

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 457 534**

51 Int. Cl.:

C12Q 1/68

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **01.06.2009 E 09770678 (2)**

97 Fecha y número de publicación de la concesión europea: **13.11.2013 EP 2297359**

54 Título: **Perfiles de expresión génica para predecir desenlaces en cáncer de mama**

30 Prioridad:

30.05.2008 US 57508 P

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

28.04.2014

73 Titular/es:

**THE UNIVERSITY OF NORTH CAROLINA AT
CHAPEL HILL (25.0%)
308 Bynum Hall Campus Box 4105
Chapel Hill, NC 27599-4105, US;
UNIVERSITY OF UTAH RESEARCH FOUNDATION
(25.0%);
WASHINGTON UNIVERSITY (25.0%) y
BRITISH COLUMBIA CANCER AGENCY BRANCH
(25.0%)**

72 Inventor/es:

**PEROU, CHARLES M.;
PARKER, JOEL S.;
MARRON, JAMES STEPHEN;
NOBEL, ANDREW;
BERNARD, PHILIP S.;
ELLIS, MATTHEW;
MARDIS, ELAINE;
NIELSEN, TORSTEN O. y
CHEANG, MAGGIE**

74 Agente/Representante:

SUGRAÑES MOLINÉ, Pedro

ES 2 457 534 T3

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

DESCRIPCIÓN

Perfiles de expresión génica para predecir desenlaces en cáncer de mama.

5 CAMPO DE LA INVENCION

La presente invención se refiere a métodos para clasificar muestras de cáncer de mama en subtipos y para evaluar el pronóstico y la respuesta a la terapia para pacientes que padecen cáncer de mama.

10 ANTECEDENTES DE LA INVENCION

El cáncer de mama es el segundo cáncer más común ente mujeres en los Estados Unidos, segundo sólo tras el cáncer de piel. Una mujer en los EE.UU. tiene una probabilidad de uno entre ocho de desarrollar cáncer de mama durante su vida, y la Sociedad americana contra el cáncer ("*American Cancer Society*") estima que se notificarán más de 178.480 casos nuevos de cáncer de mama invasivo en los Estados Unidos en 2007. El cáncer de mama es la segunda causa principal de muertes por cáncer en mujeres, con más de 40.000 muertes al año. Los métodos de detección mejorados, el cribado masivo y los avances en el tratamiento a lo largo de la última década han mejorado significativamente las perspectivas para las mujeres diagnosticadas con cáncer de mama. Hoy en día, aproximadamente el 80% de los casos de cáncer de mama se diagnostican en los estadios tempranos de la enfermedad cuando las tasas de supervivencia están en su punto más alto. Como resultado, aproximadamente el 85% de las pacientes con cáncer de mama viven al menos cinco años tras el diagnóstico. A pesar de estos avances, aproximadamente el 20% de las mujeres diagnosticadas con cáncer de mama en estadio temprano tienen un mal desenlace a diez años y padecerán reaparición de la enfermedad, metástasis o muerte dentro de este periodo de tiempo.

Una investigación significativa se ha centrado en la identificación de métodos y factores para evaluar el pronóstico de cáncer de mama y predecir la respuesta terapéutica (véase en general, Ross y Hortobagyi, eds. (2005) *Molecular Oncology of Breast Cancer* (Jones y Bartlett Publishers, Boston, MA) y las referencias citadas en el mismo). Los indicadores de pronóstico incluyen factores convencionales, tales como tamaño del tumor, estado de los ganglios y grado histológico, así como marcadores moleculares que proporcionan alguna información con respecto al pronóstico y la respuesta probable a tratamientos particulares. Por ejemplo, la determinación del estado del receptor de la hormona esteroidea estrógeno (ER) y progesterona (PgR) se ha convertido en un procedimiento de rutina en la evaluación de pacientes con cáncer de mama. Véase, por ejemplo, Fitzgibbons *et al.*, Arch. Pathol. Lab. Med. 124:966-78, 2000. Es más probable que los tumores que son positivos para el receptor de la hormona respondan a la terapia hormonal y también normalmente crecen de manera menos agresiva, dando como resultado así un mejor pronóstico para las pacientes con tumores ER+/PgR+. La sobreexpresión del receptor del factor de crecimiento epidérmico humano 2 (HER-2/neu), una proteína receptora tirosina cinasa transmembrana, se ha correlacionado con mal pronóstico de cáncer de mama (véase, por ejemplo, Ross *et al.*, The Oncologist 8:307-25, 2003), y los niveles de expresión de HER-2 en tumores de mama se usan para predecir la respuesta al anticuerpo monoclonal anti-HER-2 terapéutico trastuzumab (Herceptin®, Genentech, South San Francisco, CA).

SUMARIO DE LA INVENCION

Se proporcionan métodos para clasificar y para evaluar el pronóstico y el tratamiento de un sujeto con cáncer de mama. Los métodos incluyen predicción de subtipo de cáncer de mama usando un algoritmo supervisado entrenado para estratificar sujetos basándose en el subtipo intrínseco de cáncer de mama. El modelo de predicción se basa en el perfil de expresión génica de los genes intrínsecos enumerados en la tabla 1. En algunas realizaciones, el algoritmo es un algoritmo del centroide más cercano, similar al algoritmo del análisis de predicción de micromatriz (PAM). El algoritmo puede entrenarse basándose en datos obtenidos a partir de los perfiles de expresión génica depositados como número de registro GSE10886 en el National Center for Biotechnology Information Gene Expression Omnibus. Este modelo de predicción, denominado en el presente documento modelo de clasificación PAM50, puede usarse para predecir con precisión el subtipo intrínseco de un sujeto diagnosticado con o que se sospecha que tiene cáncer de mama.

Además se proporcionan composiciones y métodos para predecir el desenlace o la respuesta a la terapia de un sujeto diagnosticado con o que se sospecha que tiene cáncer de mama. Estos métodos son útiles para orientar o determinar las opciones de tratamiento para un sujeto que padece cáncer de mama. Los métodos de la invención incluyen además medios para evaluar perfiles de expresión génica, que incluyen micromatrices y ensayos de reacción en cadena de la polimerasa cuantitativa, así como kits que comprenden reactivos para poner en práctica los métodos de la invención.

BREVE DESCRIPCION DE LOS DIBUJOS

La figura 1 muestra desenlaces basándose en predicciones de subtipo usando el clasificador PAM50. La clasificación PAM50 para LumA, LumB, enriquecido en HER-2, de tipo basal y de tipo normal muestra significación pronóstica para 1451 pacientes a lo largo de los 5 conjuntos de prueba combinados (A), en 376 pacientes a las que

se administró sólo terapia endocrina (B) y en 701 pacientes negativas para ganglios a las que no se administró terapia sistémica adyuvante (C).

La figura 2 muestra la clasificación del riesgo para casos de prueba usando un modelo completo de subtipos intrínsecos y dos variables clínicas. (A) riesgo de puntuaciones de recidiva representado gráficamente para cada subtipo de cáncer de mama: puntuaciones de riesgo bajo $< -0,1$, puntuaciones de riesgo moderado entre $-0,1$ y $0,2$, y puntuaciones de riesgo alto $\geq 0,2$. (B) diagramas de Kaplan-Meier y significación de la puntuación de riesgo para las 1286 muestras de prueba, (C) 376 pacientes que recibieron sólo terapia endocrina adyuvante y (D) 560 pacientes que eran negativas para ganglios y no recibieron terapia sistémica adyuvante.

La figura 3 muestra la puntuación lineal para el pronóstico usando el modelo clínico de subtipos para determinar el riesgo de recidiva en 5 años. El ajuste lineal con intervalos de confianza del 95% calibra el riesgo de puntuación de recidiva. El modelo de riesgo continuo con subtipo y variables clínicas (T y N) se calibró a partir de 657 pacientes con cáncer de mama en estadio temprano positivo para ER (A) y en 1286 pacientes con enfermedad positiva para ER y negativa para ER y estadio 1-3 (B).

La figura 4 muestra la asociación de subtipo intrínseco de PAM50, determinada mediante qPCR a partir de bloques de parafina, siendo (A) la supervivencia libre de recidiva y (B) la supervivencia específica de la enfermedad entre 702 mujeres con carcinoma de mama invasivo tratado con tamoxifeno adyuvante.

La figura 5 muestra el análisis de Kaplan-Meier de supervivencia específica de la enfermedad de cáncer de mama para pacientes estratificadas en las categorías de riesgo bajo, medio y alto aplicando el algoritmo riesgo de recidiva a los datos de qPCR generados a partir de bloques de parafina. (A) ROR-S, (B) ROR-C.

La figura 6 muestra el análisis de Kaplan-Meier de supervivencia específica de la enfermedad de cáncer de mama para pacientes estratificadas en las categorías de riesgo bajo, medio y alto (tal como se definió anteriormente en el material independiente aplicando el algoritmo ROR-C a mujeres con (A) enfermedad negativa para ganglios y (B) enfermedad positiva para ganglios).

La figura 7 muestra el análisis de Kaplan-Meier de supervivencia específica de la enfermedad de cáncer de mama para pacientes estratificadas en las categorías negativo para ganglios o positivo para ganglios entre mujeres con un riesgo bajo de ROR-C (panel superior a la izquierda), mujeres con riesgo moderado de ROR-C (panel superior a la derecha); mujeres con riesgo alto de ROR-C (panel inferior a la izquierda). Para la comparación directa, todas las curvas están superpuestas en el panel inferior a la derecha. Entre las mujeres con ROR-C bajo, no hay diferencia significativa en el desenlace por estado de los ganglios.

La figura 8 muestra que un análisis del modelo ROR-C frente a la probabilidad de supervivencia, estratificado por el número de ganglios linfáticos implicados, revela buenos desenlaces independientemente de la categoría del estado de los ganglios entre pacientes con valores de ROR-C menores de 25, que tienen intervalos de confianza del 95% solapantes (indicados mediante líneas discontinuas).

La figura 9 muestra los resultados de análisis de Kaplan-Meier que se realizó por separado en cada grupo de riesgo de adyuvante, y se sometieron a prueba las diferencias en la supervivencia entre los grupos de riesgo del 90-95% y del 95-100% usando la prueba de rangos logarítmicos.

La figura 10 muestra los resultados de análisis de Kaplan-Meier que se realizó por separado en cada grupo de riesgo de adyuvante, y diferencias en la supervivencia entre (A) BCSS predicha por adyuvante del 80-90%, ROR-S bajo frente a medio/alto; (B) BCSS predicha por adyuvante del 70-80%, ROR-S bajo frente a medio/alto; y, (C) BCSS predicha por adyuvante $<70\%$, ROR-S bajo frente a medio/alto.

DESCRIPCIÓN DETALLADA DE LA INVENCION

Visión general

A pesar de los recientes avances, el reto del tratamiento del cáncer sigue siendo dirigir regímenes de tratamiento específicos a distintos tipos de tumor con diferente patogenia, y en última instancia personalizar el tratamiento tumoral con el fin de optimizar el desenlace. En particular, una vez se diagnostica a una paciente con cáncer, tal como cáncer de mama, existe una necesidad de métodos que permitan al médico predecir el transcurso esperado de la enfermedad, incluyendo la probabilidad de reaparición del cáncer, supervivencia a largo plazo de la paciente y similares y por consiguiente seleccionar las opciones de tratamiento más apropiadas.

Para los fines de la presente invención, "cáncer de mama" incluye, por ejemplo, los estados clasificados mediante biopsia o histología como patología maligna. La descripción clínica de diagnósticos de cáncer de mama es bien conocida en la técnica médica. Un experto en la técnica apreciará que cáncer de mama se refiere a cualquier tumor maligno del tejido de la mama, incluyendo, por ejemplo, carcinomas y sarcomas. Las realizaciones particulares de cáncer de mama incluyen carcinoma ductal *in situ* (DCIS), carcinoma lobular *in situ* (LCIS) o carcinoma mucinoso.

Cáncer de mama también se refiere a carcinoma ductal infiltrante (IDC) o lobular infiltrante (ILC). En la mayoría de las realizaciones de la invención, el sujeto de interés es una paciente humana que se sospecha que padece o realmente diagnosticada con cáncer de mama.

Cáncer de mama es una enfermedad heterogénea con respecto a alteraciones moleculares y composición celular. Esta diversidad crea un reto para los investigadores que tratan de desarrollar clasificaciones que son clínicamente significativas. La obtención del perfil de expresión génica mediante micromatriz ha proporcionado comprensión de la complejidad de los tumores de mama y puede usarse para proporcionar información de pronóstico más allá de parámetros patológicos convencionales (1-7).

La obtención de perfiles de expresión de cáncer de mama identifica subtipos moleculares biológica y clínicamente distintos que pueden requerir diferentes enfoques de tratamiento [van't Veer 2005][Loi 2007][Cheang 2008a]. Los subtipos intrínsecos principales de cáncer de mama denominados luminal A, luminal B, enriquecido en HER-2, de tipo basal tienen características clínicas, riesgo de recidiva y respuesta a tratamiento distintos [Sorlie 2003]. Los subtipos "intrínsecos" conocidos como luminal A (LumA), luminal B (LumB), enriquecido en HER2, de tipo basal y de tipo normal se descubrieron usando agrupamientos jerárquicos no supervisados de datos de micromatriz (1,8). Los genes intrínsecos, tal como se describe en Perou *et al.* (2000) *Nature* 406:747-752, se eligen estadísticamente para que tengan variación baja en la expresión entre repeticiones de muestras biológicas del mismo individuo y variación alta en la expresión en muestras de diferentes individuos. Por tanto, los genes intrínsecos son los genes clasificadores para la clasificación del cáncer de mama. Aunque no se usó información clínica para derivar los subtipos intrínsecos de cáncer de mama, esta clasificación ha demostrado tener significación pronóstica (1, 6, 9, 10).

Los tumores de mama del subtipo "luminal" son positivos para ER y tienen un perfil de expresión de queratina similar a las células epiteliales que revisten la luz de los conductos de la mama (Tailor-Papadimitriou *et al.* (1989) *J Cell Sci* 94:403-413; Perou *et al.* (2000) *New Technologies for Life Sciences: A Trends Guide* 67-76, cada uno de los cuales se incorpora en el presente documento como referencia en su totalidad). A la inversa, tumores negativos para ER pueden dividirse en dos subtipos principales, concretamente los que sobreexpresan (y están amplificados para ADN) HER-2 y GRB7 (enriquecidos en HER-2) y tumores "de tipo basal" que tiene un perfil de expresión similar al epitelio basal y expresan queratina 5, 6B y 17. Estos dos subtipos de tumor son agresivos y normalmente más mortales que los tumores luminales; sin embargo, existen subtipos de tumores luminales con diferentes desenlaces. Los tumores luminales con malos desenlaces comparten constantemente la característica histopatológica de ser de un grado más alto y la característica molecular de expresar altamente genes de proliferación.

La traducción de los subtipos intrínsecos en un ensayo clínico ha sido ardua debido a que un agrupamiento no supervisado es más adecuado para organizar grandes números de muestras y genes que clasificar muestras individuales usando conjuntos génicos pequeños.

Por tanto, en el presente documento se proporcionan métodos y composiciones mejorados para clasificar subtipos intrínsecos de cáncer de mama. Los métodos utilizan un algoritmo supervisado para clasificar muestras del sujeto según el subtipo intrínseco de cáncer de mama. Este algoritmo, denominado en el presente documento el modelo de clasificación PAM50, se basa en el perfil de expresión génica de un subconjunto definido de genes intrínsecos que se han identificado en el presente documento como superior para clasificar subtipos intrínsecos de cáncer de mama, y para predecir el riesgo de recidiva y/o respuesta a la terapia en un sujeto diagnosticado con cáncer de mama. El subconjunto de genes, junto con cebadores específicos para su detección, se proporcionan en la tabla 1.

En algunas realizaciones, se usan al menos aproximadamente 40 de los genes enumerados en la tabla 1 en el modelo de clasificación PAM50. En otras realizaciones, se usan al menos 41, al menos 43, al menos 44, al menos 45, al menos 46, al menos 47, al menos 48, al menos 49 o los 50 genes intrínsecos enumerados en la tabla 1 en el modelo. Los métodos dados a conocer en el presente documento no se prevén para su uso con uno o sólo algunos de los genes enumerados en la tabla 1. De hecho, es la combinación de sustancialmente todos los genes intrínsecos lo que permite la clasificación del subtipo intrínseco y pronóstico del desenlace o respuesta terapéutica al tratamiento más precisos. Por tanto, en diversas realizaciones, los métodos dados a conocer en el presente documento abarcan la obtención del perfil genético de sustancialmente todos los genes enumerados en la tabla 1. "Sustancialmente todos" puede abarcar al menos 47, al menos 48, al menos 49 o los 50 genes enumerados en la tabla 1. A menos que se especifique lo contrario, "sustancialmente todos" se refiere a al menos 49 de los genes enumerados en la tabla 1. Un experto en la técnica también entenderá que puede usarse un subconjunto de los genes enumerados en la tabla 1 para entrenar un algoritmo para predecir el subtipo de cáncer de mama o el desenlace, y que puede usarse otro subconjunto de los genes para caracterizar un sujeto individual. Preferiblemente, se usan los 50 genes para entrenar el algoritmo y se usan al menos 49 de los genes para caracterizar un sujeto.

Tabla 1. Lista de genes intrínsecos de PAM50

GEN	NÚMERO DE REGISTRO DE GENBANK REPRESENTATIVO	CEBADOR DIRECTO	SEQ ID NO:	CEBADOR INVERSO	SEQ ID NO:
ACTR3B	NM_020445 NM_001040135	AAAGATTCCTGGG ACCTGA	1	TGGGGCAGTTCTGTA TTACTTC	51
ANLN	NM_018685	ACAGCCACTTTCA GAAGCAAG	2	CGATGGTTTTGTACA AGATTTCTC	52
BAG1	NM_004323	CTGGAAGAGTTGA ATAAAGAGC	3	GCAAATCCTTGGGCA GA	53
BCL2	NM_000633	TACCTGAACCGGC ACCTG	4	GCCGTACAGTTCCAC AAAGG	54
BIRC5	NM_001012271	GCACAAAGCCATT CTAAGTC	5	GACGCTTCCTATCAC TCTATTC	55
BLVRA	BX647539	GCTGGCTGAGCAG AAAG	6	TTCCTCCATCAAGAG TTCAACA	56
CCNB1	NM_031966	CTTTCGCCTGAGCC TATTT	7	GGGCACATCCAGAT GTTT	57
CCNE1	BC035498	GGCCAAAATCGAC AGGAC	8	GGGTCTGCACAGACT GCAT	58
CDC20	BG256659	CTGTCTGAGTGCC GTGGAT	9	TCCTTGTAATGGGGA GACCA	59
CDC6	NM_001254	GTAATCACCTTCT GAGCCT	10	ACTTGGGATATGTGA ATAAGACC	60
CDCA1	NM_031423	GGAGGCGGAAGAA ACCAG	11	GGGGAAAGACAAAG TTTCCA	61
CDH3	BC041846	GACAAGGAGAATC AAAAGATCAGC	12	ACTGTCTGGGTCCAT GGCTA	62
CENPF	NM_016343	GTGGCAGCAGATC ACAA	13	GGATTTCTGTGGTGGG TTC	63
CEP55	AB091343	CCTCACGAATTGCT GAACCT	14	CCACAGTCTGTGATA AACGG	64
CXXC5	BC006428	CATGAAATAGTGC ATAGTTTGCC	15	CCATCAACATTCTCT TTATGAACG	65
EGFR	NM_005228	ACACAGAATCTAT ACCCACCAGAGT	16	ATCAACTCCCAAACG GTCAC	66
ERBB2	NM_001005862	GCTGGCTCTCACA CTGATAG	17	GCCCTTACACATCGG AGAAC	67
ESR1	NM_001122742	GCAGGGAGAGGAG TTTGT	18	GACTTCAGGGTGCTG GAC	68
EXO1	NM_130398	CCCATCCATGTGA GGAAGTATAA	19	TGTGAAGCCAGCAA TATGTATC	69
FGFR4	AB209631	CTTCTTGGACCTTG GCG	20	TATTGGGAGGCAGG AGGTTTA	70
FOXA1	NM_004496	GCTACTACGCAGA CACG	21	CTGAGTTCATGTTGC TGACC	71

FOXC1	NM_001453	GATGTTTCGAGTCA CAGAGG	22	GACAGCTACTATTCC CGTT	72
GPR160	AJ249248	TTCGGCTGGAAGG AACC	23	TATGTGAGTAAGCTC GGAGAC	73
GRB7	NM_005310	CGTGGCAGATGTG AACGA	24	AGTGGGCATCCCGTA GA	74
HSPC150 (UBE2T)	NM_014176	GGAGATCCGTCAA CTCCAAA	25	AGTGGACATGCGAG TGGAG	75
KIF2C	NM_006845	TGGGTCGTGTCAG GAAAC	26	CACCGCTGGAAACT GAAC	76
KNTC2	NM_006101	CGCAGTCATCCAG AGATGTG	27	CGTGCACATCCATGA CCTT	77
KRT14	BC042437	ACTCAGTACAAGA AAGAACCG	28	GAGGAGATGACCTT GCC	78
KRT17	AK095281	GTTGGACCAGTCA ACATCTCTG	29	GCCATAGCCACTGCC ACT	79
KRT5	M21389	TGTGGCTCATTAG GCAAC	30	CTTCGACTGGACTCT GT	80
MAPT	NM_001123066	GACTCCAAGCGCG AAAAC	31	CAGACATGTTGGTAT TGCACATT	81
MDM2	M92424	CCAACAAAATATT CATGGTTCTTG	32	AGGCGATCCTGGGA AATTAT	82
MELK	NM_014791	CCAGTAGCATTGT CCGAG	33	CCCATTTGTCTGTCT TCAC	83
MIA	BG765502	GTCTCTGGTAATGC ACACT	34	CTGATGGTTGAGGCT GTT	84
MKI67	NM_002417	GTGGAATGCCTGC TGACC	35	CGCACTCCAGCACCT AGAC	85
MLPH	NM_024101	AGGGGTGCCCTCT GAGAT	36	TCACAGGGTCAAAC TCCAGT	86
MMP11	NM_005940	CGAGATCGCCAAG ATGTT	37	GATGGTAGAGTTCCA GTGATT	87
MYBL2	BX647151	AGGCGAACACACA ACGTC	38	TCTGGTCACGCAGGG CAA	88
MYC	NM_002467	AGCCTCGAACAAT TGAAGA	39	ACACAGATGATGGA GATGTC	89
NAT1	BC013732	ATCGACTGTGTAA ACAACTAGAGAAG A	40	AGTAGCTACATCTCC AGGTTCTCTG	90
ORC6L	NM_014321	TTTAAGAGGGCAA TGGAAGG	41	CGGATTTTATCAACG ATGCAG	91
PGR	NM_000926	TGCCGCAGAACTC ACTTG	42	CATTTGCCGTCCTTC ATCG	92
PHGDH	AK093306	CCTCAGATGATGC CTATCCA	43	GCAGGTCAAAACTCT CAAAG	93
PTTG1	BE904476	CAGCAAGCGATGG CATAGT	44	AGCGGGCTTCTGTAA TCTGA	94

RRM2	AK123010	AATGCCACCGAAG CCTC	45	GCCTCAGATTTCAAC TCGT	95
SFRP1	BC036503	TCGAACTGAAGGC TATTTACGAG	46	CTGCTGAGAATCAA AGTGGGA	96
SLC39A6	NM_012319	GTCGAAGCCGCAA TTAGG	47	GGAACAAACTGCTCT GCCA	97
TMEM45 B	AK098106	CAAACGTGTGTTCT GGAAGG	48	ACAGCTCTTTAGCAT TTGTGGA	98
TYMS	BQ056428	TGCCCTGTATGATG TCAGGA	49	GGGACTATCAATGTT GGGTTCTC	99
UBE2C	BC032677	GTGAGGGGTGTCA GCTCAGT	50	CACACAGTTCAGTGC TCCACA	100

“Expresión génica” tal como se usa en el presente documento se refiere a los niveles de expresión y/o al patrón de expresión relativos de un gen. La expresión de un gen puede medirse a nivel de ADN, ADNc, ARN, ARNm, o combinaciones de los mismos. “Perfil de expresión génica” se refiere a los niveles de expresión de múltiples genes diferentes medidos para la misma muestra. Un perfil de expresión puede derivarse de una muestra biológica recogida de un sujeto en uno o más puntos de tiempo antes de, durante, o tras el diagnóstico, el tratamiento o la terapia contra el cáncer de mama (o cualquier combinación de los mismos), puede derivarse de una muestra biológica recogida de un sujeto en uno o más puntos de tiempo durante los que no hay tratamiento o terapia contra el cáncer de mama (por ejemplo, para monitorizar la progresión de la enfermedad o para evaluar el desarrollo de la enfermedad en un sujeto en riesgo de cáncer de mama), o puede recogerse de un sujeto sano. Pueden medirse perfiles de expresión génica en una muestra, tales como muestras que comprenden una variedad de tipos celulares, tejidos diferentes, órganos diferentes o fluidos (por ejemplo, sangre, orina, líquido cefalorraquídeo, sudor, saliva o suero) mediante diversos métodos que incluyen pero no se limitan a tecnologías de micromatriz y técnicas de RT-PCR cuantitativa y semicuantitativa.

Variables clínicas

El modelo de clasificación PAM50 descrito en el presente documento puede combinarse adicionalmente con información sobre variables clínicas para generar un factor pronóstico de riesgo de recidiva (ROR) continuo. Tal como se describe en el presente documento, en la técnica se describen varios de factores de cáncer de mama clínicos y pronósticos y se usan para predecir el desenlace del tratamiento y la probabilidad de reaparición de la enfermedad. Tales factores incluyen, por ejemplo, implicación de ganglios linfáticos, tamaño del tumor, grado histológico, estado del receptor de la hormona estrógeno y progesterona, niveles de HER-2 y ploidía tumoral.

En una realización, se proporciona una puntuación de riesgo de recidiva (ROR) para un sujeto diagnosticado con o que se sospecha que tiene cáncer de mama. Esta puntuación usa el modelo de clasificación PAM50 en combinación con factores clínicos de estado de los ganglios linfáticos (N) y tamaño del tumor (T). La evaluación de variables clínicas se basa en el sistema normalizado del American Joint Committee on Cancer (AJCC) para estadificar el cáncer de mama. En este sistema, se clasifica el tamaño del tumor primario en una escala de 0-4 (T0: sin evidencia de tumor primario; T1: ≤ 2 cm; T2: > 2 cm - ≤ 5 cm; T3: > 5 cm; T4: tumor de cualquier tamaño con extensión directa a la piel o la pared torácica). El estado de los ganglios linfáticos se clasifica como N0-N3 (N0: los ganglios linfáticos regionales están libres de metástasis; N1: metástasis no móvil, ganglio(s) linfático(s) axilar(es) del mismo lado; N2: metástasis al/a los ganglio(s) linfático(s) del mismo lado fijado(s) entre sí o a otras estructuras; N3: metástasis a ganglios linfáticos del mismo lado por debajo del esternón). Los métodos de identificación de pacientes con cáncer de mama y estadificación del estadio de la enfermedad se conocen bien y pueden incluir examen manual, biopsia, revisión la historia de la paciente y/o la familia, y técnicas de obtención de imágenes, tales como mamografía, obtención de imágenes por resonancia magnética (MRI) y tomografía de emisión de positrones (PET).

Usando los métodos de clasificación PAM50 de la presente invención, puede determinarse el pronóstico de una paciente con cáncer de mama independientemente de o en combinación con la evaluación de estos factores clínicos. En algunas realizaciones, combinar los métodos de clasificación del subtipo intrínseco de cáncer de mama PAM50 dados a conocer en el presente documento con evaluación de estos factores clínicos puede permitir una evaluación del riesgo más precisa. Los métodos de la invención pueden acoplarse adicionalmente con análisis de, por ejemplo, el estado de receptor de estrógeno (ER) y receptor de progesterona (PgR), y/o niveles de expresión de HER-2. Otros factores, tales como la historia clínica de la paciente, la historia familiar y el estado menopáusico, también pueden considerarse cuando se evalúa el pronóstico de cáncer de mama por medio de los métodos de la invención.

Fuente de la muestra

En una realización de la presente invención, se evalúa el subtipo de cáncer de mama a través de la evaluación de patrones, o perfiles, de expresión, de los genes intrínsecos enumerados en la tabla 1 en una o más muestras del sujeto. Para fines de descripción, el término sujeto, o muestra de sujeto, se refiere a un individuo independientemente de su estado de salud y/o de enfermedad. Un sujeto puede ser un sujeto, un participante de estudio, un sujeto control, un sujeto de examen, o cualquier otra clase de individuo del que se obtiene una muestra y se evalúa en el contexto de la invención. Por consiguiente, un sujeto puede estar diagnosticado con cáncer de mama, puede presentar uno o más síntomas de cáncer de mama, o un factor de predisposición, tal como un factor de historia familiar (genético) o médica (médico), para cáncer de mama, puede estar sometiéndose a tratamiento o terapia contra el cáncer de mama, o similares. Alternativamente, un sujeto puede estar sano con respecto a cualquiera de los factores o criterios mencionados anteriormente. Se apreciará que el término “sano” tal como se usa en el presente documento, se refiere a estado de cáncer de mama, ya que no puede definirse que el término “sano” para corresponder a cualquier evaluación o estado absoluto. Por tanto, un individuo que se define como sano con referencia a cualquier enfermedad o criterio de enfermedad especificados, puede estar diagnosticado de hecho con otra cualquiera o más enfermedades, o presentar cualquier otro o más criterios de enfermedad, incluyendo uno o más cánceres distintos de cáncer de mama. Sin embargo, los controles sanos están libres preferiblemente de cualquier cáncer.

En realizaciones particulares, los métodos para predecir subtipos intrínsecos de cáncer de mama incluyen recoger una muestra biológica que comprende una célula o tejido canceroso, tal como una muestra de tejido de mama o una muestra de tejido de tumor de mama primario. Mediante “muestra biológica” se pretende decir cualquier toma de muestras de células, tejidos o fluidos corporales en los que puede detectarse la expresión de un gen intrínseco. Los ejemplos de tales muestras biológicas incluyen, pero no se limitan a, biopsias y frotis. Los fluidos corporales útiles en la presente invención incluyen sangre, linfa, orina, saliva, aspirados de pezones, fluidos ginecológicos o cualquier otra secreción corporal o derivado de la misma. La sangre puede incluir sangre completa, plasma, suero o cualquier hemoderivado. En algunas realizaciones, la muestra biológica incluye células de mama, particularmente tejido de mama de una biopsia, tal como una muestra de tejido de tumor de mama. Las muestras biológicas pueden obtenerse de un sujeto mediante una variedad de técnicas que incluyen, por ejemplo, raspando o frotando un área, usando una aguja para aspirar células o fluidos corporales, o extrayendo una muestra de tejido (es decir, biopsia). En la técnica se conocen bien los métodos para recoger diversas muestras biológicas. En algunas realizaciones, una muestra de tejido de mama se obtiene mediante, por ejemplo, biopsia por aspiración con aguja fina, biopsia con aguja gruesa o biopsia por escisión. Pueden aplicarse disoluciones de fijación o de tinción a las células o tejidos para conservar la muestra y para facilitar el examen. Las muestras biológicas, particularmente muestras de tejido de mama, pueden transferirse a un portaobjetos de vidrio para observación con aumentos. En una realización, la muestra biológica es una muestra de tejido de mama incrustada en parafina, fijada en formalina, particularmente una muestra de tumor de mama primario. En diversas realizaciones, la muestra de tejido se obtiene de una muestra de núcleo de tejido guiado por patólogo tal como se describe en el ejemplo 4.

Obtención del perfil de expresión

En diversas realizaciones, la presente invención proporciona métodos para clasificar, pronosticar o monitorizar cáncer de mama en sujetos. En esta realización, los datos obtenidos del análisis de la expresión de genes intrínsecos se evalúan usando uno o más algoritmos de reconocimiento de patrones. Tales métodos de análisis pueden usarse para formar un modelo predictivo, que puede usarse para clasificar datos de prueba. Por ejemplo, un método de clasificación conveniente y particularmente eficaz emplea la creación de modelos de análisis estadísticos de múltiples variables, en primer lugar para formar un modelo (un “modelo matemático predictivo”) usando datos (“datos de creación de modelo”) de muestras de un subtipo conocido (por ejemplo, de sujetos que se sabe que tienen un subtipo intrínseco de cáncer de mama particular: LumA, LumB, de tipo basal, enriquecido en HER-2, o de tipo normal), y en segundo lugar para clasificar una muestra desconocida (por ejemplo, “muestra de prueba”) según el subtipo.

Se han usado ampliamente métodos de reconocimiento de patrones para caracterizar muchos tipos diferentes de problemas que varían, por ejemplo, entre lingüística, toma de huellas dactilares, química y psicología. En el contexto de los métodos descritos en el presente documento, el reconocimiento de patrones es el uso de datos estadísticos de múltiples variables, tanto paramétricos como no paramétricos, para analizar datos, y por tanto para clasificar muestras y para predecir el valor de alguna variable dependiente basándose en una variedad de mediciones observadas. Existen dos enfoques principales. Un conjunto de métodos se denomina “no supervisado” y éstos simplemente reducen la complejidad de los datos de una manera racional y también producen diagramas de representación que pueden interpretarse por el ojo humano. Sin embargo, este tipo de enfoque puede no ser adecuado para desarrollar un ensayo clínico que puede usarse para clasificar muestras derivadas de sujetos independientes de la población de muestra inicial usada para entrenar el algoritmo de predicción.

El otro enfoque se denomina “supervisado” mediante lo cual se usa un conjunto entrenado de muestras con clase o desenlace conocidos para producir un modelo matemático que entonces se evalúa con conjuntos de datos de validación independientes. En este caso, se usa un “conjunto de entrenamiento” de datos de expresión de genes

intrínsecos para construir un modelo estadístico que predice correctamente el “subtipo” de cada muestra. Este conjunto entrenado se somete entonces a prueba con datos independientes (denominados conjunto de validación o prueba) para determinar la robustez del modelo basado en ordenador. Estos modelos se denominan en ocasiones “sistemas expertos”, pero pueden basarse en una variedad de diferentes procedimientos matemáticos. Los métodos supervisados pueden usar un conjunto de datos con dimensionalidad reducida (por ejemplo, los primeros pocos componentes principales), pero normalmente usan datos no reducidos, con toda la dimensionalidad. En todos los casos, los métodos permiten la descripción cuantitativa de los límites de múltiples variables que caracterizan y separan cada subtipo en lo que se refiere a su perfil de expresión de genes intrínsecos. También es posible obtener límites de confianza sobre cualquier predicción, por ejemplo, a nivel de probabilidad para ubicarse en la bondad de ajuste (véase, por ejemplo, Kowalski *et al.*, 1986). La robustez de los modelos predictivos también puede comprobarse usando validación cruzada, dejando fuera muestras seleccionadas del análisis.

El modelo de clasificación PAM50 descrito en el presente documento se basa en el perfil de expresión génica para una pluralidad de muestras del sujeto usando los genes intrínsecos enumerados en la tabla 1. La pluralidad de muestras incluye un número suficiente de muestras derivadas de sujetos que pertenecen a cada clase de subtipo. Por “muestras suficientes” o “número representativo” en este contexto se pretende decir una cantidad de muestras derivadas de cada subtipo que es suficiente para construir un modelo de clasificación que puede distinguir de manera fiable cada subtipo de todos los demás en el grupo. Un algoritmo de predicción supervisado se desarrolla basándose en los perfiles de muestras prototipo seleccionadas objetivamente para “entrenar” el algoritmo. Las muestras se seleccionan y se clasifican en subtipos usando un conjunto de genes intrínsecos expandido según los métodos dados a conocer en la publicación de patente internacional WO 2007/061876, que se incorpora en el presente documento como referencia en su totalidad. Alternativamente, las muestras pueden clasificarse en subtipos según cualquier ensayo conocido para clasificar subtipos de cáncer de mama. Tras estratificar las muestras de entrenamiento según el subtipo, se usa un algoritmo de predicción basado en centroides para construir centroides basándose en el perfil de expresión del conjunto de genes intrínsecos descritos en la tabla 1.

