| 134     | AA621719     | NM_005496 | mantenimiento estructural SMC4 de los cromosomas 4 de tipo 1 (levadura)  |
| 135     | AA676987     |           | EST  |
| 136     | AK001402     | NM 018131 | C10orf3 fase de lectura abierta 3 del cromosoma 10   |
| 137     | AW94974 7    | NM002051  | GATA3 proteína 3 de unión a GATA   |
| 138     | AK001472     | NM_018685 | ANLN anillin, proteína de unión a actina (homólogo de restos, Drosophila)  |
| 139     | AA789233     | NM 000088 | COL1A1 colágeno, tipo I, alfa 1  |
| 140     | AF070632     |           | secuencia de ARNm del clon 24405 de Homo sapiens   |
| 141     | H04544       | NPY1R     | receptor Y1 del neuropéptido Y   |
| 142     | AIO15982     | CDCA1     | ciclo de división celular asociado 1   |
| 143     | NM 0039 79   | RAI3      | ácido retinóico inducido 3   |
| 144     | BF516445     | NM 053277 | CLIC6 canal 6 intracelular de cloruro  |
| 145     | AI361654     |           |  |
| 146     | AI077540     | NM_178530 | ADNc de FLJ38379 fis de Homo sapiens, clon FEBRA2002986.   |
| 147     | AI261804     |           | ARNm MSTP020 (MST020) de Homo sapiens, cds completo  |
| 148     | AK026559     | TPM3      | tropomiosina 3   |

| BRC N°. | Nº de acceso |           | Símbolo   | TÍTULO   |
|---------|--------------|-----------|-----------|--|
| 149     | J03473       |           | ADPRT     | ADP-ribosiltransferasa (NAD <sup>+</sup> ; poli (ADP-ribosa) polimerasa)     |
| 150     | NM 0001 87   |           | HGD       | homogentisato 1,2-dioxigenasa (homogentisato oxidasa)                        |
| 151     | L43964       |           | PSEN2     | presenilina 2 (enfermedad de Alzheimer 4)                                    |
| 152     | J05581       |           | MUC1      | mucina 1, transmembrana  |
| 153     | AA602499     | XM 379784 | GLCCI1    | transcripto 1 inducido por glucocorticoides                                  |
| 154     | U37707       |           | MPP3      | proteína de membrana, palmitoilada 3 (miembro e de la subfamilia p55 MAGUK)  |
| 155     | AB030905     |           | CBX3      | homólogo 3 de la caja cromo (homólogo HP1 gamma, Drosophila)                 |
| 156     | AL138409     | NM_198278 |           | ARNm de Homo sapiens; ADNc de DKFZp313L231 (clon de DKFZp313L231)            |
| 157     | AV756928     |           | SEC61G    | Sec61 gamma  |
| 158     | AI205684     | NM 021979 | HSPA2     | proteína 2 de 70kDa de choque térmico  |
| 159     | BE739464     | NM015161  | ARL6IP    | proteína que interacciona con el factor de tipo 6 de la ribosilación del ADP |
| 160     | AI081356     | NM_203463 | LOC253782 | proteína teórica LOC253782   |
| 161     | AA167194     |           | LOC253782 | proteína teórica LOC253782   |
| 162     | M90516       |           | GFPT1     | glutamina-fructosa-6-fosfate transaminasa 1                                  |
| 163     | AL133074     | NM_033285 | TP53INP1  | proteína 1 nuclear inducible por p53 proteína tumoral                        |
| 164     | AL137257     |           |           | clon IMAGE:5296692, ARNm de Homo sapiens                                     |
| 165     | AK025240     | NM_147128 | LOC223082 | LOC223082  |
| 166     | AJ007042     |           | WHSC1     | candidato1 del síndrome de Wolf-Hirschhorn                                   |
| 167     | U42068       |           | GRP58     | proteína regulada por glucosa, 58kDa   |
| 168     | AJ132592     |           | ZNF281    | proteína 281 de dedos de cinc  |
| 169     | W93638       |           |           | EST  |
|         | AW97739 4    | -         | C9orf12   | fase de lectura abierta 12 del cromosoma 9                                   |
| 171     | AI347925     | NM_001540 | HSPB1     | proteína 1 de choque térmico de 27kDa  |
| 172     | AK026587     |           | NET-6     | NET-6 de 4 tramos miembro de la superfamilia transmembrana 4                 |
| 173     | AI264621     |           | LASS2     | homólogo 2 de aseguramiento de longevidad LAG1 ( <i>S. cerevisiae</i> )      |
| 174     | AA767828     | XM_035527 | FLJ10980  | proteína teórica FLJ10980  |
| 175     | AU142881     | NM_018184 | FLJ10702  | proteína teórica FLJ10702  |

Tabla 4 Genes normalmente regulados negativamente en DCIS e IDC

| BCR<br>Nº | Nº de ACCESO | Símbolo   | TÍTULO  |
|-----------|--------------|-----------|---|
| 176       | X52186       | ITGB4     | integrina, beta 4   |
| 177       | NM 0062 97   | XRCC1     | reparación por rayos-X que complementa la reparación defectuosa en células 1 de hámster Chino                                     |
| 178       | X73460       | RPL3      | proteína L3 ribosomal   |
| 179       | NM_001436    | FBL       | fibrillarina  |
| 180       | X59373       | HOXD10    | caja homeo D10  |
| 181       | J04208       | MPDH2     | IMP (inosina monofosfato) dehidrogenasa 2   |
| 182       | L24203       | TRIM29    | contiene 29 motivos tripartitos   |
| 183       | LI0340       | NM_001958 | factor 1 de alargamiento de la traducción en eucariotas alfa 2  |
| 184       | J04621       | SDC2      | sindecan 2 (proteoglucano heparan sulfato 1, asociado a la superficie celular, fibroglicano)                                      |
| 185       | L08424       | ASCL1     | complejo achaete-scute de tipo 1 ( <i>Drosophila</i> )  |
| 186       | AI376713     | EST       | EST, muy similar a la proteína teórica FLJ20378 [ <i>Homo sapiens</i> ] [ <i>H.sapiens</i> ]                                      |
| 187       | AK026966     | EST       | ADNc: FLJ23313 lis de <i>Homo sapiens</i> , clon HEP11919.  |
| 188       | NM_001050    | SSTR2     | receptor 2 de somatoestatina  |
| 189       | AA632025     | EST       | EST   |
| 190       | N22918       | NM_144641 | teórica proteína FLJ32332   |
| 191       | AF272043     | ITM2C     | proteína de membrana integral 2C  |
| 192       | M58459       | RPS4Y     | proteína S4 ribosomal, unida a Y  |
| 193       | AI133697     | EST       | clon MGC: 16362 IMAGE: 3927795 de <i>Homo sapiens</i> , ARNm, cds completa  |
| 194       | AA780301     | NM_003793 | catepsina F   |
| 195       | M92843       | ZFP36     | proteína 36 de dedos de cinc, tipo C3H, homólogo (ratón)  |
| 196       | AA570186     | EST       | ADNc de longitud completa de ser humano extreme 5-PRIME del clon CS0DK007IB08 de células HeLa de <i>Homo sapiens</i> (ser humano) |
| 197       | R56906       | EST       | EST   |
| 198       | AF208860     | NM_014452 | tumor necrosis factor receptor superfamilia del receptor del factor de necrosis tumoral, miembro 21                               |
| 199       | AK025216     | TAZ       | co-activador transcripcional con motive de unión a PDZ (TAZ)  |
| 200       | AA758394     | PTPN1     | proteína tirosina fosfatasa, de tipo 1 no receptora   |

| BCR<br>Nº | Nº de ACCESO |           | Símbolo  | TÍTULO  |
|-----------|--------------|-----------|----------|---|
| 201       | AA628530     | NM_016368 | ISYNA1   | mio-inositol 1-fosfato sintasa A1   |
| 202       | AF161416     | NM_003749 | IRS2     | sustrato 2 de receptor de insulina  |
| 203       | AL045916     |           | EST      | EST   |
| 204       | AW34097 2    |           | EST      | ADNc de Homo sapiens: FLJ22864 fis, clon KAT02164.  |
| 205       | AI189414     |           | RNPC2    | región de unión a ARN (RNP1, RRM) que contiene 2  |
| 206       | AV705636     |           | EIF3S6IP | factor 3 de inicio de la traducción en eucariotas, proteína que interacciona con la subunidad 6                         |
| 207       | U28977       |           | CASP4    | caspasa 4, cisteína proteasa relacionada con apoptosis  |
| 208       | AV708528     | NM_018579 | MSCP     | proteína transportadora de solutos mitocondrial   |
| 209       | AA022956     | NM_024667 | FLJ12750 | proteína teórica FLJ12750   |
| 210       | AI928443     |           | EST      | ADNc FLJ38855 fis, de Homo sapiens, clon MESAN2010681.  |
| 211       | U14966       |           | RPL5     | proteína L5 ribosomal   |
| 212       | AI857997     |           | TPBG     | glucoproteína de trofoblastos   |
| 213       | BF697545     |           | MGP      | proteína Gla matricial  |
| 214       | AW57575 4    | NM_152309 | FLJ35564 | proteína teórica FLJ35564   |
| 215       | AI352534     | NM_001753 | CAV1     | caveolina 1, proteína caveolae, 22kDa   |
| 216       | NM_001985    |           | ETFB     | flavoproteína de transferencia de electrones, beta polipéptido  |
| 217       | AI743134     | NM_006216 | SERPINE2 | inhibidor de serina (o cisteína) proteinasa, clado E (nexina, inhibidor de tipo 1 activador de plasminógeno), miembro 2 |
| 218       | AW44470 9    | NM_001777 | CD47     | antígeno CD47 (antígeno relacionado con Rh, traductor de señal asociada a integrina)                                    |
| 219       | BF688910     | NM_001300 | COPEB    | proteína de unión al elemento promotor del núcleo   |
| 220       | AI818579     | NM_181847 | EST      | clon IMAGE:3625286, de Homo sapiens, ARNm, cds parcial  |
| 221       | S95936       |           | TF       | transferrina  |
| 222       | AF074393     |           | RPS6KA5  | proteína ribosomal S6 quinasa, 90kDa, polipéptido 5   |
| 223       | NM_000591    |           | CD14     | antígeno CD 14  |
| 224       | AK027181     | NM_031426 | IBA2     | molécula 2 adaptadora a unión de calcio ionizado  |
| 225       | X73079       |           | PIGR     | receptor polimérico de inmunoglobulina  |
| 226       | NM_001343    |           | DAB2     | fosfoproteína sensible a mitógeno, homólogo 2 deshabilitado ( <i>Drosophila</i> )                                       |

| BCR<br>Nº | Nº de ACCESO | Símbolo   | TÍTULO  |
|-----------|--------------|-----------|---|
|           |              |           | deshabilitado ( <i>Drosophila</i> )   |
| 227       | M31452       | C4BPA     | proteína de unión al componente 4 del complemento, alfa                                       |
| 228       | X07696       | KRT15     | queratina 15  |
| 229       | AFO16004     | GPM6B     | glucoproteína M6B   |
| 230       | NM_0040 78   | CSRP1     | proteína 1 rica en cisteína y glicina   |
| 231       | L36645       | EPHA4     | EphA4   |
| 232       | D78011       | DPYS      | dihidropirimidinasa   |
| 233       | W60630       | NM_032801 | molécula 3 de adhesión en unión   |
| 234       | AW956111     | D4S234E   | segmento de ADN en el cromosoma 4 (único) secuencia expresada 234                             |
| 235       | AF035752     | CAV2      | caveolina 2   |
| 236       | D37766       | LAMB3     | laminina, beta 3  |
| 237       | U66406       | EFNB3     | efrina-B3   |
| 238       | X52001       | EDN3      | endotelina 3  |
| 239       | NM_000856    | GUCY1A3   | guanilato ciclasa 1, soluble, alfa 3  |
| 240       | U60115       | FHL1      | cuatro dominios y medio LIM 1   |
| 241       | D14520       | NM_001730 | factor 5 (intestinal) de tipo Kruppel   |
| 242       | M99487       | FOLH1     | folato hidrolasa (antígeno de membrana específico de próstata) 1                              |
| 243       | U09873       | FSCN1     | homólogo 1 de fascina, proteína integrada con actina ( <i>Strongylocentrotus purpuratus</i> ) |
| 244       | AF017418     | MEIS2     | homólogo 2 (ratón) de Meis1, sitio 1 de integración viral ecotrópica mieloide                 |
| 245       | AF038540     | NM_206900 | reticulón 2   |
| 246       | AF049884     | NM_021069 | proteína ArgBP2 que interacciona con Arg/Abl  |
| 247       | NM_001122    | ADFP      | proteína relacionada con la diferenciación adiposa  |
| 248       | Y09926       | MASP2     | lectina serina proteasa 2 de unión a mannano  |
| 249       | M58297       | ZNF42     | proteína 42 de dedos de cinc (sensible a ácido retinóico específico mieloide)                 |
| 250       | AF035811     | PNUTL2    | cacahuete tipo 2 ( <i>Drosophila</i> )  |
| 251       | L22214       | ADORA1    | receptor de adenosina A1  |
| 252       | AF177775     | CES1      | carboxilesterasa 1 (serina esterasa 1 monocito/macrófago)                                     |
| 253       | U07643       | LTF       | lactotransferrina   |
| 254       | S76474       | NM_006180 | tirosina quinasa neurotrófica, receptor, tipo 2   |

| BCR<br>Nº | Nº de ACCESO |           | Símbolo  | TÍTULO  |
|-----------|--------------|-----------|----------|---|
| 255       | BE299605     | NM_012219 | MRAS     | homólogo del oncogén RAS en músculo   |
| 256       | NM_006225    |           | PLCD1    | fosfolipasa C, delta 1  |
| 257       | NM_005036    |           | PPARA    | receptor activado por el proliferador de peroxisomas, alfa  |
| 258       | M22324       |           | ANPEP    | alanil (membrana) aminopeptidasa (aminopeptidasa N, aminopeptidasa M, aminopeptidasa microsomal, CD 13, p150)                       |
| 259       | BE877416     |           | TGFBR2   | factor de crecimiento, receptor beta II (70/80kDa)  |
| 260       | BE561244     |           | RPL18A   | proteína L18a ribosomal   |
| 261       | AL048962     |           | EST      | clon IMAGE:4243767, ARNm de Homo sapiens  |
| 262       | L08895       |           | MEF2C    | factor 2 de potenciador de la transcripción de la caja MADS, polipéptido C (factor 2C potenciador de miocitos)                      |
| 263       | U48707       |           | PPP1R1A  | proteína fosfatasa 1, reguladora (inhibidora) de la subunidad 1A  |
| 264       | X56134       |           | RPLP2    | proteína grande P2 ribosomal  |
| 265       | D84239       |           | FCGBP    | fragmento Fc de la proteína de unión IgG  |
| 266       | AK026181     |           | PHLDA1   | dominio de tipo homología con pleckstrina, familia A, miembro 1   |
| 267       | K01144       |           | CD74     | antígeno CD74 (polipéptido invariante del complejo de histocompatibilidad mayor, asociado al antígeno de clase II)                  |
| 268       | U25138       |           | KCNMB1   | canal activado por calcio de conductancia grande a potasio, subfamilia M, miembro 1 beta  |
| 269       | X85337       | NM_053025 | MYLK     | miosina, polipéptido quinasa ligero   |
| 270       | D83597       |           | LY64     | homólogo del antígeno 64 de linfocitos, radioprotector 105kDa (ratón)   |
| 271       | NM_004024    |           | ATF3     | factor 3activador de la transcripción   |
| 272       | BF126636     |           | SAA1     | amieloide A1 sérico   |
| 273       | D13789       |           | MGAT3    | manosil (beta-1,4-)glucoproteína beta-1,4-N-acetilglucosaminiltransferasa   |
| 274       | L41142       |           | STAT5A   | transductor de señal y activador de la transcripción 5A   |
| 275       | AB040969     |           | KIAA1536 | proteína KIAA1536   |
| 276       | NM_002153    |           | HSD17B2  | hidroxiesteroide (17-beta) dehidrogenasa 2  |
| 277       | AV646610     | NM_001546 | ID4      | inhibidor de unión a ADN 4, protein hélice-bucle-hélice negativo dominante  |
| 278       | X03663       |           | CSF1R    | receptor 1 del factor estimulante de colonias, anteriormente homólogo del oncogén del virus del sarcoma felino (v-fms) de McDonough |

| BCR<br>Nº | Nº de ACCESO |           | Símbolo       | TÍTULO  |
|-----------|--------------|-----------|---------------|---|
| 279       | U47025       |           | PYGB          | fosforilasa, glucógeno; cerebro   |
| 280       | M81349       |           | SAA4          | amiloide A4 sérico, constitutivo  |
| 281       | AI264201     | NM_000399 | EGR2          | respuesta 2 del crecimiento temprano (homólogo a Krox-20, <i>Drosophila</i> )   |
| 282       | U18018       |           | ETV4          | gen 4 variante ets (proteína de unión al potenciador E1A, E1AF)   |
| 283       | NM_004350    |           | RUNX3         | factor 3 de la transcripción asociado a runt  |
| 284       | BF337516     |           | CRYAB         | cristalina, alfa B  |
| 285       | AF027208     |           | PROML1        | prominina de tipo 1 (ratón)   |
| 286       | D17408       |           | CNN1          | calponina 1, básica, músculo liso   |
| 287       | NM_004010    |           | DMD           | distrofina (distrofia muscular, tipos Duchenne y Becker)  |
| 288       | BF183952     |           | CSTA          | cistatina A (estefina A)  |
| 289       | M16445       |           | CD2           | antígeno CD2 (p50), receptor de eritrocitos de oveja  |
| 290       | AF055015     |           | EYA2          | homólogo 2 ausente en ojos ( <i>Drosophila</i> )  |
| 291       | AI745624     |           | ELL2          | ARN polimerasa II relacionado con ELL, factor de alargamiento   |
| 292       | AK025329     |           | DKFZP566 H073 | proteína DKFZP566H073   |
| 293       | BE745465     | NM_012427 | KLK5          | kallikreina 5   |
| 294       | AK024578     | NM_031455 | DKFZP761 F241 | proteína teórica DKFZp761F241   |
| 295       | AI870306     | XM_380171 | IRX1          | proteína 1 iroqués de la caja homeo   |
| 296       | H37853       | NM_022343 | C9orf19       | fase de lectura abierta 19 del cromosoma 9  |
| 297       | BF000047     |           | EST           | clon de ADNC ZA79C08 inserto de longitud completa de <i>Homo sapiens</i>  |
| 298       | AF126780     |           | RetSDR2       | dehidrogenasa/reductasa 2 de cadena corta retinal   |
| 299       | AI700341     |           | EST           | EST, muy similar a la proteína teórica FLJ20489 [ <i>Homo sapiens</i> ] [ <i>H.sapiens</i> ]  |
| 300       | M87770       |           | FGFR2         | factor 2 del factor del crecimiento de fibroblastos (quinasa expresada en bacterias, receptor del factor de crecimiento de queratinocitos, disostosis craneofacial 1, síndrome de Crouzon, síndrome Pfeiffer, síndrome Jackson-Weiss) |
| 301       | AA452368     | MM_144595 | FLJ30046      | proteína teórica FLJ30046   |
| 302       | NM_021200    |           | PLEKHB1       | contiene el dominio de homología de pleckstrina, familia B (evectinas) miembro 1  |
| 303       | AK026343     |           | hIAN2         | nucleótido 2 inmuno asociado humano   |
| 304       | AF251040     |           | C5orf6        | fase de lectura abierta 6 del cromosoma 5   |