En una realización, el algoritmo de predicción es la metodología del centroide más cercano relacionada con la descrita en Narashiman y Chu (2002) PNAS 99:6567-6572, que se incorpora en el presente documento como referencia en su totalidad. En la presente invención, el método calcula un centroide normalizado para cada subtipo. Este centroide es la expresión génica promedio para cada gen en cada subtipo (o “clase”) dividida entre la desviación estándar dentro de la clase para ese gen. La clasificación del centroide más cercano toma el perfil de expresión génica de una nueva muestra, y la compara con cada uno de estos centroides de clase. La predicción de subtipo se realiza calculando la correlación de rangos de Spearman de cada caso de prueba para los cinco centroides, y asignando una muestra a un subtipo basándose en el centroide más cercano.

Detección de expresión de genes intrínsecos

En el presente documento se abarca cualquier método disponible en la técnica para detectar la expresión de los genes intrínsecos enumerados en la tabla 1. Por “detectar la expresión” se pretende determinar la cantidad o presencia de un transcrito de ARN o su producto de expresión de un gen intrínseco.

Los métodos para detectar la expresión de los genes intrínsecos de la invención, es decir, obtención de perfiles de expresión génica, incluyen métodos basados en análisis de hibridación de polinucleótidos, métodos basados en la secuenciación de polinucleótidos, métodos inmunohistoquímicos y métodos basados en análisis proteómicos. Los métodos detectan generalmente productos de expresión (por ejemplo, ARNm) de los genes intrínsecos enumerados en la tabla 1. En realizaciones preferidas, se usan métodos basados en PCR, tales como PCR de transcripción inversa (RT-PCR) (Weis *et al.*, TIG 8:263-64, 1992), y métodos basados en matrices tales como micromatriz (Schena *et al.*, Science 270:467-70, 1995). Por “micromatriz” se pretende decir una disposición ordenada de elementos de matriz que pueden hibridarse, tales como, por ejemplo, sondas de polinucleótido, sobre un sustrato. El término “sonda” se refiere a cualquier molécula que puede unirse selectivamente a una biomolécula diana prevista específicamente, por ejemplo, un transcrito de nucleótido o una proteína codificada por o que corresponde a un gen intrínseco. Un experto en la técnica puede sintetizar sondas, o pueden derivarse de preparaciones biológicas apropiadas. Pueden diseñarse las sondas específicamente para marcarse. Los ejemplos de moléculas que pueden utilizarse como sondas incluyen, pero no se limitan a, ARN, ADN, proteínas, anticuerpos y moléculas orgánicas.

Muchos métodos de detección de la expresión usan ARN aislado. El material de partida es normalmente ARN total aislado de una muestra biológica, tal como un tumor o línea celular de tumor, y tejido o línea celular normales correspondientes, respectivamente. Si la fuente de ARN es un tumor primario, puede extraerse ARN (por ejemplo, ARNm), por ejemplo, de muestras de tejido congeladas o archivadas incrustadas en parafina y fijadas (por ejemplo, fijadas en formalina) (por ejemplo, muestras de núcleo de tejido guiado por patólogo).

En la técnica se conocen bien métodos generales para la extracción de ARN y se dan a conocer en libros de texto convencionales de biología molecular, que incluyen Ausubel *et al.*, ed., Current Protocols in Molecular Biology, John Wiley & Sons, Nueva York 1987-1999. Se dan a conocer métodos para la extracción de ARN a partir de tejidos incrustados en parafina, por ejemplo, en Rupp y Locker (Lab Invest. 56:A67, 1987) y De Andres *et al.* (Biotechniques 18:42-44, 1995). En particular, el aislamiento de ARN puede realizarse usando un kit de purificación, un conjunto de

tampones y proteasa de fabricantes comerciales, tales como Qiagen (Valencia, CA), según las instrucciones del fabricante. Por ejemplo, puede aislarse ARN total de células en cultivo usando mini-columnas RNeasy de Qiagen. Otros kits de aislamiento de ARN disponibles comercialmente incluyen kit de purificación de ADN y ARN completo MASTERPURE™ (Epicentre, Madison, Wis.) y kit de aislamiento de ARN en bloque de parafina (Ambion, Austin, TX). Puede aislarse el ARN total de muestras de tejido, por ejemplo, usando ARN Stat-60 (Tel-Test, Friendswood, TX). Puede aislarse ARN preparado a partir de un tumor, por ejemplo, mediante centrifugación en gradiente de densidad con cloruro de cesio. Adicionalmente, pueden procesarse fácilmente grandes números de muestras de tejido usando técnicas bien conocidas por los expertos en la técnica, tales como, por ejemplo, el procedimiento de aislamiento de ARN de etapa individual de Chomczynski (patente estadounidense n.º 4.843.155).

Puede usarse ARN aislado en ensayos de hibridación o amplificación que incluyen, pero no se limitan a, análisis de PCR y matrices de sonda. Un método para la detección de niveles de ARN implica poner en contacto el ARN aislado con una molécula de ácido nucleico (sonda) que puede hibridarse con el ARNm codificado por el gen que está detectándose. La sonda de ácido nucleico puede ser, por ejemplo, un ADNc de longitud completa, o una parte del mismo, tal como un oligonucleótido de al menos 7, 15, 30, 60, 100, 250 ó 500 nucleótidos de longitud y suficiente para hibridarse específicamente en condiciones rigurosas con un gen intrínseco de la presente invención, o cualquier ADN o ARN derivado. La hibridación de un ARNm con la sonda indica que el gen intrínseco en cuestión se está expresando.

En una realización, el ARNm se inmoviliza sobre una superficie sólida y se pone en contacto con una sonda, por ejemplo haciendo correr el ARNm aislado en un gel de agarosa y transfiriendo el ARNm del gel a una membrana, tal como nitrocelulosa. En una realización alternativa, las sondas se inmovilizan sobre una superficie sólida y el ARNm se pone en contacto con las sondas, por ejemplo, en una matriz de chip de genes de Agilent. Un experto en la técnica puede adaptar fácilmente métodos de detección de ARNm conocidos para su uso en la detección del nivel de expresión de los genes intrínsecos de la presente invención.

Un método alternativo para determinar el nivel de producto de la expresión de genes intrínsecos en una muestra implica el procedimiento de amplificación de ácido nucleico, por ejemplo, mediante RT-PCR (patente estadounidense n.º 4.683.202), reacción en cadena de la ligasa (Barany, Proc. Natl. Acad. Sci. USA 88:189-93, 1991), replicación de secuencia autosostenida (Guatelli *et al.*, Proc. Natl. Acad. Sci. USA 87:1874-78, 1990), sistema de amplificación transcripcional (Kwoh *et al.*, Proc. Natl. Acad. Sci. USA 86:1173-77, 1989), replicasa Q-Beta (Lizardi *et al.*, Bio/Technology 6:1197, 1988), replicación en círculo rodante (patente estadounidense n.º 5.854.033), o cualquier otro método de amplificación de ácido nucleico, seguido por la detección de las moléculas amplificadas usando técnicas bien conocidas por los expertos en la técnica. Estos esquemas de detección son especialmente útiles para la detección de moléculas de ácido nucleico si tales moléculas están presentes en números muy bajos.

En aspectos particulares de la invención, la expresión de genes intrínsecos se evalúa mediante RT-PCR cuantitativa. En la técnica se conocen numerosos protocolos de PCR o QPCR diferentes y se muestran a modo de ejemplo a continuación en el presente documento y pueden aplicarse directamente o adaptarse para su uso usando las composiciones descritas actualmente para la detección y/o cuantificación de los genes intrínsecos enumerados en la tabla 1. Generalmente, en la PCR, se amplifica una secuencia de polinucleótido diana mediante la reacción con al menos un cebador oligonucleotídico o par de cebadores oligonucleotídicos. El/los cebador(es) se hibridan en una región complementaria del ácido nucleico diana y una ADN polimerasa extiende el/los cebador(es) para amplificar la secuencia diana. En condiciones suficientes para proporcionar productos de la amplificación de ácido nucleico basados en polimerasa, un fragmento de ácido nucleico de un tamaño domina los productos de reacción (la secuencia de polinucleótido diana que es el producto de amplificación). Se repite el ciclo de amplificación para aumentar la concentración de la secuencia de polinucleótido diana individual. La reacción puede realizarse en cualquier termociclador usado comúnmente para PCR. Sin embargo, se prefieren cicladores con capacidades de medición de fluorescencia en tiempo real, por ejemplo, SMARTCYCLER® (Cepheid, Sunnyvale, CA), ABI PRISM 7700® (Applied Biosystems, Foster City, Calif.), ROTOR-GENE™ (Corbett Research, Sydney, Australia), LIGHTCYCLER® (Roche Diagnostics Corp, Indianapolis, Ind.), ICYCLER® (Biorad Laboratories, Hercules, Calif.) y MX4000® (Stratagene, La Jolla, Calif.).

Se prefiere la PCR cuantitativa (QPCR) (también denominada PCR en tiempo real) en algunas circunstancias ya que no sólo proporciona una medición cuantitativa, sino que también tiempo y contaminación reducidos. En algunos casos, la disponibilidad de técnicas de obtención de perfiles de expresión génica completa está limitada debido a requisitos de tejido recién congelado y equipo de laboratorio especializado, haciendo que el uso de rutina de tales tecnologías sea difícil en la práctica clínica. Sin embargo, la medición de genes de QPCR puede aplicarse a bloques de tumores clínicos incrustados en parafina, fijados en formalina convencionales, tales como los usados en bancos de tejido de archivo y muestras de patología quirúrgica de rutina (Cronin *et al.* (2007) Clin Chem 53:1084-91) [Mullins 2007][Paik 2004]. Tal como se usa en el presente documento, "PCR cuantitativa (o "QPCR en tiempo real") se refiere a la monitorización directa del progreso de amplificación de PCR mientras se produce sin necesidad de tomar muestras repetidamente de los productos de reacción. En la PCR cuantitativa, los productos de reacción pueden monitorizarse por medio de un mecanismo de señalización (por ejemplo, fluorescencia) mientras se generan y se controlan después de que la señal ascienda por encima de un nivel de fondo pero antes de que la reacción alcance una meseta. El número de ciclos requerido para lograr un nivel detectable o "umbral" de fluorescencia varía

directamente con la concentración de dianas amplificables al principio del procedimiento de PCR, permitiendo una medición de la intensidad de señal para proporcionar una medida de la cantidad de ácido nucleico diana en una muestra en tiempo real.

En otra realización de la invención, se usan micromatrices para la obtención del perfil de expresión. Las micromatrices son particularmente adecuadas para este fin debido a la reproducibilidad entre diferentes experimentos. Las micromatrices de ADN proporcionan un método para la medición simultánea de los niveles de expresión de grandes números de genes. Cada matriz consiste en un patrón reproducible de sondas de captura unidas a un soporte sólido. El ARN o ADN marcado se hibrida con las sondas complementarias en la matriz y entonces se detectan mediante exploración por láser. Se determinan las intensidades de hibridación para cada sonda en la matriz y se convierten en un valor cuantitativo que representa niveles de expresión génica relativos. Véanse, por ejemplo, las patentes estadounidenses n. ^{os} 6.040.138, 5.800.992 y 6.020.135, 6.033.860 y 6.344.316. Las matrices de oligonucleótido de alta densidad son particularmente útiles para determinar el perfil de expresión génica para un gran número de ARN en una muestra.

Técnicas para la síntesis de estas matrices usando métodos de síntesis mecánica se describen en, por ejemplo, la patente estadounidense n. ° 5.384.261. Aunque generalmente se usa una superficie de matriz plana, la matriz puede fabricarse sobre una superficie de prácticamente cualquier forma o incluso una multiplicidad de superficies. Las matrices pueden ser ácidos nucleicos (o péptidos) sobre perlas, geles, superficies poliméricas, fibras (tal como fibra óptica), vidrio, o cualquier otro sustrato apropiado. Véanse, por ejemplo, las patentes estadounidenses n. ^{os} 5.770.358, 5.789.162, 5.708.153, 6.040.193 y 5.800.992. Las matrices pueden envasarse de tal manera que se permita el diagnóstico u otra manipulación de un dispositivo todo incluido. Véanse, por ejemplo, las patentes estadounidenses n. ^{os} 5.856.174 y 5.922.591.

En una realización específica de la técnica de micromatriz, se aplican insertos amplificados por PCR de clones de ADNc a un sustrato en una matriz densa. Los genes sometidos a micromatriz, inmovilizados en el microchip, son adecuados para la hibridación en condiciones rigurosas. Pueden generarse sondas de ADNc marcadas fluorescentemente a través de la incorporación de nucleótidos fluorescentes mediante transcripción inversa de ARN extraído de tejidos de interés. Las sondas de ADNc marcadas aplicadas al chip se hibridan con especificidad a cada punto de ADN sobre la matriz. Tras lavado riguroso para eliminar sondas no unidas específicamente, se explora el chip mediante microscopía láser confocal o mediante otro método de detección, tal como una cámara de CCD. La cuantificación de hibridación de cada elemento sometido a matriz permite la evaluación de la abundancia de ARNm correspondiente.

Con fluorescencia de color doble, se hibridan por parejas a la matriz sondas de ADNc marcadas por separado generadas a partir de dos fuentes de ARN. Por tanto, la abundancia relativa de los transcritos de las dos fuentes que se corresponden con cada gen especificado se determina simultáneamente. La escala miniaturizada de la hibridación proporciona una evaluación conveniente y rápida del patrón de expresión para grandes números de genes. Se ha mostrado que tales métodos tienen la sensibilidad requerida para detectar transcritos raros, que se expresan en algunas copias por célula, y para detectar reproduciblemente al menos aproximadamente diferencias de dos veces en los niveles de expresión (Schena *et al.*, Proc. Natl. Acad. Sci. USA 93:106-49, 1996). Los análisis de micromatriz pueden realizarse mediante equipo disponible comercialmente, siguiendo los protocolos del fabricante, tales como usar la tecnología GenChip de Affymetrix o tecnología de micromatriz de inyección de tinta de Agilent. El desarrollo de métodos de micromatriz para análisis a gran escala de la expresión génica hace posible buscar de manera sistemática marcadores moleculares de clasificación del cáncer y predicción de desenlace en una variedad de tipos de tumor.

Procesamiento de datos

A menudo es útil procesar previamente los datos de expresión génica, por ejemplo, mediante tratamiento de datos que faltan, traducción, ajuste a escala, normalización, ponderación, etc. Los métodos de proyección de múltiples variables, tales como análisis del componente principal (PCA) y análisis de mínimos cuadrados parcial (PLS), son los denominados métodos sensibles de ajuste a escala. Usando el conocimiento y la experiencia anterior sobre el tipo de datos estudiados, la calidad de los datos antes de la obtención del modelo de múltiples variables puede potenciarse mediante el ajuste a escala y/o la ponderación. El ajuste a escala y/o la ponderación adecuados pueden revelar una variación importante e interesante oculta en los datos, y por tanto hacen que la obtención de modelo de múltiples variables posterior sea más eficaz. El ajuste a escala y la ponderación pueden usarse para colocar los datos en la métrica correcta, basándose en el conocimiento y la experiencia del sistema estudiado, y por tanto revelar patrones ya presentes de manera inherente en los datos.

Si es posible, los datos que faltan, por ejemplo huecos en valores de columna, deberían evitarse. Sin embargo, si es necesario, tales datos que faltan pueden sustituirse o "imputarse" con, por ejemplo, el valor medio de una columna ("imputación media"); un valor aleatorio ("imputación aleatoria"); o un valor basado en un análisis de componente principal ("imputación de componente principal").

La "traducción" de los ejes de coordenadas del descriptor puede ser útil. Los ejemplos de tales traducciones incluyen

normalización y centrado en la media. La “normalización” puede usarse para eliminar variaciones de muestra a muestra. Para datos de micromatriz, el procedimiento de normalización busca eliminar errores sistemáticos equilibrando las intensidades de fluorescencia de los dos colorantes de marcado. El sesgo del colorante puede proceder de diversas fuentes, incluyendo diferencias en la eficacia de marcado del colorante, sensibilidad al calor y a la luz, así como parámetros de exploración para explorar dos canales. Algunos métodos usados comúnmente para calcular el factor de normalización incluyen: (i) normalización global que usa todos los genes sobre la matriz; (ii) normalización de genes de mantenimiento que usa genes de mantenimiento/invariables que se expresan constantemente; y (iii) normalización de controles internos que usa una cantidad conocida de genes de control exógenos añadida durante la hibridación (Quackenbush (2002) Nat. Genet. 32 (Supl.), 496-501). En una realización, los genes intrínsecos dados a conocer en el presente documento pueden normalizarse para controlar genes de mantenimiento. Por ejemplo, los genes de mantenimiento descritos en la publicación de patente estadounidense 2008/0032293, que se incorpora en el presente documento como referencia en su totalidad, pueden usarse para la normalización. Los genes de mantenimiento a modo de ejemplo incluyen MRPL19, PSMC4, SF3A1, PUM1, ACTB, GAPD, GUSB, RPLP0 y TFRC. Un experto en la técnica entenderá que los métodos dados a conocer en el presente documento no están asociados mediante normalización a ningún gen de mantenimiento particular, y que puede usarse cualquier gen de mantenimiento adecuado conocido en la técnica.

Son posibles muchos enfoques de normalización, y a menudo pueden aplicarse en cualquiera de varios puntos en el análisis. En una realización, los datos de micromatriz se normalizan usando el método LOWESS, que es una función de normalización de suavización de la dispersión ponderada localmente global. En otra realización, los datos de qPCR se normalizan para dar la media geométrica del conjunto de múltiples genes de mantenimiento.

El “centrado en la media” también puede usarse para simplificar la interpretación. Habitualmente, para cada descriptor, se resta el valor promedio de ese descriptor para todas las muestras. De esta manera, la media de un descriptor coincide con el origen, y todos los descriptores se “centran” a cero. En “ajuste a escala de la varianza unitaria”, los datos ajustarse a escala hasta una varianza igual. Habitualmente, el valor de cada descriptor se ajusta a escala mediante $1/DE$, donde DE es la desviación estándar para ese descriptor para todas las muestras. El “ajuste a escala de Pareto” es, en algunos casos, intermedio entre el centrado en la media y el ajuste a escala de la varianza unitaria. En el ajuste a escala de Pareto, el valor de cada descriptor se ajusta a escala mediante $1/(DE)^2$, donde DE es la desviación estándar para ese descriptor para todas las muestras. De esta manera, cada descriptor tiene una varianza numéricamente igual a su desviación estándar inicial. El ajuste a escala de Pareto puede realizarse, por ejemplo, sobre los datos sin procesar o los datos centrados en la media.

El “ajuste a escala logarítmico” puede usarse para ayudar en la interpretación cuando los datos tienen un sesgo positivo y/o cuando los datos abarcan una gran amplitud, por ejemplo, varios órdenes de magnitud. Habitualmente, para cada descriptor, el valor se sustituye por el logaritmo de ese valor. En “ajuste a escala de amplitud igual”, cada descriptor se divide entre la amplitud de ese descriptor para todas las muestras. De esta manera, todos los descriptores tienen la misma amplitud, es decir, 1. Sin embargo, este método es sensible a la presencia de puntos de valores atípicos. En “ajuste a escala automática”, cada vector de datos se centra en la media y se mide a escala la varianza unitaria. Esta técnica es muy útil debido a que cada descriptor se pondera entonces igualmente, y los valores grandes y pequeños se tratan con el mismo énfasis. Esto puede ser importante para genes expresados a niveles muy bajos, pero todavía detectables.

En una realización, se recogen los datos para una o más muestras de prueba y se clasifican usando el modelo de clasificación PAM50 descrito en el presente documento. Cuando se comparan los datos de múltiples análisis (por ejemplo, comparación de los perfiles de expresión para una o más muestras de prueba con los centroides construidos a partir de las muestras recogidas y analizadas en un estudio independiente), será necesario normalizar los datos a lo largo de estos conjuntos de datos. En una realización, se usa discriminación ponderada por distancia (DWD) para combinar estos conjuntos de datos entre sí (Benito *et al.* (2004) Bioinformatics 20(1):105-114, incorporado como referencia en el presente documento en su totalidad). DWD es una herramienta de análisis de múltiples variables que puede identificar sesgos sistemáticos presentes en conjuntos de datos separados y después llevar a cabo un ajuste global para compensar estos sesgos; en esencia, cada conjunto de datos separado es una nube multidimensional de puntos de datos, y DWD toma dos nubes de puntos y cambia una de manera que solape más óptimamente a la otra.

Los métodos descritos en el presente documento pueden implementarse y/o los resultados registrarse usando cualquier dispositivo que pueda implementar los métodos y/o registrar los resultados. Los ejemplos de dispositivos que pueden usarse incluyen pero no se limitan a dispositivos computacionales electrónicos, incluyendo ordenadores de todos los tipos. Cuando los métodos descritos en el presente documento se implementan y/o registran en un ordenador, el programa informático que puede usarse para configurar el ordenador para llevar a cabo las etapas de los métodos puede contenerse en cualquier medio legible por ordenador que puede contener el programa informático. Los ejemplos de medio legible por ordenador que pueden usarse incluyen pero no se limitan a disquetes, CD-ROM, DVD, ROM, RAM, y otros dispositivos de almacenamiento informático y de memoria. El programa informático que puede usarse para configurar el ordenador para llevar a cabo las etapas de los métodos y/o registrar los resultados también puede proporcionarse por una red electrónica, por ejemplo, por internet, una intranet, u otra red.

Cálculo de riesgo de recidiva

En el presente documento se proporcionan métodos para predecir desenlace de cáncer de mama dentro del contexto del subtipo intrínseco y opcionalmente otras variables clínicas. Desenlace puede referirse a supervivencia específica de la enfermedad o global, supervivencia libre de acontecimiento o desenlace en respuesta a un tratamiento o una terapia particular. En particular, los métodos pueden usarse para predecir la probabilidad de supervivencia libre de enfermedad, a largo plazo. La "predicción de la probabilidad de supervivencia de una paciente con cáncer de mama" pretende evaluar el riesgo de que una paciente muera como resultado del cáncer de mama subyacente. "Supervivencia libre de enfermedad, a largo plazo" pretende significar que la paciente no muere de o padece una reaparición del cáncer de mama subyacente a lo largo de un periodo de al menos cinco años, o al menos diez o más años, tras el diagnóstico o tratamiento inicial.

En una realización, se predice el desenlace basándose en la clasificación de un sujeto según subtipo. Esta clasificación se basa en la obtención del perfil de expresión usando la lista de genes intrínsecos enumerados en la tabla 1. Tal como se trata en el ejemplo 1, el subtipo de tumor según el modelo PAM50 era más indicativo de respuesta a quimioterapia que la clasificación de marcador clínico convencional. Los tumores clasificados como HER2+ usando marcadores clínicos pero no enriquecidos en HER-2 usando el modelo PAM50 tenían una respuesta completa patológica menor (pCR) a un régimen de paclitaxel, 5-fluorouracilo, adriamicina y ciclofosfamida (T/FAC) que los tumores clasificados como HER2+ clínicamente y que pertenecen al subtipo de expresión enriquecido en HER-2. De manera similar, los tumores de tipo basal que no se puntuaron clínicamente como triple-negativos (ER-, PgR- y HER2-) tenían una pCR más alta en comparación con tumores triple-negativos que no eran de tipo basal mediante PAM50. Por tanto, el modelo PAM50 puede usarse para predecir con más precisión la respuesta a la quimioterapia que los marcadores clínicos convencionales.

Además de proporcionar una asignación de subtipo, el modelo bioinformático PAM50 proporciona una medición de la similitud de una muestra de prueba para los cuatro subtipos que se traduce en una puntuación de riesgo de recidiva (ROR) que puede usarse en cualquier población de pacientes independientemente del estado patológico y las opciones de tratamiento. Los subtipos intrínsecos y ROR también tiene valor en la predicción de respuesta completa patológica en mujeres tratadas con, por ejemplo, quimioterapia de taxano y antraciclina neoadyuvante [Rouzier 2005]. Por tanto, en diversas realizaciones de la presente invención, se usa un modelo de riesgo de recidiva (ROR) para predecir el desenlace. Usando estos modelos de riesgo, los sujetos pueden estratificarse en grupos de riesgo de recidiva bajo, medio y alto. El cálculo de ROR puede proporcionar información de pronóstico para orientar las decisiones de tratamiento y/o monitorizar la respuesta a terapia.

En algunas realizaciones descritas en el presente documento, el rendimiento pronóstico de los subtipos intrínsecos y/u otros parámetros clínicos definidos en PAM50 se evalúa utilizando un análisis de modelo de riesgos proporcionales de Cox, que es un método de regresión para datos de supervivencia que proporciona una estimación de la razón de riesgo y su intervalo de confianza. El modelo de Cox es una técnica estadística bien reconocida para explorar la relación entre la supervivencia de una paciente y variables particulares. Este método estadístico permite la estimación del riesgo de individuos dadas sus variables pronósticas (por ejemplo, perfil de expresión de genes intrínsecos con o sin factores clínicos adicionales, tal como se describió en el presente documento). La "razón de riesgo" es el riesgo de muerte a cualquier punto de tiempo dado para pacientes que presentan variables pronósticas particulares. Véase de manera general Spruance *et al.*, Antimicrob. Agents & Chemo. 48:2787-92, 2004.

El modelo de clasificación PAM50 descrito en el presente documento puede entrenarse para determinar el riesgo de recidiva usando distancias de subtipo (o correlaciones) solas, o usando distancias de subtipo con variables clínicas tal como se trató anteriormente. En una realización, la puntuación de riesgo para una muestra de prueba se calcula usando distancias de subtipo intrínseco solas usando la siguiente ecuación: $ROR = 0,05 \cdot \text{Basal} + 0,11 \cdot \text{Her2} - 0,25 \cdot \text{LumA} + 0,07 \cdot \text{LumB} - 0,11 \cdot \text{Normal}$, en la que las variables "Basal", "Her2", "LumA", "LumB" y "Normal" son las distancias al centroide de cada clasificador respectivo cuando se compara el perfil de expresión de una muestra de prueba con centroides construidos usando los datos de expresión génica depositados con el Gene Expression Omnibus (GEO) como número de registro GSE2845. Puede accederse a estos datos en la dirección de internet www.ncbi.nlm.nih.gov/geo; también es posible que puedan usarse otros conjuntos de datos para derivar coeficientes del modelo de Cox similares. Cuando se usa la lista de genes intrínsecos expuesta en la tabla 1 para desarrollar un modelo de predicción a partir de un conjunto de muestras distinto de las muestras usadas para derivar el conjunto de datos depositado como GSE2845, pueden usarse los métodos descritos en el ejemplo 1 o el ejemplo 3 para construir una fórmula para calcular el riesgo de recidiva a partir de este conjunto de muestras alternativo.

La puntuación de riesgo también puede calcularse usando una combinación de subtipo de cáncer de mama y las variables clínicas tamaño del tumor (T) y estado de los ganglios linfáticos (N) usando la siguiente ecuación: $ROR (\text{completo}) = 0,05 \cdot \text{Basal} + 0,1 \cdot \text{Her2} - 0,19 \cdot \text{LumA} + 0,05 \cdot \text{LumB} - 0,09 \cdot \text{Normal} + 0,16 \cdot T + 0,08 \cdot N$, de nuevo cuando se comparan perfiles de expresión de prueba con centroides construidos usando los datos de expresión génica depositados con GEO como número de registro GSE2845.

Aún en otra realización, la puntuación de riesgo para una muestra de prueba se calcula usando distancias de subtipo intrínseco solas usando la siguiente ecuación:

$$\text{ROR-S} = 0,05 \cdot \text{Basal} + 0,12 \cdot \text{Her2} + -0,34 \cdot \text{LumA} + 0,023 \cdot \text{LumB},$$

en la que las variables “Basal”, “Her2”, “LumA” y “LumB” son tal como se definieron anteriormente y los perfiles de expresión de prueba se comparan con centroides construidos usando los datos de expresión génica depositados con GEO como número de registro GSE2845.

Aún en otra realización, la puntuación de riesgo también puede calcularse usando una combinación de subtipo de cáncer de mama y la variable clínica tamaño del tumor (T) usando la siguiente ecuación (en la que las variables son tal como se describieron anteriormente):

$$\text{ROR-C} = 0,05 \cdot \text{Basal} + 0,11 \cdot \text{Her2} + -0,23 \cdot \text{LumA} + 0,09 \cdot \text{LumB} + 0,17 \cdot \text{T}.$$

Predicción de la respuesta a terapia

El cáncer de mama se trata mediante varias estrategias alternativas que pueden incluir, por ejemplo, cirugía, radioterapia, terapia hormonal, quimioterapia o alguna combinación de las mismas. Tal como se conoce en la técnica, las decisiones del tratamiento para pacientes con cáncer de mama individuales pueden basarse en la receptividad endocrina del tumor, el estado menopáusico de la paciente, la ubicación y el número de ganglios linfáticos de la paciente implicados, el estado de receptor de estrógeno y progesterona del tumor, el tamaño del tumor primario, la edad de la paciente y el estadio de la enfermedad en el diagnóstico. El análisis de una variedad de factores clínicos y ensayos clínicos ha conducido al desarrollo de recomendaciones y directrices de tratamiento para cáncer de mama en estadio temprano por el Panel de Consenso Internacional (“*International Consensus Panel*”) de la conferencia de St. Gallen (2005). Véase, Goldhirsch *et al.*, *Annals Oncol.* 16:1569-83, 2005. Las directrices recomiendan que a las pacientes se les ofrezca quimioterapia para enfermedad no sensible endocrina; terapia endocrina como terapia principal para enfermedad sensible endocrina, añadiendo quimioterapia para algunos grupos de riesgo intermedio y todos los de riesgo alto en esta categoría; y tanto quimioterapia como terapia endocrina para todas las pacientes en la categoría de respuesta endocrina incierta excepto las que estén en el grupo de riesgo bajo.

La estratificación de pacientes según el riesgo de recidiva usando el modelo PAM50 y la puntuación de riesgo dados a conocer en el presente documento proporcionan un factor de decisión de tratamiento adicional o alternativo. Los métodos comprenden evaluar el riesgo de recidiva usando el modelo de clasificación PAM50 opcionalmente en combinación con una o más variables clínicas, tales como estado de los ganglios, tamaño del tumor y estado de ER. La puntuación de riesgo puede usarse para orientar las decisiones de tratamiento. Por ejemplo, un sujeto que tiene una puntuación de riesgo bajo puede que no se beneficie de determinados tipos de terapia, mientras que un sujeto que tiene una puntuación de riesgo alto puede estar indicado para una terapia más agresiva.

Los métodos de la invención encuentran uso particular en la elección del tratamiento apropiado para pacientes con cáncer de mama en estadio temprano. La mayoría de las pacientes con cáncer de mama diagnosticadas en un estadio temprano de la enfermedad disfrutan de supervivencia a largo plazo tras cirugía y/o radioterapia sin terapia adyuvante adicional. Sin embargo, un porcentaje significativo (aproximadamente el 20%) de estas pacientes padecerá reaparición de la enfermedad o muerte, conduciendo a la recomendación clínica de que algunas o todas las pacientes con cáncer de mama en estadio temprano deben recibir terapia adyuvante. Los métodos de la presente invención encuentran uso en la identificación de esta población de mal pronóstico, de riesgo alto de pacientes con cáncer de mama en estadio temprano y en la determinación de ese modo de qué pacientes se beneficiarían de terapia continuada y/o más agresiva y monitorización estrecha tras el tratamiento. Por ejemplo, las pacientes con cáncer de mama en estadio temprano que se ha evaluado que tienen una puntuación de riesgo alto mediante los métodos dados a conocer en el presente documento pueden seleccionarse para una terapia adyuvante más agresiva, tal como quimioterapia, tras cirugía y/o tratamiento con radiación. En realizaciones particulares, los métodos de la presente invención pueden usarse junto con las directrices de tratamiento establecidas por la conferencia de St. Gallen para permitir a los médicos tomar decisiones de tratamiento del cáncer de mama más informadas.

En diversas realizaciones, el modelo de clasificación PAM50 proporciona información sobre subtipos de cáncer de mama que no puede obtenerse usando ensayos clínicos convencionales tales como inmunohistoquímica u otros análisis histológicos. Por ejemplo, los sujetos que se han clasificado como receptor de estrógeno (ER)-positivos y/o receptor de progesterona (PR)-positivos podrían estar indicados según las directrices convencionales para la terapia endocrina. Tal como se trata en el ejemplo 2, el modelo dado a conocer en el presente documento puede identificar un subconjunto de estos casos ER+/PgR+ que se clasifican como de tipo basal, lo que puede indicar la necesidad de terapia más agresiva que no estaría indicada basándose sólo en el estado de ER o PgR.

Por tanto, los métodos dados a conocer en el presente documento también encuentran uso en la predicción de la respuesta de una paciente con cáncer de mama a un tratamiento seleccionado. “Predicción de la respuesta de una paciente con cáncer de mama a un tratamiento seleccionado” pretende significar evaluar la probabilidad de que una paciente experimente un desenlace positivo o negativo con un tratamiento particular. Tal como se usa en el presente documento, “indicativo de un desenlace de tratamiento positivo” se refiere a un aumento de la probabilidad de que la paciente experimente resultados beneficiosos a partir del tratamiento seleccionado (por ejemplo, remisión completa

o parcial, reducción del tamaño del tumor, etc.). “Indicativo de un desenlace de tratamiento negativo” pretende significar un aumento de la probabilidad de que la paciente no se beneficie del tratamiento seleccionado con respecto a la progresión del cáncer de mama subyacente.

En algunas realizaciones, el tiempo relevante para evaluar el pronóstico o el tiempo de supervivencia libre de enfermedad empieza con la eliminación quirúrgica del tumor o la supresión, mitigación o inhibición del crecimiento tumoral. En otra realización, la puntuación de riesgo basada en PAM50 se calcula basándose en una muestra obtenida tras el inicio de la terapia neoadyuvante tal como terapia endocrina. La muestra puede tomarse a cualquier tiempo tras el inicio de la terapia, pero preferiblemente se obtiene tras aproximadamente un mes de modo que la terapia neoadyuvante pueda cambiarse a quimioterapia en pacientes que no responden. Se ha mostrado que un subconjunto de tumores indicados para el tratamiento endocrino antes de la cirugía no responde a esta terapia. El modelo proporcionado en el presente documento puede usarse para identificar tumores agresivos que es probable que no respondan a la terapia endocrina, incluso cuando los tumores son positivos para receptores de estrógeno y/o progesterona. En esta realización, se obtiene una puntuación de riesgo de PAM50 ponderada para la proliferación según la siguiente ecuación: $RSp = (-0,0129 \cdot Basal) + (0,106 \cdot Her2) + (-0,112 \cdot LumA) + (0,039 \cdot LumB) + (-0,069 \cdot Normal) + (0,272 \cdot Prolif)$, en la que la puntuación de proliferación (“prolif”) se asigna como la medición media de los siguientes genes (tras la normalización): CCNB1, UBE2C, BIRC5, KNTC2, CDC20, PTTG1, RRM2, MK167, TYMS, CEP55 y CDCA 1. Todas las otras variables son las mismas que las ecuaciones de RS descritas a continuación. Tal como se trata en el ejemplo 2, la evaluación de la puntuación de riesgo tras el inicio de la terapia es más predictivo del desenlace al tratamiento, al menos en una población de pacientes ER+ que están sometidos a terapia endocrina neoadyuvante.

Kits

La presente invención también proporciona kits útiles para clasificar subtipos intrínsecos de cáncer de mama y/o proporcionar información de pronóstico. Estos kits comprenden un conjunto de sondas de captura y/o cebadores específicos para los genes intrínsecos enumerados en la tabla 1, así como reactivos suficientes para facilitar la detección y/o la cuantificación del producto de la expresión de genes intrínsecos. El kit puede comprender además un medio legible por ordenador.