| BCR<br>Nº | Nº de ACCESO |           | Símbolo   | TÍTULO  |
|-----------|--------------|-----------|-----------|---|
| 305       | M87507       |           | CASP1     | caspasa 1, cisteína proteasa relacionada con apoptosis (interleucina 1, beta, convertasa) |
| 306       | M97675       |           | ROR1      | receptor 1 huérfano de tipo receptor tirosina quinasa                                     |
| 307       | NM_020549    |           | CHAT      | colina acetiltransferasa  |
| 308       | X00457       | NM_033554 | HLA-DPA1  | complejo mayor de histocompatibilidad, clase II, DP alfa 1                                |
| 309       | W72411       | NM_003722 | TP73L     | proteína tumoral de tipo p73  |
| 310       | AI769569     |           | EST       | EST   |
| 311       | K02765       |           | C3        | componente 3 de complemento   |
| 312       | AW971490     |           | FLJ14906  | proteína teórica FLJ14906   |
| 313       | AF077044     |           | RPAC2     | probablemente ortólogo de ARN polimerasa 1-3 de ratón (subunidad de 16 kDa)               |
| 314       | H70803       | NM_015278 | KIAA0790  | proteína KIAA0790   |
| 315       | AL050367     | XM_167709 | LOC221061 | proteína teórica LOC221061  |
| 316       | AK001643     | NM_018215 | FLJ10781  | proteína teórica FLJ 10781  |
| 317       | AW18227 3    |           | EST       | ADNc FLJ31517 fis, clon NT2RI2000007 de Homo sapiens.                                     |
| 318       | W67951       |           | EST       | ARNm S6 A-5 humano expresado en células de melanoma con cromosoma 6 suprimido.            |
| 319       | ALU 7605     |           | EST       | ARNm de Homo sapiens; ADNc de DKFZp564N1063 (del clon DKFZp564N1063)                      |
| 320       | AI376418     |           | EST       | ADNc de FLJ35169 fis de Homo sapiens, clon PLACE6012908.                                  |
| 321       | AA683373     |           | EST       | EST   |
| 322       | AK022877     |           | EST       | ADNc FLJ 12815 fis de Homo sapiens, clon NT2RP2002546.                                    |
| 323       | NM_002258    |           | KLRB1     | receptor de tipo lectina de linfocito citolítico subfamilia B, miembro 1                  |
| 324       | M69225       |           | BPAG1     | antígeno del pemfigoide ampolloso 1, 230/240kDa   |
| 325       | AW29957 2    | NM_015461 | EHZF      | dedo de cinc hematopoyético temprano  |
| 326       | BE044467     | NM_005737 | ARL7      | factor de ribosilación de ADP de tipo 7   |
| 327       | AA938297     | NM_017938 | FLJ20716  | proteína teórica FLJ20716   |
| 328       | AA706316     | NM_033317 | ZD52F10   | gen teórico ZD52F10   |
| 329       | AI827230     | NM_153000 | APCDD1    | poliposis coli adenomatosa regulada negativamente de tipo 1                               |
| 330       | AK000251     |           | FLJ20244  | proteína teórica FLJ20244   |
| 331       | N62352       | NM_020925 | KIAA1573  | proteína KIAA1573   |

| BCR<br>Nº | Nº de ACCESO |           | Símbolo   | TÍTULO   |
|-----------|--------------|-----------|-----------|--|
| 332       | H53164       |           | ICSBP1    | proteína 1 de unión a secuencia consenso de interferón   |
| 333       | BE394824     |           | WFDC2     | dominio 2 de núcleo de cuatro puentes disulfuro WAP  |
| 334       | AL117462     | NM_015481 | ZFP385    | probable ortólogo de la proteína de dedos de cinc de ratón 385   |
| 335       | NM_003186    |           | TAGLN     | transgelina  |
| 336       | U58514       |           | CHI3L2    | quitinasa 3 de tipo 2  |
| 337       | AB026125     |           | ART-4     | proteína ART-4   |
| 338       | AL080059     | NM_033512 | KIAA1750  | proteína KIAA1750  |
| 339       | AA747005     |           | SDCCAG43  | antígeno 43 de cáncer de colon serológicamente definido  |
| 340       | NM_005928    |           | MFGE8     | proteína 8 del factor EGF-de glóbulos de grasa de leche  |
| 341       | D62470       | NM_004796 | NRXN3     | neurexina 3  |
| 342       | N29574       |           | RAGD      | proteína Rag D   |
| 343       | K02276       |           | MYC       | homólogo del oncogén viral de la mielocitomatosis v-mic (aviar)  |
| 344       | D78611       |           | MEST      | homólogo del transcripto específico de mesodermo (ratón)   |
| 345       | NM_022003    |           | FXYD6     | FXID dominio que contiene regulador 6 de transporte de iones   |
| 346       | BF508973     |           | RPL13     | proteína L13ribosomal  |
| 347       | NM_001615    |           | ACTG2     | actina, gamma 2, músculo liso, entérico  |
| 348       | R41532       |           | EST       | ESTs, muy similar a poliproteína POL relacionada con retrovirus POL2_RATON [Contiene: Transcriptasa inversa; Endonucleasa] [M. musculus] |
| 349       | AA142875     |           | EST       | ESTs   |
| 350       | U03688       |           | CYP1B1    | citocromo P450, familia 1, subfamilia B, polipéptido 1   |
| 351       | W94363       |           | EST       | clon ZE12G01 de ADNc inserto de longitud completa de Homo sapiens  |
| 352       | W44613       |           | HSJ001348 | ADNc para el gen CO16 expresado diferencialmente   |
| 353       | AL118812     |           | EST       | ARNm de Homo sapiens; ADNc de DKFZp761G1111 (del clon DKFZp761G1111)   |
| 354       | D56064       |           | MAP2      | Proteína 2 asociada a microtúbulos   |
| 355       | BF966838     | NM_172069 | KIAA2028  | similar al dominio PH (homología con pleckstrina)  |
| 356       | AI338625     | NM_014344 | FJX1      | Caja 1 de cuatro uniones (Drosofila)   |

| BCR N° | Nº de ACCESO | Símbolo   | TÍTULO   |
|--------|--------------|-----------|--|
| 357    | AI263022     | EST       | ESTs   |
| 358    | AL050107     | NM_015472 | TAZ<br>co-activador transcripcional con motivo de unión a PDZ (TAZ)  |
| 359    | AI056364     | NM 033210 | FLJ14855<br>proteína teórica FLJ14855  |
| 360    | AI351898     | NM_032581 | DRCTNNB 1A<br>Regulado negativamente por Ctnnb1, a   |
| 361    | AV700003     |           | ARL6IP2<br>proteína 2 que interacciona con el factor 6 de tipo ribosilación de ADP                               |
| 362    | NM 0007 00   |           | ANXA1<br>anexina A1  |
| 363    | M81141       |           | HLA-DQB1<br>Complejo mayor de histocompatibilidad, clase II, DQ beta 1   |
| 364    | AI598227     | NM 024911 | FLJ23091<br>proteína teórica FLJ23091  |
| 365    | BG034740     |           | ROPN1<br>roporina, proteína 1 asociada a rofilina  |
| 366    | AB011175     |           | TBC1D4<br>Familia del dominio TBC1, miembro 4  |
| 367    | AK024449     |           | PP2135<br>proteína PP2135  |
| 368    | AW97877 0    |           | DKFZP566 A1524<br>proteína teórica DKFZp566A1524   |
| 369    | AI821113     |           | ADNc de FLJ36327 fis, clon THYMU2005748 de Homo sapiens.   |
| 370    | AI057450     |           | SLC13A2<br>Familia 13 transportadora de solutos (transportador de dicarboxilato dependiente de sodio), miembro 2 |
| 371    | X86693       |           | SPARCL1<br>SPARC de tipo 1 (mast9, hevin)  |
| 372    | AI224952     | NM 173640 | FLJ40906<br>proteína teórica FLJ40906  |
| 373    | D13639       |           | Ciclina D2   |

Tabla 5 Genes con expresión elevada en la transición de DCIS a IDC

| BCR N° | Nº de ACCESO | Símbolo   | TÍTULO   |
|--------|--------------|-----------|--|
| 374    | U74612       | FOXM1     | Caja M1 de cabeza horquillada  |
| 375    | U63743       | KIF2C     | Miembro 2C, familia quinesina  |
| 376    | D88532       | PIK3R3    | fosfoinositidina-3-quinasa, subunidad reguladora, polipéptido 3 (p55, gamma) |
| 377    | NM 00553 2   | IFI27     | interferón, proteína 27 alfa-inducible                                       |
| 378    | D14657       | KIAA0101  | Producto del gen KIAA0101  |
| 379    | AF030186     | GPC4      | Glicoproteína 4  |
| 380    | Z11566       | STMN1     | Estatimina 1/oncoproteína 18   |
| 381    | U90914       | NM_001304 | CPD<br>Carboxipeptidasa D  |
| 382    | NM 00253 4   | OAS1      | 2',5'-oligoadenilato sintetasa 1, 40/46kDa                                   |
| 383    | S67310       | BF        | Factor-B, properdina   |

| BRC Nº | Nº de ACCESO |            | Símbolo        | TÍTULO  |
|--------|--------------|------------|----------------|---|
| 384    | AA192445     | NM_020182  | TMEPAI         | transmembrana, ARN inducido por andrógenos prostáticos                    |
| 385    | AB003103     |            | PSMD12         | Proteosoma (prosoma, macropáina) subunidad 26S, no-ATPasa, 12             |
| 386    | BE878057     | NM_030796  | DKFZP564 K0822 | proteína teórica DKFZp564K0822  |
| 387    | AB003698     |            | CDC7L1         | Ciclo 7 de división celular CDC7 de tipo 1 ( <i>S. cerevisiae</i> )       |
| 388    | M91670       |            | E2-EPF         | proteína transportadora de ubiquitina                                     |
| 389    | AK023414     |            | FLJ13352       | proteína teórica FLJ13352   |
| 390    | L09235       |            | ATP6V1A1       | ATPasa, transportadora de H+, lisosomal 70kDa, V1 subunidad A, isoforma 1 |
| 391    | AF007152     |            | ABHD3          | Contiene el dominio abhidrolasa 3   |
| 392    | U33632       |            | KCNK1          | Canal de potasio, subfamilia K, miembro 1                                 |
| 393    | AA621719     | NM 0054 96 | SMC4L1         | Mantenimiento estructural de SMC4 de 4 cromosomas de tipo 1 (levadura)    |
| 394    | AF176228     |            | DNMT3B         | ADN (citosina-5)-metiltransferasa 3 beta                                  |
| 395    | H22566       | NM 0807 59 | DACH           | homólogo dachshund ( <i>Drosophila</i> )                                  |
| 396    | All85804     | NM 2124 82 | FN1            | Fibronectina 1  |
| 397    | All89477     | NM 0021 68 | IDH2           | isocitrato deshidrogenasa 2 (NADP+), mitocondrial                         |
| 398    | AA205444     |            | AP1S2          | Complejo 1 de proteína adaptadora-relacionada, subunidad sigma 2          |

Tabla 6 Genes con expresión disminuida en la transición de DCIS a IDC

| BRC Nº | Nº de ACCESO |            | Símbolo  | TÍTULO   |
|--------|--------------|------------|----------|--|
| 399    | AF070609     | NM_004172  | SLC1A3   | Familia transportadora de solutos 1 (transportador de glutamato de alta afinidad glial), miembro 3 |
| 400    | U85267       |            | DSCR1    | Gen 1 de la región crítica del síndrome de Down  |
| 401    | NM 00539 7   |            | PODXL    | De tipo podocalixina   |
| 402    | D13811       |            | AMT      | aminometiltransferasa (proteína T del sistema de escisión glicina)                                 |
| 403    | X53586       |            | ITGA6    | integrina, alfa 6  |
| 404    | L13288       |            | VEPR1    | receptor 1 del péptido intestinal vasoactivo   |
| 405    | M12125       |            | TPM2     | tropomiosina 2 (beta)  |
| 406    | M65066       | NM 0027 35 | PRKAR1B  | proteína quinasa, dependiente de AMPc, reguladora, de tipo I beta                                  |
| 407    | AJ001183     |            | SOX10    | SRY (región Y determinante del sexo)-caja 10   |
| 408    | AW241712     |            | MXI1     | proteína 1 que interacciona con MAX  |
| 409    | AL160111     |            | KIAA1649 | proteína KIAA1649  |

# ES 2 365 732 T3

| BRC Nº | Nº de ACCESO |            | Símbolo  | TÍTULO   |
|--------|--------------|------------|----------|--|
| 410    | X93920       |            | DUSP6    | Fosfatasa 6 de especificidad dual  |
| 411    | AF132734     | NM 0218 07 | SEC8     | proteína secretora SEC8  |
| 412    | AI133467     |            |          | ESTs   |
| 413    | D88153       |            | HYA22    | proteína HYA22   |
| 414    | AF014404     |            | PTE1     | acil-CoA tioestearasa peroxisomal  |
| 415    | BE907755     | NM 0133 99 | C16orf5  | Fase de lectura abierta 5 del cromosoma 16   |
| 416    | AA135341     | NM 0210 78 | GCN5L2   | control general GCN5 de la síntesis de amino-ácidos 5 de tipo 2 (levaduras)                                    |
| 417    | AL110126     |            |          | ARNm de Homo sapiens; ADNc<br>DKFZp564H1916 (del clon<br>DKFZp564H1916)  |
| 418    | BE254330     | NM 0030 45 |          | ARNm de homo sapiens; ADNc<br>DKFZp564D016 (del clon<br>DKFZp564D016)  |
| 419    | BE264353     |            | RBP1     | proteína 1 de unión a retinol, celular   |
| 420    | W75991       |            |          | Clon IMAGE:4249217, de Homo sapiens, ARNm  |
| 421    | AF091434     |            | PDGFC    | factor C de crecimiento derivado de plaquetas  |
| 422    | W67577       |            | CD74     | Antígeno CD74 (polipéptido invariante del complejo mayor de histocompatibilidad, clase II asociada a antígeno) |
| 423    | NM 00299 6   |            | CX3CL1   | ligando 1de quimocina (motivo C-X3-C)  |
| 424    | AA024459     |            |          | ESTs   |
| 425    | NM 00016 3   |            | GHR      | receptor de hormona del crecimiento  |
| 426    | AA858162     | NM 0321 60 | NCAG1    | NCAG1  |
| 427    | BE327623     |            |          | ESTs, muy similar a la proteína teórica FLJ20234 [Homo sapiens] [H.sapiens]                                    |
| 428    | BE671156     |            | MAPRE2   | proteína asociada a microtúbulos, familia RP/EB, miembro 2   |
| 429    | D12614       |            | LTA      | linfotoxina alfa (superfamilia TNF, miembro 1)   |
| 430    | L13720       |            | MGC5560  | proteína teórica MGC5560   |
| 431    | U15131       |            | ST5      | supresión de tumorigenidad 5   |
| 432    | Y00711       |            | LDHB     | lactato deshidrogenasa B   |
| 433    | AI651212     |            |          | ADNc FLJ31125 fis, clon IMR322000819 de Homo sapiens   |
| 434    | M31159       |            | IGFBP3   | proteína 3 de unión al factor de crecimiento de tipo insulina  |
| 435    | NM 01444 7   |            | HSU52521 | Arfaptina 1  |

| BRC Nº | Nº de ACCESO |            | Símbolo  | TÍTULO   |
|--------|--------------|------------|----------|--|
| 436    | AB011089     |            | TRIM2    | contiene 2 motivos tripartitos                     |
| 437    | BF969355     | NM 0026 12 | PDK4     | Piruvato deshidrogenasa quinasa, isoenzima 4       |
| 438    | AK025950     | XM 3711 14 | KIAA1695 | proteína teórica FLJ22297                          |
| 439    | D86961       | NM 0057 79 | LHFPL2   | Compañero de fusión de HMGIC de lipoma de tipo 2   |
| 440    | AK025953     |            |          | ADNc: FLJ22300 fis de Homo sapiens, clon HRC04759  |
| 441    | AJ223812     |            | CALD1    | Caldesmona 1                                       |
| 442    | R40594       |            |          | ADNc: FLJ22845 fis de Homo sapiens, clon KAIA5195. |
| 443    | AF145713     |            | SCHIP1   | proteína 1 que interacciona con schwannomina       |
| 444    | AK024966     |            | FLJ21313 | proteína teórica FLJ21313                          |
| 445    | NM 00559 6   |            | NFIB     | factor nuclear I/B                                 |
| 446    | NM 00161 3   |            | ACTA2    | Actina, alfa 2, músculo liso, aorta                |
| 447    | H03641       | XM 3763 28 | FAM13A1  | familia 13 con similitud de secuencia, miembro A1  |

Tabla 7 Genes normalmente regulados positivamente en IDC

| BRC Nº. | Nº de ACCESO |           | Símbolo  | TÍTULO  |
|---------|--------------|-----------|----------|---|
| 448     | X14420       |           | COL3A1   | colágeno, tipo III, alfa 1 (Síndrome de Ehlers-Danlos tipo IV, dominante autosómico)  |
| 449     | AF044588     |           | PRC1     | regulador de proteína de citocinesis 1  |
|         | AF161499     | ,         | HSPC150  | proteína HSPC150 similar a la enzima conjugadora de ubiquitina  |
| 451     | AA78923 3    | NM_000088 | COL1A1   | colágeno, tipo I, alfa 1  |
| 452     | U16306       |           | CSPG2    | condroitin sulfato proteoglicano 2 (versican)   |
| 453     | NM_004425    |           | ECM1     | proteína 1 de matriz extracelular   |
| 454     | NM_006855    |           | KDELR3   | receptor 3 de retención de proteínas del retículo endoplásmico KDEL (Lys-Asp-Glu-Leu)   |
| 455     | AI972071     | NM 031966 | CCNB1    | Ciclina B1  |
| 456     | AF237709     | NM 018492 | TOPK     | proteína quinasa originada en células T-LAK   |
| 457     | BE747327     |           | HIST1H1C | histona 1, H1c  |
| 458     | J03464       |           | COL1A2   | colágeno, tipo I, alfa 2  |
| 459     | AI080640     | NM 006408 | AGR2     | Homólogo anterior al gradiente 2 ( <i>Xenopus laevis</i> )  |
| 460     | AA971042     |           | RHPN1    | rofilina, proteína 1 de unión a Rho GTPasa  |
| 461     | AI419398     |           | MGC33662 | proteína teórica MGC33662   |
| 462     | AI149552     | NM_004448 |          | ESTs, Moderadamente similar al precursor erbB-2 receptor de la proteína tirosina quinasa ERB2_HUMANO (pl85erbB2) (NEU proto-oncogen) (C-erbB-2) (receptor HER2 de superficie celular de tipo tirosina quinasa) (MLN 19) [H.sapiens] |

| BRC Nº. | Nº de ACCESO |           | Símbolo | TÍTULO   |
|---------|--------------|-----------|---------|--|
| 463     | D14874       |           | ADM     | Adrenomedulina   |
| 464     | X03674       | NM 000402 | G6PD    | glucosa-6-fosfato deshidrogenasa   |
| 465     | NM_002358    |           | MAD2L1  | MAD2 sin parada mitótica-de tipo 1 (levadura)                              |
| 466     | BF214508     |           | CYCS    | citocromo c, somático  |
| 467     | BG03053 6    | NM_001067 | TOP2A   | topoisomerasa (ADN) II alfa 170 kDa  |
| 468     | X57766       |           | MMP11   | metaloproteína de matriz 11 (estromelisina 3)                              |
| 469     | AA02990 0    | NM_015170 | SULF1   | Sulfatasa 1  |
| 470     | AF053306     |           | BUB1B   | BUB1 germinación no inhibida por benzimidazoles 1 homólogo beta (levadura) |
| 471     | AF074002     |           | LGALS8  | lectina, unión a galactósido, soluble, 8 (galectina 8)                     |

Tabla 8 Genes normalmente regulados negativamente en IDC

| BCR Nº | Nº de ACCESO |               | Símbolo | TÍTULO   |
|--------|--------------|---------------|---------|--|
| 472    | NM_004484    |               | GPC3    | glipicano 3  |
| 473    | NM_006219    |               | PDC3CB  | fosfoinositida-3-quinasa, catalítica, polipéptido beta   |
| 474    | BE793000     |               | RBP1    | proteína 1 de unión a retinol, celular   |
| 475    | AL117565     | NM_033<br>027 | AXUD1   | AXIN1 regulada positivamente 1   |
| 476    | BF055342     |               | ZNF6    | proteína 6 de dedos de zinc (CMPX1)  |
| 477    | U03688       |               | CYP1B1  | citocromo P450, familia 1, subfamilia B, polipéptido 1   |
| 478    | AF038193     | NM_004<br>311 |         | Homo sapiens, clon IMAGE:3610040, ARNm   |
| 479    | X72760       | NM_002292     | LAMB2   | laminina, beta 2 (laminina S)  |
| 480    | J03817       |               | GSTM1   | glutatión S-transferasa M1   |
| 481    | M69226       |               | MAOA    | monoamina oxidasa A  |
| 482    | BF690180     | NM_006990     | WASF2   | familia de proteína WAS, miembro 2   |
| 483    | AL133600     |               | STAM2   | molécula adaptadora de traducción de señales (dominio SH3 y motivo ITAM) 2   |
| 484    | AF215981     |               | GPR2    | receptor 2 acoplado a proteína G   |
| 485    | BG149764     |               |         | Homo sapiens, clon IMAGE.5286091, ARNm, cds parcial  |
| 486    | AF067800     |               | CLECSF6 | lectina de tipo C (dependiente de calcio, dominio de reconocimiento a hidratos de carbono), superfamilia miembro 6 |
| 487    | AA713487     |               | PIK3R1  | fosfoinositida-3-quinasa, subunidad reguladora, polipéptido 1 (p85 alfa)   |
| 488    | AA828505     |               | FBXW7   | caja-F y proteína 7 de dominio WD-40 (homólogo a archipiélago, Drosophila)   |
| 489    | AK021865     |               | CKJJP-1 | proteína 1 que interacciona con CK2; proteína HQ0024c  |

| BCR Nº | Nº de ACCESO | Símbolo   | TÍTULO  |
|--------|--------------|-----------|---|
| 490    | AK001605     | FLJ10743  | proteína teórica FLJ 10743  |
| 491    | AI041186     | HSPC182   | proteína HSPC182  |
| 492    | AA873363     | NM_144650 | ADH8 alcohol dehidrogenasa 8  |
| 493    | NM 013409    | FST       | follistatina  |
| 494    | AK000322     | FLJ20315  | proteína teórica FLJ20315   |
| 495    | AB020637     | XM_290546 | KIAA0830 proteína KIAA0830  |
| 496    | AA872040     | INHBB     | inhibina, beta B (polipéptido beta activina AB)   |
| 497    | NM_004430    | EGR3      | respuesta 3 del crecimiento temprano  |
| 498    | D59989       |           | ESTs  |
| 499    | D78013       | DPYSL2    | dihidropirimidinasa de tipo 2   |
| 500    | AI081821     |           | ARNm de Homo sapiens; ADNc DKFZp313M0417 (del clon DKFZp313M0417)                           |
| 501    | AA309603     | KIAA1430  | proteína KIAA1430   |
| 502    | NM_004107    | FCGRT     | fragmento Fc de IgG, receptor, transportador, alfa  |
| 503    | AW268719     |           | ADNc de Homo sapiens FLJ32438 fis, clon SKMUS2001402.                                       |
| 504    | BF446578     | NM_145313 | LOC22100 2 producto del gen CG4853  |
| 505    | BG054844     | NM_005168 | ARHE familia del gen homólogo a ras, miembro E  |
| 506    | AF054987     |           | ALDOC aldolasa C, fructosa-bisfosfato   |
| 507    | AI052390     | FLJ20071  | dimeclina   |
| 508    | NM_004530    | MMP2      | metaloproteinasa 2 de matriz (gelatinasa A, gelatinasa 72kDa, colagenasa de tipo IV 72kDa ) |
| 509    | AF054999     | NM_001431 | EPB41L2 banda 4,1 de la proteína de membrana de eritrocitos de tipo 2                       |
| 510    | AU151591     | NM_182964 | NAV2 navegador 2 de neurona   |
| 511    | AA447744     |           | ESTs  |
| 512    | R61253       | ST6GalII  | beta-galactósido alfa-2,6-sialiltransferasa II  |

Tabla 9 Secuencias cebadoras para experimentos mediante RT-PCR semi-cuantitativa

| Nº de Acceso | Símbolo       | Cebador directo                                  | Cebador inverso                                  |
|--------------|---------------|--|--|
| AI261804     | EST           | 5'-CTGTTCTGGC TTCTGTTA TGT TCT-3' (SEC ID Nº: 1) | 5'-AGAAAATACG GTCCTCT TGT TGC-3' (SEC ID Nº: 2)  |
| AA205444     | AP1S2         | 5'-CACTGTAATG CACGAC ATTT GA-3' (SEC ID Nº: 3)   | 5'-GTTACAGCTT AGCACAA GGC ATC-3' (SEC ID Nº: 4)  |
| AA167194     | LOC25<br>3782 | 5'-ACCTCTGAGT TTGATT CCC AA-3' (SEC ID Nº: 5)    | 5'-CGAGGCTTGT AACAAATC TAC TGG-3' (SEC ID Nº: 6) |
| AA676987     | EST           | 5'-GAAACTGTAC GGGGGT TAAA GAG-3' (SEC ID Nº: 7)  | 5'-CATCAATGTG GTGAGTG ACA TCT-3' (SEC ID Nº: 8)  |

| Nº de Acceso | Símbolo | Cebador directo                                 | Cebador inverso                                  |
|--------------|---------|---|--|
| H22566       | DACH    | 5'-AAGCCCTTGG AACAGA ACAT ACT-3' (SEC ID Nº: 9) | 5'-CAGTAAACGT GGTTCTC ACA TTG-3' (SEC ID Nº: 10) |
| NMJH8492     | TOPK    | 5'-AGACCCTAAAGATCGTC CTTCTG-3' (SEC ID Nº: 13)  | 5'-GTGTTTAAGTCAGCATG AGCAG-3' (SEC ID Nº: 14)    |
| NM_002046    | GAPD    | 5'-CGACCACTT GTCAAGC TCA-3' (SEC ID Nº: 11)     | 5'-GGTTGAGCAC AGGGTAC TTT ATT-3' (SEC ID Nº: 12) |

Tabla 10 Listado de genes con expresión modificada entre el tipo bien y mal diferenciado en casos sencillos

| BCR Nº | Nº de acceso |  | Símbolo       | Título  | Valor P |
|--------|--------------|--|---------------|---|---------|
| 513    | AV729269     |  | DKFZP564 D166 | proteína que contiene supuestas repeticiones de anquirina     | 3,1E-07 |
| 514    | AI246554     |  | NDUFA8        | NADH deshidrogenasa (ubiquinona) subcomplejo 1 alfa, 8, 19kDa | 1,4E-06 |
| 515    | J04080       |  | CIS           | componente 1 del complemento, subcomponentes                  | 1,4E-05 |
| 516    | N93264       |  | EST           | clon IMAGE:4908933, de Homo sapiens ARNm                      | 1,4E-05 |
| 517    | NM_00231 8   |  | LOXL2         | lisil oxidasa de tipo 2                                       | 1,6E-05 |
| 518    | J03464       |  | COL1A2        | colágeno, tipo I, alfa 2                                      | 2,4E-05 |
| 519    | U01184       |  | FLU           | homólogo I no volador (Drosofila)                             | 2,5E-05 |
| 520    | X63556       |  | FBN1          | fibrilina 1 (síndrome de Marfan)                              | 3,8E-05 |
| 521    | X78137       |  | PCBP1         | proteína 1 de unión a polí(rC)                                | 4,6E-05 |
| 522    | AK021534     |  | EST           | ADNC FLJ11472 fis, de Homo sapiens clon HEMBA1001711.         | 6,3E-05 |
| 523    | AK024012     |  | NPD002        | proteína NPD002   | 6,3E-05 |
| 524    | AI200892     |  | BIK           | destructor de interacción con BCL2 (inductor de apoptosis)    | 9,1E-05 |
| 525    | J03040       |  | SPARC         | proteína secretada, ácida, rica en cisteína (osteonectina)    | 9,3E-05 |
| 526    | AW970143     |  | C6orf49       | fase de lectura abierta 49 del cromosoma 6                    | 1,0E-04 |
| 527    | D62873       |  | EST           | clon IMAGE:5288080, de Homo sapiens, ARNm                     | 1,2E-04 |
| 528    | D42041       |  | G2AN          | alfa glucosidasa II subunidad alfa                            | 1,2E-04 |
| 529    | AI376418     |  | EST           | ADNC FLJ35169 fis, de Homo sapiens clon PLACE6012908.         | 1,7E-04 |
| 530    | AK026744     |  | FLJ23091      | proteína teórica FLJ23091                                     | 1,8E-04 |
| 531    | AF026292     |  | CCT7          | chaperonina que contiene TCP1, subunidad 7 (eta)              | 2,0E-04 |
| 532    | Y10805       |  | HRMT1L2       | HMT1 hnRNP metiltransferasa de tipo 2 (S. cerevisiae)         | 2,1E-04 |
| 533    | L12350       |  | THBS2         | trombospondina 2  | 2,1E-04 |

| BCR Nº | Nº de acceso | Símbolo   | Título   | Valor P  |         |
|--------|--------------|-----------|--|--|---------|
| 534    | AK025706     | AMPD2     | adenosina monofosfato desaminasa 2 (isoforma L)                                      | 2,4E-04  |         |
| 535    | BE618804     | PIG11     | proteína inducida por p53  | 2,5E-04  |         |
| 536    | AV713686     | RPS29     | proteína ribosomal S29   | 2,8E-04  |         |
| 537    | M26481       | TACSTD1   | transductor 1 de señal de calcio asociado a tumores                                  | 2,8E-04  |         |
| 538    | D00099       | ATP1A1    | ATPasa, transportador Na <sup>+</sup> /K <sup>+</sup> , polipéptido alfa 1           | 2,9E-04  |         |
| 539    | AA946602     | ORMDL2    | ORMI de tipo 2 ( <i>S. cerevisiae</i> )  | 2,9E-04  |         |
| 540    | NM_00153 3   | HNRPL     | ribonucleoproteína L nuclear heterogénea   | 3,9E-04  |         |
| 541    | BG107866     | SIVA      | proteína (Siva) de unión a CD27  | 4,4E-04  |         |
| 542    | W72297       | NM_017866 | FLJ20533   | proteína teórica FLJ20533  | 4,4E-04 |
| 543    | U76992       | HTATSF1   | factor 1 específico de HTV TAT   | 4,8E-04  |         |
| 544    | AA191454     | NM_198897 | FIBP   | proteína de unión intracelular al factor del crecimiento de fibroblastos (ácido) | 4,9E-04 |
| 545    | BE903483     | RPS20     | proteína S20ribosomal  | 5,4E-04  |         |
| 546    | AJ005282     | NPR2      | receptor natriurético B/guanilato ciclase B (receptor B peptídico atrionatriurético) | 5,5E-04  |         |
| 547    | D86322       | CLGN      | calmegin   | 5,7E-04  |         |
| 548    | AA621665     | EST       | EST  | 5,8E-04  |         |
| 549    | M77349       | TGFBI     | factor del crecimiento transformante, inducido por beta, 68kDa                       | 6,3E-04  |         |
| 550    | BE176466     | ZAP3      | proteína ZAP3  | 6,6E-04  |         |
| 551    | AA776882     | NM_030795 | STMN4  | estatmina de tipo 4  | 7,1E-04 |
| 552    | AI261382     | NM_01634  | SH120  | receptor acoplado a supuesta proteína G  | 7,1E-04 |
| 553    | AB007618     | COX7A2L   | citocromo c oxidasa subunidad VII a polipéptido de tipo 2                            | 7,2E-04  |         |
| 554    | D21261       | TAGLN2    | transgelina 2  | 7,5E-04  |         |
| 555    | M68864       | LOC51035  | ORF  | 7,7E-04  |         |
| 556    | AB007836     | TGFB1I1   | factor beta 1 de crecimiento transformante inducido por transcripto 1                | 8,1E-04  |         |
| 557    | AA173339     | EST       | EST  | 8,4E-04  |         |
| 558    | D87810       | PMM1      | fosfomanomutasa 1  | 8,4E-04  |         |
| 559    | M1 5798      | NM_183356 | ASNS   | asparagina sintetasa   | 8,7E-04 |
| 560    | AW072418     | B7        | proteína B7  | 9,0E-04  |         |
| 561    | D38293       | AP3M2     | complejo 3 de proteína relacionada adaptadora, mu 2 subunidad                        | 9,5E-04  |         |

| BCR Nº | Nº de acceso | Símbolo   | Título  | Valor P                                    |         |
|--------|--------------|-----------|---|--|---------|
| 562    | NM_018950    | HLA-F     | complejo mayor de histocompatibilidad, clase I, F   | 1,0E-03                                    |         |
| 563    | NM_001219    | CALU      | calumenina  | 1,1E-03                                    |         |
| 564    | J04162       | FCGR3A    | receptor para (CD16) fragmento Fc de IgG, de baja afinidad, IIIa  | 1,1E-03                                    |         |
| 565    | U09873       | FSCN1     | homólogo 1 de fascina, proteína apareada con actina ( <i>Strongylocentrotus purpuratus</i> )                | 1,1E-03                                    |         |
| 566    | N51082       | NM_80759  | DACH  | homólogo de dacshund ( <i>Drosophila</i> ) | 1,3E-03 |
| 567    | NM_004199    | P4HA2     | procolágeno-prolina, 2-oxoglutarato 4-dioxigenasa (prolina 4-hidroxilasa), alfa polipéptido II              | 1,3E-03                                    |         |
| 568    | BE904196     | GNB1      | proteína de unión al nucleótido guanina (proteína G), beta polipéptido 1                                    | 1,3E-03                                    |         |
| 569    | L08895       | MEF2C     | factor 2 potenciador de la transcripción de la caja MADS, polipéptido C (factor 2C potenciador de miocitos) | 1,3E-03                                    |         |
| 570    | AK022670     | NM_016649 | C20orf6   | fase de lectura abierta 6 del cromosoma 20 | 1,3E-03 |
| 571    | AW157725     | POLR2F    | polimerasa (ARN) II (dirigida a ADN) polipéptido F  | 1,4E-03                                    |         |
| 572    | NM_004939    | DDX1      | polipéptido 1 de la caja DEAD/H (Asp-Glu-Ala-Asp/His)   | 1,4E-03                                    |         |
| 573    | X65463       | NM_021976 | PvXRB   | receptor X retinoide, beta                 | 1,5E-03 |
| 574    | Z68179       | LY6E      | complejo 6 de antígeno linfocitario, locus E  | 1,5E-03                                    |         |
| 575    | BF976420     | SNRPF     | polipéptido F de ribonucleoproteína pequeña nuclear   | 1,5E-03                                    |         |
| 576    | D79986       | BTF       | factor de la transcripción asociado a Bcl-2   | 1,5E-03                                    |         |
| 577    | AK001023     | NUBP2     | proteína 2 de unión a nucleótidos (homólogo MinD, <i>E. coli</i> )  | 1,6E-03                                    |         |
| 578    | BE065329     | EST       | EST   | 1,6E-03                                    |         |
| 579    | L34600       | MTIF2     | factor 2 de inicio de la traducción mitocondrial  | 1,7E-03                                    |         |
| 580    | D13630       | BZW1      | cremallera de leucina básica y dominios W2 1  | 1,7E-03                                    |         |
| 581    | X15880       | NM_001848 | COL6A1  | colágeno, tipo VI, alfa 1                  | 1,7E-03 |
| 582    | AB003723     | PIGQ      | fosfatidilinositol glicano, clase Q   | 1,7E-03                                    |         |
| 583    | L36645       | EPHA4     | EphA4   | 1,7E-03                                    |         |
| 584    | BF974358     | RPS27     | proteína ribosomal S27 (metalopanestimulina 1)  | 1,8E-03                                    |         |