En una realización de la presente invención, las sondas de captura están inmovilizadas sobre una matriz. Por “matriz” se pretende decir un soporte sólido o un sustrato con sondas peptídicas o de ácido nucleico unidas al soporte o sustrato. Las matrices comprenden normalmente una pluralidad de diferentes sondas de captura que están acopladas a una superficie de un sustrato en diferentes ubicaciones conocidas. Las matrices de la invención comprenden un sustrato que tiene una pluralidad de sondas de captura que pueden unirse específicamente a un producto de expresión de genes intrínsecos. El número de sondas de captura sobre el sustrato varía con el fin para el que la matriz esté prevista. Las matrices pueden ser matrices de baja densidad o matrices de alta densidad y pueden contener 4 o más, 8 o más, 12 o más, 16 o más, 32 o más direcciones, pero como mínimo comprenderán sondas de captura para los 50 genes intrínsecos enumerados en la tabla 1.

Se describen técnicas para la síntesis de estas matrices usando métodos de síntesis mecánica en, por ejemplo, la patente estadounidense n.º 5.384.261, incorporada en el presente documento como referencia en su totalidad para todos los fines. La matriz puede fabricarse sobre una superficie de prácticamente cualquier forma o incluso una multiplicidad de superficies. Las matrices pueden ser sondas (por ejemplo, sondas de unión a ácido nucleico) sobre perlas, geles, superficies poliméricas, fibras tales como fibras ópticas, vidrio o cualquier otro sustrato apropiado, véanse las patentes estadounidenses n.ºs 5.770.358, 5.789.162, 5.708.153, 6.040.193 y 5.800.992, cada una de las cuales se incorpora por el presente documento en su totalidad para todos los fines. Las matrices pueden envasarse de una manera tal como para permitir el diagnóstico u otra manipulación en el dispositivo. Véase, por ejemplo, las patentes estadounidenses n.ºs 5.856.174 y 5.922.591 incorporadas en el presente documento como referencia.

En otra realización, el kit comprende un conjunto de cebadores oligonucleotídicos suficientes para la detección y/o cuantificación de cada uno de los genes intrínsecos enumerados en la tabla 1. Los cebadores oligonucleotídicos pueden proporcionarse en forma liofilizada o reconstituida, o pueden proporcionarse como un conjunto de secuencias de nucleótidos. En una realización, los cebadores se proporcionan en un formato de microplaca, en el que cada conjunto de cebadores ocupa un pocillo (o múltiples pocillos, como es el caso de replicados) en la microplaca. La microplaca puede comprender además suficientes cebadores para la detección de uno o más genes de mantenimiento tal como se trata a continuación. El kit puede comprender además reactivos e instrucciones suficientes para la amplificación de productos de expresión a partir de los genes enumerados en la tabla 1.

Con el fin de facilitar el acceso directo, por ejemplo, para comparación, revisión, recuperación y/o modificación, los perfiles de expresión/distintivos moleculares se registran normalmente en una base de datos. Lo más normalmente, la base de datos es una base de datos relacional a la que puede accederse mediante un dispositivo informático, aunque pueden usarse otros formatos, por ejemplo, archivos indexados accesibles manualmente de perfiles de expresión tales como fotografías, lecturas de imágenes analógicas o digitales, hojas de cálculo, etc. Independientemente de si los patrones de expresión registrados inicialmente son de naturaleza analógica o digital, los patrones de expresión, perfiles de expresión (patrones de expresión colectivos) y distintivos moleculares

(patrones de expresión correlacionados) se almacenan digitalmente y se accede a los mismos por medio de una base de datos. Normalmente, la base de datos se compila y se mantiene en una instalación central, estando disponible el acceso de manera local y/o remota.

- 5 El artículo “un” y “una” se usan en el presente documento para referirse a uno o más de uno (es decir, al menos uno) del objeto gramatical del artículo. A modo de ejemplo, “un elemento” significa uno o más elementos.

10 A lo largo de toda la memoria descriptiva la palabra “comprendiendo”, o variaciones tales como “comprende” o “que comprende”, se entenderá que implican la inclusión de un elemento, número entero o etapa, o grupo de elementos, números enteros o etapas establecidos, pero no la exclusión de cualquier otro elemento, número entero o etapa, o grupo de elementos, números enteros o etapas.

Los siguientes ejemplos se ofrecen a modo de ilustración y no a modo de limitación:

15 PARTE EXPERIMENTAL

Ejemplo 1.

Métodos

20 *Muestras y datos clínicos:*

25 Las cohortes de pacientes para los conjuntos de entrenamiento y prueba consistieron en muestras con datos ya del dominio público (Loi *et al.* (2007) J. Clin. Oncol. 25:1239-1246; va de Vijver *et al.* (2002) N Engl J Med 247:1999-2009; Wang *et al.* (2005) Lancet 365:671-679; Ishvina *et al.* (2006) Cancer Res 66:10292-10301; y Hess *et al.* (2006) J Clin Oncol 24:4236-4244, cada de uno de los cuales se incorpora como referencia en su totalidad) y tejidos recién congelados e incrustados en parafina fijados en formalina (FFPE) recogidos según los protocolos aprobados por la junta de revisión institucional en las instituciones respectivas.

30 Se obtuvo un conjunto de entrenamiento de 189 muestras de tumor de mama y 29 muestras normales como tejidos recién congelados y FFPE según los protocolos de la IRB aprobados en la Universidad de Carolina del Norte en Chapel Hill, Universidad de Utah, Universidad Thomas Jefferson y Universidad de Washington. El conjunto de entrenamiento, del que se había obtenido el perfil de expresión génica mediante micromatriz y qRT-PCR, tenía una mediana de seguimiento de 49 meses y representa pacientes tratadas de manera heterogénea según las normas asistenciales indicadas en su estadio, estado de ER y HER2. Se obtuvo el perfil de expresión génica de un conjunto de prueba de 279 cánceres de mama con supervivencia específica de la enfermedad a largo plazo a partir de FFPE mediante qRT-PCR. Los datos clínicos para el conjunto de entrenamiento y de prueba sometidos a ensayo mediante qRT-PCR se proporcionan en las tablas 2 y 3.

40 *Extracción de ácido nucleico:*

45 Se purificó el ARN total de muestras congeladas frescas para micromatriz usando el kit RNeasy Midi de Qiagen según el protocolo del fabricante (Qiagen, Valencia CA). Se determinó la integridad del ARN usando un bioanizador Agilent 2100 (Agilent Technologies, Palo Alto, CA). Se usó el kit de parafina para ARN de alta pureza (Roche Applied Science, Indianápolis, IN) para extraer ARN de tejidos FFPE (punzones de 2X10 micrómetros o 1,5 mm) para qRT-PCR. Se eliminó el ADN contaminante usando ADNasa Turbo (Ambion, Austin, TX). Se evaluó el rendimiento del ARN total usando el espectrofotómetro NanoDrop ND-1000 (NanoDrop Technologies, Inc., Rockland, DE).

50 *Transcripción inversa y PCR cuantitativa en tiempo real:*

55 Se sintetizó el ADNc de primera cadena a partir de 1,2 µg de ARN total usando la transcriptasa inversa Superscript III (Kit 1st Strand; Invitrogen, Carlsbad, CA) y una mezcla de hexámeros aleatorios y cebadores específicos de genes. Se mantuvo la reacción a 55°C durante 60 minutos y luego a 70°C durante 15 minutos. Se lavó el ADNc en una columna de purificación de PCR QIAquick (Qiagen Inc., Valencia, CA) y se almacenó a -80°C en Tris 25 mM, EDTA 1 mM hasta su uso adicional. Cada 5 µl de reacción de PCR incluían 1,25 ng (0,625 ng/µl) de ADNc procedente de muestras de interés o 10 ng (5 ng/µl) para referencia, 2 pmol de cebadores tanto anteriores como posteriores y 2X mezcla maestra de SYBR Green I de LightCycler 480 (Roche Applied Science, Indianápolis, IN). Cada serie contenía un único gen del que se había obtenido el perfil por duplicado para las muestras de prueba, la muestra de referencia y el control negativo. El ADNc de muestra de referencia estaba compuesto por una contribución igual de ARN total de referencia humano (Stratagene, La Jolla, CA) y las líneas celulares de mama MCF7, ME16C y SKBR3. Se realizó la amplificación por PCR con el instrumento LightCycler 480 (Roche Applied Science, Indianápolis, IN) usando una etapa de desnaturalización inicial (95°C, 8 minutos) seguida por 45 ciclos de desnaturalización (95°C, 4 segundos), apareamiento (56°C, 6 segundos con transición de 2,5°C/s) y extensión (72°C, 6 segundos con transición de 2°C/s). Se adquirió la fluorescencia (530 nm) del colorante SYBR Green I de ADNbc cada ciclo tras la etapa de extensión. Se determinó la especificidad de la PCR mediante el análisis de la curva

de fusión tras la amplificación, se enfriaron las muestras hasta 65°C y se calentaron lentamente a de 2°C/s a 99°C mientras se monitorizaba continuamente la fluorescencia (10 adquisiciones/1°C). Se determinó el número relativo de copias para cada gen a partir de un conjunto de calibrador intraserial a 10 ng y usando una eficacia de PCR de 1,9. Se normalizó cada uno de los genes del clasificador PAM50 a la media geométrica de 5 genes de mantenimiento.

Micromatriz:

Se realizaron aislamiento, etiquetado e hibridaciones de ARN total en micromatrices 1Av2 humanas de Agilent o matrices 22k humanas de Agilent diseñadas a medida usando el protocolo descrito en Hu *et al* (6). Todos los datos de las micromatrices se han depositado en GEO (dirección de internet www.ncbi.nlm.nih.gov/geo/) con el número de registro de GSE10886. En la tabla 4 se facilitan fuentes para todos los conjuntos de datos de entrenamiento y de prueba de micromatriz.

Procesamiento previo de datos de micromatriz:

Los datos de micromatriz para el conjunto de entrenamiento (189 muestras) se extrajeron de la base de datos de micromatrices de la Universidad de Carolina del Norte (UNC). Se normalizaron al nivel inferior las intensidades de señal sin procesar de ambos canales mediante chip y se excluyeron las sondas del análisis de datos si no tenían una intensidad de señal de al menos 30 en ambos canales para al menos el 70% de los experimentos. Los datos normalizados para este conjunto se han colocado en GEO (GSE10886). Se obtuvo el centro mediano del conjunto de entrenamiento y se asignaron símbolos de genes usando la anotación proporcionada por el fabricante. Los símbolos de genes duplicados se agruparon obteniendo el promedio en cada muestra.

Se descargaron datos normalizados para todos los conjuntos de prueba de GEO (GSE2845, GSE6532, GSE4922, GSE2034, GSE10886) o de los datos públicamente disponibles en la dirección de internet bioinformatics.mdanderson.org/pubdata (véase la tabla 5). Se transformaron logarítmicamente todas las medidas de intensidad (razones para los datos del NKL). Antes del cálculo del centroide más cercano, se obtuvo el centro mediano de los conjuntos de datos de Hess *et al.* (bioinformatics.mdanderson.org/pubdata), van de Vijver *et al.* (GSE2845), y Wang *et al.* (GSE2034) para minimizar los efectos de plataforma. El ajuste de esta forma supone una toma de muestras relativamente similar de la población como conjunto de entrenamiento. Los conjuntos de datos de Loi *et al.* (GSE6532) e Ivshina *et al.* (GSE4922) se enriquecieron en gran medida para las muestras de ER+ en relación con el conjunto de entrenamiento, por tanto puede violarse la suposición subyacente para estos conjuntos. En estos dos casos, los genes en el conjunto de entrenamiento se centraron a la mediana de las muestras de ER+ (en contraposición a la mediana para todas las muestras). Al igual que con el conjunto de entrenamiento, se asignaron símbolos de genes usando la anotación proporcionada por el fabricante y se agruparon los símbolos de genes duplicados obteniendo el promedio en cada muestra.

Identificación de genes y muestras de subtipo intrínseco prototipo:

Se usó inicialmente un conjunto de genes "intrínsecos" expandido, compuesto principalmente por genes encontrados en 4 estudios previos (1, 6, 9, 11), para identificar muestras de tumor prototipo. Se representó la clase de tipo normal usando "valores normales" verdaderos a partir de mamoplastia de reducción o tejido no implicado a simple vista. Se analizaron 189 tumores de mama a través de 1906 genes "intrínsecos" mediante agrupamiento jerárquico (obtención de centro mediano por característica/gen, correlación de Pearson, agrupamiento promedio)(12) y se analizó el dendograma de muestras usando "SigClust"(13). El algoritmo SigClust identifica estadísticamente grupos significativos/únicos sometiendo a prueba la hipótesis nula de que un grupo de muestras procede de un único grupo, donde un grupo se caracteriza como una distribución normal de múltiples variables. Se ejecutó SigClust en cada nodo del dendograma comenzando en la raíz y deteniéndose cuando la prueba ya no era significativa ($p > 0,001$).

Reducción del conjunto de genes usando muestras prototipo y qRT-PCR:

122 cánceres de mama de 189 mujeres de las que se había obtenido el perfil mediante qRT-PCR y micromatriz tenían perfiles prototipo tal como se determina mediante SigClust (tabla 2). Se derivó un conjunto de genes minimizado de estas muestras prototipo usando los datos de qRT-PCR para 161 genes que pasaban los criterios de rendimiento de FFPE establecidos en Mullins *et al* (14). Se emplearon varios métodos de minimización incluyendo los "N" mejores datos estadísticos de la prueba de la t para cada grupo (15), las mejores puntuaciones de índice de agrupamiento (16) y los genes restantes tras la "contracción" de los datos estadísticos de la prueba de la t modificados (17). Se usó validación cruzada (10% al azar omitido en cada uno de 50 ciclos) para evaluar la robustez de los conjuntos de genes minimizados. Se seleccionó el método de la prueba de la t de "N" debido a que tiene el menor error de CV.

Predicción de subtipo de muestra.

Se compararon conjuntos de genes minimizados para determinar la reproducibilidad de la clasificación a través de 3 métodos de predicción basados en centroide: Análisis de predicción de micromatriz (PAM) (17), un centroide más cercano simple (6), y clasificación de centroide más cercano (ClANC) (18). La predicción de subtipos se realizó

calculando la correlación de rangos de Spearman de cada caso de prueba para cinco centroides (LumA, LumB, enriquecido en HER-2, de tipo basal y de tipo normal) y se asignó la clase basándose en el centroide más cercano. Se construyeron los centroides tal como se describe para el algoritmo PAM (17) usando los datos proporcionados en GSE10886; sin embargo, no se usó “contracción” y se usó la correlación de rangos de Spearman para la medida de la distancia. Se seleccionó este método como clasificador debido a su reproducibilidad de predicciones de subtipo a partir de conjuntos de genes grandes y minimizados. Se usó el clasificador de 50 genes final (en adelante denominado PAM50) para realizar predicciones de subtipo en 6 conjuntos de datos de micromatriz y 1 conjunto de datos de qRT-PCR (tabla 4). El conjunto de datos de Hess *et al* (19) no tiene datos de desenlace y se evalúa basándose en marcadores clínicos, subtipos y respuesta de neo-adyuvante.

Pronóstico usando datos de subtipo clínicos y moleculares:

Se evaluó la significación pronóstica de la clasificación de subtipo intrínseco junto con variables clínicas convencionales (tamaño del tumor (T), estado de los ganglios (N) y estado de ER) usando análisis de una variable y de múltiples variables con el tiempo hasta la recidiva (es decir cualquier acontecimiento) como criterio de valoración. Se realizaron pruebas de razón de verosimilitud para comparar modelos de datos clínicos disponibles, datos de subtipo y variables clínicas y moleculares combinadas. Se realizaron análisis de supervivencia categóricos usando una prueba de rangos logarítmicos y se visualizaron con diagramas de Kaplan-Meier.

Modelos de riesgo de desarrollo con datos clínicos y moleculares:

Se entrenaron modelos para determinar predicciones de riesgo de recidiva (ROR) usando subtipo solo y subtipo con información clínica. En ambos casos, se fijó un modelo de Cox de múltiples variables usando regresión de Ridge para el subconjunto no tratado de la cohorte de NKI295 (20). Se asignó una puntuación de riesgo a cada caso de prueba usando correlación para el subtipo solo (ROR; modelo 1) o usando un modelo completo con correlación de subtipo y dos variables clínicas (ROR (completo); modelo 2):

$$(1) \text{ ROR} = 0,05 \cdot \text{Basal} + 0,11 \cdot \text{Her2} + -0,25 \cdot \text{LumA} + 0,07 \cdot \text{LumB} + -0,11 \cdot \text{Normal}$$

$$(2) \text{ ROR (completo)} = 0,05 \cdot \text{Basal} + 0,1 \cdot \text{Her2} + -0,19 \cdot \text{LumA} + 0,05 \cdot \text{LumB} + -0,09 \cdot \text{Normal} + 0,16 \cdot \text{T} + 0,08 \cdot \text{N}$$

La suma de los coeficientes del modelo de Cox es la puntuación de “riesgo de recidiva” para cada paciente. Con el fin de clasificar muestras en grupos de riesgo específicos, se eligieron umbrales del conjunto de entrenamiento del NKI que no requerían que la muestra de LumA estuviera en el grupo de alto riesgo y que la muestra de tipo basal no estuviera en el grupo de bajo riesgo. Se determinaron umbrales del conjunto de entrenamiento y permanecieron sin cambios cuando se evaluaron los casos de prueba. Se compararon las predicciones para el subtipo solo y modelos combinados usando el índice C (dirección de internet lib.stat.cmu.edu/S/Harrell/Design.html). Se realizó el análisis mediante SiZer para caracterizar la relación entre la puntuación de ROR y la supervivencia libre de recidiva (21). Los intervalos de confianza del 95% para la puntuación de ROR son versiones locales de intervalos de confianza binomiales, calculándose el tamaño de muestra local a partir de un estimador de densidad tipo núcleo gaussiano, basado en la elección de Sheather-Jones de anchura de ventana (22).

Resultados

Creación de un nuevo modelo de subtipo basado en genes y muestras prototipo:

Ha habido numerosos estudios que han analizado interacciones entre subtipos intrínsecos de cáncer de mama y pronóstico (1, 6, 9), alteraciones genéticas (23) y respuesta a fármacos (24). El fin de los métodos descritos en el presente documento fue normalizar y validar una clasificación para los subtipos intrínsecos para fines clínicos y de investigación. “SigClust” identificó objetivamente cinco subtipos de mama intrínsecos a partir de datos de micromatriz agrupados. Se usaron entonces estos prototipos para derivar un conjunto mínimo de 50 genes (PAM50). Finalmente, se seleccionó el mejor método de clasificación y se usó con PAM50 para predecir subtipos en múltiples conjuntos de prueba a partir de datos de micromatriz y qRT-PCR. De los 5 estudios de micromatriz con datos de desenlace (tabla 4), la cohorte de UNC tuvo significativamente peores desenlaces que las otras. Las predicciones de subtipo en un conjunto de prueba de micromatriz combinado mostró significación pronóstica en todas las pacientes, en pacientes a las que se administró tratamiento endocrino solo y en pacientes negativas para ganglios que no recibieron terapia sistémica adyuvante (figura 1).

Se evaluaron los factores pronóstico clínicos y moleculares de supervivencia en análisis de una variable y de múltiples variables en 1451 pacientes (tabla 5). En el análisis de una variable, se encontró que todos los subtipos de LumA, LumB y enriquecido en HER-2 eran significativos, como lo eran las variables clínicas ER, T y N. Los subtipos de LumA y enriquecido en HER-2 y las variables clínicas también fueron significativas en los análisis de múltiples variables, lo que sugiere que el modelo más amplio debe incluir información clínica y de subtipo. Al someter a prueba esta hipótesis se reveló que el modelo combinado representa significativamente más variación en supervivencia que cualquiera de las variables clínicas o de subtipo solas ($p < 0,0001$ para ambas pruebas).

Distribución de subtipos biológicos a través de tumores positivos para ER y negativos para ER:

De todos los tumores positivos para ER en el conjunto de prueba de micromatriz combinado, el 73% eran luminales (A y B), el 10% estaban enriquecidos en HER2 y el 5% eran de tipo basal (tabla 6). A la inversa, cuando se consideraban los tumores negativos para ER, aproximadamente el 13% eran luminales (A y B), el 31% estaban enriquecidos en HER-2 y el 48% eran de tipo basal. Los tumores identificados como de subtipo de tipo normal estaban divididos casi por igual entre tumores positivos para ER (11%) y negativos para ER (8%). Por tanto, aunque la representación de subtipos cambió notablemente en distribución dependiendo del estado de ER, todos los subtipos estaban representados en ambas categorías de positivos para ER y negativos para ER. Los diagramas de desenlaces para los subtipos en casos positivos para ER solo fueron significativos para la supervivencia libre de recidiva y siguieron las mismas tendencias que se observaron cuando se consideraba toda la enfermedad de mama invasiva.

Subtipos y respuesta a tratamiento neoadyuvante con T/FAC:

El estudio de Hess *et al.* que realizaron micromatrices en tumores para pacientes a las que se administró un régimen de paclitaxel, 5-fluorouracilo adriamicina y ciclofosfamida (T/FAC) (19) permitió la investigación de la relación entre los subtipos de PAM50, marcadores clínicos y cómo se relaciona cada uno con una respuesta patológica completa (pCR). Para el estado de HER2, el 64% de los tumores que eran positivos para HER2 mediante el ensayo clínico (FISH+ y/o IHC 3+, denominado *HER2+clin*) se clasificaron en el subtipo de expresión enriquecido en HER-2, estando el resto de *HER2+clin* asociado principalmente con los subtipos luminales. Los tumores que eran *HER2+clin* pero no del subtipo de expresión enriquecido en HER-2 tenían una tasa de pCR baja (16%) frente a los que eran *HER2+clin* y subtipo de expresión enriquecido en HER-2 (52%).

Otra distinción clínica relevante es la clasificación de tumores “triple negativos” (ER-, PgR- y HER2-), de los cuales el 65% se denominaron de tipo basal por PAM50, denominándose el resto enriquecido en HER-2 (15%), LumA (4%), LumB (4%) y de tipo normal (12%). La clasificación de PAM50 de tipo basal parece ser superior al triple negativo clínico con respecto a la tasa de pCR porque los tumores de tipo basal que no se clasificaron como triples negativos tuvieron un 50% de pCR en comparación con los tumores triples negativos que no eran de tipo basal por PAM50 (22% de pCR, tabla 7).

Predicción de riesgo basada en subtipo biológico:

Se desarrolló un clasificador de riesgo supervisado para predecir desenlaces dentro del contexto de los subtipos intrínsecos y variables clínicas. Se seleccionó una cohorte no tratada del conjunto de datos de micromatriz de NKI para entrenar el modelo de riesgo de recidiva (ROR) y seleccionar puntos de corte. Se validaron dos modelos de Cox (uno basado en el subtipo solo y otro basado en el subtipo, el tamaño del tumor y el estado de los ganglios) usando el conjunto de prueba de micromatriz combinado. Excluyendo las variables clínicas, el modelo sólo de subtipo funcionó bien en la estratificación de pacientes en grupos de riesgo de recidiva bajo, medio y alto (índice de $c=0,65$ [0,61-0,69]); sin embargo, el modelo completo (subtipo, tamaño del tumor, estado de los ganglios) funcionó mejor (índice de $c=0,70$ [0,66- 0,74]) y, en la práctica, el estadio es un parámetro que es necesario tener en cuenta (figura 2). La figura 3 muestra la probabilidad de supervivencia libre de recidiva a los 5 años representada gráficamente como una escala lineal continua usando el modelo completo.

Se aplicó el clasificador PAM50, sometido a ensayo mediante qRT-PCR, a una cohorte tratada de manera heterogénea archivada entre 1976 y 1995. Las clasificaciones de subtipo siguieron las mismas tendencias de supervivencia que las observadas en los datos de micromatriz y la puntuación de ROR fue significativa para predicciones de recidiva a largo plazo. Este conjunto de muestras de edad anciana también se puntuó para marcadores clínicos convencionales (ER y HER2) mediante inmunohistoquímica (IHC) y se comparó con la prueba basada en expresión génica. El análisis de ESR1 y ERBB2 mediante la expresión génica mostró alta sensibilidad y especificidad en comparación con el ensayo de IHC.

Discusión

Se desarrolló el clasificador PAM50 usando un conjunto de muestras y genes derivado estadísticamente y se validó a través de múltiples cohortes y plataformas con el intento de suministrar una prueba de diagnóstico clínico para los subtipos intrínsecos de cáncer de mama. Los conjuntos de prueba grandes y diversos permitieron la evaluación del rendimiento del ensayo en un nivel de población y en relación con marcadores moleculares convencionales. Un hallazgo importante a partir de estos análisis es que todos los subtipos intrínsecos están presentes dentro de ambos subconjuntos de tumor positivos para ER y negativos para ER clínicamente definidos, mostrando las designaciones de subtipo en las pacientes positivas para ER significación pronóstica. Por tanto, los subtipos moleculares no son simplemente otro método de clasificación basado en el estado de ER.

También hubo otros hallazgos importantes relativos a subtipos individuales. Por ejemplo, algunos de los tumores clasificados en el subtipo de expresión enriquecido en HER-2 no eran *HER2+clin*, lo que sugiere la presencia de un subtipo no basal negativo para ER que no se dirige por la amplificación génica de HER2. También se encontró que

aproximadamente el 10% de los cánceres de mama se clasificaron como de tipo normal y pueden ser o bien positivos para ER o bien negativos para ER y tienen un pronóstico intermedio. Puesto que estos tumores se predecían por entrenamiento en tejido de mama normal, la clase de tipo normal puede ser un artefacto de tener un alto porcentaje de “contaminación” normal en la muestra de tumor. Otras posibilidades son que éstos son tumores de tipo basal que crecen lentamente que carecen de alta expresión de los genes de proliferación, o son un posible nuevo subtipo que se ha denominado tumores bajos en claudina (25). Son necesarios análisis detallados histológicos, inmunohistoquímicos y de expresión génica adicionales en estos casos para resolver estas cuestiones.

Las discrepancias entre marcadores moleculares de subtipo y convencionales tienen implicaciones terapéuticas importantes. Por ejemplo, una paciente con un tumor de subtipo de tipo basal que se clasificó positivo para ER o PgR se trataría probablemente mediante terapia endocrina y no sería elegible para protocolos que se dirigen a desarrollar terapias específicas de tipo basal (por ejemplo regímenes que contienen platino). Estos análisis del conjunto de datos de Hess *et al.* (19) mostraron que ninguna paciente con el subtipo LumA tenía una pCR cuando se les administró un régimen neoadyuvante agresivo, mientras que la tasa de pCR de los tumores de tipo basal fue del 59%. Además, ha habido debate sobre si el fenotipo triple negativo (ER-, PR-, HER2-) es el mismo que en el subtipo 26 de expresión de tipo basal. Un estudio de micromatriz tisular reciente de 3744 tumores confirmó el mal pronóstico de los casos triples negativos, pero también reveló que los tumores que carecen de todos los marcadores no se comportaban igual que los que eran positivos para uno o dos marcadores de tipo basal (es decir CK5/6 o HER1) (27). De acuerdo con la idea de que el diagnóstico de tipo basal debe realizarse independientemente del estado clínico de ER y PgR, se encontró una respuesta terapéutica más alta a T/FAC en las pacientes identificadas como de tipo basal pero no triples negativas (50%) frente a las identificadas como triples negativas pero no de tipo basal (22%). Esto sugiere que la designación de subtipo de tipo basal puede demostrar en última instancia ser superior a la definición de triple negativo en la identificación de tumores con un alto grado de sensibilidad a quimioterapia.

Proporcionar una clasificación de subtipo absoluta es algo artificial ya que los tumores no existen como entidades biológicas diferenciadas. La clasificación de tumores en grupos de riesgo bajo-medio-alto basándose en la distancia a cada centroide de subtipo (es decir, el modelo de ROR) fue un intento para tratar esta cuestión y produjo segregación de supervivencia significativa. Esto fue cierto cuando se combinaban todos los casos de prueba, o tras estratificación en cohortes habiéndose administrado sólo terapia endocrina, o sin tratamiento adyuvante sistémico. Uno de los principales beneficios del factor pronóstico de ROR es la identificación de pacientes LumA que tienen un riesgo de recidiva muy bajo y para las que es improbable el beneficio de la quimioterapia adyuvante. En este contexto, el factor pronóstico de ROR basado en subtipos proporciona información similar a la puntuación de reaparición de OncotypeDx para pacientes positivas para ER, negativas para ganglios (4, 5). Sin embargo, el ensayo basado en PAM50 proporciona una puntuación de riesgo de recidiva para todas las pacientes, incluyendo las que tienen enfermedad negativa para ER.

En resumen, este factor pronóstico de subtipo y clasificador de ROR identifica eficazmente características moleculares en tumores de mama que son importantes para pronóstico y tratamiento. El ensayo de qRT-PCR puede realizarse usando tejidos de mama archivados, que serán útiles para ensayos clínicos prospectivos y estudios retrospectivos.

Tabla 2. Datos de subtipo y clínicos para muestras prototipo de conjuntos de entrenamiento de micromatriz/qRT-PCR

ID de la paciente	Registro de GEO	Asignación de subtipo (SigClust)	Nombre de qPCR	Edad en el diagn óstico	Etnia	pT*	pN^	M	% de grado	Supervivencia global	Estado vital	ER (IHC)**	PR (IHC)**	HER2 (estado)^
1	GSM275694	De tipo basal	BR000161BP E_UU	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
3	GSM140985	De tipo basal	BR000572BP E_UU	45	AA	3	0	0	3	NA	NA	0	0	NA
7	GSM140999	De tipo basal	BR010235BP E_UU	36	AA	1	0	0	3	49	0	0	0	0
10	GSM275782	De tipo basal	BR010532BP E_UU	47	AA	3	1	0	3	14	1	0	0	0
11	GSM80221	De tipo basal	BR020018BP E_UU	55	C	2	0	0	3	31	0	0	0	0
13	GSM141099	Enriquecido en HER2	BR020306BP E_UU	42	C	4	2	1	3	22	1	1	1	1
14	GSM141102	LumB	BR020439BP E_UU	53	C	4	1	1	3	16	1	1	0	0
15	GSM275783	LumA	BR020464BP E_UU	44	C	2	0	0	1	2	0	1	1	0
16	GSM141105	De tipo basal	BR020578BP E_UU	76	AA	4	0	0	3	5	1	0	0	1
18	GSM141110	De tipo basal	BR030459BP E_UU	30	C	3	0	0	3	37	0	0	0	0
19	GSM275771	LumA	BR030584BP E_UU	54	C	1	1	NA	NA	NA	0	1	1	0
21	GSM141114	LumB	BR040114BP E_UU	56	C	2	0	0	2	16	0	1	1	0
22	GSM141117	LumB	BR040182BP E_UU	88	C	2	1	0	3	16	0	1	1	1

23	GSM141121	Enriquecido en HER2	BR040269BP E_UU	46	C	2	0	0	0	3	17	0	0	0	1
28	GSM34523	De tipo basal	PB00205PE_ UU	39	C	NA	NA	1	3	5	1	0	0	0	0
29	GSM52895	LumA	PB00284PE_ UU	34	C	1	0	0	1	54	0	1	1	0	0
30	GSM34565	De tipo basal	PB00297PE_ UU	55	AA	2	0	0	3	55	0	0	0	0	0
31	GSM34481	Enriquecido en HER2	PB00311PE_ UU	47	C	2	1	0	3	50	0	1	1	0	0
32	GSM34497	Enriquecido en HER2	PB00314PE_ UU	50	C	3	1	0	3	52	0	0	0	1	1
33	GSM34527	De tipo basal	PB00334PE_ UU	50	AA	1	0	0	3	54	0	0	0	0	0
35	GSM34544	Enriquecido en HER2	PB00376PE_ UU	50	AA	2	0	0	3	49	0	0	0	0	0
37	GSM34549	LumA	PB00441PE_ UU	83	C	1	0	0	2	14	0	1	1	0	0
38	GSM34528	Enriquecido en HER2	PB00455PE_ UU	52	AA	3	1	0	2	46	0	0	0	1	1
39	GSM52884	LumA	PB00479PE_ UU	50	NA	2	0	0	NA	NA	0	1	1	0	0
41	GSM50157	De tipo basal	UB00028PE_ UU	46	C	1	0	0	3	59	0	0	0	0	0
42	GSM34437	De tipo basal	UB00029PE_ UU	59	C	2	0	0	3	59	0	0	0	0	0
43	GSM34431	Enriquecido en HER2	UB00037PE_ UU	42	C	1	1	0	3	58	0	0	1	0	0
44	GSM34548	LumA	UB00038PE_ UU	50	C	1	0	0	2	57	0	1	1	0	0

47	GSM34428	LumA	UB00044PE_ UU	49	C	2	1	0	2	59	0	1	1	0
50	GSM34557	LumA	UB00056PE_ UU	63	C	1	1	0	2	56	0	1	1	0
53	GSM34532	Enriquecido en HER2	UB00060PE_ UU	72	C	3	3	0	3	49	0	0	0	1
56	GSM34450	De tipo basal	UB00067PE_ UU	80	C	1	1	0	3	38	1	0	0	0
57	GSM34451	LumA	UB00069PE_ UU	40	C	1	0	0	2	7	0	NA	NA	0
58	GSM34452	De tipo basal	UB00071PE_ UU	60	C	1	0	0	3	50	0	0	0	0
60	GSM141079	LumA	UB00081LPE_ UU	65	C	NA	1	0	2	44	0	1	1	0
61	GSM141081	LumA	UB00082PE_ UU	43	C	1	1	0	1	40	0	1	1	0
63	GSM141084	LumB	UB00088PE_ UU	69	C	2	2	0	2	38	1	1	1	1
64	GSM141085	LumA	UB00091PE_ UU	77	C	1	0	0	2	36	0	1	1	0
65	GSM141088	LumA	UB00099PE_ UU	50	C	3	1	0	2	35	0	1	1	0
66	GSM141070	De tipo basal	UB00100PE_ UU	49	C	1	0	0	3	34	0	NA	NA	0
67	GSM141071	De tipo basal	UB00110PE_ UU	76	C	2	0	0	3	31	0	0	0	0
68	GSM141072	De tipo basal	UB00116PE_ UU	67	C	2	0	NA	3	34	0	0	0	1
69	GSM141073	Enriquecido en HER2	UB00117PE_ UU	72	otro	NA	0	0	3	31	0	0	0	1

72	GSM275802	De tipo basal	WU00328-16563PE_UU	59	C	2	0	0	3	82	0	0	0	0
73	GSM275803	Enriquecido en HER2	WU00431-16439PE_UU	73	C	4	1	0	3	44	1	0	0	1
74	GSM275800	Enriquecido en HER2	WU00441-19793PE_UU	49	C	3	1	0	2	27	1	1	1	1
75	GSM275804	LumA	WU00509-19794PE_UU	57	C	2	1	0	3	51	0	0	0	0
76	GSM275805	Enriquecido en HER2	WU00531-19795PE_UU	75	C	2	0	0	1	81	0	1	0	0
78	GSM275807	LumB	WU00556-21032PE_UU	46	C	2	1	0	2	88	0	1	1	0
82	GSM275810	De tipo basal	WU00899-18760PE_UU	47	AA	2	0	0	3	83	0	1	1	0
86	GSM275813	De tipo basal	WU01407-16456PE_UU	39	AA	2	0	0	3	80	0	0	0	0
88	GSM275815	Enriquecido en HER2	WU01500-18755PE_UU	88	C	1	0	0	3	70	0	1	0	0
89	GSM275816	Enriquecido en HER2	WU01502-16455PE_UU	74	C	2	2	0	3	88	0	1	0	0
90	GSM275817	Enriquecido en HER2	WU01511-19773PE_UU	50	AA	1	1	0	3	82	0	0	0	1
91	GSM275818	LumA	WU01520-21957PE_UU	58	C	2	1	0	2	77	NA	1	1	NA
92	GSM275819	Enriquecido en HER2	WU01540-14690PE_UU	46	C	3	1	0	3	20	1	NA	NA	0
93	GSM275799	LumB	WU01576-19797PE_UU	64	O	2	1	0	3	56	0	1	1	0
95	GSM275821	LumB	WU01587-16348PE_UU	72	C	1	0	1	2	82	0	1	0	0

96	GSM275822	LumB	WU01613-16349PE_UU	78	C	2	2	0	3	17	0	1	1	0
97	GSM275823	De tipo basal	WU01680-16347PE_UU	47	C	2	0	0	3	77	0	0	0	1
99	GSM275825	De tipo basal	WU01790-16344PE_UU	32	AA	3	1	0	3	15	0	0	0	1
101	GSM275792	De tipo basal	WU01887-16342PE_UU	73	AA	2	0	0	3	78	0	0	1	0
104	GSM275829	De tipo basal	WU02104-16341PE_UU	57	C	2	1	0	3	51	0	0	0	0
105	GSM275830	De tipo basal	WU02132-18761PE_UU	57	C	3	1	0	3	76	0	0	0	0
107	GSM275832	Enriquecido en HER2	WU02338-21961PE_UU	42	C	2	1	0	3	59	1	1	1	1
108	GSM275833	De tipo basal	WU02390-16330PE_UU	46	C	2	0	1	3	9	1	0	0	0
109	GSM275834	De tipo basal	WU02455-14693PE_UU	44	C	3	0	0	3	15	1	0	0	0
110	GSM275797	Enriquecido en HER2	WU02468-21279PE_UU	63	AA	1	2	0	3	72	0	NA	NA	1
113	GSM275795	Enriquecido en HER2	WU02769-16337PE_UU	73	C	1	0	0	2	69	0	1	0	1
114	GSM275837	De tipo basal	WU02771-14694PE_UU	43	C	3	0	0	3	16	1	0	0	1
116	GSM275839	De tipo basal	WU02843-19762PE_UU	46	C	1	0	0	3	75	0	0	0	1
118	GSM275841	De tipo basal	WU02948-16566PE_UU	44	C	1	0	0	3	62	0	0	0	0
120	GSM275842	Enriquecido en HER2	WU03064-16462PE_UU	74	AA	2	0	0	3	70	0	1	1	1

121	GSM275843	De tipo basal	WU03292-16446PE_UU	50	AA	3	0	0	3	79	0	0	0	0
123	GSM275791	Enriquecido en HER2	WU03456-16361PE_UU	52	AA	1	0	0	3	67	0	1	1	0
125	GSM275846	LumB	WU03535-16451PE_UU	82	C	2	0	0	3	60	0	1	1	0
126	GSM275796	Enriquecido en HER2	WU03653-16448PE_UU	49	C	1	2	0	3	102	0	0	0	1
127	GSM275847	De tipo basal	WU03661-16447PE_UU	53	AA	4	1	0	3	3	1	1	1	0
128	GSM275788	LumA	WU03662-16452PE_UU	75	AA	2	0	0	3	45	0	0	0	0
129	GSM275848	De tipo basal	WU03685-16502PE_UU	42	AA	2	0	0	3	65	0	0	0	0
131	GSM275850	De tipo basal	WU03714-21262PE_UU	66	AA	1	0	0	3	72	0	1	1	0
132	GSM275793	Enriquecido en HER2	WU03721-16570PE_UU	29	C	2	1	0	3	13	0	1	0	1
134	GSM275852	De tipo basal	WU03791-16497PE_UU	61	C	1	1	0	3	62	0	0	0	0
135	GSM275853	De tipo basal	WU03831-21959PE_UU	51	AA	2	1	0	3	68	0	0	0	1
139	GSM275857	De tipo basal	WU03885-16469PE_UU	52	AA	2	1	0	3	26	1	1	1	0
140	GSM275858	Enriquecido en HER2	WU03946-14842PE_UU	72	C	2	1	0	2	15	NA	0	1	1
141	GSM275789	De tipo basal	WU04000-16466PE_UU	49	AA	1	2	0	3	24	1	NA	NA	0
144	GSM275861	Enriquecido en HER2	WU04038-16465PE_UU	51	AA	2	1	0	2	62	0	1	1	1

146	GSM275863	De tipo basal	WU04327-19803PE_UU	73	C	1	0	0	2	69	0	1	0	1
147	GSM275864	LumB	WU04532-16463PE_UU	75	AA	2	1	0	3	53	0	1	0	1
148	GSM275865	De tipo basal	WU04834-16461PE_UU	42	AA	2	0	0	3	65	0	0	0	0
149	GSM275866	De tipo basal	WU04952-19753PE_UU	64	AA	2	1	0	3	62	0	0	0	0
152	GSM275872	LumA	WU05094-16580PE_UU	29	C	1	1	0	1	60	0	1	1	0
153	GSM275873	Enriquecido en HER2	WU05118-19759PE_UU	54	C	1	0	0	3	64	0	1	0	1
155	GSM275875	Enriquecido en HER2	WU05162-21960PE_UU	43	C	2	0	0	3	95	1	NA	NA	NA
156	GSM275876	De tipo basal	WU05191-14791PE_UU	51	AA	4	1	1	2	46	1	NA	NA	1
157	GSM275877	Enriquecido en HER2	WU05196-16573PE_UU	59	C	1	0	0	2	56	0	0	0	0
158	GSM275878	Enriquecido en HER2	WU05207-16473PE_UU	72	AA	4	1	0	3	59	0	1	0	1
159	GSM275879	De tipo basal	WU05215-16503PE_UU	79	AA	2	0	0	3	48	0	0	0	0
160	GSM275880	Enriquecido en HER2	WU05337-14835PE_UU	51	C	2	1	0	3	14	1	0	0	0
161	GSM275881	Enriquecido en HER2	WU05415-16357PE_UU	52	C	2	0	0	3	48	0	1	1	1
163	GSM275883	De tipo basal	WU05478-16356PE_UU	41	C	2	0	0	3	54	0	0	0	0
165	GSM275885	De tipo basal	WU05641-16354PE_UU	40	AA	2	0	0	3	54	0	0	0	0

167	GSM275887	De tipo basal	WU05991-14687PE_UU	70	AA	4	1	0	3	19	1	0	0	0
168	GSM275888	De tipo basal	WU06036-16352PE_UU	42	AA	2	0	0	3	50	0	0	0	NA
169	GSM275889	De tipo basal	WU06397-19805PE_UU	64	C	2	0	0	3	54	0	0	0	0
170	GSM275890	Enriquecido en HER2	WU06398-19767PE_UU	34	C	1	1	0	3	40	0	1	1	1
171	GSM275891	Enriquecido en HER2	WU06416-19781PE_UU	74	C	1	0	0	3	51	0	0	0	0
172	GSM275892	LumaA	WU06545-18758PE_UU	69	C	3	0	0	3	49	0	1	1	0
173	GSM275893	Enriquecido en HER2	WU06559-14689PE_UU	42	AA	4	2	0	3	35	1	0	0	0
174	GSM275894	De tipo basal	WU06580-16575PE_UU	44	AA	2	0	0	3	48	0	0	0	0
175	GSM275895	LumB	WU06611-16475PE_UU	46	C	2	0	0	3	51	0	1	1	1
177	GSM275897	De tipo basal	WU06857-15260PE_UU	40	AA	2	1	0	3	11	1	0	0	0
178	GSM275898	LumaA	WU07407-19770PE_UU	66	C	2	1	0	3	5	0	1	1	0
180	GSM275900	LumaA	WU07509-19782PE_UU	57	AA	1	1	0	2	22	0	1	1	0
182	GSM275902	LumaA	WU07512-14793PE_UU	83	C	1	NA	0	2	21	0	1	1	0
184	GSM275904	De tipo basal	WU07558-16584PE_UU	41	AA	1	0	0	3	29	0	0	0	0
185	GSM275905	De tipo basal	WU07589-16582PE_UU	70	AA	2	1	0	3	19	0	0	0	0

187	GSM275907	Luma	WU07791- 19784PE_UU	78	C	1	0	0	2	53	0	1	1	1
188	GSM275908	Luma	WU07805- 19777PE_UU	41	C	2	1	0	3	58	0	1	1	0
189	GSM275909	De tipo basal	WU08626- 16506PE_UU	34	AA	4	1	0	3	37	0	0	0	0
190	GSM34464	De tipo basal	NA	61	NA	2	1	0	3	13	1	0	NA	NA
191	GSM140992	De tipo basal	NA	29	NA	2	2	0	3	NA	1	0	NA	NA
192	GSM50148	De tipo basal	NA	51	C	4	2	0	NA	74	1	0	NA	NA
193	GSM34562	De tipo basal	NA	50	C	1	0	0	3	29	0	1	0	1
194	GSM52896	LumB	NA	79	NA	2	2	0	3	15	0	1	NA	1
195	GSM141067	De tipo normal	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
196	GSM80240	De tipo normal	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA		
197	GSM34547	De tipo normal	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
198	GSM34483	De tipo normal	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
199	GSM34482	De tipo normal	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
200	GSM140990	De tipo normal	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
201	GSM140991	De tipo normal	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

202	GSM275777	De tipo normal	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
203	GSM275778	De tipo normal	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
204	GSM275781	De tipo normal	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
205	GSM275780	De tipo normal	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
206	GSM275779	De tipo normal	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

*estadio patológico del tumor: T1≤2 cm, T2>2 cm-5 cm, T3>5 cm, NA=no evaluado
 ^estadio patológico de ganglio: N0=sin ganglios positivos, N1=ganglios axilares positivos, NA=no evaluado
 % de grado histológico: 0=grados 1 y 2, 1=grado 3
 **inmunohistoquímica: 0=de ausencia de tinción a tinción moderada, 1=fuerte tinción en la mayoría de las células cancerosas
 ^inmunohistoquímica e hibridación *in situ* fluorescente: 0=negativo mediante IHC (0, 1) o 2+ mediante IHC y negativo mediante FISH, 1=3+ mediante IHC o 2+ mediante IHC y positivo mediante FISH

Tabla 3. Datos de subtipo y clínicos para conjunto de prueba de qRT-PCR

ID de la paciente	Predicción de subtipo	pT*	pN^	% de grado	Supervivencia global	Supervivencia libre de recidiva	Cualquier recidiva***	DSS**	ER (IHC)^^^	PR (IHC)^^^	HER2 (IHC)^^^
1001	LumB	2	1	1	2,3232877	0,690411	1	1	1	1	0
1002	Enriquecido en HER2	2	NA	1	19,512329	2,9753425	1	1	0	0	0
1003	Enriquecido en HER2	1	1	1	7,0410959	5,4	1	1	0	NA	1
1004	LumB	2	0	0	NA	NA	NA	1	1	1	0
1005	Enriquecido en HER2	1	0	0	0,939726	0,3671233	1	1	0	NA	NA
1006	LumB	2	1	0	0,7178082	0,5178082	1	2	1	1	0
1007	LumB	NA	NA	0	3,3287671	1,7616438	1	1	1	1	0
1008	Enriquecido en HER2	2	NA	0	0,8849315	0,7068493	1	1	1	0	0
1009	De tipo basal	3	1	0	5,9917808	0,8246575	1	1	0	0	0
1010	LumA	2	1	NA	12,273973	9,0712329	1	1	1	1	NA
1011	Enriquecido en HER2	2	1	1	3,2027397	1,2493151	1	1	0	0	1
1012	De tipo normal	1	1	0	25,435616	22,709589	1	1	1	1	0
1013	LumB	2	1	0	NA	NA	NA	1	1	0	0

1014	LumB	2	0	0	0	16,654795	16,654795	0	2	1	1	0
1015	LumA	2	0	0	0	4,5150685	3,8821918	1	1	1	0	0
1016	De tipo basal	2	0	1	1	2,0383562	1,6712329	1	1	0	0	0
1017	LumA	2	1	0	0	10,331507	5,9315068	1	1	1	0	0
1018	LumB	2	0	0	0	22,230137	21,89589	1	1	1	1	0
1019	LumB	2	1	0	0	3,0931507	1,4520548	1	1	1	1	0
1020	LumA	2	1	0	0	4,8630137	4,8630137	0	2	1	1	0
1021	LumA	1	0	0	0	6,7972603	4,9671233	1	2	1	1	0
1022	LumB	2	1	0	0	3,9150685	1,4164384	1	1	1	0	0
1023	LumA	2	1	0	0	25,945205	25,945205	0	3	1	1	0
1024	De tipo basal	3	NA	1	1	2,4438356	1,7753425	1	1	0	0	0
1025	LumA	1	1	0	0	2,8767123	0,0027397	1	1	1	0	1
1026	LumA	1	NA	0	0	8,0821918	8,0821918	0	2	1	1	0
1027	LumB	1	1	0	0	25,778082	25,778082	0	3	1	1	0
1028	LumA	2	1	0	0	9,2520548	8,9753425	1	1	1	1	0

1029	De tipo basal	2	0	1	3,4410959	1,9726027	1	1	0	0	0	0
1030	Enriquecido en HER2	2	1	0	2,9232877	1,709589	1	1	1	1	1	1
1031	LumA	1	1	0	2,9616438	2,8958904	1	1	1	1	0	0
1032	LumA	2	0	0	4,509589	0,8465753	1	1	1	1	1	0
1033	LumB	1	NA	1	10,312329	9,9780822	1	1	NA	NA	NA	NA
1034	LumB	1	0	1	15,19726	15,19726	0	2	1	1	1	0
1035	De tipo basal	1	1	1	25,339726	25,339726	0	3	1	1	1	0
1036	LumA	2	1	0	3,4465753	1,460274	1	1	1	1	NA	0
1037	De tipo basal	1	0	0	11,958904	11,958904	0	2	0	0	0	0
1038	De tipo basal	2	1	1	2,4849315	2,2082 L 92	1	1	0	0	NA	0
1039	LumA	2	NA	0	8,539726	6,8136986	1	1	1	1	1	0
1040	De tipo basal	2	0	0	25,090411	25,090411	0	3	0	0	0	0
1041	LumA	2	1	0	3,7369863	1,7643836	1	1	NA	NA	NA	0
1042	De tipo basal	1	NA	0	2,1780822	0,9561644	1	1	1	1	1	1
1043	LumB	2	1	0	2,5452055	0,5232877	1	1	1	1	1	0

1044	LumA	1	1	0	2,630137	0,7315068	1	1	1	1	1	0
1045	De tipo basal	2	1	1	1,4109589	1,060274	1	1	0	0	0	0
1046	De tipo basal	2	1	1	24,835616	24,835616	0	3	0	0	0	0
1047	De tipo basal	2	0	1	14,873973	14,536986	1	1	0	0	0	0
1048	Enriquecido en HER2	1	1	1	2,3917808	1,5506849	1	1	1	1	1	0
1049	LumA	2	1	0	19,339726	19,339726	0	2	1	1	0	0
1050	LumB	2	1	0	13,605479	13,605479	0	2	1	1	0	NA
1051	LumB	2	1	0	2,4191781	1,4520548	1	1	1	1	1	0
1052	De tipo basal	2	1	0	12,073973	11,739726	1	1	NA	NA	NA	0
1053	De tipo basal	2	1	0	0,3506849	0,0027397	1	1	0	0	0	0
1054	De tipo basal	1	0	1	24,449315	24,449315	0	3	0	0	0	0
1055	LumB	2	0	0	6,1589041	4,1643836	1	1	1	1	0	0
1056	LumB	2	NA	0	1,2575342	0,0109589	1	1	1	1	1	0
1057	LumA	2	0	0	7,7753425	5,6630137	1	1	1	1	1	0
1058	De tipo basal	1	1	1	24,323288	24,323288	0	3	0	0	NA	NA

1059	LumB	2	1	0	5,9863014	5,0876712	1	1	1	1	1	0
1060	LumB	1	1	1	24,115068	24,115068	0	3	1	1	1	0
1061	De tipo basal	2	1	1	4,7972603	4,7972603	0	2	0	0	0	0
1062	De tipo basal	3	1	1	24,084932	24,084932	0	3	0	0	0	0
1063	De tipo basal	2	0	0	24,019178	24,019178	0	3	1	0	0	0
1064	De tipo basal	2	0	1	22,791781	22,791781	0	3	0	NA	NA	NA
1065	Enriquecido en HER2	2	0	0	9,5150685	9,5150685	0	NA	0	0	0	0
1066	De tipo basal	1	0	1	20,945205	20,945205	0	3	0	0	0	0
1067	De tipo normal	2	0	0	10,917808	10,917808	0	2	NA	NA	1	1
1068	Enriquecido en HER2	2	0	0	6,7013699	2,9726027	1	1	0	NA	0	0
1069	LumB	3	0	0	11,912329	11,912329	0	2	1	1	1	0
1070	LumB	2	0	0	20,731507	17,008219	1	3	1	1	1	0
1071	LumB	1	NA	0	4,0191781	1,2493151	1	2	1	1	1	0
1072	LumA	2	1	0	12,441096	12,441096	0	2	1	1	1	0
1073	Enriquecido en HER2	3	0	0	20,660274	20,660274	0	3	NA	NA	NA	1

1074	LumA	1	NA	0	4,7835616	4,4465753	1	1	1	0	0
1075	LumA	2	0	0	2,7534247	2,4246575	1	1	1	1	0
1076	LumA	2	1	0	20,539726	20,539726	0	3	1	NA	0
1077	LumA	1	1	0	20,408219	12,328767	1	3	1	1	0
1078	De tipo normal	2	NA	0	2,6630137	1,2493151	1	1	1	NA	0
1079	Enriquecido en HER2	1	1	0	NA	NA	NA	1	0	0	1
1080	De tipo basal	2	1	0	NA	NA	NA	3	1	1	0
1081	De tipo normal	2	1	0	NA	NA	NA	1	1	1	0
1082	Enriquecido en HER2	2	0	1	NA	NA	NA	1	1	1	1
1083	LumB	2	0	0	NA	NA	NA	2	1	1	NA
1084	LumA	1	0	0	NA	NA	NA	3	NA	NA	0
1085	De tipo basal	3	0	0	NA	NA	NA	1	1	0	0
1086	De tipo basal	1	0	1	19,969863	19,969863	0	3	0	0	0
1087	De tipo basal	1	0	0	19,657534	19,657534	0	3	0	0	0
1088	LumA	1	0	0	16,238356	3,9616438	1	1	1	1	0

1089	Enriquecido en HER2	2	0	1	19,506849	19,506849	0	3	1	1	1	1
1090	LumA	1	0	0	8,5945205	6,9041096	1	1	0	0	0	0
1091	LumA	1	0	0	19,432877	2,6082192	1	3	1	1	1	0
1092	De tipo basal	1	0	NA	19,405479	19,405479	0	3	0	0	NA	NA
1093	LumB	2	0	0	19,408219	16,756164	1	3	1	1	NA	0
1094	LumA	1	NA	0	19,358904	19,358904	0	3	1	1	1	0
1095	LumB	2	0	0	19,353425	19,353425	0	3	1	1	0	0
1096	LumA	2	0	0	13,345205	13,345205	0	2	1	1	1	0
1097	Enriquecido en HER2	2	0	0	8,3890411	5,290411	1	2	1	1	0	0
1098	Enriquecido en HER2	1	0	0	6,5863014	3,8767123	1	1	0	0	NA	1
1099	LumB	2	0	1	3,7780822	3,4438356	1	1	1	1	0	1
1100	Enriquecido en HER2	1	0	0	19,20274	19,20274	0	3	1	1	0	1
1101	De tipo basal	2	0	1	19,186301	19,186301	0	3	0	0	0	1
1102	De tipo basal	1	0	0	5,8410959	4,7589041	1	1	0	0	NA	NA
1103	LumB	1	0	0	17,734247	3,6876712	1	2	1	1	NA	NA

1104	LumA	2	0	1	9,7917808	9,7917808	0	2	1	NA	0
1105	De tipo basal	1	0	1	19,106849	19,106849	0	3	0	NA	0
1106	De tipo basal	2	0	1	19,09863	19,09863	0	3	0	NA	NA
1107	LumA	1	0	0	19,871233	19,871233	0	3	1	1	0
1108	De tipo basal	1	0	1	19,808219	19,808219	0	3	1	0	NA
1109	LumB	2	0	0	19,791781	1,6794521	1	3	1	1	0
1110	Enriquecido en HER2	2	0	0	16,778082	16,778082	0	2	1	0	0
1111	LumA	1	0	0	19,789041	19,789041	0	3	1	1	0
1112	LumB	2	0	0	3,5068493	3,3643836	1	1	1	1	0
1113	LumB	2	0	1	2,3150685	0,9369863	1	2	1	0	0
1114	De tipo basal	2	0	1	3,3232877	3,3232877	1	1	0	NA	0
1115	Enriquecido en HER2	2	0	1	19,986301	19,986301	0	3	1	1	0
1116	De tipo basal	2	0	1	NA	NA	NA	2	0	0	0
1117	Enriquecido en HER2	1	0	1	19,758904	19,758904	0	3	1	NA	1
1118	LumB	1	0	1	19,717808	19,717808	0	3	1	1	0

1119	LumB	1	0	0	0	4,4054795	2,460274	1	1	1	NA	0
1120	LumB	2	0	1	1	5,3342466	4,8794521	1	1	1	1	0
1121	LumA	1	0	0	0	11,531507	11,531507	0	2	1	1	0
1122	De tipo normal	1	0	0	0	15,424658	14,824658	1	2	NA	NA	NA
1123	LumB	2	0	0	0	3,8876712	3,8794521	1	1	1	1	NA
1124	De tipo basal	1	0	0	0	19,252055	19,252055	0	3	0	0	0
1125	De tipo normal	1	0	0	0	19,205479	19,205479	0	3	1	1	0
1126	LumB	1	0	0	0	20,005479	20,005479	0	3	1	1	0
1127	LumA	1	0	0	0	19,950685	19,950685	0	3	1	NA	0
1128	De tipo normal	2	1	0	0	19,931507	2,2657534	1	3	1	NA	0
1129	LumA	1	0	0	0	19,849315	4,509589	1	3	1	1	0
1130	De tipo basal	NA	0	1	1	8,0630137	8,0630137	0	NA	0	NA	NA
1131	De tipo basal	2	1	1	1	1,030137	0,0520548	1	1	0	0	1
1132	De tipo normal	1	0	0	0	19,608219	19,608219	0	3	1	1	0
1133	Enriquecido en HER2	1	1	1	1	19,550685	19,550685	0	3	1	1	0

1134	De tipo basal	2	1	0	2,0712329	1,230137	1	1	NA	NA	NA
1135	LumA	1	1	0	19,449315	19,449315	0	3	1	1	0
1136	LumB	1	1	0	5,0520548	4,4684932	1	1	1	1	0
1137	LumA	2	1	0	19,331507	17,657534	1	3	1	1	0
1138	LumA	1	1	0	19,331507	19,331507	0	3	1	1	0
1139	De tipo basal	1	0	1	19,046575	19,046575	0	3	NA	0	0
1140	Enriquecido en HER2	1	0	0	18,917808	18,917808	0	3	1	0	0
1141	LumB	3	NA	0	5,5205479	1,8410959	1	1	1	1	0
1142	LumA	1	0	0	18,649315	18,649315	0	NA	1	NA	0
1143	LumA	1	0	0	4,8876712	4,8876712	0	1	1	1	0
1144	De tipo basal	NA	1	0	2,9972603	2,9123288	1	1	NA	NA	0
1145	LumA	1	0	0	18,753425	18,753425	0	3	1	1	NA
1146	LumA	3	NA	0	3,1917808	0,0027397	1	1	1	1	0
1147	De tipo basal	1	0	1	10,79726	10,79726	0	2	1	0	0
1148	LumB	1	1	0	13,542466	4,6027397	1	2	1	1	0

1149	LumB	2	1	0	18,715068	18,715068	0	3	1	1	0
1150	Enriquecido en HER2	2	1	1	2,0931507	2,0931507	1	1	0	NA	1
1151	LumB	NA	NA	1	18,641096	18,641096	0	3	1	1	0
1152	LumA	1	0	0	18,621918	18,621918	0	3	1	1	0
1153	LumA	NA	NA	0	1,460274	1,460274	0	NA	1	NA	0
1154	LumA	2	0	0	7,3479452	7,3479452	0	2	1	1	0
1155	LumA	1	1	0	6,939726	6,939726	1	2	1	1	0
1156	De tipo normal	2	0	0	18,452055	18,452055	0	3	0	0	0
1157	Enriquecido en HER2	2	1	1	3,7671233	3,7671233	1	1	NA	0	NA
1158	LumA	1	1	0	8,060274	8,060274	1	1	1	NA	0
1159	Enriquecido en HER2	3	NA	1	0,9863014	0,9863014	1	3	1	NA	1
1160	Enriquecido en HER2	2	0	1	17,969863	17,969863	0	3	1	0	NA
1161	LumB	2	1	0	4,1452055	4,1452055	0	2	1	0	0
1162	LumA	1	1	0	17,909589	17,909589	0	3	1	1	0
1163	LumA	2	1	0	9,3972603	9,3972603	0	2	1	1	0

1164	LumA	2	0	0	0	7,8109589	7,8109589	0	2	1	1	0
1165	Enriquecido en HER2	2	1	0	0	NA	NA	NA	1	0	0	NA
1166	De tipo basal	1	0	1	1	17,378082	17,378082	0	3	0	0	0
1167	Enriquecido en HER2	NA	NA	0	0	17,071233	17,071233	0	3	1	0	0
1168	LumA	1	0	NA	NA	17,161644	17,161644	0	3	1	1	0
1169	De tipo basal	3	NA	1	1	10,742466	6,5315068	1	1	0	NA	0
1170	De tipo basal	1	NA	1	1	NA	NA	NA	1	0	0	0
1171	Enriquecido en HER2	1	1	1	1	4,8438356	2,7506849	1	2	1	1	1
1172	LumB	1	NA	0	0	7,3972603	7,3972603	0	2	1	1	NA
1173	LumB	1	0	1	1	16,934247	3,4219178	1	3	1	1	0
1174	LumA	1	0	0	0	16,90137	16,90137	0	3	1	1	0
1175	LumB	2	1	0	0	16,882192	16,882192	0	3	1	1	0
1176	Enriquecido en HER2	2	NA	1	1	9,1232877	9,1232877	0	2	1	1	1
1177	De tipo basal	1	1	1	1	2,0986301	0,6547945	1	1	0	0	0
1178	Enriquecido en HER2	2	0	1	1	2,1534247	1,7506849	1	1	0	0	0

1179	De tipo basal	1	NA	1	0,0493151	0,0027397	1	1	0	0	0	0
1180	LumB	1	0	0	8,0328767	4,7917808	1	1	1	0	0	0
1181	LumA	1	0	0	7,8383562	7,8383562	0	2	1	NA	0	0
1182	Enriquecido en HER2	3	1	0	16,706849	16,706849	0	3	0	0	0	0
1183	De tipo basal	2	1	0	3,3835616	1,0547945	1	1	1	1	NA	NA
1184	De tipo basal	2	0	1	16,547945	16,547945	0	3	0	0	0	1
1185	Enriquecido en HER2	2	0	0	16,520548	16,520548	0	3	1	1	1	1
1186	De tipo normal	2	1	0	1,7506849	1,7506849	0	2	1	NA	0	0
1187	Enriquecido en HER2	3	NA	0	1,7150685	0,0082192	1	1	NA	0	0	1
1188	De tipo basal	2	1	1	16,479452	1,8219178	1	3	0	0	0	0
1189	De tipo normal	2	0	0	11,153425	7,0520548	1	1	NA	NA	0	0
1190	LumA	2	0	0	16,287671	16,287671	0	3	1	1	0	0
1191	LumB	2	0	0	16,345205	16,345205	0	3	1	1	0	0
1192	LumB	1	0	0	16,227397	16,227397	0	3	1			
1193	LumA	NA	NA	0	6,2821918	6,2821918	0	2	1	1	0	0

1194	LumA	3	NA	0	4,2767123	1,2849315	1	1	1	1	0
1195	LumA	2	0	0	9,7479452	9,7479452	0	2	1	1	NA
1196	LumB	1	NA	0	7,0684932	7,0684932	0	2	1	0	0
1197	LumA	1	0	0	5,8027397	5,8027397	0	2	1	0	0
1198	LumA	2	1	0	1,0383562	1,0383562	0	2	1	NA	0
1199	LumB	2	0	0	7,8520548	7,5178082	1	1	1	1	0
1200	LumA	2	0	0	15,863014	15,863014	0	3	1	1	0
1201	De tipo normal	1	0	0	15,693151	15,693151	0	3	1	0	0
1202	LumB	1	NA	1	0,030137	0,030137	0	2	1	0	0
1203	Enriquecido en HER2	2	1	0	10,046575	10,046575	0	2	0	0	0
1204	LumA	2	0	0	15,210959	15,210959	0	3	1	1	0
1205	LumA	1	NA	0	15,189041	10,890411	1	3	1	1	0
1206	LumA	2	0	0	15,169863	15,169863	0	3	1	NA	NA
1207	LumA	1	0	NA	2,2794521	0,9726027	1	1	1	NA	0
1208	Enriquecido en HER2	1	0	1	8,0712329	6,0547945	1	1	0	0	1

1209	De tipo basal	2	1	0	4,1287671	3,8630137	1	1	1	0	0
1210	LumB	1	0	0	15,035616	15,035616	0	3	1	1	0
1211	LumB	3	NA	0	1,9780822	0,9123288	1	1	1	1	0
1212	De tipo basal	1	1	0	15,032877	15,032877	0	3	1	1	0
1213	Enriquecido en HER2	2	1	0	15	15	0	3	0	0	0
1214	De tipo basal	1	0	1	14,967123	14,967123	0	3	0	NA	0
1215	De tipo normal	1	1	0	12,073973	12,073973	0	2	1	0	NA
1216	LumA	2	0	1	8,8	8,7835616	1	2	1	1	0
1217	LumA	1	1	0	14,89863	14,375342	1	3	1	1	0
1218	LumA	2	1	0	7,6767123	6,4410959	1	1	1	NA	0
1219	LumA	3	1	0	4,1890411	1,6493151	1	1	1	1	0
1220	De tipo basal	3	1	0	2,5150685	0,4109589	1	1	0	0	0
1221	De tipo normal	1	0	1	14,852055	14,852055	0	3	0	NA	NA
1222	LumA	1	0	0	14,772603	14,772603	0	3	NA	NA	0
1223	De tipo basal	3	1	1	1,5589041	1,3260274	1	2	0	0	0

1224	De tipo normal	1	0	0	14,709589	14,709589	0	3	NA	NA	0
1225	De tipo basal	2	1	0	13,29589	14,69589	1	3	1	1	0
1226	LumA	3	NA	0	0,0410959	2,5232877	1	1	1	0	0
1227	LumA	1	0	0	14,578082	14,578082	0	3	1	0	0
1228	LumB	2	0	0	14,572603	14,572603	0	3	1	1	0
1229	Enriquecido en HER2	2	0	1	14,613699	14,613699	0	3	0	0	0
1230	LumA	2	0	0	2,0109589	2,0109589	0	2	1	1	0
1231	LumA	1	0	0	14,542466	14,542466	0	3	1	0	0
1232	Enriquecido en HER2	1	0	1	14,534247	14,534247	0	3	1	1	0
1233	LumB	1	NA	0	4,0849315	4,9123288	1	1	1	1	0
1234	LumB	NA	1	0	5,4876712	5,4876712	0	2	1	NA	0
1235	LumA	1	0	0	14,641096	14,641096	0	3	1	1	0
1236	LumA	2	0	0	14,520548	14,520548	0	3	1	1	0
1237	LumA	2	0	0	5,0109589	7,3780822	1	2	1	1	0
1238	Enriquecido en HER2	3	NA	1	NA	NA	NA	1	1	0	0

1239	Enriquecido en HER2	2	1	0	14,465753	14,465753	0	3	1	NA	NA
1240	De tipo basal	3	1	0	14,438356	14,438356	0	3	1		
1241	LumA	1	0	0	14,421918	14,421918	0	3	1	1	0
1242	LumA	2	0	0	14,419178	14,419178	0	3	1	NA	NA
1243	LumB	1	0	0	14,408219	14,408219	0	3	1	1	0
1244	LumB	1	1	0	10,013699	10,013699	0	2	1	1	0
1245	LumA	2	0	0	14,383562	14,383562	0	3	1	1	0
1246	LumB	2	0	0	4,5643836	4,5643836	0	2	1	1	0
1247	LumB	1	0	0	14,312329	1,4739726	1	3	1	1	0
1248	De tipo normal	1	1	0	14,249315	14,249315	0	3	NA	NA	0
1249	LumA	3	0	0	13,49863	13,49863	0	2	1	0	0
1250	De tipo basal	NA	0	1	14,235616	14,235616	0	3	0	NA	0
1251	LumB	1	0	0	14,945205	14,945205	0	3	1	0	0
1252	De tipo normal	1	0	0	1,6547945	1,2493151	1	1	NA	NA	0
1253	LumA	2	1	0	4,2164384	4,2164384	0	2	1	0	0

1254	De tipo normal	1	0	0	0	12,526027	12,526027	0	3	1	NA	0
1255	LumB	2	1	0	0	12,463014	12,463014	0	3	1	1	0
1256	De tipo basal	2	1	0	0	3,7205479	1,7452055	1	1	1	1	NA
1257	De tipo normal	1	0	0	0	12,427397	12,427397	0	3	1	NA	0
1258	LumA	2	0	0	0	12,372603	12,372603	0	3	1	NA	0
1259	LumA	1	1	0	0	12,328767	12,328767	0	3	1	NA	NA
1260	De tipo basal	1	1	0	0	12,29589	2,5945205	1	3	1	1	0
1261	LumA	2	0	0	0	12,312329	12,312329	0	3	1	NA	0
1262	De tipo normal	1	0	0	0	12,180822	12,180822	0	3	1	NA	0
1263	De tipo basal	1	0	0	0	12,2	12,2	0	3	NA	NA	NA
1264	Enriquecido en HER2	1	0	1	1	3,6	2,0684932	1	1	1	1	0
1265	LumA	2	1	0	0	12,2	12,2	0	3	1	1	0
1266	De tipo normal	1	1	NA	NA	11,857534	11,857534	0	3	1	NA	0
1267	Enriquecido en HER2	2	0	1	1	11,186301	11,186301	0	3	0	0	0
1268	De tipo normal	1	1	NA	NA	11,073973	11,073973	0	3	1	NA	0

1269	De tipo normal	3	0	0	0	10,969863	10,969863	0	3	1	1	0
1270	LumA	2	NA	0	0	10,920548	10,920548	0	3	1	1	0
1271	LumA	2	1	0	0	10,79726	10,79726	0	3	1	0	0
1272	LumB	1	0	0	0	10,668493	10,668493	0	3	1	0	0
1273	LumA	2	1	0	0	10,167123	10,167123	0	3	1	1	NA
1274	LumA	1	0	0	0	9,6821918	9,6821918	0	3	1	NA	0
1275	De tipo normal	2	0	0	0	9,5917808	9,5917808	0	3	1	1	0
1276	De tipo normal	1	0	0	0	9,6082192	9,6082192	0	3	1	1	0
1277	De tipo basal	2	0	1	1	9,5287671	9,5287671	0	3	0	NA	0
1278	De tipo normal	1	0	0	0	6,5835616	6,2547945	1	1	1	1	NA
1279	De tipo basal	2	1	0	0	9,3643836	9,3643836	0	3	NA	NA	0

*tamaño del tumor: T1≤2 cm, T2>2 cm-5 cm, T3>5 cm, NA=no evaluado
 ^estado de los ganglios: 0=negativo para ganglios, 1= positivo para ganglios, NA=estado de los ganglios desconocido
 % de grado histológico de Nottingham: 0=grados 1 y 2, 1=grado 3, NA=desconocido
 ***cualquier supervivencia libre de recidiva: 0=libre de recidiva, 1=recidiva
 **supervivencia específica de enfermedad: 1=muerte por cáncer de mama CA, 2=muerte por causa distinta a cáncer de mama, 3=con vida, NA=desconocida
 ^i)inmunohistoquímica:

Biomarcador	0	1	NA
ER	< 1% de núcleos positivos	≥ 1% de núcleos positivos	No interpretable
PR	< 1% de núcleos positivos	≥ 1% de núcleos positivos	No interpretable
Enriquecido en HER2	Expresión negativa o débil	Expresión fuerte	No interpretable