# ES 2 365 732 T3

| BCR Nº | Nº de acceso |           | Símbolo  | Título  | Valor P |
|--------|--------------|-----------|----------|---|---------|
| 585    | AA747449     |           | HBP2     | proteína 2 de interacción con huntingtina   | 1,9E-03 |
| 586    | AA283813     |           | FLJ12150 | proteína teórica FLJ12150   | 2,0E-03 |
| 587    | L38995       | NM_003321 | TUFM     | factor de alargamiento de la traducción Tu, mitocondrial  | 2,0E-03 |
| 588    | N67293       |           | EST      | ADNc FLJ11997 fis, Homo sapiens, clon HEMBB1001458.   | 2,1E-03 |
| 589    | AB014549     |           | KIAA0649 | producto del gen KIAA0649   | 2,1E-03 |
| 590    | D38305       |           | TOB1     | transductor de ERBB2,1  | 2,2E-03 |
| 591    | L40391       | NM_006827 | TMP21    | proteína de tráfico transmembrana   | 2,2E-03 |
| 592    | H28960       |           | EST      | ESTs  | 2,2E-03 |
| 593    | U86753       |           | CDC5L    | ciclo de división celular CDC5 de tipo 5 (S. pombe)   | 2,3E-03 |
| 594    | AI143226     |           | BLP1     | proteína 1 de tipo BBP  | 2,3E-03 |
| 595    | M57730       |           | EFNA1    | efrina-A1   | 2,3E-03 |
| 596    | AI928868     |           | UBR1     | ubiquitina proteína ligasa E3 componente n-recognin 1   | 2,3E-03 |
| 597    | AF077044     |           | RPAC2    | probablemente ortólogo de polimerasa 1-3 de ARN de ratón (subunidad 16 kDa)                                 | 2,3E-03 |
| 598    | AF097431     |           | LEPRE1   | proteoglicano enriquecido con leucina prolina (leprecan) 1  | 2,4E-03 |
| 599    | NM_00435 0   |           | RUNX3    | factor 3 de la transcripción relacionado con runt   | 2,4E-03 |
| 600    | AL162047     |           | NCOA4    | coactivador 4 receptor nuclear  | 2,5E-03 |
| 601    | BF915013     |           | EST      | ADNc de FLJ37302 fis, Homo sapiens, clon BRAMI2016009.  | 2,5E-03 |
| 602    | Z37166       |           | BAT1     | transcrito 1 asociado a HLA-B   | 2,5E-03 |
| 603    | M81349       |           | SAA4     | amilóide A4 sérico, constitutivo  | 2,6E-03 |
| 604    | AL137338     | NM_007214 | SEC63L   | proteína SEC63  | 2,6E-03 |
| 605    | AI745624     |           | ELL2     | ELL-relacionada con ARN polimerasa II, factor de alargamiento   | 2,6E-03 |
| 606    | BG167522     |           | HSPC016  | proteína teórica HSPC016  | 2,6E-03 |
| 607    | U58766       |           | TSTA3    | antígeno de trasplante específico de tejidos P35B   | 2,7E-03 |
| 608    | J04474       | NM_000709 | BCKDHA   | cadena ramificada ceto ácido deshidrogenasa E1, polipéptido alfa (enfermedad de la orina de jarabe de arce) | 2,7E-03 |
| 609    | HI 5977      | NM_021116 | EST      | ADNc FLJ30781 fis de Homo sapiens, clon FEBRA2000874.   | 2,8E-03 |
| 610    | AL049339     | NM_001304 | CPD      | carboxipeptidasa D  | 2,8E-03 |

| BCR Nº | Nº de acceso |           | Símbolo   | Título   | Valor P |
|--------|--------------|-----------|-----------|--|---------|
| 611    | AL133555     | NM_080821 | C20orf108 | fase de lectura abierta 108 del cromosoma 20   | 2,9E-03 |
| 612    | AW662518     |           | FLJ10876  | proteína teórica FLJ10876  | 2,9E-03 |
| 613    | BE883507     | NM_003663 | CGGBP1    | proteína 1 de unión a repeticiones triplete CGG  | 2,9E-03 |
| 614    | BE797472     |           | RPL17     | proteína L17 ribosomal   | 3,0E-03 |
| 615    | U41371       |           | SF3B2     | factor 3b de corte y empalme, subunidad 2, 145kDa  | 3,0E-03 |
| 616    | L39068       |           | DHPS      | desoxihipusina sintasa   | 3,1E-03 |
| 617    | NM_004517    |           | ILK       | quinasa unida a integrina  | 3,1E-03 |
| 618    | U14972       |           | RPS10     | proteína S10 ribosomal   | 3,2E-03 |
| 619    | U61500       |           | TMEM1     | proteína 1 transmembrana   | 3,3E-03 |
| 620    | NM_002719    |           | PPP2R5C   | proteína fosfatasa 2, subunidad reguladora B (B56), isoforma gamma                             | 3,3E-03 |
| 621    | AF053233     |           | VAMP8     | proteína 8 de membrana asociada a vesículas (endobrevina)                                      | 3,3E-03 |
| 622    | NM_002822    | NM_198974 | PTK9      | proteína tirosina quinase 9 PTK9   | 3,3E-03 |
| 623    | U16996       |           | DUSP5     | fosfatasa 5 de especificidad dual  | 3,3E-03 |
| 624    | AV705747     | NM_006276 | SFRS7     | factor de corte y empalme, rico en arginina-serina 7, 35kDa                                    | 3,3E-03 |
| 625    | AF178984     |           | IER5      | respuesta inmediata temprana 5   | 3,3E-03 |
| 626    | Z29093       |           | DDR1      | familia del receptor de dominio discoidina, miembro 1  | 3,3E-03 |
| 627    | AB024536     |           | ISLR      | superfamilia de inmunoglobulina que contiene repeticiones ricas en leucina                     | 3,3E-03 |
| 628    | BF791601     |           | EMP2      | proteína 2 de membrana epitelial   | 3,3E-03 |
| 629    | AF061737     |           | SPC18     | complejo de señal peptidasa (18kD)   | 3,3E-03 |
| 630    | AB002386     |           | EZH1      | Potenciador del homólogo 1 de zeste ( <i>Drosophila</i> )                                      | 3,5E-03 |
| 631    | AA634090     |           | EST       | Homo sapiens, Similar a la ribonucleoproteína A1 nuclear heterogénea, clón IMAGE:2900557, ARNm | 3,5E-03 |
| 632    | AK023674     |           | FLJ13612  | probablemente ortólogo de proteína de unión a calcio expresada neuronalmente                   | 3,6E-03 |
| 633    | D13626       |           | GPR105    | receptor 105 acoplado a proteína G   | 3,7E-03 |
| 634    | AK026849     | XM_371844 | TSPYL     | de tipo TSPY   | 3,8E-03 |
| 635    | Y18643       |           | METTL1    | metiltransferasa de tipo 1   | 3,9E-03 |
| 636    | AF176699     |           | FBXL4     | caja F y proteína 4 con repeticiones ricas en leucina  | 3,9E-03 |

# ES 2 365 732 T3

| BCR Nº | Nº de acceso | Símbolo   | Título  | Valor P  |         |
|--------|--------------|-----------|---|--|---------|
| 637    | NM_00397 7   | AIP       | proteína que interacciona con el receptor de hidrocarburos aromáticos                             | 3,9E-03  |         |
| 638    | AK000498     | HARS      | histidil-ARNt sintetasa   | 4,0E-03  |         |
| 639    | U05237       | NM_004459 | FALZ  | antígeno de Alzheimer fetal  | 4,0E-03 |
| 640    | BF696304     | NM_032832 | FLJ14735  | proteína teórica FLJ14735  | 4,0E-03 |
| 641    | X14420       | COL3A1    | colágeno, tipo III, alfa 1 (síndrome de Ehlers-Danlos de tipo IV, dominante autosómico)           | 4,1E-03  |         |
| 642    | BE796098     | NDUFS8    | NADH deshidrogenasa (ubiquinona) proteína 8 Fe-S, 23kDa (NADH-coenzima Q reductasa)               | 4,3E-03  |         |
| 643    | X60221       | ATP5F1    | ATP sintasa, transportadora de H <sup>+</sup> , complejo F0 mitocondrial, subunidad b, isoforma 1 | 4,4E-03  |         |
| 644    | AA135341     | NM_021078 | GCN5L2  | control general de GCN5 de la síntesis de aminoácidos 5 de tipo 2 (levadura) | 4,6E-03 |
| 645    | AF009368     | CREB3     | proteína 3 de unión a elemento sensible a AMPc (luman)  | 4,7E-03  |         |
| 646    | BF970013     | SPC12     | peptidasa de señal 12kDa  | 4,7E-03  |         |
| 647    | W45522       | ATPIF1    | factor 1 inhibidor de ATPasa  | 4,7E-03  |         |
| 648    | AI733356     | NM_006306 | EST   | ADNc FLJ31746 fis, Homo sapiens, clon NT2RI2007334.                          | 4,8E-03 |
| 649    | AW117927     | EIF3S9    | factor 3 de inicio de la traducción en eucariotas, subunidad 9 eta, 116kDa                        | 4,8E-03  |         |
| 650    | AF275798     | NM_012073 | CCT5  | TCP1 contenido chaperonina, subunidad 5 (epsilon)                            | 5,0E-03 |
| 651    | AI937126     | WTAP      | proteína asociada a tumor 1 de Wilms  | 5,0E-03  |         |
| 652    | AK024891     | NM_203463 | LOC25378 2  | proteína teórica LOC253782   | 5,1E-03 |
| 653    | D13629       | KTN1      | quinectina 1 (receptor de quinesina)  | 5,2E-03  |         |
| 654    | AI682994     | AHCYL1    | S-adenosilhomocisteína hidrolasa de tipo 1  | 5,3E-03  |         |
| 655    | BF980325     | NM_005742 | ATPasa, transportadora de H <sup>+</sup> , lisosomal 42kDa, isoforma 2 subunidad C V1             | 5,3E-03  |         |
| 656    | AI378996     | NM_005381 | NCL   | nucleolina   | 5,3E-03 |
| 657    | D88153       | HYA22     | proteína HIA22  | 5,3E-03  |         |
| 658    | S67310       | BF        | factor B, properdina  | 5,4E-03  |         |
| 659    | AW438585     | EST       | clon IMAGE:5273745 Homo sapiens, ARNm   | 5,4E-03  |         |
| 660    | M12267       | OAT       | ormitina aminotransferasa (atrofia girada)  | 5,5E-03  |         |

| BCR Nº | Nº de acceso | Símbolo       | Título  | Valor P |
|--------|--------------|---------------|---|---------|
| 661    | AB001636     | DDX15         | polipéptido 15 caja de DEAD/H (Asp-Glu-Ala-Asp/His)   | 5,7E-03 |
| 662    | D13315       | GLOI          | glioxalasa I  | 5,9E-03 |
| 663    | AF244931     | WDR10         | WD con dominio 10 repetido  | 5,9E-03 |
| 664    | AL050094     | EDH3B         | isocitrato dehidrogenasa 3 (NAD <sup>+</sup> ) beta   | 6,0E-03 |
| 665    | AK022881     | KIAA1272      | proteína KIAA1272   | 6,0E-03 |
| 666    | AI720096     | RPL29         | proteína L29 ribosomal  | 6,1E-03 |
| 667    | Y12781       | TBL1X         | transducina (beta) de tipo IX-unida   | 6,2E-03 |
| 668    | AI014538     | NM_138384     | proteína teórica BC004409   | 6,2E-03 |
| 669    | NM_02098 7   | ANK3          | anquirina 3, nódulo de Ranvier (anquirina G)  | 6,3E-03 |
| 670    | NM_00438 7   | NKX2-5        | locus 5 ( <i>Drosophila</i> ) relacionado con el factor de la transcripción NK2                 | 6,3E-03 |
| 671    | J03817       | GSTM1         | glutatión S-transferasa M1  | 6,3E-03 |
| 672    | BF435769     | EST           | ESTs, muy similar a la proteína teórica FLJ20378 [ <i>Homo sapiens</i> ] [ <i>H.sapiens</i> ]   | 6,5E-03 |
| 673    | AL390147     | DKFZp547 D065 | proteína teórica DKFZp547D065   | 6,5E-03 |
| 674    | AA961412     | NM_003333     | producto 1 de fusión a proteína ribosomal de resto A-52 de ubiquitina                           | 6,6E-03 |
| 675    | NM_00270 2   | POU6F1        | dominio POU, clase 6, factor 1 de la transcripción  | 6,6E-03 |
| 676    | M58050       | MCP           | proteína del cofactor de membrana (CD46, antígeno de reactividad cruzada trofoblasto-linfocito) | 6,6E-03 |
| 677    | NM_00129 3   | CLNS1A        | canal de cloruro, sensible a nucleótidos, 1A  | 6,7E-03 |
| 678    | BF213049     | COX7A2        | citocromo c oxidasa subunidad VIIa polipéptido 2 (hígado)                                       | 6,7E-03 |
| 679    | AF236056     | GOLPH2        | fosfoproteína 2 de golgi  | 6,7E-03 |
| 680    | U79285       | NM_021079     | N-miristoiltransferasa 1  | 6,8E-03 |
| 681    | AB027196     | RNF1O         | proteína de dedo de anillo 10   | 6,9E-03 |
| 682    | AA036952     | FLJ30973      | proteína teórica FLJ30973   | 7,0E-03 |
| 683    | AW732157     | NM_052963     | topoisomerasa mitocondrial I  | 7,1E-03 |
| 684    | AL049319     | NM_032804     | proteína teórica FLJ14547   | 7,3E-03 |
| 685    | BE613161     | EST           | ADNc FLJ37042 fis, <i>Homo sapiens</i> , clon BRACE2011947.                                     | 7,3E-03 |
| 686    | U28749       | HMGA2         | grupo AT-hook 2 de elevada movilidad  | 7,3E-03 |
| 687    | BF793677     | MGC49942      | proteína teórica MGC49942   | 7,4E-03 |
| 688    | BG032216     | NM_017746     | proteína teórica FLJ20287   | 7,4E-03 |

| BCR Nº | Nº de acceso | Símbolo   | Título  | Valor P                                      |         |
|--------|--------------|-----------|---|--|---------|
| 689    | AL449244     | PP2447    | proteína teórica PP2447   | 7,5E-03                                      |         |
| 690    | AK024103     | EST       | ADNc FLJ14041 fis de Homo sapiens, clon HEMBA1005780.   | 7,5E-03                                      |         |
| 691    | U17838       | PRDM2     | PR que contiene 2 dominios, con dominio ZNF   | 7,5E-03                                      |         |
| 692    | D86479       | NM_001129 | AEBP1   | proteína 1 de unión AE                       | 7,5E-03 |
| 693    | D50420       | NHP2L1    | proteína 2 de cromosoma no-histona NHP2 de tipo 1 ( <i>S. cerevisiae</i> )                          | 7,5E-03                                      |         |
| 694    | D87258       | PRSS11    | proteasa, serina, 11 (unión a IGF)  | 7,5E-03                                      |         |
| 695    | BF434108     | NM_014187 | HSPC171   | proteína HSPC171                             | 7,6E-03 |
| 696    | NM_00070 5   | ATP4B     | ATPasa, intercambiadora de H <sup>+</sup> /K <sup>+</sup> , polipéptido beta                        | 7,7E-03                                      |         |
| 697    | AF077599     | SBB103    | proteína teórica SBBI03   | 7,7E-03                                      |         |
| 698    | NM_00153 0   | HIF1A     | factor 1 inducible por hipoxia, subunidad alfa (factor de transcripción básico hélice-bucle-hélice) | 7,8E-03                                      |         |
| 699    | AB023204     | EPB41L3   | banda 4.1 de la proteína de membrana de eritrocitos de tipo 3                                       | 7,8E-03                                      |         |
| 700    | AA253194     | NM_022121 | PIGPC1  | proteína PIGPC1 inducida por p53             | 7,9E-03 |
| 701    | BE502341     | NM_139177 | C17orf26  | fase de lectura abierta 26 y de cromosoma 17 | 7,9E-03 |
| 702    | AL050265     | TARDBP    | proteína de unión a ADN TAR   | 8,0E-03                                      |         |
| 703    | AK001643     | NM_018215 | FLJ10781  | proteína teórica FLJ10781                    | 8,3E-03 |
| 704    | BG179412     | COX7B     | citocromo c oxidasa subunidad VIIb  | 8,6E-03                                      |         |
| 705    | X03212       | KRT7      | queratina 7   | 8,8E-03                                      |         |
| 706    | L07033       | HMGCL     | 3-hidroximetil-3-metilglutaril-coenzima A liasa (hidroximetilglutaricaciduria)                      | 9,0E-03                                      |         |
| 707    | M19383       | ANXA4     | anexina A4  | 9,0E-03                                      |         |
| 708    | NM_00127 3   | CHD4      | proteína 4 de unión a ADN helicasa dominio cromo  | 9,1E-03                                      |         |
| 709    | NM_00446 1   | FARSL     | fenilalanina de tipo ARNt sintetasa   | 9,1E-03                                      |         |
| 710    | AI192880     | CD44      | antígeno CD44 (función buscadora y sistema de grupo sanguíneo Indian)                               | 9,1E-03                                      |         |
| 711    | AF038961     | MPDU1     | manosa-P-dolichol utilización defectuosa 1  | 9,5E-03                                      |         |
| 712    | U67322       | C20orf18  | fase de lectura abierta 18 del cromosoma 20   | 9,5E-03                                      |         |
| 713    | AA521017     | EST       | EST   | 9,5E-03                                      |         |
| 714    | AA811043     | NM_003730 | RNASE6P L   | precursor ribonucleasa 6                     | 9,9E-03 |
| 715    | AA536113     | TMEPAI    | ARN inducido por andrógeno prostático, transmembrana  | 9,9E-03                                      |         |

| BCR Nº | Nº de acceso |  | Símbolo    | Título                     | Valor P |
|--------|--------------|--|------------|----------------------------|---------|
| 716    | BF973104     |  | LOC20172 5 | proteína teórica LOC201725 | 9,9E-03 |
| 717    | NM 00029 3   |  | PHKB       | fosforilasa quinasa, beta  | 9,9E-03 |
| 718    | NM 00054 8   |  | TSC2       | esclerosis tuberosa 2      | 1,0E-02 |

Tabla 11 Listado de genes con expresión modificada entre tumores nódulo-positivos y nódulo-negativos

| BCR Nº | Nº de ACCESO |            | Símbolo     | TÍTULO  | P-valor | + o - |
|--------|--------------|------------|-------------|---|---------|-------|
| 719    | BF686125     |            | UBA52       | ubiquitina A-52 resto del producto de fusión de proteína ribosomal 1  | 8.1E-09 | -     |
| 720    | AA634090     |            |             | Homo sapiens, Similar a ribonucleoproteína A1 nuclear heterogénea, clón IMAGE:2900557, ARNm                   | 1.4E-07 |       |
| 721    | L00692       |            | CEACAM 3    | Molécula 3 de adhesión celular relacionada con antígeno carcinoembriónico                                     | 4.2E-07 | -     |
| 722    | AW954403     | NM_04781   | VAMP3       | Proteína 3 de membrana asociada a vesículas (celubrevina)   | 2.2E-06 | +     |
| 723    | AA865619     |            | C21orf97    | Fase de lectura abierta 97 del cromosoma 21   | 2.6E-06 | -     |
| 724    | W74502       | NM_032350  | MGC11257    | Proteína teórica MGC11257   | 2.4E-05 | +     |
| 725    | NM 00209 4   |            | GSPT1       | Transición 1 de la fase G1 a S  | 2.7E-05 | +     |
| 726    | T55178       |            | KIAA1040    | Proteína KIAA1040   | 3.2E-05 | -     |
| 727    | L36983       |            | DNM2        | Dinamina 2  | 4.1E-05 | +     |
| 728    | Z21507       |            | EEF1D       | Factor 1 delta de alargamiento de la traducción en eucariotas (proteína de intercambio de nucleótido guanina) | 5.2E-05 |       |
| 729    | AI581728     | NM 0055 07 | CFL1        | cofilina 1 (no muscular)  | 8.0E-05 | +     |
| 730    | NM 00129 3   |            | CLNS1A      | Canal de cloruro, sensible a nucleótido, 1A   | 9.0E-05 | +     |
| 731    | BF680847     |            | SENP2       | Proteasa específica de sentrina   | 9.0E-05 | +     |
| 732    | AF100743     |            | NDUFS3      | NADH dehidrogenasa (ubiquinona) proteína 3 Fe-S, 3,30 kDa (NADH-coenzima Q reductasa)                         | 9.8E-05 | +     |
| 733    | NM 00496 0   |            | FUS         | fusión, derivada de liposarcoma tumoral t (12;16)   | 9.8E-05 | -     |
| 734    | AK023975     | NM 0159 34 | NOP5/NO P58 | Proteína nucleolar NOP5/NOP58   | 1.3E-04 | +     |
| 735    | AF083245     |            | PSMD13      | proteosoma (prosoma, macropainina) subunidad 26S, no-ATPasa, 13   | 1.5E-04 | +     |
| 736    | AA129776     |            | SUOX        | sulfato oxidasa   | 1.8E-04 | +     |
| 737    | U55766       | NM 0070 43 | HRB2        | Proteína 2 de unión inversa a HIV-1   | 2.0E-04 | +     |
| 738    | BF526092     |            | LOC1544 67  | Proteína teórica BC003515   | 2.1E-04 | +     |
| 739    | BF677579     | XM 3707 54 | THTPA       | Tiamina trifosfatasa  | 2.3E-04 | +     |

| BCR Nº | Nº de ACCESO |            | Símbolo   | TÍTULO   | P-valor | + o - |
|--------|--------------|------------|-----------|--|---------|-------|
| 740    | X98260       |            | ZRF1      | Factor 1 relacionado con zuotina                                 | 2.3E-04 | +     |
| 741    | BE440010     |            | LOC51255  | Proteína teórica LOC51255  | 2.7E-04 | +     |
| 742    | AF007165     | NM 0210 08 | DEAF1     | Factor 1 autorregulador epidérmico deformado (Drosofila)         | 2.7E-04 | +     |
| 743    | X78687       |            | NEU1      | sialidasa 1 (sialidasa lisosomal)                                | 3.0E-04 | +     |
| 744    | AW965200     |            |           | Homo sapiens, clon IMAGE: 5286019, ARNm                          | 3.1E-04 | -     |
| 745    | AK023240     |            | UGCGL1    | UDP-glucosa ceramida glucosiltransferasa de tipo 1               | 3.1E-04 | +     |
| 746    | M95712       |            | BRAF      | Homólogo B1 del oncogén viral del sarcoma murino v-raf           | 3.7E-04 | +     |
| 747    | L38995       | NM 0033 21 | TUFM      | Factor de alargamiento de la traducción Tu, mitocondrial         | 3.9E-04 | +     |
| 748    | AW014268     |            | FLJ10726  | Proteína teórica FLJ10726  | 4.2E-04 | +     |
| 749    | D49547       |            | DNAJB1    | Homólogo DnaJ (Hsp40), subfmalia B, miembro 1                    | 4.4E-04 | +     |
| 750    | BE466450     |            | AP4S1     | Complejo 4 de proteína relacionada adaptadora, subunidad sigma 1 | 4.5E-04 | +     |
| 751    | AB007944     |            | KIAA047 5 | Producto del gen KIAA0475  | 4.9E-04 | -     |
| 752    | AF034091     |            | MRPL40    | Proteína ribosomal mitocondrial L40                              | 5.1E-04 | +     |