Tabla 4: Datos fuente para conjuntos de datos de qRT-PCR y micromatriz

Autor	Muestras	Plataforma	Registros de GEO (u otra disponibilidad)	Uso en clasificación de subtipo	Uso en predicción de riesgo	Número N-terapia sistémica sin adyuvante	Número de terapia endocrina sólo	% de ER+
Parker <i>et al</i>	189	qRT-PCR		-	-	0	0	54%
Parker <i>et al</i>	279	qRT-PCR		Prueba	Prueba	0	0	62%
Parker <i>et al</i>	544	Agilent Custom, 1A, 1Av2	GSE10886	Común a qRT-PCR para entrenamiento (189); otros en Prueba (355)	355 en Prueba	31	27	56%
Hess <i>et al</i>	133	Affymetrix U133A	bioinformatics.mds.rockefeller.edu/pubdata	Prueba	Prueba	0	0	62%
Ivshina <i>et al</i>	289	Affymetrix U133A	GSE4922	Prueba	Prueba	142	66	86%
Loi <i>et al</i>	414	Affymetrix U133A & U133+2	GSE6532	Prueba	Prueba	137	277	89%
van de Vijver <i>et al</i>	295	Agilent	GSE2845	Prueba	Sin tratar para entrenamiento (165); otros en prueba	165	20	76%
Wang <i>et al</i>	286	Affymetrix U133A	GSE2034	Prueba	Prueba	286	0	73%

Tabla 5. Análisis de múltiples variables y de una variable usando 1451 muestras de un conjunto de prueba de micromatriz combinado con datos clínicos

Variable	Una variable		Múltiples variables* (subtipo)		Múltiples variables* (clínico)		Múltiples variables*†‡ (subtipo + clínico)	
	Coeficiente	Valor de p	Coeficiente	Valor de p	Coeficiente	Valor de p	Coeficiente	Valor de p
De tipo basal	0,14	0,25	0,12	5,10E-01	-		-0,11	5,50E-01
Enriquecido en HER	0,62	1,00E-08	0,53	1,60E-03	-		0,35	4,00E-02
LumA	-0,94	1,00E-22	-0,67	6,20E-05	-		-0,64	1,60E-04
LumB	0,42	5,60E-06	0,3	5,50E-02	-		0,24	1,30E-01
Estado de ER	-0,47	1,80E-06	-	-	-0,5	5,50E-07	-0,37	3,00E-03
Tamaño del tumor	0,62	3,50E-12	-	-	0,54	6,10E-09	0,47	5,30E-07
Estado de los ganglios	0,37	2,80E-05	-	-	0,24	1,10E-02	0,19	5,00E-02
*Clase de tipo normal usada como estado de referencia ^Las variables significativas están en cursiva † p = 4e-10 (mediante la prueba de razón de verosimilitud) para comparación con el modelo de subtipo ‡ = 2e-13 (mediante la prueba de razón de verosimilitud) para comparación con el modelo clínico								

5

Tabla 6. Distribución de subtipos intrínsecos mediante estado de ER

Conjunto de prueba	Estado de ER	N.º de muestras	% de LumA	% de LumB	% de enriquecido en HER2	% de tipo basal	% de tipo normal
UNC	Positivo para ER	137	44%	35%	7%	4%	9%
	Negativo para ER	107	7%	5%	19%	51%	18%
Hess <i>et al</i>	Positivo para ER	82	44%	32%	10%	1%	13%
	Negativo para ER	51	2%	2%	41%	51%	4%
Ivshina <i>et al</i>	Positivo para ER	211	42%	29%	11%	8%	9%
	Negativo para ER	34	9%	15%	35%	38%	3%
Loi <i>et al</i>	Positivo para ER	349	39%	38%	8%	7%	8%
	Negativo para ER	45	18%	9%	33%	27%	13%
van de Vijver <i>et al</i>	Positivo para ER	225	39%	31%	14%	4%	12%

	Negativo para ER	70	1%	0%	31%	64%	3%
Wang <i>et al</i>	Positivo para ER	209	35%	33%	11%	8%	13%
	Negativo para ER	77	5%	3%	29%	57%	6%

Tabla 7. Tasas de respuesta completa patológica a T/FAC para subtipos basadas en PAM50 y clasificación triple negativa

5

Clasificación	RD	pCR
De tipo basal	11 (41%)	16 (59%)
Enriquecido en HER2	17 (59%)	12 (41%)
LumA	36 (100%)	0 (0%)
LumB	22 (82%)	5 (18%)
De tipo normal	13 (93%)	1 (7%)
Triple negativo	13 (50%)	13 (50%)
Cualquier positivo	82 (80%)	20 (20%)
Triple negativo/basal	6 (35%)	11 (65%)
Triple negativo/no basal	7 (78%)	2 (22%)
No triple negativo/basal	4 (50%)	4 (50%)
No triple negativo/no basal	78 (83%)	16 (17%)
*Los porcentajes se calculan en total por clasificación		

Bibliografía:

1. Sorlie T, Perou CM, Tibshirani R, *et al*: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 98:10869-74, 2001
2. van't Veer LJ, Dai H, van de Vijver MJ, *et al*: Gene expression profiling predicts clinical outcome of breast cancer. Nature 415:530-6, 2002
3. van't Veer LJ, Paik S, Hayes DF: Gene expression profiling of breast cancer: a new tumor marker. J Clin Oncol 23:1631-5, 2005
4. Paik S, Shak S, Tang G, *et al*: A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 351:2817-26, 2004
5. Paik S, Tang G, Shak S, *et al*: Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J Clin Oncol 24:3726-34, 2006
6. Hu Z, Fan C, Oh DS, *et al*: The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 7:96, 2006
7. Loi S, Haibe-Kains B, Desmedt C, *et al*: Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. J Clin Oncol 25:1239-46, 2007
8. Perou CM, Sorlie T, Eisen MB, *et al*: Molecular portraits of human breast tumours. Nature 406:747-52, 2000
9. Sorlie T, Tibshirani R, Parker J, *et al*: Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100:8418-23, 2003
10. Fan C, Oh DS, Wessels L, *et al*: Concordance among gene-expression based predictors for breast cancer. N Engl J Med 355:560-9, 2006
11. Perreard L, Fan C, Quackenbush JF, *et al*: Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay. Breast Cancer Res 8:R23, 2006
12. Eisen MB, Spellman PT, Brown PO, *et al*: Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95:14863-8, 1998
13. Yufeng L, Hayes DL, Nobel A, *et al*: Statistical significance of clustering for high dimension low sample size data.

Journal of the American Statistical Association, en prensa

14. Mullins M, Perreard L, Quackenbush JF, *et al*: Agreement in breast cancer classification between microarray and quantitative reverse transcription PCR from fresh-frozen and formalin-fixed, paraffin-embedded tissues. Clin Chem 53:1273-9, 2007
15. Storey JD, Tibshirani R: Statistical methods for identifying differentially expressed genes in DNA microarrays. Methods Mol Biol 224:149-57, 2003
16. Dudoit S, Fridlyand J: A prediction-based resampling method for estimating the number of clusters in a dataset. Genome Biol 3:RESEARCH0036, 2000
17. Tibshirani R, Hastie T, Narasimhan B, *et al*: Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 99:6567-72, 2002
18. Dabney AR: Classification of microarrays to nearest centroids. Bioinformatics 21:4148-54, 2005
19. Hess KR, Anderson K, Symmans WF, *et al*: Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. J Clin Oncol 24:4236-44, 2006
20. van de Vijver MJ, He YD, van't Veer LJ, *et al*: A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347:1999-2009, 2002
21. Chaudhuri P, Marron JS: SiZer for Exploration of Structures in Curves. Journal of the American Statistical Association 94:807-823, 1999
22. Sheather SJ, Jones MC: A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. Journal of the Royal Statistical Society 53:683-690, 1991
23. Neve RM, Chin K, Fridlyand J, *et al*: A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 10:515-27, 2006
24. Rouzier R, Pusztai L, Delaloge S, *et al*: Nomograms to predict pathologic complete response and metastasis-free survival after preoperative chemotherapy for breast cancer. J Clin Oncol 23:8331-9, 2005
25. Herschkowitz JI, Simin K, Weigman VJ, *et al*: Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol 8:R76, 2007
26. Rakha E, Ellis I, Reis-Filho J: Are triple-negative and basal-like breast cancer synonymous? Clin Cancer Res 14:618; author reply 618-9, 2008
27. Cheang MC, Voduc D, Bajdik C, *et al*: Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. Clin Cancer Res 14:1368-76, 2008

Ejemplo 2.

Introducción y datos de antecedentes

- Esta tecnología también cubre el uso del clasificador de subtipo intrínseco basado en PAM50 como distintivo predictivo y de pronóstico en la práctica de terapia endocrina neoadyuvante. Las pacientes postmenopáusicas con cáncer de mama en estadio 2 y 3, positivo para ER y/o PgR pueden tratarse con un agente endocrino, normalmente un inhibidor de aromatasa o tamoxifeno, antes de la cirugía para mejorar los desenlaces clínicos, es decir para promover el uso de cirugía conservadora de mama o para mejorar la operabilidad en el entorno de un tumor que ha invadido los tejidos circundantes a la mama. Una prueba predictiva para aumentar la confianza con que una paciente individual responderá a terapia endocrina neoadyuvante es un avance significativo.

Sumario

- La puntuación de riesgo ponderado por proliferación y subtipo intrínseco basado en PAM50, cuando se aplica a muestras de cánceres de mama ER+ recogidas tras iniciar el tratamiento con un agente endocrino, pueden usarse para predecir la respuesta a la terapia endocrina neoadyuvante y determinar el pronóstico para pacientes con cáncer de mama ER+ que se someterán a terapia a largo plazo con un agente endocrino. Se aplicó un modelo de expresión génica de pronóstico entrenado con muestras tumorales tomadas antes del tratamiento (puntuación de riesgo ponderada por proliferación de PAM50, descrito en otra parte en el presente documento) a muestras tomadas tras el inicio de la terapia endocrina neoadyuvante. Este enfoque es único porque los estudios previos sobre la interacción

de perfiles de expresión génica y pronóstico sólo han examinado muestras previas al tratamiento y nunca han aplicado estos modelos a muestras tras el tratamiento. Las propiedades de pronóstico y predictivas del modelo de pronóstico ponderado por proliferación y subtipo intrínseco de PAM50 en muestras iniciales se comparan con los mismos modelos aplicados a muestras tomadas un mes tras el inicio de la terapia endocrina neoadyuvante. La aplicación del modelo de riesgo de recidiva ponderado por proliferación y subtipo intrínseco de PAM50 a las muestras de un mes de tratamiento identifica de manera precisa tumores agresivos que no responden a tratamiento endocrino neoadyuvante o adyuvante. Las pacientes con estos tumores deben derivarse inmediatamente a tratamientos neoadyuvantes alternativos, tales como quimioterapia, porque estos tumores malos se comportan como enfermedad agresiva que no responde a la terapia endocrina. Se estableció un alto grado de correlación entre el marcador de proliferación Ki67 y la puntuación de riesgo de PAM50 ponderada por proliferación que apoya la afirmación de que la puntuación de riesgo ponderada por proliferación de PAM50 tiene propiedades de pronóstico. Sin embargo, estas propiedades de pronóstico están notablemente potenciadas cuando se aplica el análisis a muestras recogidas de tumores que se han expuesto a un agente endocrino. En la práctica, esto puede lograrse fácilmente recetando un agente endocrino durante algunas semanas antes de la cirugía definitiva o volviendo a tomar muestras de un tumor de manera temprana en el transcurso de un tratamiento endocrino neoadyuvante con el fin de identificar tumores que no responden.

Metodología: La evidencia para apoyar estas afirmaciones surge de un ensayo de fase 2 patrocinado por el National Cancer Institute de terapia neoadyuvante con el inhibidor de aromataza letrozol (subvención del NCI n.º R01 CA095614). La elegibilidad para el ensayo requirió mujeres postmenopáusicas con cánceres de mama en estadio 2 y 3 positivos para ER y/o PgR. Las pacientes recibieron 4 meses de terapia y luego se sometieron a cirugía. Se obtuvieron muestras de tumor congeladas en el nivel inicial, al mes y en la cirugía. Se analizaron las muestras mediante corte congelado y se extrajo el ARN usando metodologías convencionales de muestras ricas en tumor y se sometió a análisis de expresión génica usando las matrices 1x44K de Agilent. Se normalizaron los datos al conjunto de datos usado para entrenar el clasificador PAM50 (métodos descritos anteriormente) y se produjeron dos lecturas: una clasificación intrínseca (LumA, LumB, enriquecido en HER2, de tipo basal y de tipo normal) y una puntuación de riesgo de PAM50 ponderada por proliferación. El objetivo de este estudio fue correlacionar los desenlaces de terapia endocrina neoadyuvante con la puntuación de riesgo ponderada por proliferación y clasificación intrínseca derivada tanto de la muestra de nivel inicial como de la muestra durante el tratamiento tomada al mes.

Resultados:

La puntuación de riesgo ponderada por proliferación y subtipo intrínseco de PAM50 mostró cambios marcados al mes tras la terapia (tabla 8). La mayoría de las transiciones se produjeron en el grupo de LumB, cambiando la mayoría a LumA, pero el 16% permaneció en la categoría de LumB pese al tratamiento. Por el contrario, la mayoría de los tumores LumA permanecieron LumA tras la terapia. Estas transiciones se debieron a la supresión del agrupamiento de proliferación en el grupo de LumB puesto que la puntuación de riesgo ponderada por proliferación de PAM50 mostró cambios similares, convirtiéndose la mayoría de los tumores de tipo de alto riesgo (68%) en de riesgo intermedio o bajo en las muestras durante el tratamiento. La estrecha correlación con inmunohistoquímica de Ki67 destaca adicionalmente esta conclusión. La correlación entre los valores de Ki67 de nivel inicial y la puntuación de riesgo ponderada por proliferación de PAM50 fue alta ($P=2,8 \times 10^{-8}$). De manera similar, los valores de Ki67 al mes y la puntuación de proliferación de PAM50 al mes también estuvieron estrechamente correlacionados ($P=3,8 \times 10^{-10}$). Sin embargo, aunque el subtipo de puntuación de riesgo ponderada por proliferación de PAM50 de nivel inicial sólo mostró una correlación muy débil con los valores de Ki67 de fin del estudio ($P=0,04$), hubo una estrecha correlación entre la puntuación de riesgo ponderada por proliferación de PAM50 al mes y los valores de Ki67 de fin del estudio (la mayoría de los cuales se obtuvieron en la cirugía de 4 a 6 meses más tarde ($P=6,8 \times 10^{-11}$)). Esta última observación apoya fuertemente la afirmación de que puede usarse una prueba basada en PAM50 en tratamiento temprano para predecir si las muestras quirúrgicas finales tendrán características de biomarcador favorables, tales como una baja tasa de proliferación.

Para determinar las correlaciones clínicas asociadas con estos cambios inducidos por terapia endocrina en la puntuación de riesgo y subtipo intrínseco de cáncer de mama, se examinaron cuatro criterios de valoración: respuesta clínica (criterios RECIST), tamaño T patológico (T1 frente a más alto (como evidencia para el descenso del estadio TNM patológico con el tratamiento), valores de Ki67 dicotomizados (se considera que los tumores que muestran un valor logarítmico natural de Ki67 de 1 o menos muestran un perfil favorable) y acontecimientos de recidiva. La puntuación de riesgo o subtipo de nivel inicial no mostró capacidad convincente para predecir ninguno de estos criterios de valoración lo que, en lo que se refiere a la recidiva, es probablemente una función del pequeño tamaño de muestra en este ensayo (tabla 9). Por el contrario, y pese al pequeño tamaño de muestra, el subtipo intrínseco de PAM50 al mes (tabla 10) mostró relaciones estadísticamente significativas con respuesta clínica ($P=0,01$), valor de Ki67 favorable al fin del tratamiento ($P=0,0003$) y recidiva ($P=0,009$). Estas fuertes relaciones estuvieron conducidas por el desenlace extremadamente malo asociado con los tumores que se designaban o bien “no luminales” o bien luminales B en las muestras en tratamiento. La puntuación de riesgo ponderada por proliferación de PAM50 tenía propiedades similares. La puntuación de riesgos ponderada por proliferación de PAM50 de nivel inicial no predijo los desenlaces con neoadyuvante o a largo plazo de manera muy eficaz (tabla 11). Sin embargo, los tumores que se designaron de alto riesgo al mes mostraron correlaciones significativas con malos desenlaces en los cuatro criterios de valoración examinados, es decir mala respuesta clínica ($P=0,02$), descenso del

estadio TNM patológico bajo ($p=0,02$), valor de Ki67 desfavorable al fin del tratamiento ($P=0,0001$) y recidiva ($p=0,001$) (tabla 12).

5 Por tanto, puede usarse la aplicación de la puntuación de riesgo y subtipo intrínseco basados en PAM50 para muestras de tumor recogidas de cánceres de mama ER+ primarios que se someten a tratamiento prequirúrgico con un agente para los siguientes fines:

1) Predicción de ausencia de respuesta a la terapia endocrina neoadyuvante

10 2) Determinación del pronóstico para pacientes con cáncer de mama ER+ que se someten posteriormente a tratamiento endocrino adyuvante.

Tabla 8. Grupo de riesgo ponderado por proliferación y subtipo de PAM50 que cambia al mes tras el tratamiento.

Categoría de cambio	Número	Porcentaje
Cambios de subtipo intrínseco de PM50		
De LumA a LumA	18	31,0
De LumA a LumB	1	1,7
De LumA a no Lum	0	0
De LumB a LumA	29	50,0
De LumB a LumB	6	10,3
De LumB a no Lum	1	1,7
De no Lum a no Lum	1	1,7
De no Lum a LumA	0	0
De no Lum a LumB	2	3,4
Total	58	100
Puntuación de riesgo de PAM50 ponderada por proliferación		
De baja a baja	5	8,6
De baja a media	1	1,7
De baja a alta	0	0
De media a baja	7	12,1
De media a media	12	20,7
De media a alta	1	1,7
De alta a baja	11	19
De alta a media	14	24,1
De alta a alta	7	12,1
Total	58	100

Tabla 9. Interacciones entre designaciones de subtipo intrínseco de PAM50 *de nivel inicial* y desenlaces de terapia endocrina neoadyuvante.

Subtipo o puntuación en el nivel inicial	Criterio de valoración de fin del estudio	Número/total	% de desenlaces favorables	Valor de P en la interacción
Subtipo	Respuesta clínica CR+PR frente a SD+PD			0,54
LumA		28/76	60,71	
LumB		42/76	69,05	

No Lum†		6/76	50,00	
	Tamaño del tumor patológico* ≤2 cm frente a >2 cm			0,29
LumA		29/78	37,79	
LumB		43/78	48,84	
No Lum†		6/78	16,67	
	Ki67 en log normal# ≤ log 1,0 frente a >1,0			0,03
LumA		30/29	66,67	
LumB		43/79	37,21	
No Lum†		6/79	33,33	
	Recidiva Sí frente a No			0,262
LumA		30/78	90,00	
LumB		42/78	90,4762	
No Lum†		6/78	66,67	
* Puesto que todas las pacientes tenían enfermedad en estadio clínico 2 ó 3, el estadio patológico del tumor uno y cirugía se consideraron como evidencia de descenso del estadio TNM satisfactorio. Se supone que los tumores que evolucionaron durante la terapia y se sometieron a quimioterapia neoadyuvante tienen un tamaño T patológico mayor de 2 cm al final del estudio. # Ki67 de fin del estudio se define como o bien la muestra quirúrgica o bien el valor al mes si la paciente evolucionó con terapia endocrina neoadyuvante y se sometió a quimioterapia o no se sometió a cirugía. †No luminal se refiere a muestras designadas de tipo basal o enriquecidas en HER2. No se incluye de tipo normal en este análisis porque se supone que estas muestras no contienen suficientes células tumorales para la subtipificación adecuada.				

Tabla 10. Interacciones entre designaciones de subtipo intrínseco de PAM50 *al mes con tratamiento* y desenlaces de terapia endocrina neoadyuvante.

5

Subtipo de PAM50 al mes	Criterio de valoración de fin del estudio	Número/total	% de desenlaces favorables	Valor de P en la interacción
Subtipo	Respuesta clínica CR+PR frente a SD + PD			0,01
LumA		45/56	75,56	
LumB		9/56	44,44	
No Lum†		2/56	0	
	Tamaño del tumor patológico* ≤2 cm frente a >2 cm			0,41
LumA		46/57	47,83	
LumB		9/57*	22,22	

No Lum†		2/57	50,00	
	Ki67 en log normal# ≤ log 1,0 frente a >1,0			0,0003
LumA		47/58	61,70	
LumB		9/58	0	
No Lum†		2/58	0	
	Recidiva Sí frente a No			0,009
LumA		45/53	93,62	
LumB		7/53	57,14	
No Lum†		2/53	50,00	
* Puesto que todas las pacientes tenían enfermedad en estadio clínico 2 ó 3, el estadio patológico del tumor uno y cirugía se consideraron como evidencia de descenso del estadio TNM satisfactorio. Se supone que los tumores que evolucionaron durante la terapia y se sometieron a quimioterapia neoadyuvante tienen un tamaño T patológico mayor de 2 cm al final del estudio. # Ki67 de fin del estudio se define como o bien la muestra quirúrgica o bien el valor al mes si la paciente evolucionó con terapia endocrina neoadyuvante y se sometió a quimioterapia o no se sometió a cirugía. †No luminal se refiere a muestras designadas de tipo basal o enriquecidas en HER2. No se incluye de tipo normal en este análisis porque se supone que estas muestras no contienen suficientes células tumorales para la subtipificación adecuada.				

Tabla 11. Interacciones entre designaciones de puntuación de riesgo ponderado por proliferación de PAM50 de nivel inicial y desenlaces de terapia endocrina neoadyuvante

5

Puntuación de riesgo, con proliferación en nivel inicial	Criterio de valoración de fin del estudio	Número/total	% de desenlaces favorables	Valor de P en la interacción †
	Respuesta clínica CR+PR frente a SD + PD			0,4573
Baja		9/76	44,44	
Media		28/76	67,79	
Alta		39/76	66,67	
	Tamaño del tumor patológico* ≤2 cm frente a >2 cm			1,0
Baja		9/78	44,44	
Media		29/78	41,38	
Alta		37/78	42,50	
	Ki67 en log normal# ≤ log 1,0 frente a >1,0			0,03431
Baja		9/79	77,78	
Media		30/79	56,67	

Alta		40/79	35,00	
	Recidiva Sí frente a No			0,1191
Baja		9/74	77,78	
Media		29/74	96,67	
Alta		36/74	84,62	
* Puesto que todas las pacientes tenían enfermedad en estadio clínico 2 ó 3, el estadio patológico del tumor uno y cirugía se consideraron como evidencia de descenso del estadio TNM satisfactorio. Se supone que los tumores que evolucionaron durante la terapia y se sometieron a quimioterapia neoadyuvante tienen un tamaño T patológico mayor de 2 cm al final del estudio. # Ki67 de fin del estudio se define como o bien la muestra quirúrgica o bien el valor al mes si la paciente evolucionó con terapia endocrina neoadyuvante y se sometió a quimioterapia o no se sometió a cirugía. †No luminal se refiere a muestras designadas de tipo basal o enriquecidas en HER2. No se incluye de tipo normal en este análisis porque se supone que estas muestras no contienen suficientes células tumorales para la subtipificación adecuada.				

Tabla 12. Interacciones entre puntuación de riesgo ponderada por proliferación de PAM50 al mes con tratamiento y desenlaces de terapia endocrina neoadyuvante

5

Puntuación de riesgo ponderada por proliferación de PAM50 al mes	Criterio de valoración de fin del estudio	Número/total	% de desenlaces favorables	Valor de P en la interacción
	Respuesta clínica CR+PR frente a SD + PD			0,02
Baja		21/56	80,95	
Media		27/56	70,37	
Alta		8/56	25,00	
	Tamaño del tumor patológico* ≤2 cm frente a >2 cm			0,02
Baja		23/57	47,83	
Media		26/57	53,85	
Alta		8/57	0	
	Ki67 en log normal# ≤ log 1,0 frente a >1,0			0,0001
Baja		23/58	78,26	
Media		27/58	40,74	
Alta		8/58	0	
	Recidiva Sí frente a No			0,001
Baja		23/56	95,65	
Media		27/56	92,59	

Alta		6/56	33,33	
* Puesto que todas las pacientes tenían enfermedad en estadio clínico 2 ó 3, el estadio patológico del tumor uno y cirugía se consideraron como evidencia de descenso del estadio TNM satisfactorio. Se supone que los tumores que evolucionaron durante la terapia y se sometieron a quimioterapia neoadyuvante tienen un tamaño T patológico mayor de 2 cm al final del estudio. # Ki67 de fin del estudio se define como o bien la muestra quirúrgica o bien el valor al mes si la paciente evolucionó con terapia endocrina neoadyuvante y se sometió a quimioterapia o no se sometió a cirugía. †No luminal se refiere a muestras designadas de tipo basal o enriquecidas en HER2. No se incluye de tipo normal en este análisis porque se supone que estas muestras no contienen suficientes células tumorales para la subtipificación adecuada.				

Ejemplo 3.

Se realizó un análisis de riesgo de recidiva en las muestras descritas en el ejemplo 1, excepto en que se eliminó la clase de tipo normal del modelo. Se representó la clase de tipo normal usando “valores normales” verdaderos a partir de mamoplastia de reducción o tejido no implicado a simple vista. Por tanto, esta clase se ha eliminado de todos los análisis de desenlace y esta clasificación se considera una medida de control de calidad. Los métodos no descritos a continuación son idénticos a los métodos descritos en el ejemplo 1.

Métodos

Modelos pronósticos y predictivos usando datos de subtipo molecular y clínico:

Se usaron análisis de una variable y de múltiples variables para determinar la significación de subtipos intrínsecos (LumA, LumB, enriquecido en HER2 y de tipo basal) en pacientes no tratadas y en pacientes que recibieron quimioterapia neoadyuvante. Para el pronóstico, se compararon los subtipos con variables clínicas convencionales (T, N, estado de ER y grado histológico), con tiempo hasta la recidiva (es decir, cualquier acontecimiento) como criterio de valoración. Se compararon los subtipos con el grado y los marcadores moleculares (ER, receptor de progesterona (PR), HER2) para la predicción en la práctica de neoadyuvante puesto que no es aplicable la estadificación patológica. Se realizaron pruebas de razón de verosimilitud para comparar modelos de datos clínicos, datos de subtipo y variables moleculares y clínicas combinadas disponibles. Se realizaron análisis de supervivencia categóricos usando una prueba de rangos logarítmicos y se visualizó con diagramas de Kaplan-Meier.

Modelos de riesgo de desarrollo con datos clínicos y moleculares

Se entrenó el modelo de riesgo de subtipo con un modelo de Cox de múltiples variables usando ajuste de regresión de Ridge para el subconjunto no tratado, negativo para ganglios de la cohorte de van de Vijver *et al.* (2002). Se asignó una puntuación de ROR a cada caso de prueba usando correlación con el subtipo solo (1) (ROR-S) o usando correlación de subtipo junto con tamaño del tumor (2) (ROR-C):

$$(1) \text{ ROR-S} = 0,05 \cdot \text{Basal} + 0,12 \cdot \text{Her2} + -0,34 \cdot \text{LumA} + 0,023 \cdot \text{LumB}$$

$$(2) \text{ ROR-C} = 0,05 \cdot \text{Basal} + 0,11 \cdot \text{Her2} + -0,23 \cdot \text{LumA} + 0,09 \cdot \text{LumB} + 0,17 \cdot \text{T}$$

La suma de los coeficientes para el modelo de Cox es la puntuación de ROR para cada paciente. Para clasificar las muestras en grupos de riesgo específicos, se eligieron umbrales del conjunto de entrenamiento tal como se describe en el ejemplo 1. Se realizó un análisis mediante SiZer para caracterizar la relación entre la puntuación de ROR y la supervivencia libre de recidiva. Los IC del 95% para la puntuación de ROR son versiones locales de IC binomiales, calculándose el tamaño de muestra local a partir de un estimador de densidad tipo núcleo gaussiano, basado en la elección de Sheather-Jones de anchura de ventana.

Comparación de modelos de predicción de recidiva

Se compararon cuatro modelos para predicción de recidiva: (1) un modelo de variables clínicas solas (tamaño del tumor, grado y estado de ER), (2) ROR-S, (3) ROR-C, y (4) un modelo que combina subtipo, tamaño del tumor y grado. Se eligió el índice C para comparar la fuerza de los diversos modelos. Para cada modelo, se estimó el índice C de 100 aleatorizaciones de la cohorte no tratada a dos tercios de conjunto de entrenamiento y un tercio de conjunto de prueba. Se calculó el índice C para cada conjunto de prueba para formar la estimación de cada modelo y se compararon las estimaciones del índice C en todos los modelos usando la prueba de la *t* de dos muestras.

Resultados

Modelos de riesgo de recidiva para pronóstico en cáncer de mama negativo para ganglios

Se sometieron a prueba modelos de Cox usando el subtipo intrínseco solo y junto con variables clínicas. La tabla 13 muestra los análisis de múltiples variables de estos modelos en una cohorte independiente de pacientes no tratadas (véase el ejemplo 1). En el modelo A, se encontró que los subtipos, el tamaño del tumor (T1 o mayor) y el grado histológico eran factores significativos para ROR. Se encontró que la gran mayoría de los tumores de tipo basal (95,9%) eran de grado medio o alto, y por tanto, en el modelo B, que es un análisis sin grado, el tipo basal se convierte en significativo. El modelo C muestra la significación de los subtipos en la población negativa para ganglios. Todos los modelos que incluyeron el subtipo y las variables clínicas fueron significativamente mejores que cualquiera de las variables clínicas solas ($P < 0,0001$) o el subtipo solo ($P < 0,0001$). Se entrenó un clasificador de recidiva para predecir desenlaces dentro del contexto de los subtipos intrínsecos y variables clínicas. Se seleccionó una cohorte de tratamiento no sistémico, negativa para ganglios ($n=141$) del conjunto de datos de micromatriz de Vijver *et al.* (2002) para entrenar el modelo de ROR y para seleccionar puntos de corte. Hubo una clara mejoría en la reducción con subtipo (ROR-S) en relación con el modelo de variables clínicas disponibles sólo (véase Parker *et al.* (2009) *J Clin Oncol* 27(8):1160-1167). Una combinación de variables clínicas y subtipo (ROR-C) también es una mejora significativa con respecto a cualquier factor pronóstico individual. Sin embargo, la información sobre el grado no mejoró significativamente el índice C en el modelo combinado, lo que indica que el valor pronóstico del grado se ha sustituido por la información proporcionada por el modelo de subtipo intrínseco. Cuando se usa ROR-C para ROR en un conjunto de prueba pronóstico de pacientes no tratadas negativas para ganglios, sólo el grupo de LumA contenía pacientes con bajo riesgo y la distinción de tres clases de riesgo bajo, medio y alto fue pronóstico. Además, las puntuaciones de RORC tienen una relación lineal con probabilidad de recidiva a los 5 años.

Tabla 13. Modelos de supervivencia libre de recidiva (no tratada)

Variable	Modelo A		Modelo B		Modelo C	
	Razón de riesgo	p	Razón de riesgo	p	Razón de riesgo	p
De tipo basal*	1,33	0,33	1,79	0,3	1,58	0,066
Enriquecido en HER*	2,53	0,00012	3,25	<0,0001	2,9	<0,0001
LumB*	2,43	<0,0001	2,88	<0,0001	2,54	<0,0001
Estado de ER†	0,83	0,38	0,83	0,34	0,83	0,32
Tamaño del tumor‡	1,36	0,034	1,43	0,012	1,57	0,001
Estado de los ganglios§	1,75	0,035	1,72	0,041	-	-
Grado histológico	1,4	0,0042	-	-	-	-
Completo frente a subtipo¶		<0,0001		<0,0001		<0,0001
Completo frente a clínico#		<0,0001		<0,0001		<0,0001
*Clase luminal A usada como estado de referencia en análisis de múltiples variables. †Razones de riesgo para ER usando marcador positivo en el numerador. ‡Tamaño ≤ 2 cm frente a > 2 cm. §Cualquier ganglio positivo. Grado codificado como una variable ordinal con tres niveles. ¶Los valores de P significativos indican predicción mejorada en relación con subtipo solo. #Los valores de P significativos indican predicción mejorada en relación con datos clínicos solos.						

Subtipos y predicción de respuesta a tratamiento neoadyuvante con TIFAC

El estudio de Hess *et al.* (2006) que realizó micromatriz sobre tumores de pacientes tratadas con T/FAC permitió la investigación de la relación entre los subtipos y marcadores clínicos y cómo se relaciona cada uno con pCR>. La tabla 14 muestra los análisis de múltiples variables de los subtipos junto con marcadores moleculares clínicos (ER, PR, HER2) y o bien con (modelo A) o bien sin (modelo B) grado histológico. Las únicas variables significativas en el contexto de este estudio fueron los subtipos intrínsecos. Se encontró un 94% de sensibilidad y un 97% de valor predictivo negativo para identificar pacientes que no responden a quimioterapia cuando se usa el modelo de ROR-S para predecir pCR. La relación entre puntuaciones de riesgo alto y una probabilidad más alta de pCR es compatible con la conclusión de que los tumores positivos para ER indolentes (LumA) responden menos a quimioterapia. Sin embargo, a diferencia de ROR para pronóstico, parece alcanzarse una meseta para el ROR frente a probabilidad la

de pCR, lo que confirma la presencia de resistencia a quimioterapia significativa entre los tumores de riesgo más alto.

Tabla 14. Modelos de respuesta neoadyuvante

Variable	Modelo A		Modelo B		Modelo C	
	Razón de probabilidades	P	Razón de probabilidades	P	Razón de probabilidades	P
De tipo basal*	1,33	0,3	1,79	0,3	1,58	0,066
Enriquecido en HER*	2,53	0,00012	3,25	<0,0001	2,9	<0,0001
LumB*	2,43	<0,0001	2,88	<0,0001	2,54	<0,0001
Estado de ER†	0,83	0,38	0,83	0,34	0,83	0,32
Estado de PR†	1,36	0,034	1,43	0,012	1,57	0,001
Grado histológico‡	1,4	0,0042	-	-	-	-
Completo frente a subtipo§		<0,0001		<0,0001		<0,0001
Completo frente a clínico		<0,0001		<0,0001		<0,0001
*Clase luminal A usada como estado de referencia en análisis de múltiples variables. †Razones de riesgo para ER, PR y HER2 son marcadores positivos en el numerador. ‡Grado codificado como una variable ordinal con tres niveles. §Los valores de P significativos indican predicción mejorada en relación con subtipo solo. Los valores de P significativos indican predicción mejorada en relación con datos clínicos solos.						

Ejemplo 4.

En este estudio, se usó qRT-PCR y puntos de corte previamente establecidos (véase el ejemplo 1) para evaluar el valor pronóstico del clasificador PAM50 en el grupo de mujeres común, clínicamente importante que son positivas para receptor de estrógeno y se trataron con tamoxifeno como su única terapia sistémica adyuvante. A diferencia de la mayoría de informes previos, esta cohorte del estudio tratada homogéneamente incluye una gran proporción de pacientes positivas para ganglios linfáticos. El seguimiento a largo plazo detallado disponible permite la evaluación no sólo de la supervivencia libre de recidiva, sino también del riesgo de muerte específica por enfermedad de cáncer de mama, en comparación con todos los factores de riesgo clinicopatológicos convencionales.

Métodos

Pacientes:

La cohorte del estudio se deriva de pacientes femeninas con cáncer de mama invasivo, recién diagnosticadas en la provincia de la Columbia Británica (*British Columbia*) en el periodo entre 1986 y 1992. Se cortó tejido en diversos hospitales alrededor de la provincia, se congeló y se envió al laboratorio de receptor de estrógeno (ER) central en el Hospital de Vancouver (*Vancouver Hospital*); en este estudio se usa la parte del material recibido que se fijó en formalina y se incrustó en parafina como referencia histológica. La información clínica vinculada a las muestras incluye edad, histología, grado, tamaño del tumor, número de ganglios axilares implicados, invasión linfática o vascular, estado de ER mediante el método DCC, tipo de terapia sistémica adyuvante local e inicial, fechas de diagnóstico, primera reaparición local, regional o distante, fecha y causa de la muerte. Las características de esta cohorte de pacientes se han descrito previamente en detalle en un estudio basado en la población que valida el modelo de pronóstico ADJUVANT! [Olivotto 2005], y se usaron los mismos bloques fuente para montar micromatrices de tejido que se han caracterizado para la expresión de ER [Cheang 2006] y HER2 [Chia 2008]. Para este estudio, se seleccionaron pacientes que tenían tumores positivos para ER mediante inmunohistoquímica, y recibieron tamoxifeno como su única terapia sistémica adyuvante. Durante el periodo de tiempo en el que estas pacientes recibieron su tratamiento, las directrices provinciales recomendaban tamoxifeno adyuvante para mujeres posmenopáusicas, con tumores positivos para ER que tenían algunas características de riesgo alto presentes tales como invasión linfovascular. Se trató principalmente a pacientes similares sin características de riesgo alto sin terapia sistémica adyuvante. En la mayoría de los casos, se ofreció sólo quimioterapia a mujeres premenopáusicas.