Tabla 12 Información Histoclínica

| ID        | Edad durante la operación | Estado meno-páusico | T | N | M | Estado | Tipo HistoLógico | Infiltrado linfocítico | Angioinvasión | ER | PgR |
|-----------|---------------------------|---------------------|---|---|---|--------|------------------|------------------------|---------------|----|-----|
| MMK010003 | 51                        | pre-                | 2 | 1 | 0 | 2      | a3               | 3                      | 0             | +  | +   |
| MMK010004 | 47                        | pre                 | 2 | 1 | 0 | 2      | a1               | 0                      | 0             | +  | +   |
| MMK010005 | 44                        | pre                 | 2 | 0 | 0 | 2      | a1               | 1                      | 0             | +  | +   |
| MMK010013 | 45                        | pre                 | 2 | 1 | 0 | 2      | a1               | 1                      | 0             | -  | -   |
| MMK010016 | 44                        | pre                 | 2 | 0 | 0 | 2      | a2               | 0                      | 0             | -  | -   |
| MMK010025 | 46                        | pre                 | 2 | 0 | 0 | 2      | a1               | 0                      | 0             | +  | +   |
| MMK010031 | 29                        | pre                 | 2 | 2 | 0 | 3      | a3               | 3                      | 0             | -  | -   |
| MMK010037 | 62                        | post                | 0 | 0 | 0 | 0      | 1a               | 0                      | 0             | +  | +   |
| MMK010042 | 47                        | pre                 | 2 | 1 | 0 | 2      | a3               | 1                      | 2             | +  | +   |
| MMK010086 | 42                        | pre                 | 2 | 0 | 0 | 2      | A1               | 0                      | 0             | +  | +   |
| MMK010102 | 51                        | pre                 | 2 | 1 | 0 | 3      | A2               | 3                      | 0             | +  | +   |
| MMK010110 | 39                        | pre                 | 2 | 0 | 0 | 2      | a1               | 2                      | 0             | -  | -   |

# ES 2 365 732 T3

| ID         | Edad durante la operación | Estado meno-páusico | T | N | M | Estado | Tipo HistoLógico | Infiltrado linfocítico | Angioinvasión | ER | PgR |
|------------|---------------------------|---------------------|---|---|---|--------|------------------|------------------------|---------------|----|-----|
| MMK010129  | 52                        | pre                 | 2 | 2 | 0 | 3      | A1               | 2                      | 0             | -  | -   |
| MMK010135  | 41                        | pre                 | 2 | 0 | 0 | 2      | a1               | 0                      | 0             | +  | +   |
| MMK010138  | 38                        | pre                 | 2 | 0 | 0 | 2      | a1               | 0                      | 0             | +  | +   |
| MMK010145  | 51                        | pre                 | 2 | 1 | 0 | 2      | a3               | 0                      | 0             | +  | +   |
| MMK010147  | 49                        | pre                 | 2 | 1 | 0 | 2      | a1               | 1                      | 0             | +  | +   |
| MMK010149  | 35                        | pre                 | 2 | 0 | 0 | 2      | a3               | 1                      | 0             | -  | -   |
| MMK010175  | 38                        | pre                 | 2 | 0 | 0 | 2      | a3               | 0                      | 0             | +  | +   |
| MMK010178  | 51                        | pre                 | 0 | 0 | 0 | 0      | 1a               | 0                      | 0             | +  | +   |
| MMK010207  | 40                        | pre                 | 2 | 0 | 0 | 2      | a1               | 0                      | 0             | +  | +   |
| MMK010214  | 42                        | pre                 | 2 | 1 | 0 | 2      | a1               | 0                      | 0             | -  | -   |
| MMK010247  | 48                        | pre                 | 2 | 1 | 0 | 2      | a2               | 3                      | 0             | -  | -   |
| MMK010252  | 52                        | pre                 | 2 | 1 | 0 | 2      | a2               | 0                      | 0             | -  | -   |
| MMK010255  | 47                        | pre                 | 2 | 0 | 0 | 2      | a2               | 0                      | 0             | -  | -   |
| MMK010302  | 46                        | pre                 | 2 | 1 | 0 | 2      | a2               | 2                      | 1             | -  | -   |
| MMK010304  | 48                        | pre                 | 2 | 1 | 0 | 2      | a3               | 1                      | 0             | +  | +   |
| MMK010326  | 53                        | post                | 0 | 0 | 0 | 0      | 1a               | 0                      | 0             | -  | -   |
| MMK010327  | 43                        | pre                 | 2 | 1 | 0 | 2      | a1               | 1                      | 1             | +  | +   |
| MMK010341  | 42                        | pre                 | 2 | 1 | 0 | 2      | a1               | 2                      | 0             | +  | +   |
| MMK010370  | 46                        | pre                 | 2 | 1 | 0 | 2      | a3               | 2                      | 0             | +  | +   |
| MMK010397  | 38                        | pre                 | 2 | 1 | 0 | 2      | a3               | 3                      | 2             | +  | +   |
| MMK010411  | 46                        | pre                 | 2 | 0 | 0 | 2      | a1               | 0                      | 0             | +  | +   |
| MMK010431  | 50                        | pre                 | 2 | 0 | 0 | 2      | a3               | 0                      | 0             | -  | -   |
| MMK010435  | 49                        | pre                 | 2 | 1 | 0 | 2      | a3               | 0                      | 0             | +  | +   |
| MMKO 10453 | 49                        | pre                 | 2 | 1 | 0 | 2      | a3               | 3                      | 0             | +  | +   |
| MMK010471  | 42                        | pre                 | 2 | 1 | 0 | 2      | a1               | 3                      | 0             | -  | -   |
| MMKO 10473 | 40                        | pre                 | 2 | 1 | 0 | 2      | a2               | 0                      | 0             | -  | -   |
| MMK010478  | 38                        | pre                 | 2 | 2 | 0 | 3      | a2               | 0                      | 0             | +  | +   |
| MMKO 10491 | 46                        | pre                 | 2 | 0 | 0 | 2      | a3               | 1                      | 0             | +  | +   |
| MMK010497  | 44                        | pre                 | 0 | 0 | 0 | 0      | 1a               | 0                      | 0             | -  | +   |
| MMKO 10500 | 45                        | pre                 | 2 | 0 | 0 | 2      | a1               | 0                      | 0             | +  | +   |
| MMK010502  | 51                        | pre                 | 2 | 0 | 0 | 2      | A2               | 0                      | 0             | -  | -   |
| MMKO 10508 | 51                        | pre                 | 2 | 1 | 0 | 2      | a2               | 0                      | 0             | -  | -   |

# ES 2 365 732 T3

| ID        | Edad durante la operación | Estado meno-páusico | T | N | M | Estado | Tipo HistoLógico | Infiltrado linfocítico | Angioinvasión | ER | PgR |
|-----------|---------------------------|---------------------|---|---|---|--------|------------------|------------------------|---------------|----|-----|
| MMK010521 | 21                        | pre                 | 2 | 0 | 0 | 2      | a1               | 1                      | 1             | -  | -   |
| MMK010552 | 49                        | pre                 | 2 | 0 | 0 | 2      | a2               | 0                      | 0             | -  | -   |
| MMK010554 | 51                        | pre                 | 2 | 0 | 0 | 2      | a3               | 2                      | 0             | +  | +   |
| MMK010571 | 45                        | pre                 | 2 | 1 | 1 | 4      | a3               | 3                      | 0             | +  | +   |
| MMK010591 | 40                        | pre                 | 0 | 0 | 0 | 0      | 1a               | 0                      | 0             | -  | +   |
| MMK010613 | 37                        | pre                 | 0 | 0 | 0 | 0      | 1a               | 0                      | 0             | -  | 4-  |
| MMK010623 | 39                        | pre                 | 2 | 1 | 0 | 2      | a1               | 3                      | 0             | +  | +   |
| MMK010624 | 39                        | pre                 | 2 | 1 | 0 | 2      | a1               | 3                      | 0             | +  | +   |
| MMK010626 | 48                        | pre                 | 2 | 0 | 0 | 2      | a1               | 1                      | 1             | -  | -   |
| MMK010631 | 41                        | pre                 | 2 | 0 | 0 | 2      | a1               | 0                      | 0             | +  | +   |
| MMK010640 | 35                        | pre                 | 0 | 0 | 0 | 0      | 1a               | 0                      | 0             | +  | +   |
| MMK010644 | 47                        | pre                 | 2 | 2 | 0 | 2      | a3               | 3                      | 0             | +  | +   |
| MMK010646 | 37                        | pre                 | 2 | 1 | 0 | 2      | a3               | 1                      | 0             | +  | +   |
| MMK010660 | 46                        | pre                 | 2 | 0 | 0 | 2      | a1               | 0                      | 0             | -  | -   |
| MMK010671 | 45                        | pre                 | 2 | 0 | 0 | 2      | a1               | 0                      | 0             | -  | -   |
| MMK010679 | 68                        | post                | 0 | 0 | 0 | 0      | 1a               | 0                      | 0             | +  | +   |
| MMK010680 | 58                        | post                | 0 | 0 | 0 | 0      | 1a               | 0                      | 0             | -  | +   |
| MMK010709 | 33                        | pre                 | 2 | 0 | 0 | 2      | a3               | 0                      | 2             | -  | -   |
| MMK010711 | 51                        | pre                 | 0 | 0 | 0 | 0      | 1a               | 0                      | 0             | -  | +   |
| MMK010724 | 40                        | pre                 | 2 | 1 | 0 | 2      | a3               | 3                      | 2             | +  | +   |
| MMK010744 | 41                        | pre                 | 0 | 0 | 0 | 0      | 1a               | 0                      | 0             | +  | +   |
| MMK010758 | 40                        | pre                 | 2 | 1 | 0 | 2      | a1               | 0                      | 1             | +  | +   |
| MMK010760 | 42                        | pre                 | 2 | 0 | 0 | 2      | a1               | 0                      | 0             | +  | +   |
| MMK010762 | 50                        | pre                 | 2 | 1 | 0 | 2      | a3               | 3                      | 1             | +  | +   |
| MMK010769 | 33                        | pre                 | 2 | 0 | 0 | 2      | a2               | 0                      | 0             | -  | -   |
| MMK010772 | 45                        | pre                 | 2 | 1 | 0 | 2      | a3               | 2                      | 0             | -  | -   |
| MMK010779 | 46                        | pre                 | 2 | 1 | 0 | 2      | a2               | 0                      | 1             | -  | -   |
| MMK010780 | 31                        | pre                 | 2 | 0 | 0 | 2      | a2               | 0                      | 0             | -  | -   |
| MMK010781 | 44                        | pre                 | 2 | 0 | 0 | 2      | a3               | 0                      | 2             | +  | +   |
| MMK010794 | 52                        | pre                 | 2 | 1 | 0 | 2      | a3               | 2                      | 1             | +  | +   |
| MMK010818 | 51                        | pre                 | 2 | 0 | 0 | 2      | a1               | 0                      | 2             | +  | +   |
| MMK010835 | 42                        | pre                 | 0 | 0 | 0 | 0      | 1a               | 0                      | 0             | +  | +   |

| ID        | Edad durante la operación | Estado meno-páusico | T | N | M | Estado | Tipo HistoLógico | Infiltrado linfocítico | Angioinvasión | ER | PgR |
|-----------|---------------------------|---------------------|---|---|---|--------|------------------|------------------------|---------------|----|-----|
| MMK010846 | 47                        | pre                 | 2 | 0 | 0 | 2      | a1               | 0                      | 0             | +  | +   |
| MMK010858 | 42                        | pre                 | 2 | 1 | 0 | 2      | a3               | 2                      | 3             | +  | +   |
| MMK010864 | 52                        | pre                 | 2 | 1 | 0 | 2      | a1               | 0                      | 1             | -  | -   |
| MMK010869 | 45                        | pre                 | 2 | 0 | 0 | 2      | a1               | 0                      | 1             | -  | -   |
| MMK010903 | 47                        | pre                 | 2 | 0 | 0 | 2      | a1               | 0                      | 0             | +  | +   |

Tabla 13

|           |   |              |
|-----------|---|--------------|
| Si1-F     | 5'-CACCGAACGATATAAGCCAGCCTCAAGAGAGGC<br>TGGCTTTATATCGTTC-3'   | SEC ID Nº 23 |
| Si1-R     | 5' -AAAAGAACGATATAAGCCAGCCTCTCTGAAGG<br>CTGGCTTTATATCGTTC-3'  | SEC ID Nº 24 |
| Si1-Tiana | 5' -GAACGAT ATAAAGCC AGCC-3'                                  | SEC ID Nº 25 |
| Si3-F     | 5'-CACCTGGATGAATCATACCAGATTCAAGAGATCT<br>GGTATGATTATCCAG-3'   | SEC ID Nº 26 |
| Si3-R     | 5'-AAAACTGGATGAATCATACCAGATCTCTTGAATCT<br>GGTATGATTATCCAG-3'  | SEC ID Nº 27 |
| Si3-Diana | 5'-CTGGATGAATCATACCAGA-3'                                     | SEC ID Nº 28 |
| Si4-F     | 5'-CACCGTGTGGCTTGCCTAAATAATTCAAGAGATTA<br>TTTACGCAAGCCACAC-3' | SEC ID Nº 29 |
| Si4-R     | 5'-AAAAGTGTGGCTTGCCTAAATAATCTCTTGAATTATTT<br>ACGCAAGCCACAC-3' | SEC ID Nº 30 |

El análisis de expresión de genes del cáncer de mama descrito en la presente memoria, obtenido mediante una combinación de disección de captura por láser y micromatriz de ADNc pangenómico ha identificado genes específicos como dianas para la prevención y terapia contra el cáncer. Basándose en la expresión de un subconjunto de estos genes expresados diferencialmente, la presente invención proporciona marcadores de diagnóstico moleculares para identificar y detectar el cáncer de mama.

5 Los procedimientos descritos en la presente memoria también son útiles en la identificación de dianas moleculares adicionales para prevención, diagnóstico y tratamiento del cáncer de mama. Los datos indicados en la presente memoria junto con un mayor entendimiento del cáncer de mama, facilitan el desarrollo de nuevas estrategias de

10 diagnóstico y proporcionan indicios para la identificación de dianas moleculares para fármacos terapéuticos y agentes preventivos. Dicha información contribuye a un entendimiento más profundo de la tumorigénesis de mama y proporciona indicadores para desarrollar nuevas estrategias para el diagnóstico, tratamiento y finalmente la prevención del cáncer de mama.

15 Adicionalmente, aunque la invención se ha descrito con detalle y con referencia a realizaciones específicas de las mismas, deberá entenderse que la descripción anterior es de naturaleza ejemplar y explicativa y pretende ilustrar la invención y sus realizaciones preferidas.

## LISTADO DE SECUENCIAS

<110> ONCOTHERAPY SCIENCE, INC.  
THE UNIVERSITY OF TOKYO

<120> MÉTODO DE DIAGNÓSTICO DEL CÁNCER DE MAMA

5 <130> ONC-AO306P2

<150> US 60/505.571  
<151> 24-09-2003

<160> 31

<170> PatentIn versión 3. 1

10 <210> 1  
<211> 23  
<212> ADN  
<213> Artificial

<220>

15 <223> Secuencia cebadora sintetizada artificialmente

<400> 1  
**ctgttctggc ttcgttatgt tct** 23

<210>2  
<211> 23  
20 <212> ADN  
<213> Artificial

<220>

<223> Secuencia cebadora sintetizada artificialmente

<400> 2  
**agaaaaatacg gtcctcttgt tgc** 23

25 <210> 3  
<211> 22  
<212> ADN  
<213> Artificial

30 <220>

<223> Secuencia cebadora sintetizada artificialmente

<400> 3  
**cactgtaatg cacgacattt ga** 22

35 <210> 4  
<211> 23  
<212> ADN  
<213> Artificial

<220>

<223> Secuencia cebadora sintetizada artificialmente

40 <400> 4  
**gttacagctt agcacaaggc atc** 23

<210> 5  
<211> 22  
45 <212> ADN  
<213> Artificial

<220>

<223> Secuencia cebadora sintetizada artificialmente

|    |  |           |
|----|--|-----------|
|    | <400> 5  |           |
|    | <b>acctctgagt ttgatttccc aa</b>                      | <b>22</b> |
|    | <210> 6  |           |
|    | <211> 23   |           |
| 5  | <212> ADN  |           |
|    | <213> Artificial                                     |           |
|    | <220>  |           |
|    | <223> Secuencia cebadora sintetizada artificialmente |           |
|    | <400> 6  |           |
| 10 | <b>cgaggcgtt aacaatctac tgg</b>                      | <b>23</b> |
|    | <210> 7  |           |
|    | <211> 23   |           |
|    | <212> ADN  |           |
|    | <213> Artificial                                     |           |
| 15 | <220>  |           |
|    | <223> Secuencia cebadora sintetizada artificialmente |           |
|    | <400> 7  |           |
|    | <b>gaaactgtac ggggttaaa gag</b>                      | <b>23</b> |
|    | <210> 8  |           |
| 20 | <211> 23   |           |
|    | <212> ADN  |           |
|    | <213> Artificial                                     |           |
|    | <220>  |           |
|    | <223> Secuencia cebadora sintetizada artificialmente |           |
| 25 | <400> 8  |           |
|    | <b>catcaatgtg gtgagtgaca tct</b>                     | <b>23</b> |
|    | <210> 9  |           |
|    | <211> 23   |           |
|    | <212> ADN  |           |
| 30 | <213> Artificial                                     |           |
|    | <220>  |           |
|    | <223> Secuencia cebadora sintetizada artificialmente |           |
|    | <400> 9  |           |
|    | <b>aagcccttgg aacagaacat act</b>                     | <b>23</b> |
| 35 | <210> 10   |           |
|    | <211> 23   |           |
|    | <212> ADN  |           |
|    | <213> Artificial                                     |           |
|    | <220>  |           |
| 40 | <223> Secuencia cebadora sintetizada artificialmente |           |
|    | <400> 10   |           |
|    | <b>cagtaaacgt ggttctcaca ttg</b>                     | <b>23</b> |
|    | <210> 11   |           |
|    | <211> 20   |           |
| 45 | <212> ADN  |           |
|    | <213> Artificial                                     |           |
|    | <220>  |           |
|    | <223> Secuencia cebadora sintetizada artificialmente |           |