Preparación de ARN:

Se aisló ARN de núcleos de tejido guiado por patólogo. En resumen, se revisaron secciones de H y E de cada bloque por un patólogo. Se seleccionaron áreas que contenían carcinoma de mama invasivo representativo y se señalaron con un círculo en el bloque fuente. Usando una aguja de biopsia de 1,0 mm, se extrajeron al menos dos núcleos de tumor del área señalada con un círculo. Se recuperó ARN usando el kit High Pure ARN Paraffin (Roche Applied Science, Indianápolis EN), se eliminó el ADN con ADNasa Turbo (Ambion, Austin TX) y se avaluó el rendimiento de ARN usando un espectrofotómetro ND-1000 (NanoDrop Technologies, Rockland A).

qRT-PCR:

Se realizó la síntesis de ADNc usando una mezcla de hexámeros al azar y cebadores específicos de genes, y se realizó la qPCR con el instrumento LightCycler 480 de Roche tal como se describió anteriormente [Mullins 2007]. Cada placa de 384 pocillos contenía muestras por duplicado (2,5 ng de ADNc por reacción) y un calibrador por triplicado (10 ng de ADNc por reacción). Se consideró que una muestra de tumor era de calidad insuficiente si faltaba cualquiera de los controles de referencia (*ACTB*, *PSMC4*, *RPLP0*, *MRPL19* o *SF3A1*). La PCR fue técnicamente satisfactoria para los 50 genes discriminadores en el 73% de los casos, y para 49 de los 50 en otro 15% de los casos. Para evaluar la tolerancia de los resultados del ensayo PAM50 para la información génica que faltaba, se evaluaron los valores de ROR-C en los datos tras la eliminación simulada al azar de un número creciente de genes. La pérdida de un gen dio como resultado un cambio de 0-2 unidades en la puntuación de riesgo, que se corresponde con un 1% de aumento/disminución de la supervivencia específica de la enfermedad a 10 años.

Asignación del subtipo biológico a las muestras clínicas:

Se construyeron centroides de expresión génica que se corresponden con subtipos luminal A, luminal B, enriquecido en HER-2, de tipo basal y de tipo normal usando el panel de 50 genes intrínsecos tal como se describió en el ejemplo 1 y en Parker *et al.* (2007 J. Clin. Oncol. 27(8):1160-7, que se incorpora en el presente documento como referencia en su totalidad). Se asignaron las muestras a un subtipo intrínseco basándose en la distancia al centroide más cercano calculada mediante la correlación de rangos de Spearman, por investigadores ciegos a los datos de desenlace.

Relación del subtipo PAM50 con el desenlace clínico:

Se realizaron análisis estadísticos usando SPSS v16.0 y R v2.8.0. Se realizó un análisis de una variable del subtipo de tumor frente a supervivencia libre de recidiva distante de cáncer de mama y supervivencia específica de la enfermedad de cáncer de mama mediante análisis de Kaplan-Meier, con prueba de rangos logarítmicos para determinar la significación. Se realizó un análisis de múltiples variables frente a los parámetros clínicos convencionales de tamaño del tumor, estado de los ganglios (% de ganglios positivos con respecto al total examinado), grado histológico, edad de la paciente y estado de HER2 (basándose en núcleos adyacentes del mismo bloque fuente, montados en micromatrices de tejido y sometidos a inmunotinción y análisis FISH usando protocolos equivalentes clínicos [Chia 2008]). Se desarrollaron modelos de regresión de Cox [Cox 1984] para estimar las razones de riesgo ajustadas de los subtipos de cáncer de mama asignados a qPCR [Truong 2005]. Sólo se incluyeron casos con información para todas las covariables en el análisis. Se usaron gráficos suavizados de residuos de Schoenfeld ponderados para evaluar asunciones de riesgo proporcionales [Grambsch 1994].

Relación de la puntuación de riesgo de recidiva (ROR) con el desenlace clínico:

Se entrenó el algoritmo de puntuación de ROR (ROR-S que incorpora la correlación de una muestra con los subtipos luminal A, luminal B, enriquecido en HER2 y de tipo basal; ROR-C que incorpora esta información más tamaño del tumor) y se validó en tres series de cáncer de mama perfilados con micromatriz y una perfilada con qPCR. Se asignaron puntos de corte de estratificación del riesgo en el conjunto de entrenamiento de tal manera que ninguna paciente luminal A estuviera en la categoría de riesgo alto, y ninguna paciente de tipo basal estuviera en la categoría de riesgo bajo. Se realizaron análisis de regresión de Cox y Kaplan-Meier como anteriormente.

Resultados

A partir de muestras quirúrgicas que se habían fijado en formalina e incrustado en parafina 15-20 años antes, se extrajeron núcleos de tumores de áreas identificadas por el patólogo de carcinoma de mama invasivo para 991 casos. Tras la extracción del ARN, 815 muestras proporcionaron al menos 1,2 µg de ARN total a una concentración de al menos 25 ng/µl, y se procedió al análisis mediante PCR. El molde era de calidad técnicamente suficiente (basándose en controles de genes de mantenimiento internos) para qRT-PCR en 806. Entre estos casos, un total de 711 muestras proporcionaron datos cuantitativos de qRT-PCR de alta calidad para al menos 49 de los genes discriminadores de PAM50, y se incluyeron en análisis clínicos y de supervivencia posteriores. Las características clínicas para estos 711 pacientes se presentan en la tabla 15.

Tabla 15

Parámetro clínico		Serie de TAM completa	Luminal A	Luminal B	Her2	Basal	Normal
Tamaño de la muestra	N	711	329	312	58	3	9
Edad (en años)	Mediana [RIC]	67	67	68	66	68	66
Premenopáusica	Sí	18	9	7	2	0	0
	No	678	315	297	56	3	7
	Desconocido/embarazada	15	5	8	0	0	2
Cirugía	Mastectomía completa	428	187	196	36	3	6
	Mastectomía parcial	274	139	111	21	0	3
	Otros	9	3	5	1	0	0
Disección del ganglio axilar	Sí	675	308	298	57	3	9
	No	36	21	14	1	0	0
Radioterapia de la mama/pared del pecho	Sí	372	180	153	34	0	5
	No	339	149	159	24	3	4
Tamoxifeno adyuvante	Sí	711	329	312	58	3	9
	No	0	0	0	0	0	0
Quimioterapia adyuvante	Sí	0	0	0	0	0	0
	No	711	329	312	58	3	9
Tamaño del tumor (cm)	Mediana [RIC]	2,2	2,0	2,5	2,5	2,5	3,0
Estadio T (clínico)	T0/IS	0	0	0	0	0	0
	T1	298	155	113	24	3	3
	T2	346	147	169	27	0	3
	T3	18	10	5	3	0	0
	T4	28	9	15	1	0	3
	TX	21	8	10	3	0	0
N.º de ganglios positivos	0	199	83	91	18	0	7
	1-3	328	162	139	24	1	2
	4-9	111	49	51	10	1	0
	+10	26	8	16	2	0	0

	Desconocido	47	27	15	4	1	0
Grado	Grado 1: bien diferenciado	24	20	2	1	0	1
	Grado 2: moderadamente diferenciado	306	169	119	13	0	5
	Grado 3: escasamente diferenciado	338	117	173	43	2	3
	Desconocido	43	23	18	1	1	0
Subtipo histológico	NOS ductal	642	289	288	54	3	8
	Lobular	54	30	19	4	0	1
	Mucinoso	7	4	3	0	0	0
	Tubular	5	5	0	0	0	0
	Medular	2	1	1	0	0	0
	Apocrino	1	0	1	0	0	0
Invasión linfovascular	Sí	444	184	215	39	1	5
	No	230	122	84	18	2	4
	Desconocido	37	23	13	1	0	0
Estado del receptor de estrógeno clínico (DCC)	Perdido	6	4	2	0	0	0
	Negativo (0-9 fmol/g)	9	3	2	4	0	0
	Positivo (>10 fmol/mg)	696	322	308	54	3	9
ER inmunohistoquímico	Negativo	0	0	0	0	0	0
	Positivo	711	329	312	58	3	9

Basándose en el centroe de PAM50 más cercano, se asignó un total de 329 (46,3%) de estos casos positivos para ER clínicamente como subtipos de cáncer de mama intrínsecos luminal A, 312 (43,8%) como luminal B, 58 (8,2%) como enriquecido en HER-2, 3 (0,4%) como de tipo basal y 9 (1,3%) como de tipo normal mediante la expresión génica (tabla 13). Para los nueve casos asignados como de tipo normal, se revisó la histología, usando los núcleos de micromatriz de tejido tomados de la misma área del bloque fuente. En ocho de estos nueve casos, estaban ausentes células cancerosas invasivas viables o eran poco comunes en un núcleo inmediatamente adyacente, lo que concuerda con el perfil de expresión de tipo normal que representa una toma de muestras de tumor inadecuada. Por tanto se excluyeron los casos de tipo normal de análisis adicionales.

El subtipo biológico intrínseco era fuertemente pronóstico mediante el análisis de Kaplan-Meier (figura 4). En la población de la Columbia Británica en el tiempo de la adquisición de muestras para este estudio, muchos pacientes con un perfil de riesgo clínicamente bajo no recibieron terapia sistémica adyuvante [Olivotto 2005]. Por el contrario, los que recibieron tamoxifeno adyuvante que son los sujetos en este estudio comprendían un grupo de riesgo clínico superior, con tasas de supervivencia libre de recidiva distante a 10 años global del 62% y tasas de supervivencia específica de la enfermedad de cáncer de mama del 72%. Los que se determinó que tenían un perfil luminal A mediante el ensayo PAM50 tenían un desenlace significativamente mejor (supervivencia libre de recidiva a 10 años al 74%, supervivencia específica de la enfermedad = 83%) que tumores luminal B, enriquecido en HER-2 o de tipo basal.

Todos los casos en este estudio eran positivos para receptor de estrógeno mediante inmunohistoquímica evaluada centralmente [Cheang 2006], y el 98,7% también era positivo mediante ensayo bioquímico recubierto con dextrano-carbón clínico. A pesar de esto, el panel de qPCR de PAM50 asignó el 10% de casos a subtipos no lumbinales, la

mayoría a enriquecido en HER-2, tal como se observó previamente cuando se consultan conjuntos de datos publicados para la expresión de los genes de PAM50 (ejemplo 2).

Para esta cohorte de mujeres clínicamente positivas para receptor de estrógeno, tratadas uniformemente con tamoxifeno como su única terapia sistémica adyuvante, se construyó un modelo de Cox de múltiples variables para someter a prueba el valor independiente del subtipo PAM50 frente a la edad de la paciente y los factores clinicopatológicos convencionales de tamaño del tumor, estado de los ganglios, grado histológico y expresión de HER2 (tabla 16). El subtipo biológico intrínseco permaneció significativo en el modelo de múltiples variables, al igual que el estado de los ganglios y el tamaño del tumor, pero el grado y el estado de HER2 clínico, significativos en el análisis de una variable en esta cohorte, no contribuyó con información de pronóstico independiente significativa para supervivencia o bien libre de recidiva o bien específica de la enfermedad en el modelo de múltiples variables incorporando el resultado de PAM50.

Tabla 16: análisis de una variable y múltiples variables de modelo de Cox incorporando el subtipo biológico PAM50 para supervivencia libre de recidiva y específica de la enfermedad de cáncer de mama entre (A) 604 mujeres con cáncer de mama positivo para ER, tratado con tamoxifeno con datos completos para todas las covariables para supervivencia libre de recidiva y (B) supervivencia específica de la enfermedad de cáncer de mama (BCDSS; excluye 2 casos con causa de la muerte desconocida).

Criterio de valoración clínico	supervivencia libre de recidiva de una variable		supervivencia libre de recidiva de múltiples variables	
	razón de riesgo (IC del 95%)	valor de p	razón de riesgo (IC del 95%)	valor de p
edad (continuo)	1,00 (0,990-1,02)	0,53	0,996 (0,981-1,01)	0,62
grado (1 ó 2) frente a 3	1,45 (1,12-1,89)	0,0047	1,11 (0,846-1,46)	0,45
porcentaje de ganglios positivos				
0 frente a (>0 a <25%)	1,66 (1,15-2,39)	0,0070	1,76 (1,22-2,55)	0,0028
0 frente a ≥ 25%	2,98 (2,10-4,22)	7,3E-10	2,85 (2,00-4,06)	6,3E-9
tamaño del tumor ≤ 2 cm frente a > 2 cm	2,02 (1,55-2,65)	2,5E-7	1,71 (1,30-2,24)	1,3E-4
HER2 (IHC) {0, 1 ó 2 + FISH negativo} frente a {2 + FISH positivo ó 3 +}	1,52(1,04-2,23)	0,032	1,24 (0,813-1,88)	0,32
Subtipo PAM50				
luminal A frente a luminal B	1,73 (1,31-2,28)	1,0E-4	1,62 (1,22-2,16)	9,2E-4
luminal A frente a enriquecido en HER2	1,86 (1,18-2,92)	0,0074	1,53 (0,929-2,52)	0,095
luminal A frente a de tipo basal	76,4 (9,79-597)	3,5E-5	62,5 (7,87-496)	9,2E-5

B.

Criterio de valoración clínico	BCDSS de una variable		BCDSS de múltiples variables	
	razón de riesgo (IC del 95%)	valor de p	razón de riesgo (IC del 95%)	valor de p
edad (continuo)	1,02 (0,999-1,03)	0,069	1,01 (0,988-1,02)	0,56
grado (1 ó 2) frente a 3	1,43 (1,07-1,91)	0,015	1,05 (0,775-1,42)	0,76
porcentaje de ganglios positivos				

0 frente a (>0 a <25%)	1,56 (1,03-2,37)	0,034	1,68 (1,11-2,56)	0,015
0 frente a $\geq 25\%$	3,22 (2,19-4,73)	2,4E-9	3,04 (2,06-4,48)	2,3E-8
tamaño del tumor ≤ 2 cm frente a > 2 cm	2,29 (1,69-3,10)	8,0E-8	1,90 (1,40-2,58)	4,3E-5
HER2 (IHC) {0, 1 ó 2 + FISH negativo} frente a {2 + FISH positivo ó 3 +}	1,54 (1,01-2,35)	0,043	1,19 (0,755-1,86)	0,46
Subtipo PAM50				
luminal A frente a luminal B	2,05 (1,50-2,80)	6,0E-6	1,90 (1,37-2,62)	1,0E-4
luminal A frente a enriquecido en HER2	2,20 (1,33-3,64)	0,0021	1,85 (1,07-3,20)	0,028
luminal A frente a de tipo basal	104 (13,1-832)	1,2E-5	91,1 (11,2-743)	2,5E-5

Puede calcularse una puntuación de riesgo de recidiva (ROR) a partir del panel de qPCR de PAM50. Las puntuaciones tanto de ROR-S (basada sólo en la obtención del subtipo molecular a partir del panel de PAM50) como de ROR-C (que combina información del subtipo y del tamaño del tumor) son altamente pronósticas en una población tratada de manera homogénea con tamoxifeno adyuvante, para una serie que contiene grandes números de casos positivos para ganglios, y para el criterio de valoración de supervivencia específica a cáncer de mama (figura 5).

Tal como se muestra en la figura 6, el algoritmo de ROR-C no sólo es altamente pronóstico entre pacientes negativas para ganglios, sino que revela diferencias incluso más amplias en supervivencia específica de la enfermedad entre pacientes positivas para ganglios. El algoritmo identifica el 16% de pacientes positivas para ER clínicamente (tratadas con tamoxifeno adyuvante pero sin quimioterapia) quienes, a pesar de ser positivas para ganglios, se clasifican como de riesgo bajo, y estas mujeres tienen una tasa de supervivencia específica de la enfermedad a 10 años del 89%.

Como variable continua, ROR-C tiene una interacción significativa con el porcentaje de ganglios linfáticos positivos, e interacción significativa límite con el estadio ganglionar (tabla 17). El estadio ganglionar es un factor pronóstico significativo entre pacientes con valores de ROR-C de moderados a altos ($>23,5$), pero entre pacientes con puntuaciones de ROR-C bajas, los desenlaces son buenos independientemente del estado de los ganglios (figura 7 y figura 8).

Tabla 17: Prueba de interacción entre la puntuación de ROR-C derivada de PAM50 y derivada del tamaño del tumor, expresada como variable continua, y estado de los ganglios linfáticos axilares (A) expresado como % de ganglios positivos o (B) categorizado por el estadio ganglionar (en el que el grupo de referencia es negativo para ganglios, N cat2 = 1-3 ganglios axilares implicados, y N cat3 = 4 o más ganglios axilares implicados). El modelo en la tabla 17A usa la proporción de ganglios positivos y la interacción es significativa. El modelo en la tabla 17B usa un estado de los ganglios de 3 niveles (N-, 1-3 pos, >3 pos) y la interacción está en el límite.

Tabla 17A

Variable	Sólo efectos principales		Interacción	
	Riesgo	valor de p	Riesgo	valor de p
ROR-C	1,75	1,60E-11	1,73	8,8E-11
% de ganglios pos.	1,56	2,50E-10	1,43	0,000017
Interacción			1,17	0,043
Completo frente a reducido				0,04

Tabla 17B

Variable	Sólo efectos principales		Interacción	
	Riesgo	valor de p	Riesgo	valor de p
ROR-C	1,77	6,20E-11	1,52	0,018
N cat2	1,8	9,40E-03	1,73	0,022
N cat3	3,88	1,20E-08	3,15	1,4E-05
ROR*N cat2			1,08	0,71
ROR*N cat3			1,62	0,061
Completo frente a reducido				0,11

Como ROR-C incluye información del tamaño del tumor, para evaluar si el algoritmo de ROR da información de pronóstico adicional independiente más allá de los parámetros clínicos convencionales (incluyendo tamaño del tumor) en esta población de pacientes, se sometieron a prueba los modelos de Cox que incorporan ROR-S (tabla 18). Independientemente de si el criterio de valoración es supervivencia libre de recidiva o específica de la enfermedad, o si ROR-S está incluido como variable continua o categórica, sigue siendo significativo, mientras que el grado y el estado de HER2 clínico no son significativos en análisis de múltiples variables que incluyen la información derivada de qPCR.

Tabla 18: Análisis de múltiples variables de modelo de Cox que incorpora la puntuación de ROR-S para supervivencia específica de la enfermedad de cáncer de mama entre mujeres con cáncer de mama tratadas con tamoxifeno, positivas para ER y datos completos para todas las covariables. (A) categorías de riesgo definidas para ROR-S, usando criterios de valoración especificados previamente. (B) ROR-S como variable continua.

A.

Criterio de valoración clínico	Supervivencia libre de recidiva (N=613)		Supervivencia específica de la enfermedad (N=611)	
	razón de riesgo (IC del 95%)	valor de p	razón de riesgo (IC del 95%)	valor de p
edad (continuo)	0,995 (0,980-1,01)	0,56	1,00 (0,988-1,02)	0,56
grado (1 ó 2) frente a 3	1,03 (0,785-1,36)	0,81	1,00 (0,738-1,36)	1,0
porcentaje de ganglios positivos				
0 frente a (>0 a <25%)	1,79 (1,24-2,58)	0,0016	1,74 (1,16-2,63)	0,0081
0 frente a $\geq 25\%$	2,87 (2,02-4,08)	4,4E-9	3,10 (2,10-4,57)	1,3E-8
tamaño del tumor ≤ 2 cm frente a > 2 cm	1,70 (1,30-2,23)	1,2E-4	1,92 (1,42-2,61)	2,8E-5
HER2 (IHC) {0, 1 ó 2 + FISH negativo} frente a {2 + FISH positivo ó 3 +}	1,14 (0,760-1,72)	0,52	1,10 (0,701-1,74)	0,67
ROR-S (categorizado)				
bajo frente a medio	2,00 (1,39-2,87)	1,9E-4	2,21 (1,45-3,36)	2,1E-4
bajo frente a alto	2,68 (1,63-4,41)	1,0E-4	3,25 (1,86-5,67)	3,4E-5

B.

Criterio de valoración clínico	Supervivencia libre de recidiva (N=613)		Supervivencia específica de la enfermedad (N=611)	
	razón de riesgo (IC del 95%)	valor de p	razón de riesgo (IC del 95%)	valor de p
edad (continuo)	0,997 (0,982-1,01)	0,71	1,01 (0,989-1,02)	0,48
grado (1 ó 2) frente a 3	1,06 (0,808-1,40)	0,66	1,02 (0,749-1,38)	0,92
porcentaje de ganglios positivos				
0 frente a (>0 a <25%)	1,77 (1,23-2,53)	0,0021	1,71 (1,13-2,58)	0,011
0 frente a ≥ 25%	2,87 (2,02-4,06)	3,4E-9	3,12 (2,12-4,59)	8,5E-9
tamaño del tumor ≤ 2 cm frente a > 2 cm	1,70 (1,30-2,23)	1,2E-4	1,92 (1,41-2,60)	3,0E-5
HER2 (IHC) {0, 1 ó 2 + FISH negativo} frente a {2 + FISH positivo ó 3 +}	1,05 (0,699-1,59)	0,80	0,986 (0,628-1,55)	0,95
ROR-S (continuo)	1,02 (1,01-1,03)	7,3E-5	1,02 (1,01-1,03)	1,0E-5

- Los casos en esta serie se han evaluado previamente mediante inmunohistoquímica para ER, PR, HER2, citoqueratina 5/6, receptor del factor de crecimiento epidérmico y Ki67 [Cheang 2008][Cheang 2009], permitiendo que se asignara la obtención de subtipos intrínsecos mediante una definición inmunohistoquímica sustituta. Como todos los casos en esta serie son positivos para ER mediante inmunohistoquímica, todos se asignaron como o bien luminal A (si eran negativos para HER2 y de Ki67 bajo) o bien luminal B (si eran positivos para HER2 o de Ki67 alto). La disponibilidad de las asignaciones de la obtención de subtipos de qPCR permite una comparación con la asignación inmunohistoquímica en el mismo material, frente al desenlace de la paciente en esta cohorte tratada de manera homogénea. Un total de 606 casos tenían datos inmunohistoquímicos y de qPCR suficientemente completos para la asignación a un subtipo luminal mediante ambos métodos. Entre estos, se asignaron 255 como luminal A y 193 como luminal B mediante ambos métodos, mientras que se asignaron 99 como luminal A mediante inmunotinción pero como luminal B mediante qPCR, y 59 como luminal B mediante inmunotinción pero como luminal A mediante qPCR, para una concordancia del 74%, kappa = 0,48. Cuando los resultados eran discordantes, sólo los casos asignados como luminal B mediante PCR tenían un desenlace significativamente peor que los asignados de manera concordante como luminal A. En los análisis de múltiples variables entre estos casos, la asignación tanto inmunohistoquímica como con PAM50 son factores pronóstico independientemente significativos para determinar la supervivencia libre de recidiva, mientras que el grado y el estado de HER2 se salen del modelo (tabla 19). Para la supervivencia específica de la enfermedad, PAM50 es significativo mientras que la inmunohistoquímica está en el límite. La magnitud del riesgo identificado es superior con la asignación de qPCR para ambos criterios de valoración. En un modelo de regresión de Cox por etapas que incorpora asignación tanto inmunohistoquímica como de qPCR, sólo qPCR permanece significativo.
- Tabla 19: análisis de múltiples variables del modelo de Cox para casos luminales, comparando la información de pronóstico de la obtención de subtipos intrínsecos mediante inmunohistoquímica frente a qPCR de PAM50.

A. Supervivencia libre de recidiva (N = 606)

Criterio de valoración clínico	Subtipo inmunohistoquímico		Subtipo de qPCR de PAM50	
	razón de riesgo (IC del 95%)	valor de p	razón de riesgo (IC del 95%)	valor de p
edad (continuo)	0,992 (0,98-1,01)	0,36	0,990 (0,97-1,01)	0,26
grado (1 ó 2) frente a 3	1,18 (0,89-1,57)	0,24	1,12 (0,84-1,49)	0,43
porcentaje de ganglios positivos				

0 frente a (>0 a <25%)	1,66 (1,11-2,48)	0,014	1,68 (1,12-2,50)	0,012
0 frente a $\geq 25\%$	2,86 (1,95-4,19)	7,2E-8	2,93 (2,00-4,30)	3,8E-8
tamaño del tumor ≤ 2 cm frente a > 2 cm	1,80 (1,34-2,42)	8,6E-5	1,81 (1,35-2,42)	7,4E-5
HER2 (IHC) {0, 1 ó 2 + FISH negativo} frente a {2 + FISH positivo ó 3 +}	1,21 (0,74-1,99)	0,45	1,30 (0,81-2,09)	0,27
Luminal B frente a luminal A	1,38 (1,02-1,86)	0,035	1,61 (1,20-2,16)	0,0014

B. Supervivencia específica de la enfermedad de cáncer de mama (N = 605; excluye una muerte de causa indeterminada)

Criterio de valoración clínico	Subtipo inmunohistoquímico		Subtipo de qPCR de PAM50	
	razón de riesgo (IC del 95%)	valor de p	razón de riesgo (IC del 95%)	valor de p
edad (continuo)	1,00 (0,98-1,02)	0,67	1,00 (0,98-1,02)	0,89
grado (1 ó 2) frente a 3	1,14 (0,83-1,55)	0,42	1,05 (0,77-1,44)	0,74
porcentaje de ganglios positivos				
0 frente a (>0 a <25%)	1,44 (0,92-2,26)	0,106	1,50 (0,96-2,34)	0,077
0 frente a $\geq 25\%$	2,79 (1,84-4,23)	1,2E-6	2,88 (1,90-4,38)	5,8E-7
tamaño del tumor ≤ 2 cm frente a > 2 cm	2,07 (1,48-2,89)	1,8E-5	2,06 (1,48-2,87)	1,7E-5
HER2 (IHC) {0, 1 ó 2 + FISH negativo} frente a {2 + FISH positivo ó 3 +}	1,27 (0,75-2,55)	0,38	1,29 (0,78-2,13)	0,32
Luminal B frente a luminal A	1,38 (0,99-1,93)	0,060	1,89 (1,36-2,62)	1,5E-4

5

Resultados de las predicciones de Adjuvant!

Se realizó una comparación del desenlace predicho por el modelo Adjuvant! con el desenlace predicho por el modelo de ROR en una cohorte de pacientes con cáncer de mama. Esta cohorte consiste en 806 pacientes diagnosticadas con cáncer de mama invasivo, positivo para receptor de estrógeno, entre las fechas de 1986 y 1992. Todas las pacientes tenían cirugía primaria y terapia sistémica adyuvante con tamoxifeno sólo; a ninguna de éstas pacientes se les trató con quimioterapia. Se usó el modelo pronóstico Adjuvant! Para calcular la probabilidad de supervivencia específica a cáncer de mama (BCSS) a 10 años usando las características clinicopatológicas convencionales de edad de la paciente, tamaño del tumor, grado histológico, invasión linfovascular y número de ganglios linfáticos positivos. Todas las pacientes eran positivas para ER, y el riesgo de muerte por cáncer de mama se ajustó para la terapia con tamoxifeno adyuvante.

De las 806 pacientes, 748 tenían suficientes datos clinicopatológicos para obtener una estimación de Adjuvant! de BCSS. Las 58 pacientes restantes tenían o bien datos de tamaño del tumor que faltaban o de ganglios linfáticos que faltaban. La BCSS predicha por Adjuvant! media era del 73,7%. Esto corresponde con la BCSS observada del 73,2%. Entonces se dividió la cohorte en subgrupos basándose en la BCSS predicha por Adjuvant! a 10 años (tabla 20).

25

Tabla 20

Categoría de riesgo	BCSS a 10 años predicha por Adjuvant!	N	Número de acontecimientos
1	90-100%	122	16
2	80-90%	164	32
3	70-80%	168	60
4	<70%	292	121

La BCSS observada a 10 años, para cada uno de los grupos de riesgo de Adjuvant! era similar a la BCSS predicha por Adjuvant! (tabla 21). Una excepción notable es el grupo de riesgo más bajo, en el subgrupo con una BCSS predicha por Adjuvant! del 90-100%, la BCSS observada a 10 años es del 89%. Por consiguiente, parece que Adjuvant! sobreestima la supervivencia de este grupo de riesgo bajo. Esto concuerda con el estudio de validación de Adjuvant! usando la base de datos BCOU (Olivotto *et al.* (2005)). En este estudio de validación, se encontró que Adjuvant! subestimaba las muertes por cáncer de mama en un 4,9% en el subgrupo con enfermedad T1N0. En el subgrupo de pacientes con una BCSS predicha del 90-100%, 87 de 122 pacientes tenían cáncer de mama T1N0.

Tabla 21

Grupo de riesgo de Adjuvant!	BCSS predicha por Adjuvant! media a 10 años	BCSS observada a 10 años
90-100%	94%	89%
80-90%	85%	83%
70-80%	76%	75%
<70%	58%	61%

Entonces se usó ROR-S usando datos de qRT-PCR a partir de 50 genes para separar cada grupo de Adjuvant! en riesgo bajo frente a medio/alto (tabla 22). Debido al tamaño relativamente pequeño de cada grupo, se combinaron los grupos de riesgo medio y alto para mejorar la potencia estadística. Además, debido a que los subgrupos de riesgo de Adjuvant! ya están definidos usando factores clínicos, se aplicó ROR-S (en lugar de ROR-C).

Tabla 22

Grupo de riesgo de Adjuvant!	ROR-S bajo	ROR-S medio o alto
90-100%	47	75
80-90%	56	108
70-80%	43	125
<70%	58	234

Entonces se realizó el análisis de Kaplan-Meier por separado en cada grupo de riesgo de Adjuvant!, y se sometieron a prueba las diferencias en la supervivencia entre los grupos de ROR-S de riesgo bajo frente a medio/alto usando la prueba de rangos logarítmicos (tabla 23). Se observó que ROR-S podía aislar un subgrupo de riesgo bajo en cada uno de los grupos de riesgo de Adjuvant!. Se encontraron diferencias estadísticamente significativas en BCSS para pacientes de riesgo bajo frente a de riesgo medio/alto, en todos los subgrupos excepto para el grupo del 90-100%.

Tabla 23

Grupo de riesgo de Adjuvant!	BCSS observada	BCSS para ROR-S de riesgo bajo	BCSS para ROR-S de riesgo medio/alto	Prueba de rangos logarítmicos de ROR-S
90-100%	89%	93%	85%	p = 0,058
80-90%	83%	92%	78%	p = 0,020
70-80%	75%	95%	68%	p = 0,005
<70%	61%	71%	58%	P = 0,009

En este grupo de riesgo bajo, ROR-S no es lo bastante estadísticamente significativo ($p = 0,058$). Sin embargo, este grupo proporciona algunas pruebas convincentes de que ROR-S añade información de pronóstico adicional a Adjuvant!. Se muestra en la figura 9 un análisis de Kaplan-Meier del mismo grupo pero con BCSS predicha por Adjuvant! del 90-95% frente a del 95-100%.

En los grupos de riesgo intermedio, ROR-S funciona bien en la identificación de pacientes de riesgo bajo frente a de riesgo superior. Tanto en el grupo del 80-90% (figura 10A) como en el grupo del 70-80% (figura 10B), el ROR-S identifica subgrupos con BCSS a 10 años $> 90\%$. Este es un resultado importante ya que ROR-S identifica tradicionalmente pacientes de riesgo alto a los que les va bien sin quimioterapia.

En el subgrupo de riesgo muy alto identificado por Adjuvant! (BCSS predicho $<70\%$), ROR-S todavía puede identificar distintos grupos pronósticos (figura 10C).

Discusión

Estudios anteriores han establecido que los distintivos biológicos intrínsecos característicos de los subtipos luminal A, luminal B, enriquecido en HER2 y de tipo basal están presentes y tienen significación pronóstica en cohortes de cáncer de mama de múltiples instituciones diferentes, con perfiles obtenidos con diversas plataformas de micromatrices de expresión génica [Calza 2006][Kapp 2006][Hu 2006][Fan 2006]. Con el fin de identificar estos subtipos en muestras de patología fijadas en formalina, incrustadas en parafina convencionales, se desarrolló una prueba de PCR con transcriptasa inversa cuantitativa [Mullins 2007] que identificaba estos subtipos basándose en un panel de 50 genes.

El análisis notificado en el presente documento consiste exclusivamente en pruebas basadas en qPCR, aplicadas a una serie de bloques de parafina relativamente antiguos (15-20 años) con seguimiento largo y detallado, permitiendo el análisis no sólo de la supervivencia libre de recidiva, sino también de la supervivencia específica de la enfermedad de cáncer de mama. El presente estudio consiste en mujeres con cáncer de mama positivo para receptor de estrógeno que recibieron terapia hormonal (tamoxifeno) como su único tratamiento adyuvante, un grupo de importancia clínica particular y relevancia contemporánea. Se determinaron de manera central el estado de receptor de estrógeno y el estado de HER2. El 70% de estas mujeres eran positivas para ganglios en la presentación, y en la práctica actual se les recomendaría habitualmente recibir quimioterapia adyuvante. La asignación del subtipo PAM50 tal como se determinó mediante PCR es altamente pronóstica en estas mujeres. El subtipo sigue siendo significativo en análisis de múltiples variables, mientras que el grado y el estado de HER2 clínico no. Los hallazgos usando el criterio de valoración sustituto empleado comúnmente de supervivencia libre de recidiva se mantienen todos para la supervivencia específica de la enfermedad de cáncer de mama.

Aunque las pacientes de esta cohorte se trataron hace más de 20 años, los hallazgos de este estudio siguen siendo relevantes para el tratamiento de pacientes con cáncer de mama con un riesgo moderado de recidiva. Tales pacientes pueden obtener un beneficio significativo de terapia hormonal adyuvante pero la adición adicional de quimioterapia puede tener efectos modestos (mejora del 2-5% en supervivencia libre de recidiva a 10 años). Aunque la decisión de continuar con la quimioterapia adyuvante es una decisión individual tomada por la paciente y el oncólogo de consulta, una mejora de la pronosticación facilitará la toma de la decisión terapéutica.

Se desarrolló una puntuación de riesgo de recidiva y se validó en datos de micromatrices de pacientes negativas para ganglios que no revivieron terapia sistémica adyuvante (ejemplo 2), frente al criterio de valoración de supervivencia libre de recidiva. Se muestra que este algoritmo predice la respuesta completa patológica en un conjunto de datos de ensayo clínico T/FAC neoadyuvante publicado de 133 pacientes y, en su formato de qPCR, predice la supervivencia libre de recidiva en una cohorte de 279 mujeres tratadas de manera homogénea con cáncer de mama. Las puntuaciones de ROR generadas mediante qPCR a partir de muestras en bloques de parafina también son pronósticas en mujeres tratadas con tamoxifeno, positivas para estrógeno, en los subconjuntos tanto negativo para ganglios como positivo para ganglios. ROR-C identifica un grupo de pacientes de riesgo bajo entre las que incluso el estado de los ganglios no es un factor pronóstico, y que por tanto pueden no requerir enfoques de tratamiento habitualmente reservados para pacientes positivas para ganglios que incluyen, por ejemplo, regímenes de quimioterapia de tercera generación y radiación de la pared del pecho.

Muy pocos casos (1,3%) se clasifican como de tipo normal usando el ensayo de PCR, en comparación con el 12% cuando se aplica el clasificador PAM50 a datos de micromatrices de ADN de conjuntos grandes de cánceres de mama primarios. Los análisis de micromatrices de ADN utilizan muestras tumorales homogeneizadas que, a pesar de la disección macroscópica para enriquecer el tumor, todavía pueden contener cantidades significativas de tejido de mama normal. Por el contrario, el ensayo de qPCR de PAM50 se realiza en un núcleo de tejido guiado por patólogo, basándose en la identificación microscópica directa de un área representativa de tumor puro en el bloque de fuente. Esta diferencia probablemente explica la frecuencia mucho más baja de perfiles de tipo normal obtenidos usando el método de qPCR de PAM50 aplicado a bloques de parafina. La revisión de la histología, tal como se representa en núcleos de micromatrices de tejido extraídos del tejido inmediatamente adyacente, concuerda con la representación tumoral inadecuada responsable de un perfil de tipo normal en ocho de los nueve casos de tipo normal.

Tal como se indicó anteriormente basándose en la consulta de conjuntos de datos publicados con el clasificador PAM50, este ensayo identifica subtipos biológicos negativos para ER entre mujeres positivas para ER clínicamente incluso en entornos en los que el tumor es positivo mediante ensayos tanto de inmunohistoquímica como de unión a ligando. Por lo menos el 10% de los casos se reasignan a subtipos no luminales, y estas mujeres tratadas con tamoxifeno tenían desenlaces malos, compatibles con una realidad biológica de independencia hormonal. Las mediciones clínicas del estado de ER y HER2, por sí mismas, pueden estratificar pacientes con cáncer de mama en subgrupos pronósticos y predictivos [Hayes 2007]. Sin embargo, basándose en mediciones de genes individuales (ER, PR) para asignar el pronóstico y el tratamiento del cáncer de mama se corre el riesgo de no sólo los problemas de mediciones individuales de falsos positivos y negativos, sino también la posibilidad de que una biología subyacente al tumor sea independiente de hormonas (a pesar de que un miembro de la ruta esté expresándose a nivel proteico). A este respecto, la información proporcionada por la medición simultánea de 50 genes, incluyendo otros en la ruta de respuesta a estrógenos junto con marcadores positivos de otros subtipos biológicos, es probable que sea un reflejo más preciso de la biología tumoral subyacente [Oh 2006].

Se han vinculado paneles sustitutos de inmunohistoquímica más grandes a criterios de referencia de perfil de expresión y pueden proporcionar más información que la simple medición de ER, PR y HER2 [Cheang 2008b][Cheang 2009]. Se aplican fácilmente paneles de anticuerpos limitados a bloques de parafina convencionales, y pueden añadir información de pronóstico significativa más allá de factores de riesgo clinicopatológicos convencionales [Ross 2008]. En este estudio, se realizó una comparación directa de un panel de seis inmunotinciones establecidas (ER, PR, HER2, Ki67, citoqueratina 5/6 y receptor del factor de crecimiento epidérmico) frente al ensayo de qPCR de 50 genes, usando los mismos bloques fuente. Cada método añade información de pronóstico significativa más allá de factores convencionales. Sin embargo, en este conjunto de pacientes positivas para ER clínicamente, hubo muchas asignaciones discrepantes a un subtipo biológico intrínseco, y el enfoque de qPCR era mejor en la predicción del desenlace en estos casos.

En el análisis de múltiples variables que incorpora los principales factores de riesgo clínicos, el grado ya no es significativo cuando se incluye el subtipo PAM50 o ROR. En comparación con otros distintivos tales como la puntuación de reaparición y los índices de grado genómico [Paik 2004][Lvshina 2006][Sotiriou 2006], el PAM50 también tiene la ventaja de diferenciar casos de alto riesgo en subtipos luminal B, enriquecido en HER2 y de tipo basal, que es probable que respondan de manera diferente a las opciones de terapia sistémica (por ejemplo, hormonal, anti-HER2 y regímenes de quimioterapia con antraciclina frente a sin antraciclina). El ensayo también es más fácil de realizar, ya que no requiere tejido congelado [Glas 2006] ni microdissección manual de secciones de corte [Paik 2004] y puede aplicarse fácilmente a bloques de parafina convencionales incluyendo tejidos de archivo tales como los de ensayos clínicos. Sin embargo, el ensayo puede realizarse en estos tipos de muestras si se desea. Debido a que el ensayo PAM50 se diseñó para reflejar las características principales de la biología subyacente del cáncer de mama, en lugar de optimizarse frente al desenlace en una población particular, es particularmente probable que se extrapole bien en otras cohortes de pacientes, y siga siendo predictivo [Rouzier 2005]. En este estudio, se demostró por primera vez que el ensayo de qPCR de PAM50 tenía capacidad pronóstica independiente y significativa entre mujeres positivas para receptor de estrógeno, tratadas con tamoxifeno, ya fuesen positivas para ganglios o negativas para ganglios. El ensayo identifica hasta un 10% de casos que se determinó clínicamente que eran positivos para ER (mediante ensayo de inmunohistoquímica y de unión a ligando) que se encuentran en grupos de riesgo alto negativos para ER, sustituye el grado y el estado de HER2 en modelos pronósticos de múltiples variables y es superior a la obtención de subtipos mediante inmunohistoquímica y clasificadores del riesgo clínico.

Bibliografía

Calza *et al.* (2006) Breast cancer Res. 8:R34.

Cheang *et al.* (2006) J Clin Oncol. Dec 20; 24(36):5637-44.

Cheang *et al.* (2008a) Annu Rev Pathol. 3:67-97.

Cheang MC *et al.* (2008b) Clin Cancer Res. 14(5):1368-76.

Cheang MC *et al.* (2009) J Natl Cancer Inst. 101(10):736-50.

5 Chia S *et al.* (2008) J Clin Oncol. 26(35):5697-704.

Cox y Oakes (1984) Analysis of Survival Data. Chapman & Hall (Londres, Inglaterra).

Cronin M *et al.* (2007) Clin Chem 53:1084-91.

10 Fan *et al.* (2006) N. Engl. J. Med. 355:560-69.

Glas *et al.* (2006) BMC Genomics 7:278.

15 Grambsch y Therneau (1994) Biometrika 81(3):515-26.

Hayes *et al.* (2007) N Engl J Med. 357(15):1496-506.

Hu *et al.* (2006) BMC Genomics 7:96.

20 Kapp *et al.* (2006) BMC Genomics 7:231.

Loi *et al.* (2007) J Clin Oncol. 25(10):1239-46.

25 Mullins *et al.* (2007) Clin Chem. 53(7):1273-9.

Oh DS *et al.* (2006) J Clin Oncol. 24(11):1656-64.

Olivotto *et al.* (2005) J Clin Oncol 23:2716-25.

30 Paik (2004) N. Engl. J. Med. 351:2817-26.

Parker *et al.* (2009) J Clin Oncol. 27(8):1160-1167.

35 Ross *et al.* (2008) Clin Cancer Res. 14(20):6602-9.

Rouzier R *et al.* (2005) Clin Cancer Res. 11(16):5678-85.

Sorlie *et al.* (2003) Proc Natl Acad Sci USA. 100(14):8418-23.

40 Sotiriou *et al.* (2006) J Natl Cancer Inst 98:262-272.

Tibshirani *et al.* (2002) Proc Natl Acad Sci USA. 99(10):6567-72.

45 Truong *et al.* (2005) Cancer 103(10):2006-14.

van't Veer *et al.* (2005) J Clin Oncol. 23(8):1631-5.

50 Todas las publicaciones y solicitudes de patente mencionadas en la memoria descriptiva son indicativas del nivel de los expertos en la técnica a la que esta invención se refiere. Todas las publicaciones y solicitudes de patente se incorporan en el presente documento como referencia en el mismo grado que si se indicara individualmente que cada publicación o solicitud de patente individual se incorpora como referencia.

55 Aunque la invención anterior se ha descrito con algún detalle a modo de ilustración y ejemplo para fines de claridad de comprensión, será obvio que pueden ponerse en práctica determinados cambios y modificaciones dentro del alcance de las reivindicaciones adjuntas.

LISTA DE SECUENCIAS

60 <110> Charles M. Perou Philip Bernard Matthew Ellis Joel S. Parker James Stephen Marron Andrew Nobel Elaine Mardis Torsten O. Nielsen Maggie Cheang

<120> PERFILES DE EXPRESIÓN GÉNICA PARA PREDECIR DESENLACES EN CÁNCER DE MAMA

65 <130> 35052/372870

<150> 61/057.508
 <151> 2008-05-30
 <160> 100
 5 <170> FastSEQ para Windows versión 4.0
 <210> 1
 <211> 19
 10 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleotídico
 15 <400> 1
 aaagattcct gggacctga 19
 <210> 2
 20 <211> 21
 <212> ADN
 <213> Secuencia artificial
 <220>
 25 <223> Cebador oligonucleotídico
 <400> 2
 acagccactt tcagaagcaa g 21
 30 <210> 3
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 35 <220>
 <223> Cebador oligonucleotídico
 <400> 3
 ctggaagagt tgaataaaga 22
 40 <210> 4
 <211> 18
 <212> ADN
 <213> Secuencia artificial
 45 <220>
 <223> Cebador oligonucleotídico
 <400> 4
 50 tacctgaacc ggcacctg 18
 <210> 5
 <211> 20
 <212> ADN
 55 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleotídico
 60 <400> 5
 gcacaaagcc attctaagtc 20
 <210> 6
 <211> 17
 65 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico
 5 <400> 6
 gctggctgag cagaaag 17

 <210> 7
 <211> 19
 10 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Cebador oligonucleotídico
 15
 <400> 7
 ctttcgcctg agcctattt 19

 <210> 8
 <211> 18
 20 <212> ADN
 <213> Secuencia artificial

 <220>
 25 <223> Cebador oligonucleotídico

 <400> 8
 ggccaaaatc gacaggac 18
 30 <210> 9
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 35 <220>
 <223> Cebador oligonucleotídico

 <400> 9
 ctgtctgagt gccgtggat 19
 40
 <210> 10
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 45
 <220>
 <223> Cebador oligonucleotídico

 <400> 10
 50 gtaaatcacc ttctgacct 20

 <210> 11
 <211> 18
 <212> ADN
 55 <213> Secuencia artificial

 <220>
 <223> Cebador oligonucleotídico

 <400> 11
 60 ggaggcggaa gaaaccag 18

 <210> 12
 <211> 24
 65 <212> ADN
 <213> Secuencia artificial

	<220>		
	<223> Cebador oligonucleotídico		
5	<400> 12 gacaaggaga atcaaaagat cagc	24	
	<210> 13		
	<211> 17		
10	<212> ADN		
	<213> Secuencia artificial		
	<220>		
	<223> Cebador oligonucleotídico		
15	<400> 13 gtggcagcag atcaca	17	
	<210> 14		
20	<211> 20		
	<212> ADN		
	<213> Secuencia artificial		
	<220>		
25	<223> Cebador oligonucleotídico		
	<400> 14 cctcacgaat tgctgaactt	20	
30	<210> 15		
	<211> 23		
	<212> ADN		
	<213> Secuencia artificial		
35	<220>		
	<223> Cebador oligonucleotídico		
	<400> 15 catgaaatag tgcatagttt gcc	23	
40	<210> 16		
	<211> 25		
	<212> ADN		
	<213> Secuencia artificial		
45	<220>		
	<223> Cebador oligonucleotídico		
	<400> 16 acacagaatc tatacccacc agagt	25	
50	<210> 17		
	<211> 20		
	<212> ADN		
55	<213> Secuencia artificial		
	<220>		
	<223> Cebador oligonucleotídico		
60	<400> 17 gctggctctc aactgatag	20	
	<210> 18		
	<211> 18		
65	<212> ADN		
	<213> Secuencia artificial		

<220>
 <223> Cebador oligonucleotídico

5 <400> 18
 gcagggagag gagttgt 18

<210> 19
 <211> 23
 10 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

15 <400> 19
 cccatccatg tgaggaagta taa 23

<210> 20
 20 <211> 17
 <212> ADN
 <213> Secuencia artificial

<220>
 25 <223> Cebador oligonucleotídico

<400> 20
 ctcttgac cttggcg 17

30 <210> 21
 <211> 17
 <212> ADN
 <213> Secuencia artificial

35 <220>
 <223> Cebador oligonucleotídico

<400> 21
 gctactacgc agacacg 17

40 <210> 22
 <211> 19
 <212> ADN
 <213> Secuencia artificial

45 <220>
 <223> Cebador oligonucleotídico

<400> 22
 50 gatgttcgag tcacagagg 19

<210> 23
 <211> 17
 <212> ADN
 55 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

60 <400> 23
 ttcggctgga aggaacc 17

<210> 24
 <211> 18
 65 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico
 5 <400> 24
 cgtggcagat gtgaacga 18
 <210> 25
 <211> 20
 10 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleotídico
 15 <400> 25 ggagatccgt caactccaaa 20
 <210> 26
 <211> 18
 20 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleotídico
 25 <400> 26
 tgggtcgtgt caggaaac 18
 <210> 27
 <211> 20
 30 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleotídico
 35 <400> 27
 cgcagtcatc cagagatgtg 20
 <210> 28
 <211> 21
 40 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleotídico
 45 <400> 28
 actcagtaca agaaagaacc g 21
 50 <210> 29
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 55 <220>
 <223> Cebador oligonucleotídico
 <400> 29
 60 gttggaccag tcaacatctc tg 22
 <210> 30
 <211> 18
 <212> ADN
 65 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

5 <400> 30
 tgtggctcat taggcaac 18

10 <210> 31
 <211> 18
 <212> ADN
 <213> Secuencia artificial

15 <220>
 <223> Cebador oligonucleotídico

15 <400> 31
 gactccaagc gcgaaaac 18

20 <210> 32
 <211> 24
 <212> ADN
 <213> Secuencia artificial

25 <220>
 <223> Cebador oligonucleotídico

25 <400> 32
 ccaacaaaat attcatggtt ctg 24

30 <210> 33
 <211> 18
 <212> ADN
 <213> Secuencia artificial

35 <220>
 <223> Cebador oligonucleotídico

35 <400> 33
 ccagtagcat tgtccgag 18

40 <210> 34
 <211> 19
 <212> ADN
 <213> Secuencia artificial

45 <220>
 <223> Cebador oligonucleotídico

50 <400> 34
 gtctctggta atgcacact 19

50 <210> 35
 <211> 18
 <212> ADN
 <213> Secuencia artificial

55 <220>
 <223> Cebador oligonucleotídico

60 <400> 35
 gtggaatgcc tgctgacc 18

65 <210> 36
 <211> 18
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

 <400> 36
 5 aggggtgccc tctgagat 18

 <210> 37
 <211> 18
 <212> ADN
 10 <213> Secuencia artificial

 <220>
 <223> Cebador oligonucleotídico

 <400> 37
 15 cgagatcgcc aagatgtt 18

 <210> 38
 <211> 18
 20 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Cebador oligonucleotídico

 <400> 38
 25 aggccaacac acaacgtc 18

 <210> 39
 30 <211> 19
 <212> ADN
 <213> Secuencia artificial

 <220>
 35 <223> Cebador oligonucleotídico

 <400> 39
 agcctcgaac aattgaaga 19

 <210> 40
 40 <211> 27
 <212> ADN
 <213> Secuencia artificial

 <220>
 45 <223> Cebador oligonucleotídico

 <400> 40
 50 atcgactgtg taaacaacta gagaaga 27

 <210> 41
 <211> 20
 <212> ADN
 <213> Secuencia artificial

 <220>
 55 <223> Cebador oligonucleotídico

 <400> 41
 60 tttaagaggg caatggaagg 20

 <210> 42
 <211> 18
 <212> ADN
 65 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

5 <400> 42
 tgccgcagaa ctacttg 18

10 <210> 43
 <211> 20
 <212> ADN
 <213> Secuencia artificial

15 <220>
 <223> Cebador oligonucleotídico

15 <400> 43
 cctcagatga tgcctatcca 20

20 <210> 44
 <211> 19
 <212> ADN
 <213> Secuencia artificial

25 <220>
 <223> Cebador oligonucleotídico

25 <400> 44
 cagcaagcga tggcatagt 19

30 <210> 45
 <211> 17
 <212> ADN
 <213> Secuencia artificial

35 <220>
 <223> Cebador oligonucleotídico

35 <400> 45
 aatgccaccg aagcctc 17

40 <210> 46
 <211> 23
 <212> ADN
 <213> Secuencia artificial

45 <220>
 <223> Cebador oligonucleotídico

50 <400> 46
 tcgaactgaa ggctatttac gag 23

50 <210> 47
 <211> 18
 <212> ADN
 <213> Secuencia artificial

55 <220>
 <223> Cebador oligonucleotídico

60 <400> 47
 gtcgaagccg caattagg 18

65 <210> 48
 <211> 20
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

 <400> 48
 5 caaacgtgtg ttctggaagg 20

 <210> 49
 <211> 20
 <212> ADN
 10 <213> Secuencia artificial

 <220>
 <223> Cebador oligonucleotídico

 <400> 49
 15 tgccctgtat gatgtcagga 20

 <210> 50
 <211> 20
 20 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Cebador oligonucleotídico
 25
 <400> 50
 gtgaggggtg tcagctcagt 20

 <210> 51
 30 <211> 22
 <212> ADN
 <213> Secuencia artificial

 <220>
 35 <223> Cebador oligonucleotídico

 <400> 51
 tggggcagtt ctgtattact tc 22

 <210> 52
 40 <211> 24
 <212> ADN
 <213> Secuencia artificial

 <220>
 45 <223> Cebador oligonucleotídico

 <400> 52
 50 cgatggtttt gtacaagatt tctc 24

 <210> 53
 <211> 17
 <212> ADN
 <213> Secuencia artificial
 55
 <220>
 <223> Cebador oligonucleotídico

 <400> 53
 60 gcaaatacctt gggcaga 17

 <210> 54
 <211> 20
 <212> ADN
 65 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

5 <400> 54
 gccgtacagt tccacaaagg 20

10 <210> 55
 <211> 22
 <212> ADN
 <213> Secuencia artificial

15 <220>
 <223> Cebador oligonucleotídico

15 <400> 55
 gacgcttcct atcactctat tc 22

20 <210> 56
 <211> 22
 <212> ADN
 <213> Secuencia artificial

25 <220>
 <223> Cebador oligonucleotídico

25 <400> 56
 ttctccatc aagagtcaa ca 22

30 <210> 57
 <211> 18
 <212> ADN
 <213> Secuencia artificial

35 <220>
 <223> Cebador oligonucleotídico

35 <400> 57
 gggcacatcc agatgttt 18

40 <210> 58
 <211> 19
 <212> ADN
 <213> Secuencia artificial

45 <220>
 <223> Cebador oligonucleotídico

50 <400> 58
 gggctctgcac agactgcat 19

50 <210> 59
 <211> 20
 <212> ADN
 <213> Secuencia artificial

55 <220>
 <223> Cebador oligonucleotídico

60 <400> 59
 tccttgaat ggggagacca 20

65 <210> 60
 <211> 23
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

 <400> 60
 5 acttgggata tgtgaataag acc 23

 <210> 61
 <211> 20
 <212> ADN
 10 <213> Secuencia artificial

 <220>
 <223> Cebador oligonucleotídico

 <400> 61
 15 ggggaaagac aaagttcca 20

 <210> 62
 <211> 20
 20 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Cebador oligonucleotídico
 25
 <400> 62
 actgtctggg tccatggcta 20

 <210> 63
 30 <211> 18
 <212> ADN
 <213> Secuencia artificial

 <220>
 35 <223> Cebador oligonucleotídico

 <400> 63
 ggatttcgtg gtgggttc 18

 <210> 64
 40 <211> 20
 <212> ADN
 <213> Secuencia artificial

 <220>
 45 <223> Cebador oligonucleotídico

 <400> 64
 50 ccacagtctg tgataaacgg 20

 <210> 65
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 55
 <220>
 <223> Cebador oligonucleotídico

 <400> 65
 60 ccatcaacat tctcttatg aacg 24

 <210> 66
 <211> 20
 <212> ADN
 65 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

5 <400> 66
 atcaactccc aaacggtcac 20

<210> 67
 <211> 20
 <212> ADN
 10 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

15 <400> 67
 gcccttacac atcgagaaac 20

<210> 68
 <211> 18
 20 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

25 <400> 68
 gacttcaggg tgctggac 18

<210> 69
 <211> 22
 30 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

35 <400> 69
 tgtgaagcca gcaatatgta tc 22

<210> 70
 <211> 21
 40 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

<400> 70
 50 tattgggagg caggaggttt a 21

<210> 71
 <211> 20
 <212> ADN
 <213> Secuencia artificial

55 <220>
 <223> Cebador oligonucleotídico

<400> 71
 60 ctgagttcat gttgctgacc 20

<210> 72
 <211> 19
 <212> ADN
 65 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

 <400> 72
 5 gacagctact attcccgtt 19

 <210> 73
 <211> 21
 <212> ADN
 10 <213> Secuencia artificial

 <220>
 <223> Cebador oligonucleotídico

 <400> 73
 15 tatgtgagta agctcggaga c 21

 <210> 74
 <211> 17
 20 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Cebador oligonucleotídico
 25
 <400> 74
 agtgggcatc ccgtaga 17

 <210> 75
 30 <211> 19
 <212> ADN
 <213> Secuencia artificial

 <220>
 35 <223> Cebador oligonucleotídico

 <400> 75
 agtggacatg cgagtggag 19

 <210> 76
 40 <211> 18
 <212> ADN
 <213> Secuencia artificial

 <220>
 45 <223> Cebador oligonucleotídico

 <400> 76
 50 caccgctgga aactgaac 18

 <210> 77
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 55
 <220>
 <223> Cebador oligonucleotídico

 <400> 77
 60 cgtgcacatc catgacctt 19

 <210> 78
 <211> 17
 <212> ADN
 65 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

5 <400> 78
 gaggagatga ccttgcc 17

<210> 79
 <211> 18
 <212> ADN
 10 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

15 <400> 79
 gccatagcca ctgccact 18

<210> 80
 <211> 17
 20 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

25 <400> 80
 cttcgactgg actctgt 17

<210> 81
 30 <211> 23
 <212> ADN
 <213> Secuencia artificial

<220>
 35 <223> Cebador oligonucleotídico

<400> 81
 cagacatgtt ggtattgcac att 23

40 <210> 82
 <211> 20
 <212> ADN
 <213> Secuencia artificial

45 <220>
 <223> Cebador oligonucleotídico

<400> 82
 50 aggcgatcct gggaaattat 20

<210> 83
 <211> 19
 <212> ADN
 <213> Secuencia artificial

55 <220>
 <223> Cebador oligonucleotídico

<400> 83
 60 cccatttgtc tgtcttcac 19

<210> 84
 <211> 18
 <212> ADN
 65 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

5 <400> 84
 ctgatgggtg aggctgtt 18

<210> 85
 <211> 19
 <212> ADN
 10 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

15 <400> 85
 cgcaactccag cacctagac 19

<210> 86
 <211> 21
 20 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

25 <400> 86
 tcacagggtc aaactccag t 21

<210> 87
 <211> 21
 30 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

35 <400> 87
 gatggtagag ttccagtgt t 21

40 <210> 88
 <211> 18
 <212> ADN
 <213> Secuencia artificial

45 <220>
 <223> Cebador oligonucleotídico

<400> 88
 50 tctggtcacg cagggcaa 18

<210> 89
 <211> 20
 <212> ADN
 <213> Secuencia artificial

55 <220>
 <223> Cebador oligonucleotídico

<400> 89
 60 acacagatga tggagatgtc 20

<210> 90
 <211> 25
 <212> ADN
 65 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

 <400> 90
 5 agtagctaca tctccaggtt ctctg 25

 <210> 91
 <211> 21
 <212> ADN
 10 <213> Secuencia artificial

 <220>
 <223> Cebador oligonucleotídico

 15 <400> 91
 cggattttat caacgatgca g 21

 <210> 92
 <211> 19
 20 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Cebador oligonucleotídico
 25
 <400> 92
 catttgccgt ccttcacg 19

 <210> 93
 30 <211> 20
 <212> ADN
 <213> Secuencia artificial

 <220>
 35 <223> Cebador oligonucleotídico

 <400> 93
 gcaggtcaaa actctcaaag 20

 40 <210> 94
 <211> 20
 <212> ADN
 <213> Secuencia artificial

 45 <220>
 <223> Cebador oligonucleotídico

 <400> 94
 50 agcgggcttc tgtaatctga 20

 <210> 95
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 55
 <220>
 <223> Cebador oligonucleotídico

 <400> 95
 60 gcctcagatt tcaactcg 19

 <210> 96
 <211> 21
 <212> ADN
 65 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico

 <400> 96
 5 ctgctgagaa tcaaagtgga a 21

 <210> 97
 <211> 19
 <212> ADN
 10 <213> Secuencia artificial

 <220>
 <223> Cebador oligonucleotídico

 15 <400> 97
 ggaacaaaact gctctgcca 19

 <210> 98
 <211> 22
 20 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Cebador oligonucleotídico
 25
 <400> 98
 acagctcttt agcatttgga ga 22

 <210> 99
 30 <211> 23
 <212> ADN
 <213> Secuencia artificial

 <220>
 35 <223> Cebador oligonucleotídico

 <400> 99
 gggactatca atgttgga ctc 23

 40 <210> 100
 <211> 21
 <212> ADN
 <213> Secuencia artificial

 45 <220>
 <223> Cebador oligonucleotídico

 <400> 100
 50 cacacagttc actgtccac a 21

REIVINDICACIONES

1. Método de clasificación de un subtipo intrínseco de cáncer de mama en una muestra de prueba que comprende:

a) obtener un perfil de expresión génica para cada una de una pluralidad de muestras de cáncer de mama de entrenamiento que se han clasificado según el subtipo intrínseco de cáncer de mama, en el que cada uno de los subtipos intrínsecos luminal A (LumA), luminal B (LumB), de tipo basal (basal), enriquecido en HER2 (HER2) y de tipo normal (normal) está representado en la pluralidad de muestras de cáncer de mama, y en el que el perfil de expresión génica se basa en la expresión de al menos 40 de los genes intrínsecos enumerados en la tabla 1;

b) usar un algoritmo para construir perfiles de expresión génica o "centroides" para cada uno de los subtipos intrínsecos de cáncer de mama en el conjunto de entrenamiento;

c) obtener un perfil de expresión génica para dicha muestra de prueba, en el que el perfil de expresión génica se basa en la expresión de los mismos al menos 40 genes intrínsecos enumerados en la tabla 1 tal como se usan en la etapa a);

d) comparar el perfil de expresión génica de la muestra de prueba con los centroides construidos en la etapa (b);

e) calcular la distancia del perfil de expresión génica de la muestra de prueba a cada uno de los centroides; y,

f) asignar la muestra de prueba a uno de los subtipos intrínsecos de cáncer de mama basándose en el centroide más cercano.

2. Método de predicción de respuesta a la terapia neoadyuvante y/o adyuvante en un sujeto que tiene cáncer de mama que comprende clasificar dicho sujeto según el método según la reivindicación 1, en el que el subtipo de tumor intrínseco es indicativo de la respuesta a dicha terapia, prediciendo después la respuesta.

3. Método según la reivindicación 2, en el que la terapia es terapia endocrina neoadyuvante, y el subtipo intrínseco se predice a partir de una muestra recogida de dicho sujeto tras el inicio de dicha terapia endocrina neoadyuvante.

4. Método según la reivindicación 3, en el que la muestra se recoge al menos un mes tras el inicio de dicha terapia endocrina neoadyuvante.

5. Método de predicción de desenlace en un sujeto de prueba que tiene cáncer de mama que comprende:

a) obtener un perfil de expresión génica para cada una de una pluralidad de muestras de cáncer de mama de entrenamiento que se han clasificado según el subtipo intrínseco de cáncer de mama, en el que cada uno de los subtipos intrínsecos luminal A (LumA), luminal B (LumB), de tipo basal (basal), enriquecido en HER2 (HER2) y de tipo normal (normal) está representado en la pluralidad de muestras de cáncer de mama, en el que se conoce el desenlace en dichas muestras, y en el que el perfil de expresión génica se basa en la expresión de al menos 40 de los genes intrínsecos enumerados en la tabla 1;

b) usar un algoritmo para construir centroides para cada uno de los subtipos intrínsecos de cáncer de mama en el conjunto de entrenamiento;

c) calcular una puntuación de riesgo para cada muestra de cáncer de mama de entrenamiento ajustando un modelo a los perfiles de expresión génica obtenidos en la etapa (a), en el que el modelo representa la similitud de cada perfil de expresión génica con los centroides construidos en la etapa (b);

d) estratificar las puntuaciones de riesgo obtenidas en la etapa (c) en múltiples grupos de riesgo;

e) obtener un perfil de expresión génica a partir de una muestra de cáncer de mama de dicho sujeto de prueba, en el que el perfil de expresión génica se basa en la expresión de los mismos al menos 40 genes intrínsecos enumerados en la tabla 1 tal como se usan en la etapa a);

f) comparar el perfil de expresión génica de la muestra de prueba con los centroides construidos en la etapa (b);

g) calcular la distancia del perfil de expresión génica de la muestra de prueba a cada uno de los

centroides;

h) calcular una puntuación de riesgo usando el modelo de la etapa (c) y la distancia a cada centroide obtenido en la etapa (f); y

i) determinar en qué grupo de riesgo se encuentra dicho sujeto de prueba comparando la puntuación de riesgo obtenida en la etapa (h) con los grupos de riesgo asignados en la etapa (d), en el que la asignación a un grupo de riesgo más bajo indica un desenlace más favorable, y la asignación a un grupo de riesgo más alto indica un desenlace menos favorable.

6. Método según la reivindicación 5, en el que

(i) la puntuación de riesgo se calcula usando una suma ponderada de variables independientes; o

(ii) el desenlace es supervivencia específica a cáncer de mama, supervivencia libre de acontecimiento, o respuesta a la terapia; o

(iii) el algoritmo se entrena usando los datos de expresión génica obtenidos en la reivindicación 5; o

(iv) dicho modelo correlaciona además el desenlace con una o más variables clínicas.

7. Método según la reivindicación 6(iv), en el que dichas variables clínicas se seleccionan del grupo que consiste en tamaño del tumor, estado de los ganglios, grado histológico, estado del receptor de la hormona estrógeno, estado del receptor de la hormona progesterona, niveles de HER-2 y ploidía tumoral.

8. Método según la reivindicación 6(ii), en el que el desenlace es respuesta a la terapia y se mide mediante uno o más criterios de valoración de terapia.

9. Método según la reivindicación 8, en el que dicho uno o más criterios de valoración de terapia se seleccionan de respuesta clínica (criterios RECIST), tamaño del tumor patológico, valor de Ki67 dicotomizado o acontecimiento de recidiva.

10. Método según la reivindicación 1 ó 5, en el que

(i) dicho algoritmo es un algoritmo del centroide más cercano; o

(ii) el perfil de expresión génica para la muestra de prueba se obtiene usando reacción en cadena de la polimerasa cuantitativa (qPCR) en tiempo real y transcripción inversa con cebadores específicos para cada uno de dichos genes intrínsecos; o

(iii) el perfil de expresión génica para la muestra de prueba se obtiene usando análisis de micromatriz con sondas específicas para un producto de expresión de cada uno de dichos genes intrínsecos; o

(iv) los datos obtenidos a partir de los perfiles de expresión génica para las muestras de entrenamiento y el perfil de expresión génica para la muestra de prueba se procesan por medio de métodos de normalización antes del análisis.

11. Método según la reivindicación 1, en el que los centroides se construyen a partir de los datos de expresión génica depositados como número de registro GSE10886 en el National Center for Biotechnology Information Gene Expression Omnibus.

12. Método según la reivindicación 10(iv), en el que dicho procesamiento comprende normalización para dar un conjunto de genes de mantenimiento.

13. Método según la reivindicación 12, en el que dichos genes de mantenimiento se seleccionan de MRPL19, PSMC4, SF3A1, PUMI, ACTB, TFRC, GUSB, RPLPO y/o GAPD.

14. Kit que consiste en reactivos suficientes para la medición de los niveles de expresión de entre 40 y 50 de los genes intrínsecos enumerados en la tabla 1.