|    |  |    |
|----|--|----|
|    | <400> 11   |    |
|    | <b>cgaccacattt gtcaagctca</b>  | 20 |
|    | <210> 12   |    |
|    | <211> 23   |    |
| 5  | <212> ADN  |    |
|    | <213> Artificial   |    |
|    | <220>  |    |
|    | <223> Secuencia cebadora sintetizada artificialmente                           |    |
|    | <400> 12   |    |
| 10 | <b>ggttgagcac aggtaacttt att</b>   | 23 |
|    | <210> 13   |    |
|    | <211> 23   |    |
|    | <212> ADN  |    |
|    | <213> Artificial   |    |
| 15 | <220>  |    |
|    | <223> Secuencia cebadora sintetizada artificialmente                           |    |
|    | <400> 13   |    |
|    | <b>agaccctaaa gatcgccctt ctg</b>   | 23 |
|    | <210> 14   |    |
| 20 | <211> 23   |    |
|    | <212> ADN  |    |
|    | <213> Artificial   |    |
|    | <220>  |    |
|    | <223> Secuencia cebadora sintetizada artificialmente                           |    |
| 25 | <400> 14   |    |
|    | <b>gtgttttaag tcagcatgag cag</b>   | 23 |
|    | <210> 15   |    |
|    | <211> 51   |    |
|    | <212> ADN  |    |
| 30 | <213> Artificial   |    |
|    | <220>  |    |
|    | <223> Una secuencia de oligonucleótidos para ARNip sintetizada artificialmente |    |
|    | <400> 15   |    |
|    | <b>tcccgcgccc tttgttaggat tcgttcaaga gacgaatcct acaaaggcgcc c</b>              | 51 |
| 35 | <210> 16   |    |
|    | <211> 51   |    |
|    | <212> ADN  |    |
|    | <213> Artificial   |    |
|    | <220>  |    |
| 40 | <223> Una secuencia de oligonucleótidos para ARNip sintetizada artificialmente |    |
|    | <400> 16   |    |
|    | <b>aaaaggcgccgc tttgttaggat tcgtctcttg aacgaatcct acaaaggcgcc c</b>            | 51 |
|    | <210> 17   |    |
|    | <211> 51   |    |
| 45 | <212> ADN  |    |
|    | <213> Artificial   |    |
|    | <220>  |    |
|    | <223> Una secuencia de oligonucleótidos para ARNip sintetizada artificialmente |    |

|    |  |    |
|----|--|----|
|    | <400> 17   |    |
|    | tccccgtacg cggaatactt cgattcaaga gatcgaagta ttccgcgtac g                       | 51 |
|    | <210> 18   |    |
|    | <211> 51   |    |
| 5  | <212> ADN  |    |
|    | <213> Artificial   |    |
|    | <220>  |    |
|    | <223> Una secuencia de oligonucleótidos para ARNip sintetizada artificialmente |    |
|    | <400> 18   |    |
| 10 | aaaacgtacg cggaatactt cgatctttg aatcgaagta ttccgcgtac g                        | 51 |
|    | <210> 19   |    |
|    | <211> 21   |    |
|    | <212> ADN  |    |
|    | <213> Artificial   |    |
| 15 | <220>  |    |
|    | <223> Una secuencia cebadora para RT-PCR sintetizada artificialmente.          |    |
|    | <400> 19   |    |
|    | atggaaatcc catcaccatc t  | 21 |
| 20 | <210> 20   |    |
|    | <211> 23   |    |
|    | <212> ADN  |    |
|    | <213> Artificial   |    |
|    | <220>  |    |
|    | <223> Una secuencia cebadora para RT-PCR sintetizada artificialmente           |    |
| 25 | <400> 20   |    |
|    | ggttgagcac agggtacttt att  | 23 |
|    | <210> 21   |    |
|    | <211> 20   |    |
|    | <212> ADN  |    |
| 30 | <213> Artificial   |    |
|    | <220>  |    |
|    | <223> Una secuencia cebadora para RT-PCR sintetizada artificialmente           |    |
|    | <400> 21   |    |
|    | gccttcatca tccaaacatt  | 20 |
| 35 | <210> 22   |    |
|    | <211> 20   |    |
|    | <212> ADN  |    |
|    | <213> Artificial   |    |
|    | <220>  |    |
| 40 | <223> Una secuencia cebadora para RT-PCR sintetizada artificialmente           |    |
|    | <400> 22   |    |
|    | ggcaaataatg tctgccttgt   | 20 |
|    | <210> 23   |    |
|    | <211> 51   |    |
| 45 | <212> ADN  |    |
|    | <213> Artificial   |    |
|    | <220>  |    |
|    | <223> Una secuencia de oligonucleótidos para ARNip sintetizada artificialmente |    |

|    |  |    |
|----|--|----|
|    | <400> 23   |    |
|    | caccgaacga tataaagcca gccttcaaga gaggctggct ttatatcggtt c                      | 51 |
|    | <210> 24   |    |
| 5  | <211> 51   |    |
|    | <212> ADN  |    |
|    | <213> Artificial   |    |
|    | <220>  |    |
|    | <223> Una secuencia de oligonucleótidos para ARNip sintetizada artificialmente |    |
|    | <400> 24   |    |
| 10 | aaaagaacga tataaagcca gcctctcttg aaggctggct ttatatcggtt c                      | 51 |
|    | <210> 25   |    |
|    | <211> 19   |    |
|    | <212> ADN  |    |
|    | <213> Artificial   |    |
| 15 | <220>  |    |
|    | <223> Una secuencia diana para ARNip sintetizada artificialmente               |    |
|    | <400> 25   |    |
|    | gaacgatata aagccagcc   | 19 |
|    | <210> 26   |    |
| 20 | <211> 51   |    |
|    | <212> ADN  |    |
|    | <213> Artificial   |    |
|    | <220>  |    |
|    | <223> Una secuencia de oligonucleótidos para ARNip sintetizada artificialmente |    |
| 25 | <400> 26   |    |
|    | caccctggat gaatcataacc agattcaaga gatctggat gattcatcca g                       | 51 |
|    | <210> 27   |    |
|    | <211> 51   |    |
| 30 | <212> ADN  |    |
|    | <213> Artificial   |    |
|    | <220>  |    |
|    | <223> Una secuencia de oligonucleótidos para ARNip sintetizada artificialmente |    |
|    | <400> 27   |    |
|    | aaaactggat gaatcataacc agatcttttg aatctggat gattcatcca g                       | 51 |
| 35 | <210> 28   |    |
|    | <211> 19   |    |
|    | <212> ADN  |    |
|    | <213> Artificial   |    |
|    | <220>  |    |
| 40 | <223> Una secuencia diana para ARNip sintetizada artificialmente               |    |
|    | <400> 28   |    |
|    | ctggatgaat cataccaga   | 19 |
|    | <210> 29   |    |
|    | <211> 51   |    |
| 45 | <212> ADN  |    |
|    | <213> Artificial   |    |
|    | <220>  |    |

|       |  |    |
|-------|--|----|
| <223> | Una secuencia de oligonucleótidos para ARNip sintetizada artificialmente       |    |
| <400> | 29   |    |
|       | caccgtgtgg cttgcgtaaa taattcaaga gattatttac gcaagccaca c                       | 51 |
| 5     | <210> 30   |    |
|       | <211> 51   |    |
|       | <212> ADN  |    |
|       | <213> Artificial   |    |
|       | <220>  |    |
|       | <223> Una secuencia de oligonucleótidos para ARNip sintetizada artificialmente |    |
| 10    | <400> 30   |    |
|       | aaaagtgtgg cttgcgtaaa taatctcttg aattatttac gcaagccaca c                       | 51 |
|       | <210> 31   |    |
|       | <211> 19   |    |
|       | <212> ADN  |    |
| 15    | <213> Artificial   |    |
|       | <220>  |    |
|       | <223> Una secuencia diana para ARNip sintetizada artificialmente               |    |
|       | <400> 31   |    |
|       | gtgtggcttg cgtaaataa   | 19 |
| 20    |  |    |

### REIVINDICACIONES

- 5        1.        Un procedimiento *in vivo* de diagnóstico de carcinoma ductal invasivo (IDC) o una predisposición para desarrollar IDC en un sujeto, que comprende determinar un nivel de expresión de un gen asociado al cáncer de mama en una muestra biológica derivada de un paciente, en el que un aumento en dicho nivel de expresión de la muestra en comparación con un nivel de control normal de dicho gen indica que dicho sujeto padece o está en riesgo de desarrollar IDC, en el que dicho gen asociado al cáncer de mama es BRC nº 456.
- 10      2.        El procedimiento de la reivindicación 1, en el que dicho nivel de expresión de la muestra es al menos un 10% superior a dicho nivel de control normal.
- 10      3.        El procedimiento de la reivindicación 1, en el que el nivel de expresión del gen se determina mediante un procedimiento seleccionado del grupo que consiste en:
- (a) detectar el ARNm de BRC nº 456,
  - (b) detectar la proteína codificada por BRC nº 456, y
  - (c) detectar una actividad quinasa de la proteína codificada por BRC nº 456.
- 15      4.        El procedimiento de la reivindicación 3, en el que dicha detección se realiza en una matriz de ADN.
- 15      5.        El procedimiento de la reivindicación 1, en el que dicha muestra biológica derivada del paciente comprende
- (a) una célula epitelial;
  - (b) una célula IDC; o
  - (c) una célula epitelial de tejido que se sabe o se sospecha que es un IDC.
- 20      6.        Un procedimiento para la exploración de un compuesto para el tratamiento o prevención de IDC, comprendiendo dicho procedimiento las etapas de:
- (a) poner en contacto un compuesto de ensayo con el polipéptido codificado por BRC nº 456; y
  - (b) detectar la actividad de unión entre dicho polipéptido y dicho compuesto de ensayo.
- 25      7.        Un procedimiento para la exploración de un compuesto para el tratamiento o prevención de IDC, comprendiendo dicho método las etapas de:
- 25      poner en contacto un compuesto candidato con una célula que exprese BRC nº 456; y
- 25      determinar el nivel de expresión de BRC nº 456.
- 25      8.        El procedimiento de la reivindicación 7, en el que dicha célula es una célula IDC.
- 25      9.        Un procedimiento para la exploración de un compuesto para el tratamiento o prevención de IDC, comprendiendo dicho procedimiento las etapas de:
- (a) poner en contacto un compuesto de ensayo con el polipéptido codificado por BRC nº 456;
  - (b) detectar la actividad quinasa del polipéptido de la etapa (a).
- 30      10.      El uso de una secuencia de nucleótidos complementaria a la secuencia codificante de BRC nº 456 para la preparación de una composición farmacéutica para el tratamiento o prevención de IDC en un sujeto.
- 35      11.      El uso de una composición de ARNip, que reduzca la expresión de BRC nº 456 para la preparación de una composición farmacéutica para el tratamiento o prevención de IDC en un sujeto.
- 35      12.      El uso de la reivindicación 11, en el que el ARNip comprende la cadena con sentido que comprende una secuencia de nucleótidos seleccionada del grupo que consiste en las secuencias de nucleótidos de la SEC ID Nº: 25, 28 y 31 como la secuencia diana.
- 40      13.      Una composición farmacéutica que comprende el ARNip contra BRC nº 456, en el que dicho ARNip comprende la cadena con sentido que comprende una secuencia de nucleótidos seleccionada del grupo que consiste en las secuencias de nucleótidos de la SEC ID Nº: 25, 28 y 31 como la secuencia diana para su uso en el tratamiento o prevención de IDC.
- 45      14.      Una molécula bicatenaria que comprende una cadena con sentido y una cadena antisentido, en el que la cadena con sentido comprende una secuencia de ribonucleótidos correspondiente a una secuencia diana que consiste en la SEC ID Nº: 25, 28 y 31 y en el que la cadena antisentido comprende una secuencia de ribonucleótidos

que es complementaria a dicha cadena con sentido, en el que dicha cadena con sentido y dicha cadena antisentido hibridan entre sí para formar dicha molécula bicatenaria para su uso en el tratamiento o prevención de IDC.

15. La molécula bicatenaria de la reivindicación 14, que tiene la fórmula general seleccionada del grupo que consiste en:

- 5      (a) gaacgauauaaagccagcc-[B]-ggcuggccuuauaucguuc,  
          (b) cuggaugaaaucauaccaga-[B]-ucugguaugauucauccag, y  
          (c) guguggcuugcguaaauaa-[B]-uuauuuacgcaagccacac,

en el que [B] es una secuencia en bucle de ribonucleótidos que consiste en 3 a 23 nucleótidos.

Fig. 1

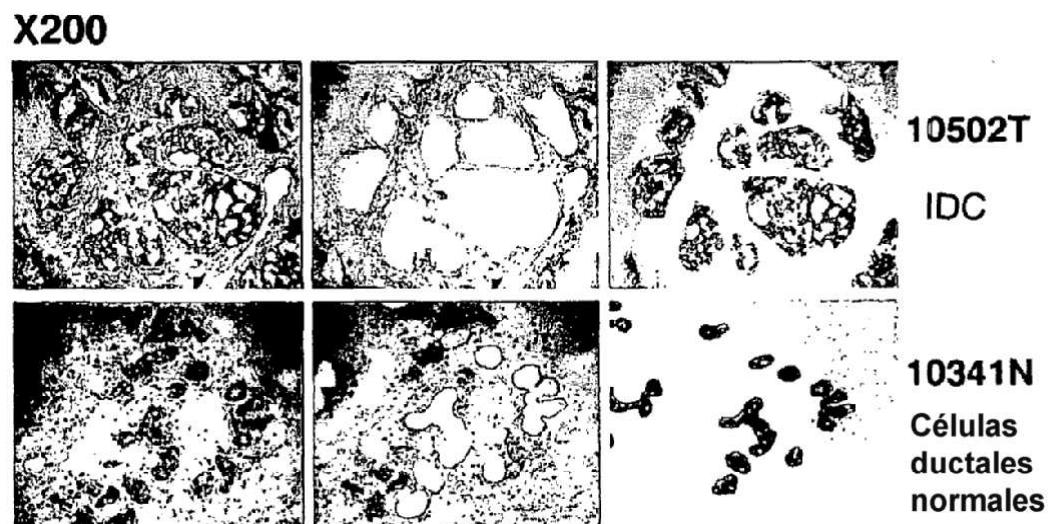
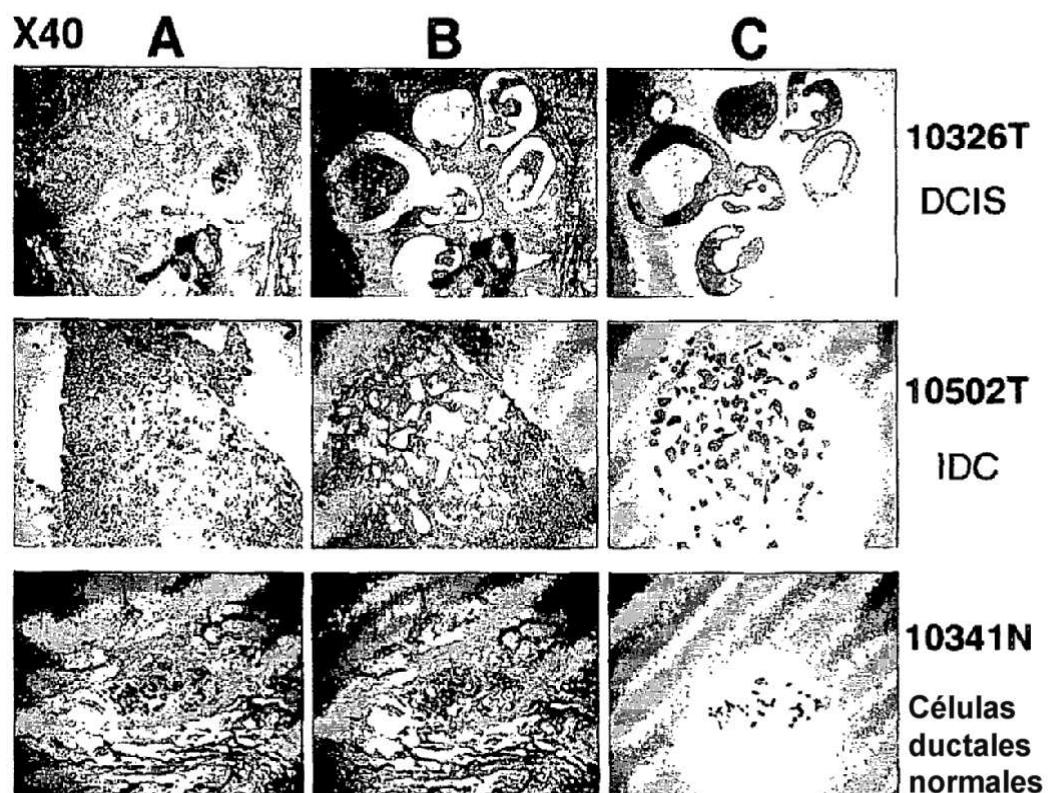


Fig. 2-1

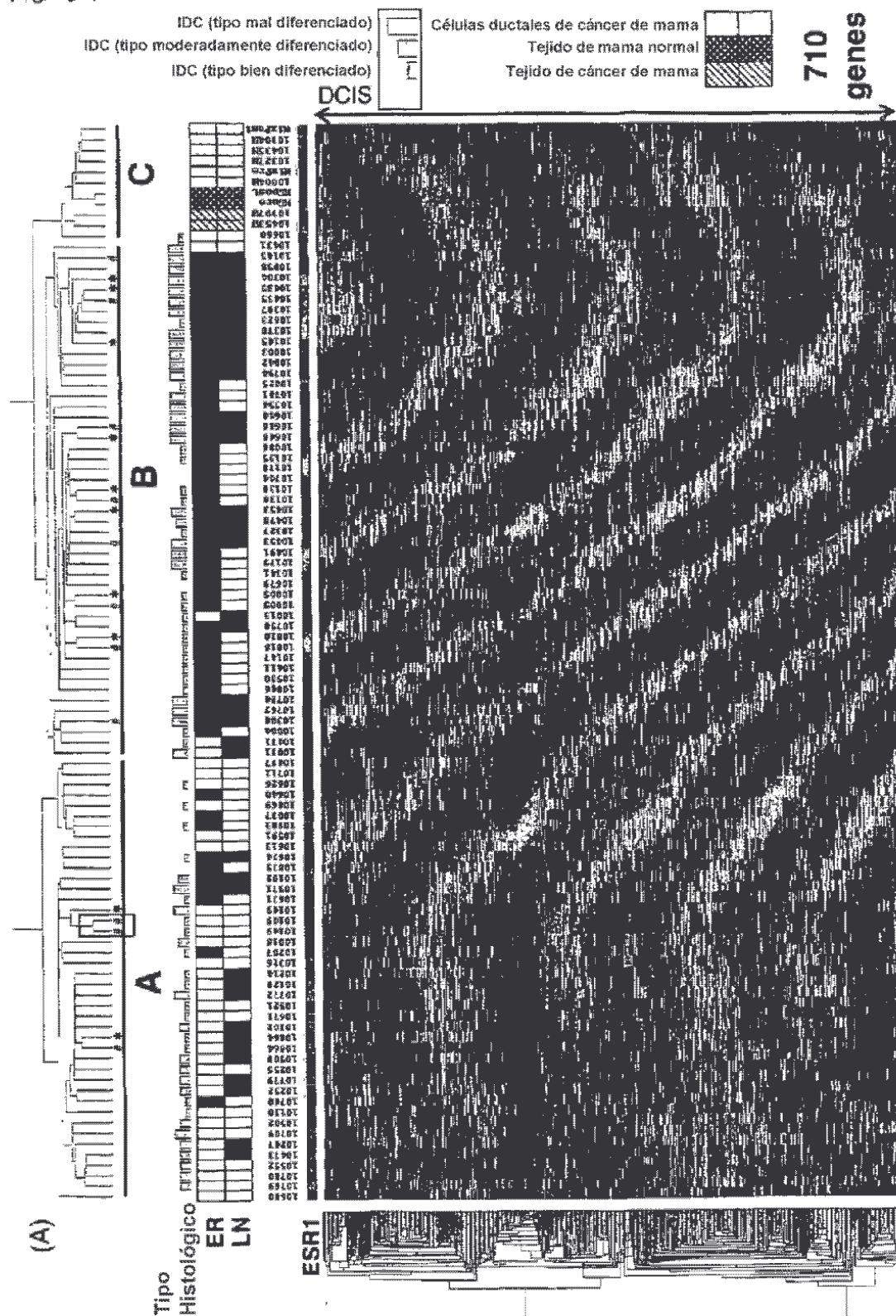


Fig. 2-2

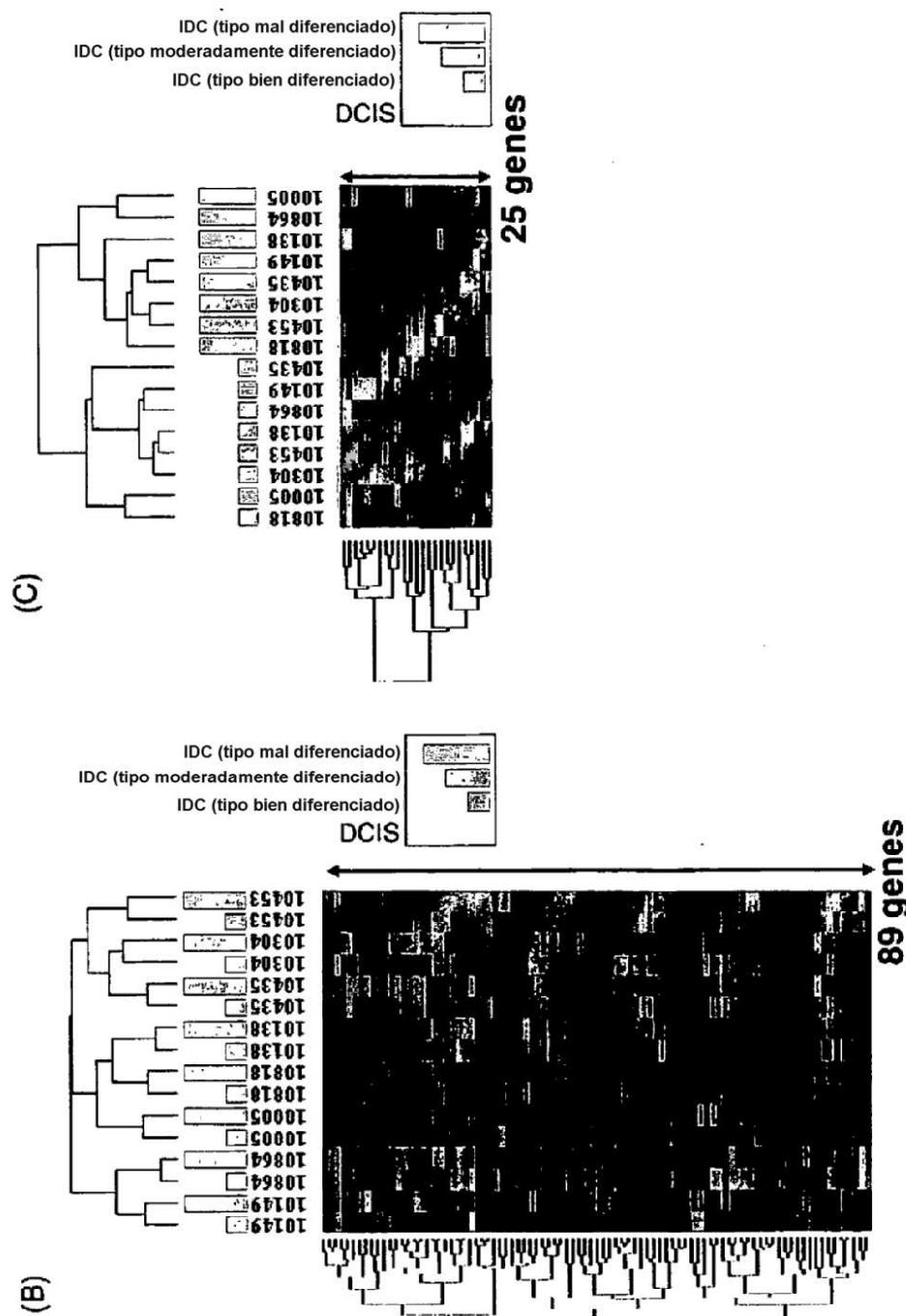


Fig. 3

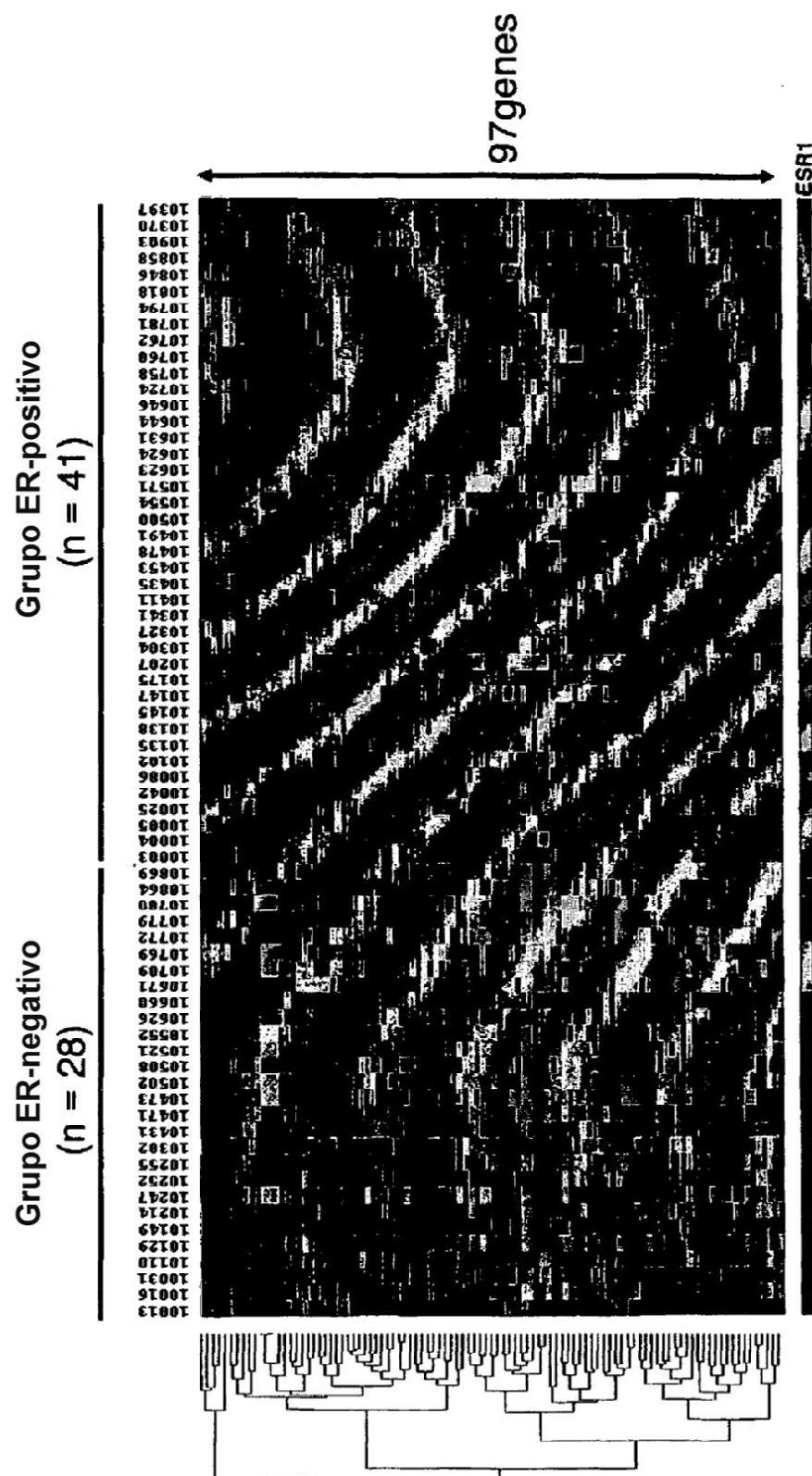


Fig. 4

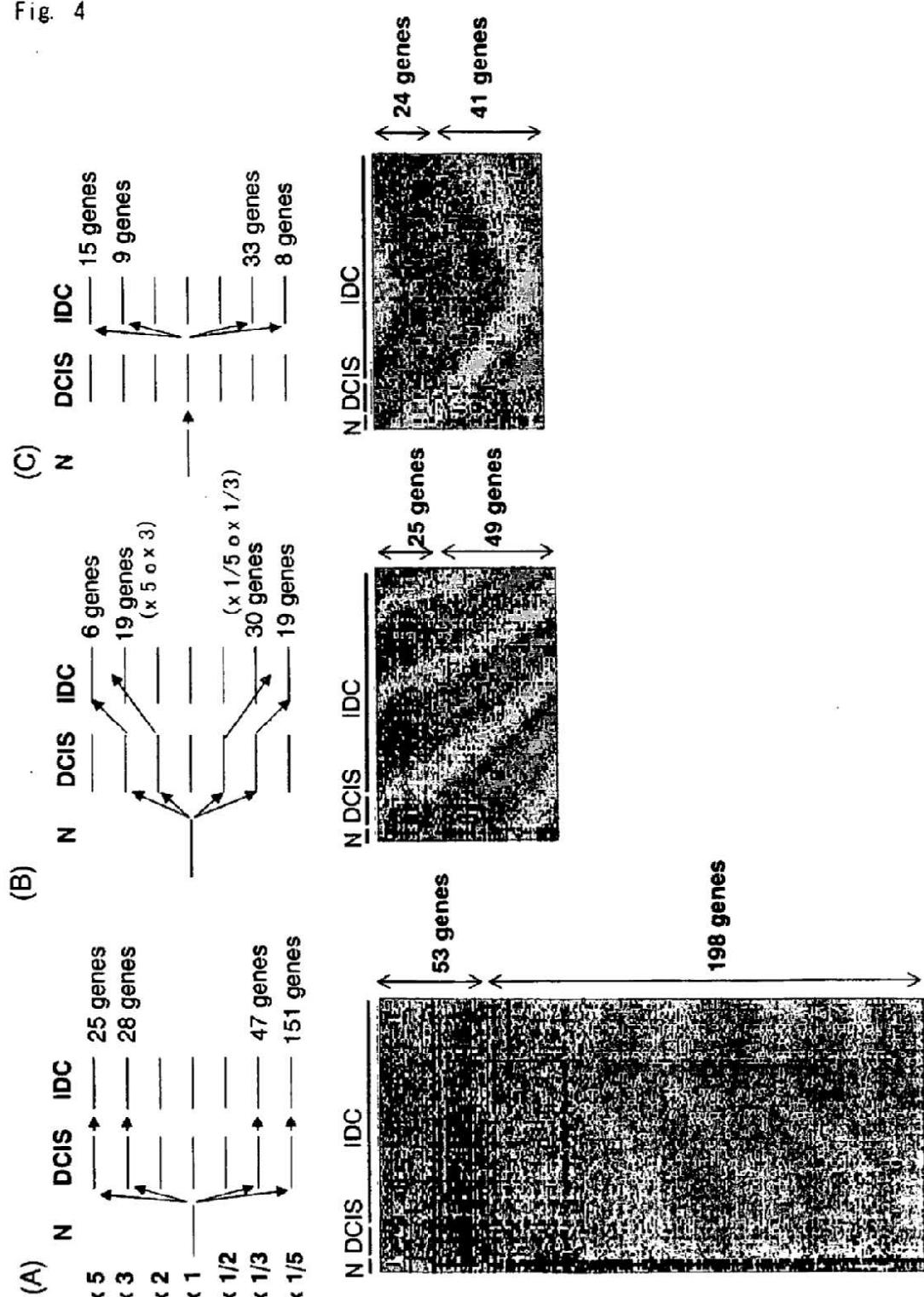


Fig. 5

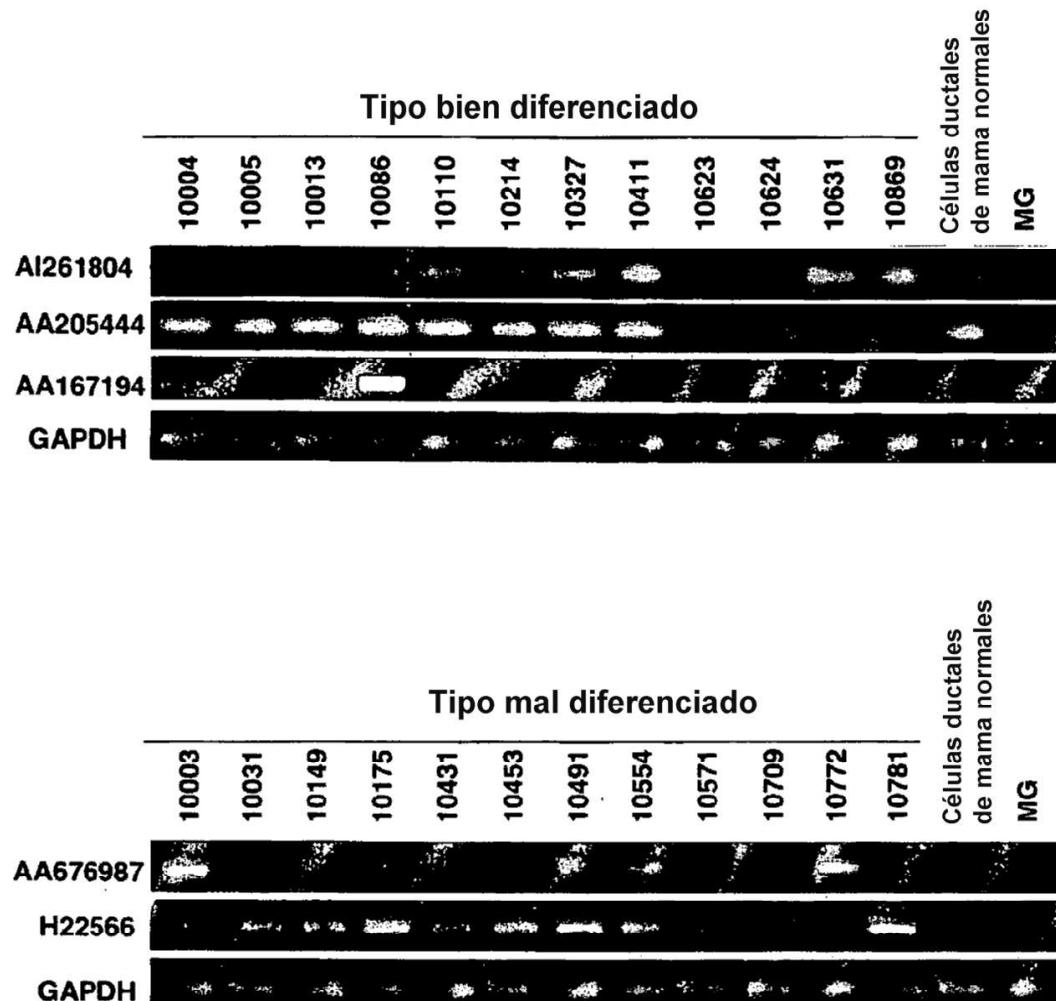


Fig. 6

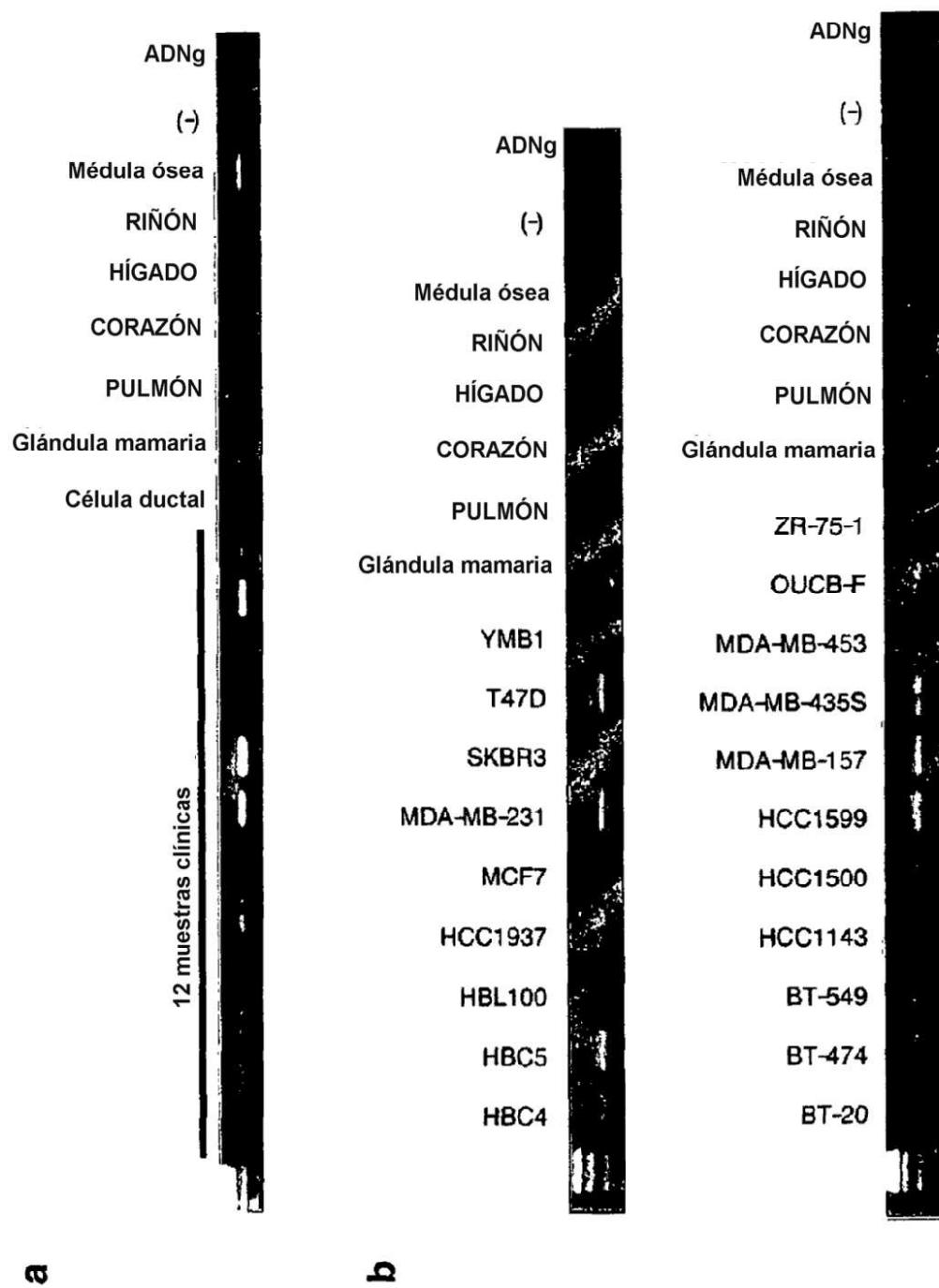
**a****b**

Fig. 7

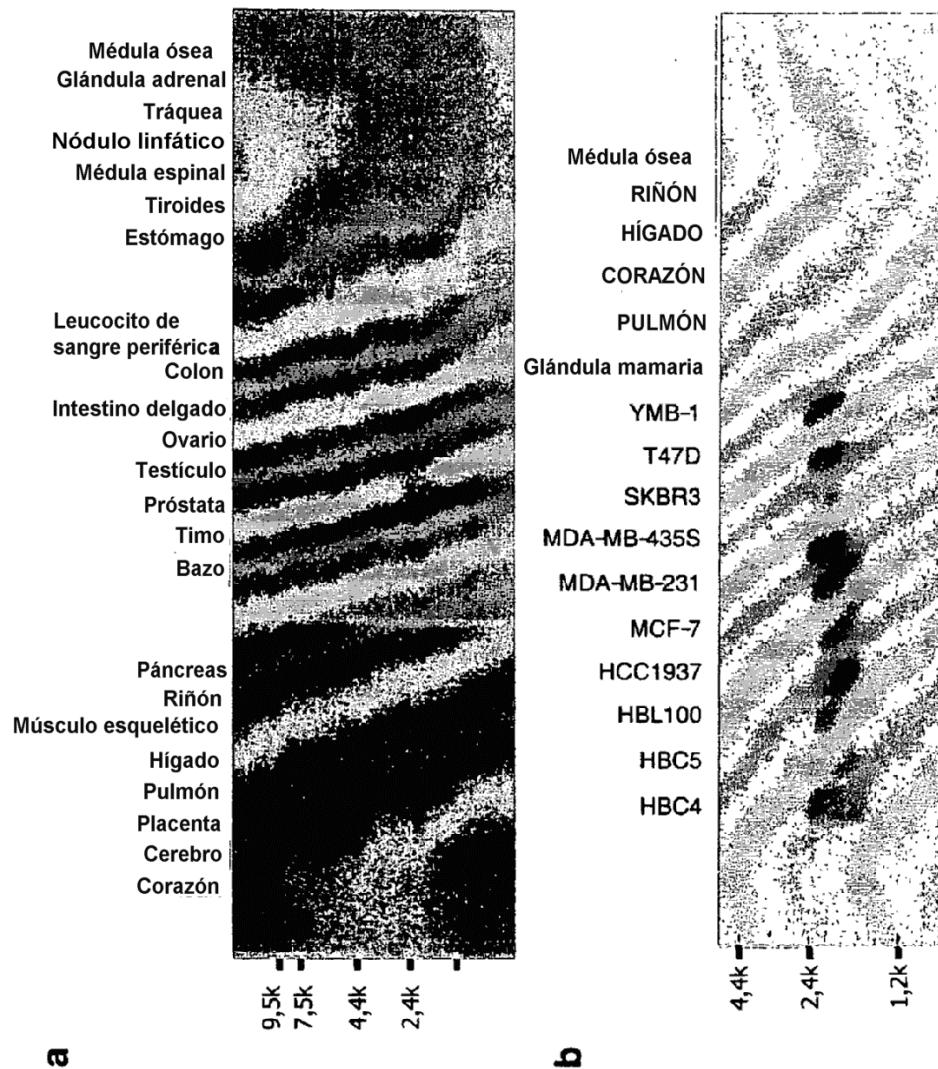


Fig. 8-1

a

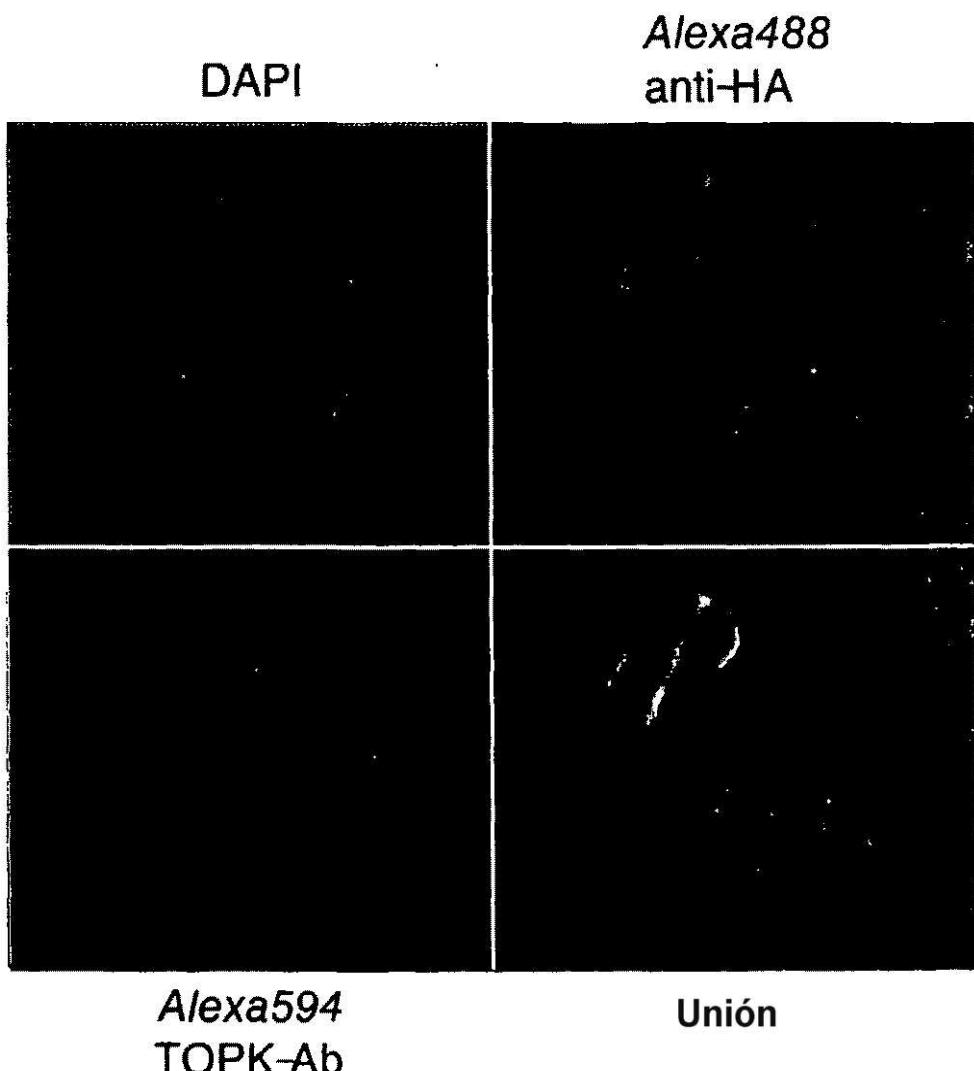


Fig. 8-2

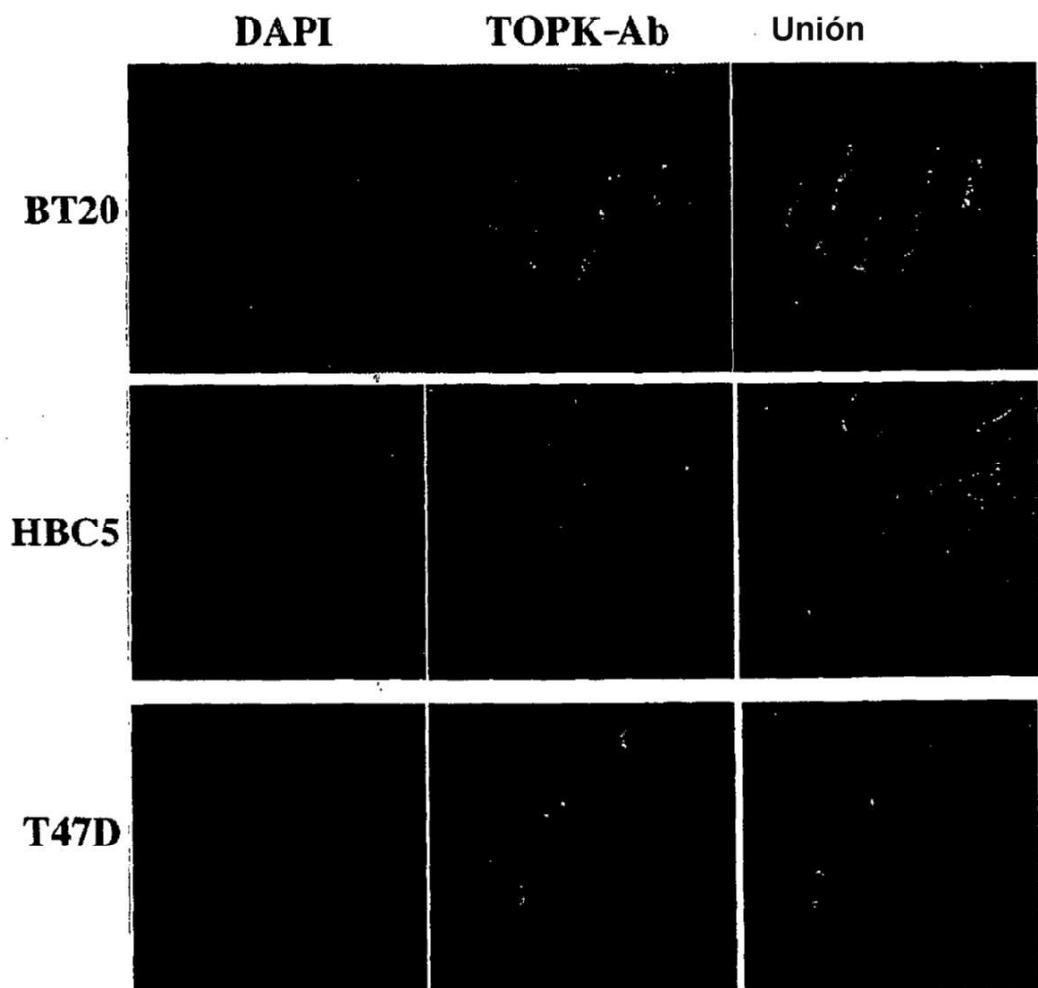
**b**

Fig. 9

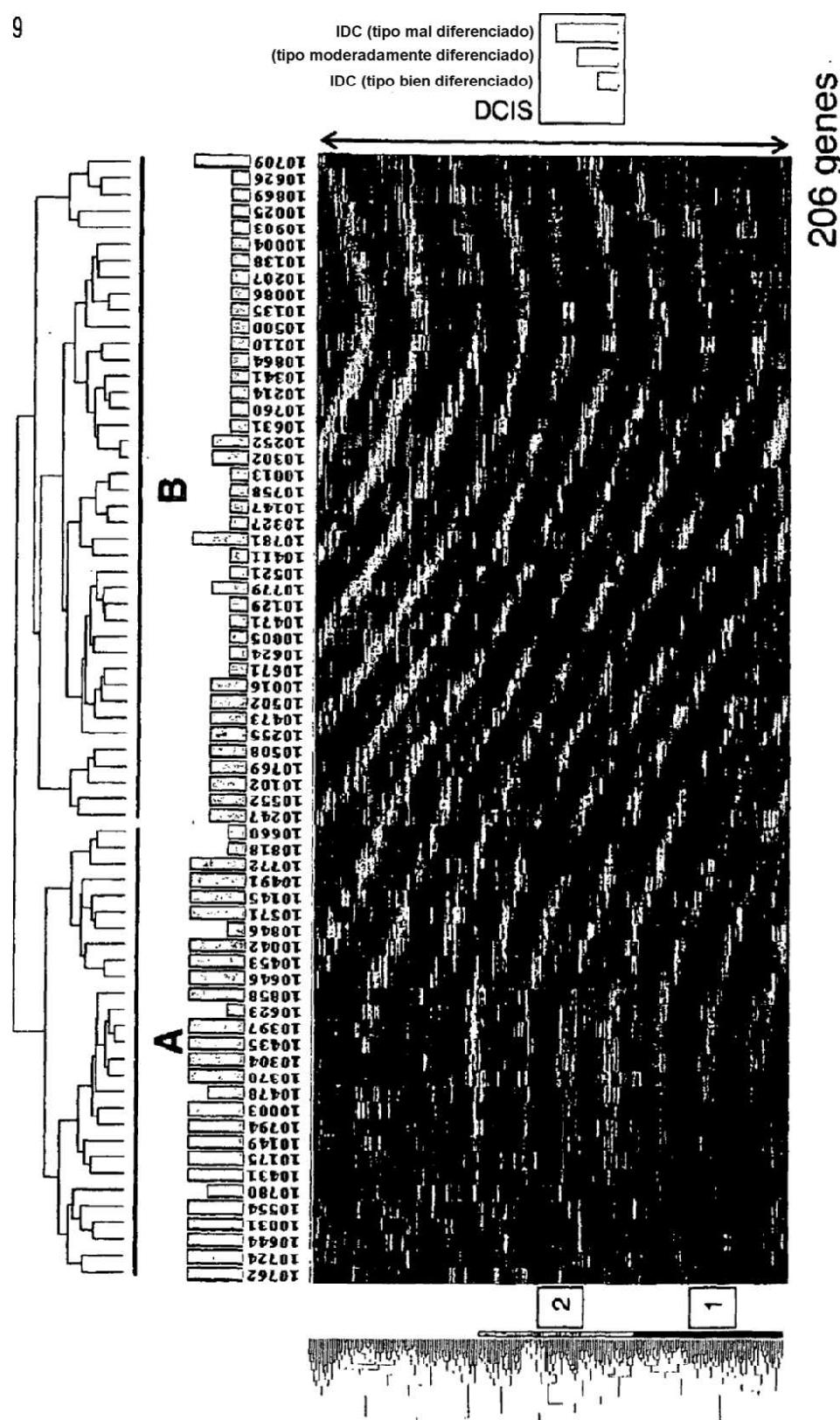


Fig. 10-1

## (B) Metástasis en nódulos linfáticos

(A)

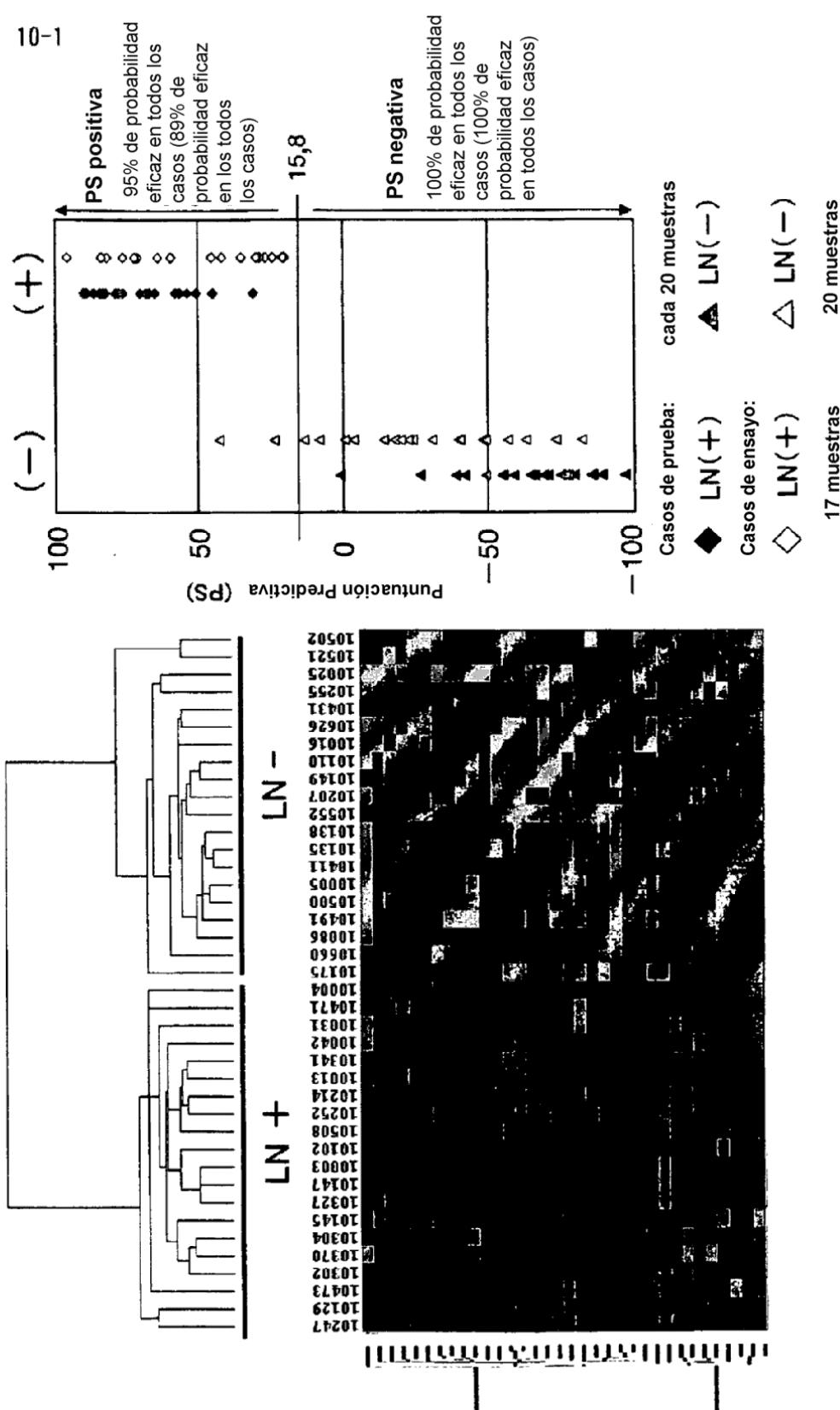


Fig. 10-2

(C)

|    |          | Estado en<br>nódulo linfático |     | Total |
|----|----------|-------------------------------|-----|-------|
|    |          | (+)                           | (-) |       |
| PS | Positivo | 37                            | 2   | 39    |
|    | Negativo | 0                             | 38  | 38    |
|    | Total    | 37                            | 40  | 77    |

(ensayo  $\chi^2$  para independencia)

|    |          | Estado en<br>nódulo linfático |     | Total |
|----|----------|-------------------------------|-----|-------|
|    |          | (+)                           | (-) |       |
| PS | Positivo | 17                            | 2   | 19    |
|    | Negativo | 0                             | 18  | 18    |
|    | Total    | 17                            | 20  | 37    |

 $P<0,001$ (ensayo  $\chi^2$  para independencia)