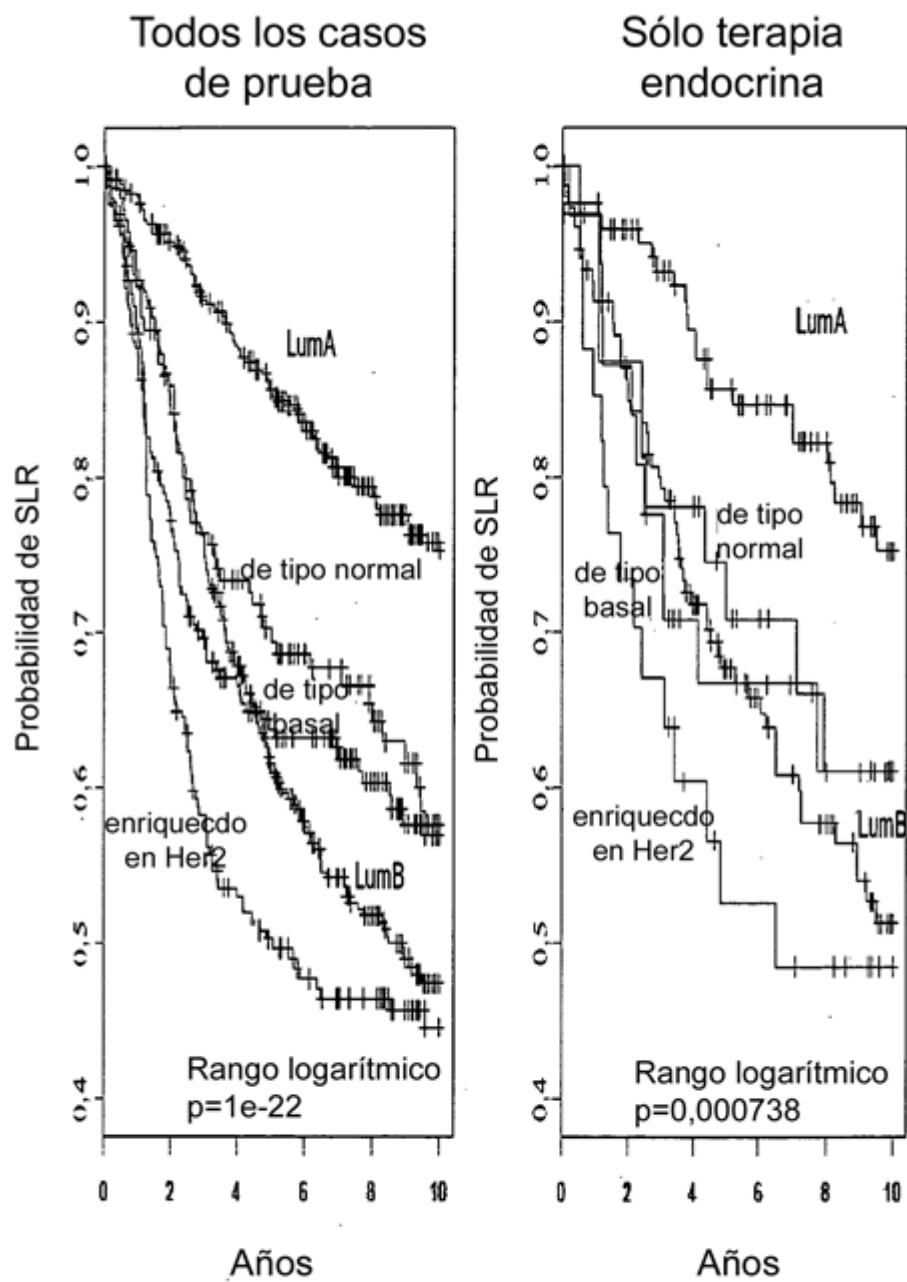


FIG. 1

Negativo para ganglios,
sin terapia sistémica adyuvante

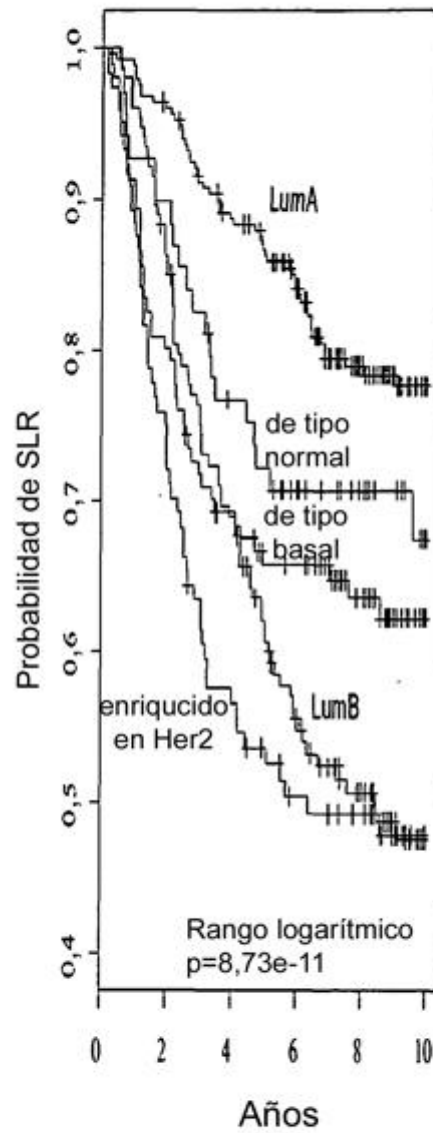


FIG. 1
(continuación)

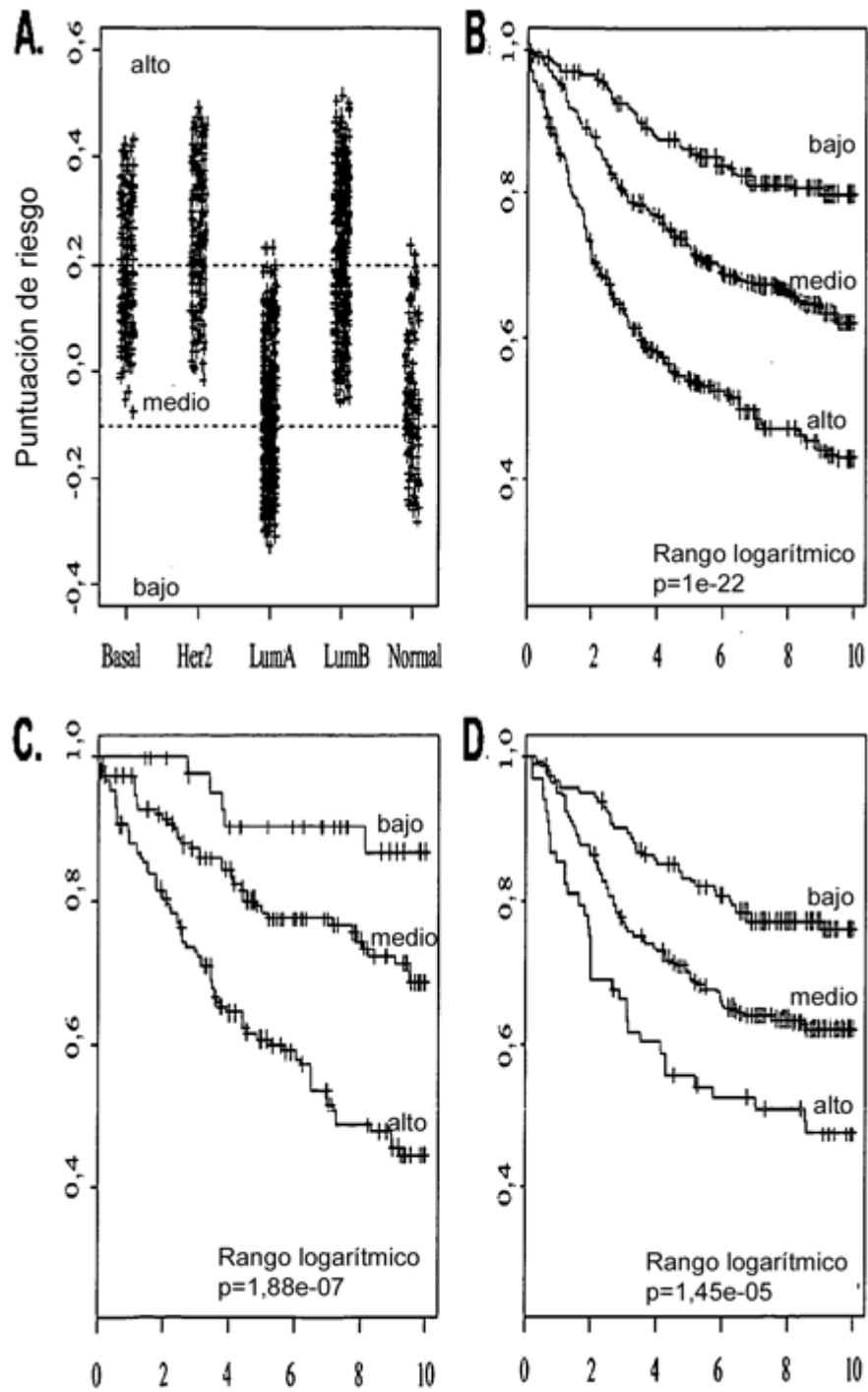


FIG. 2

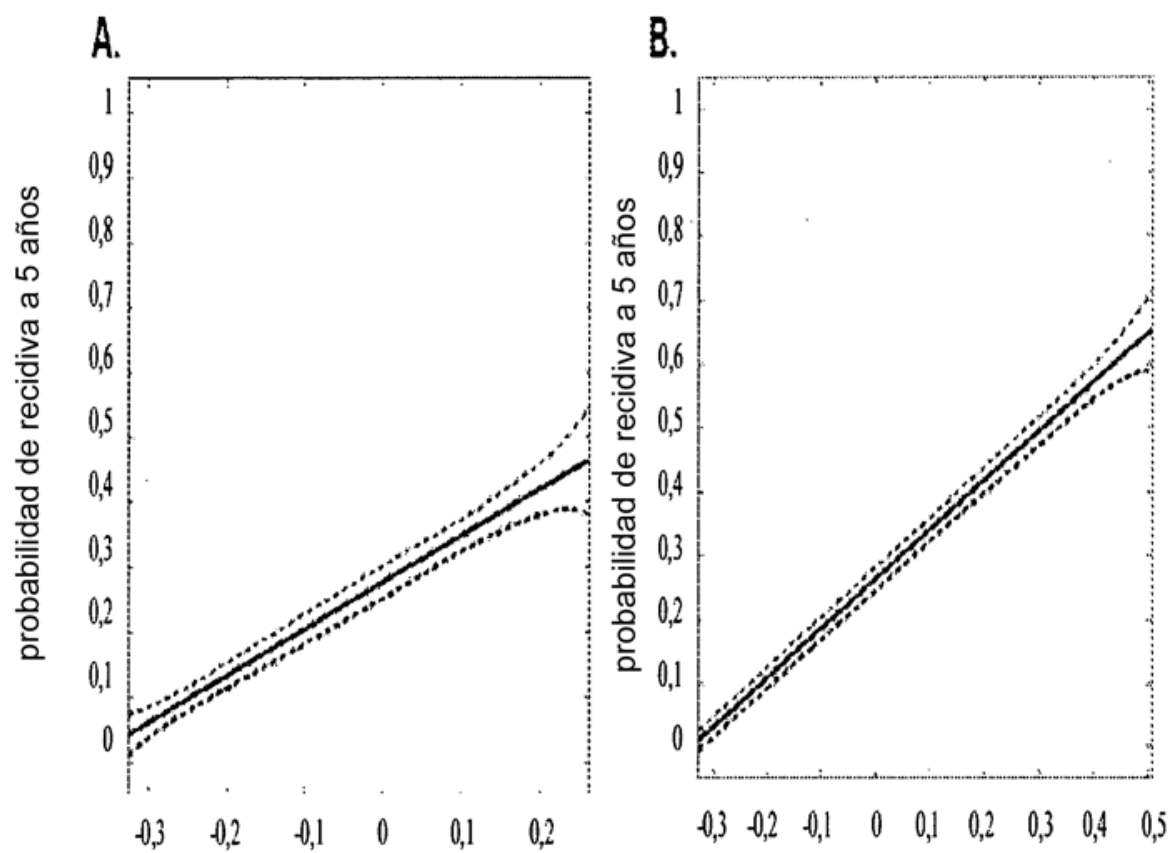


FIG. 3

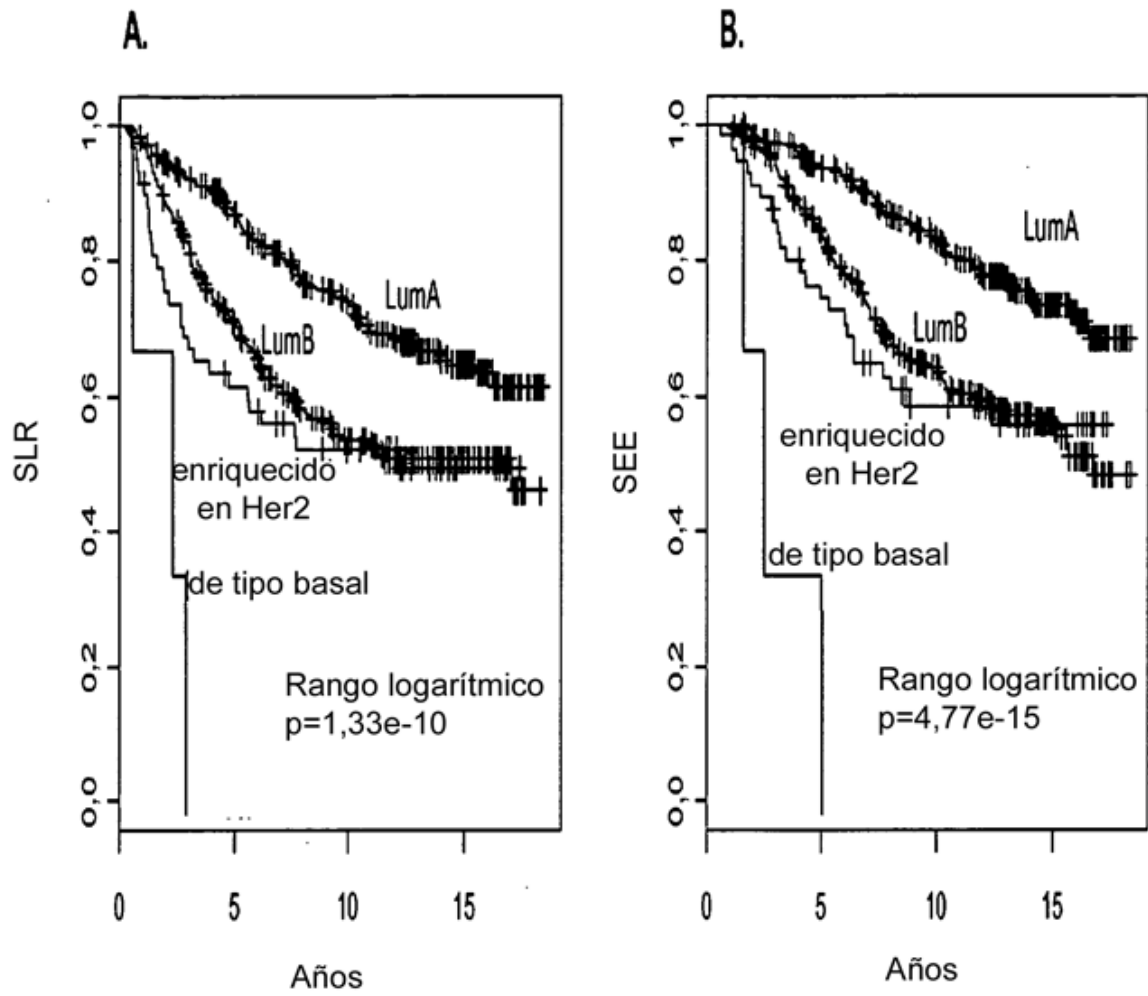
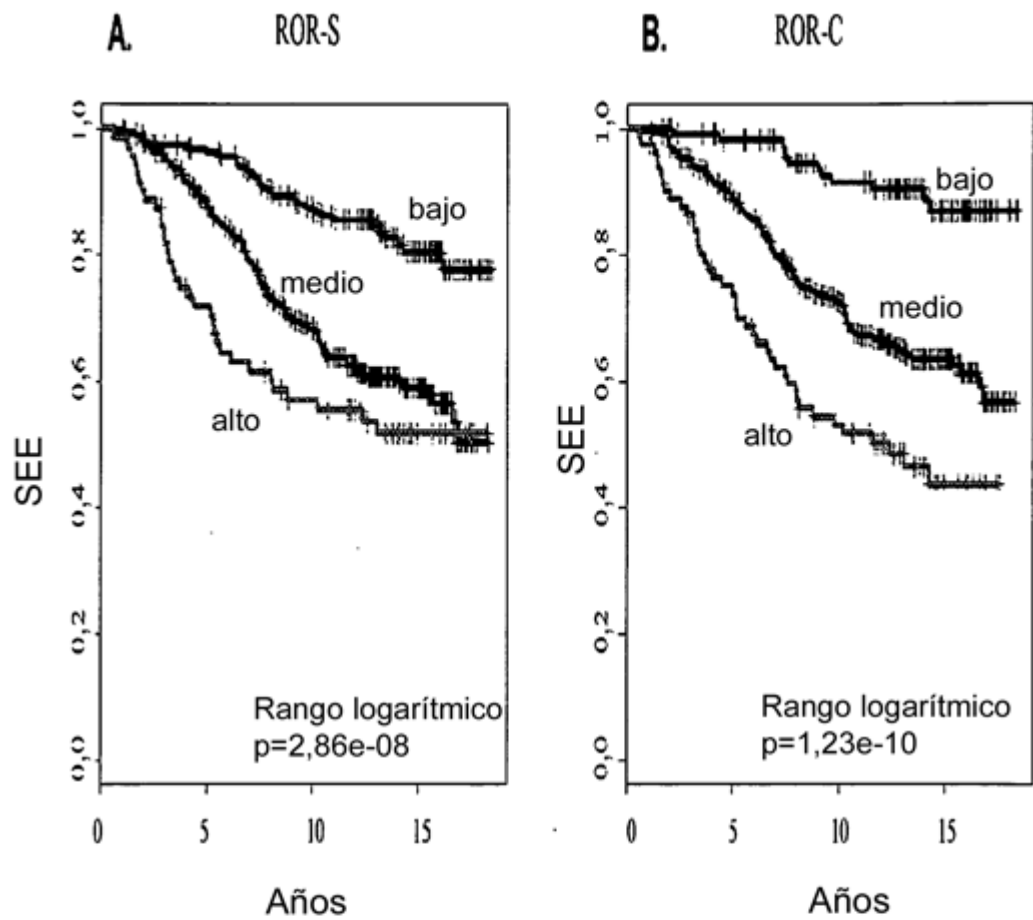
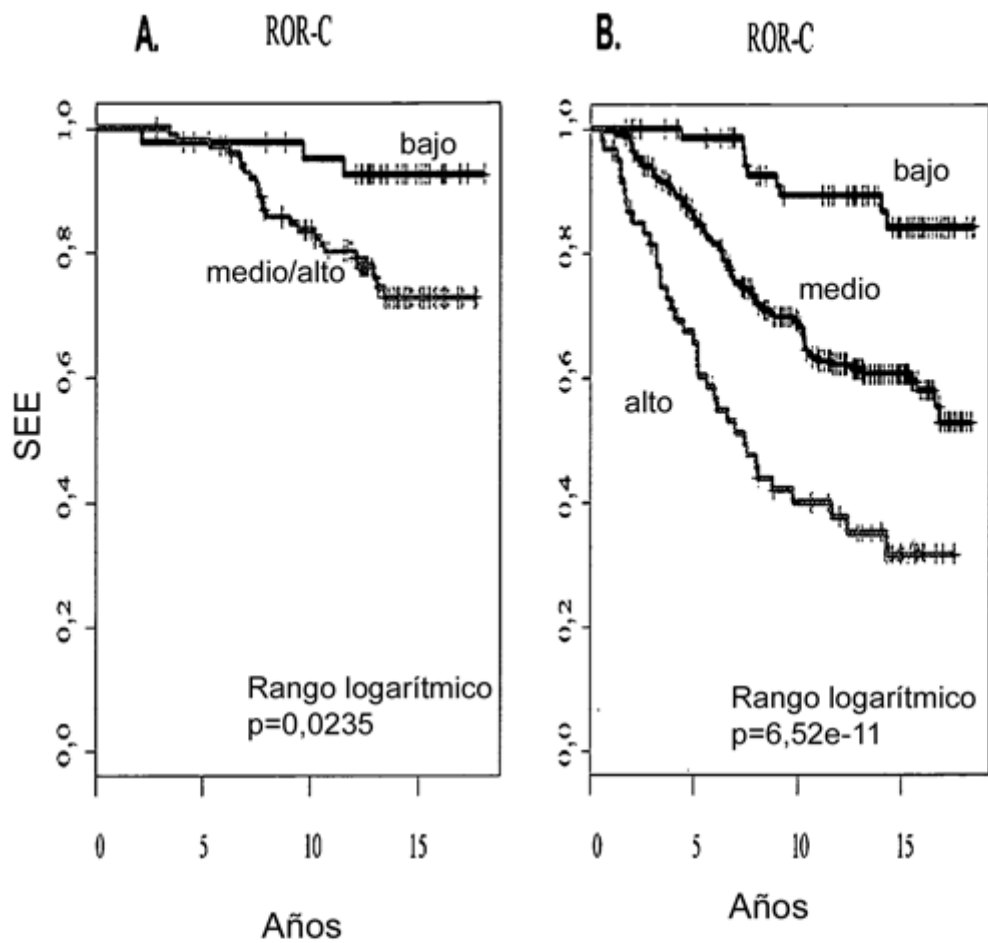


FIG. 4



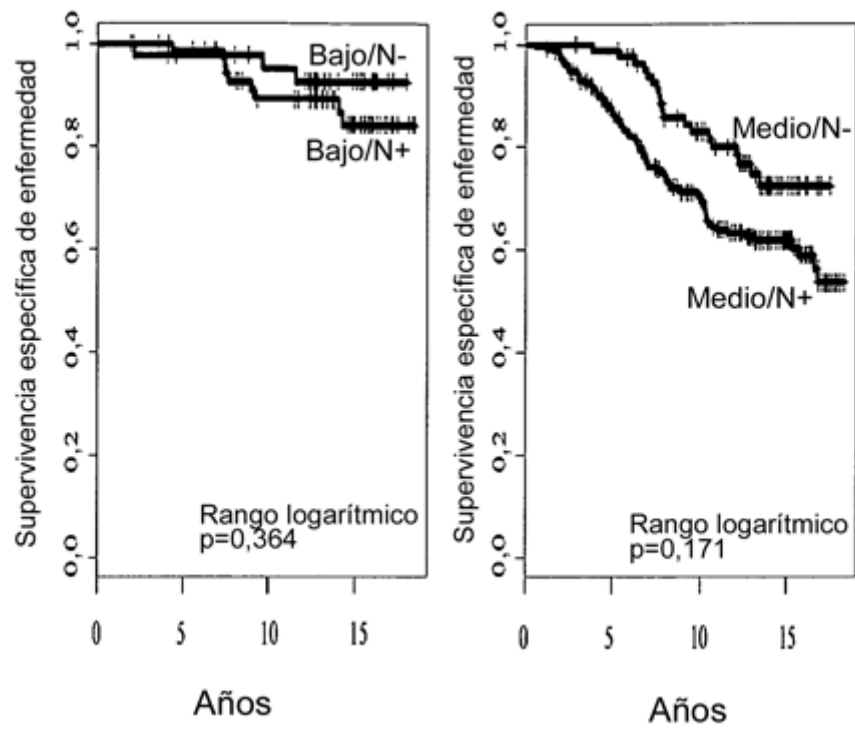
IC del 95% acontecimientos/N a 10 años			IC del 95% acontecimientos/N a 10 años		
Bajo	82-92%	31/93	Bajo	87-97%	12/124
Medio	63-74%	117/330	Medio	68-78%	108/338
Alto	45-69%	33/72	Alto	43-65%	42/83

FIG. 5



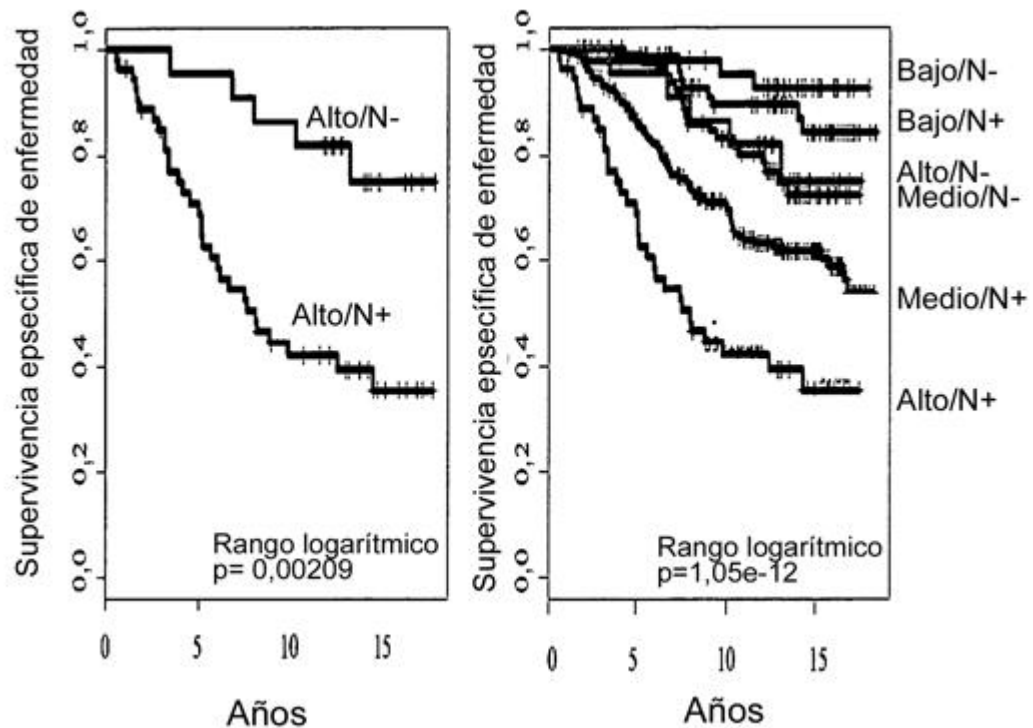
	IC del 95% acontecimientos/N a 10 años		IC del 95% acontecimientos/N a 10 años	
Bajo	89-100%	3/43	Bajo	82-97% 9/81
Medio/alto	75-91%	24/105	Medio	63-75% 89/255
			Alto	29-55% 37/61

FIG. 6



	IC del 95% a 10 años	acontecimientos/ N		IC del 95% a 10 años	acontecimientos/ N
Bajo/N-	89-100%	3/43	Medio/N-	73-91%	19/83
Bajo/N+	82-97%	9/80	Medio/N+	64-76%	84/245

FIG. 7



IC del 95% a 10 años			IC del 95% a 10 años		
		acontecimientos/ N			acontecimientos/ N
Alto/N-	67-100%	5/22	Bajo/N-	89-100%	3/43
Alto/N+	31-59%	31/54	Bajo/N+	82-97%	9/80
			Medio/N-	73-91%	19/83
			Medio/N+	64-76%	84/245
			Alto/N-	67-100%	5/22
			Alto/N+	31-59%	31/54

FIG. 7 (continuación)

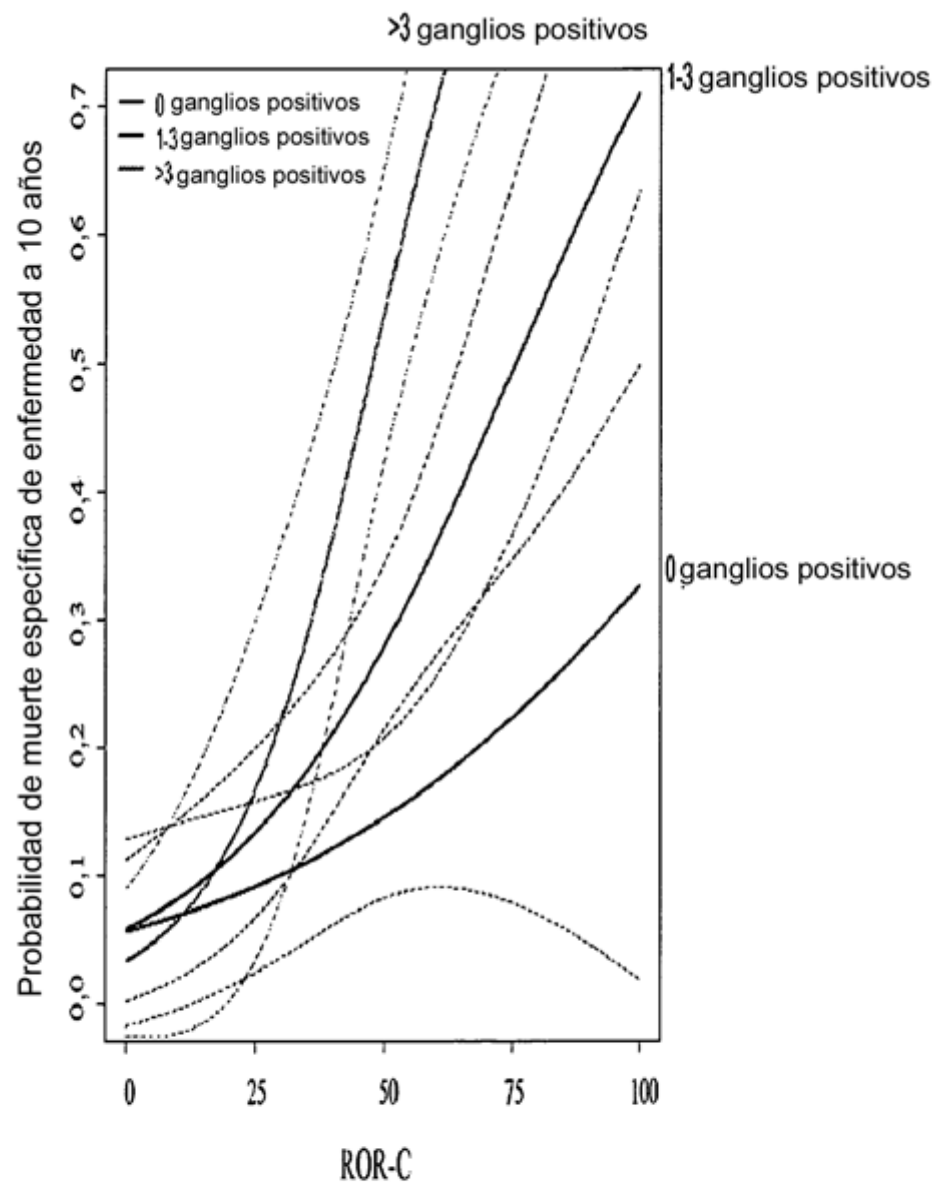


FIG. 8

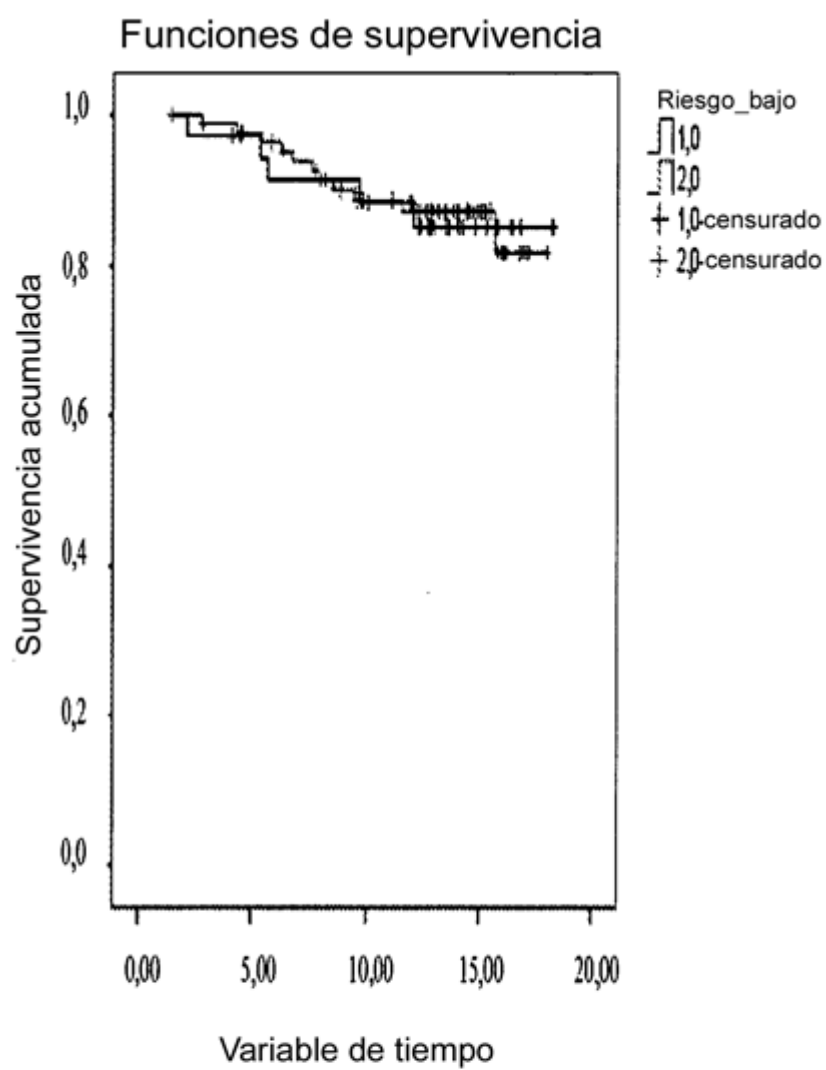


FIG. 9

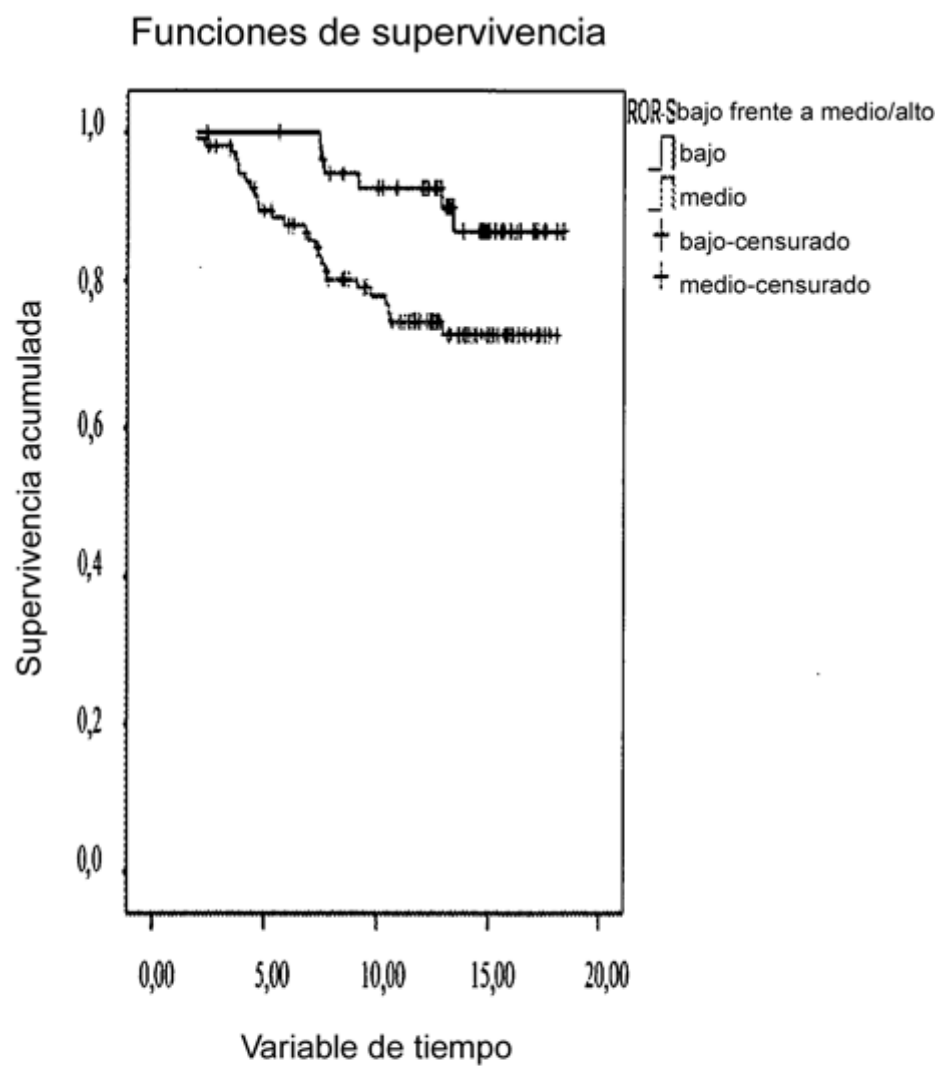


FIG. 10A

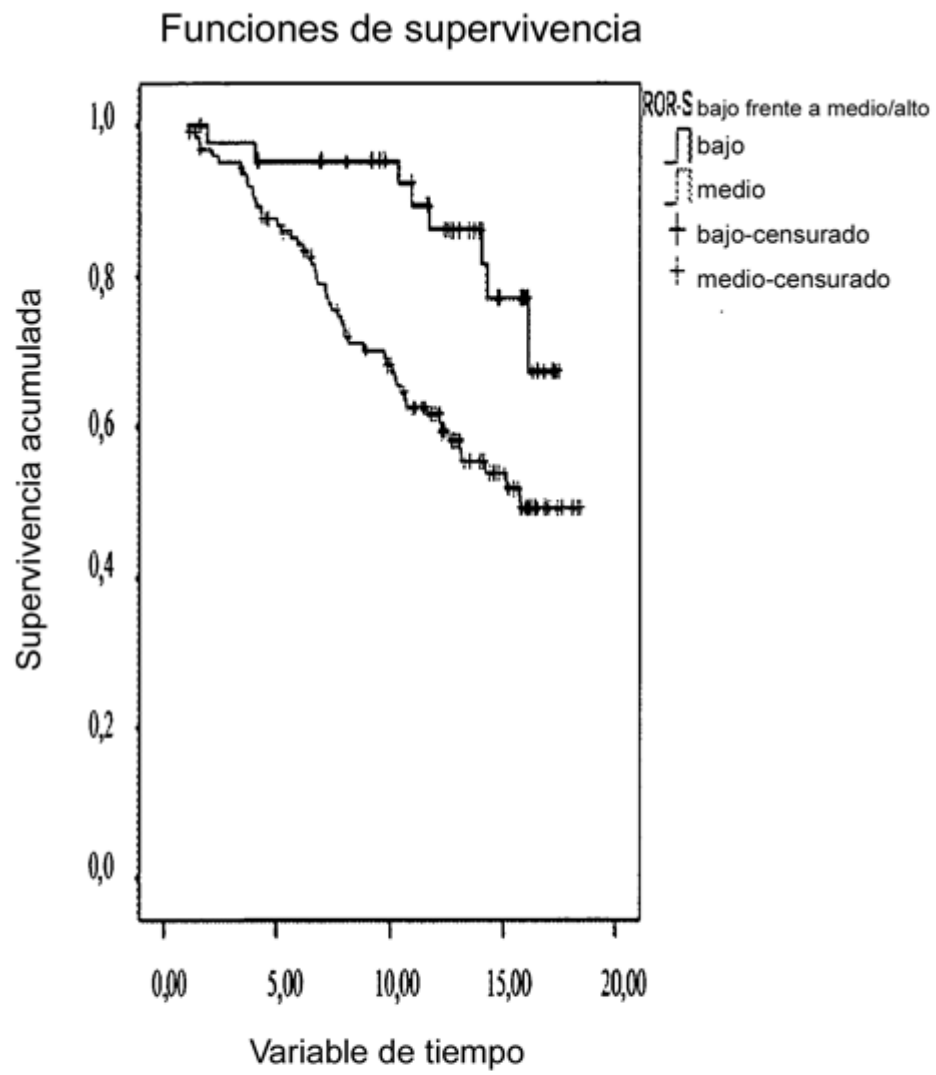


FIG. 10B

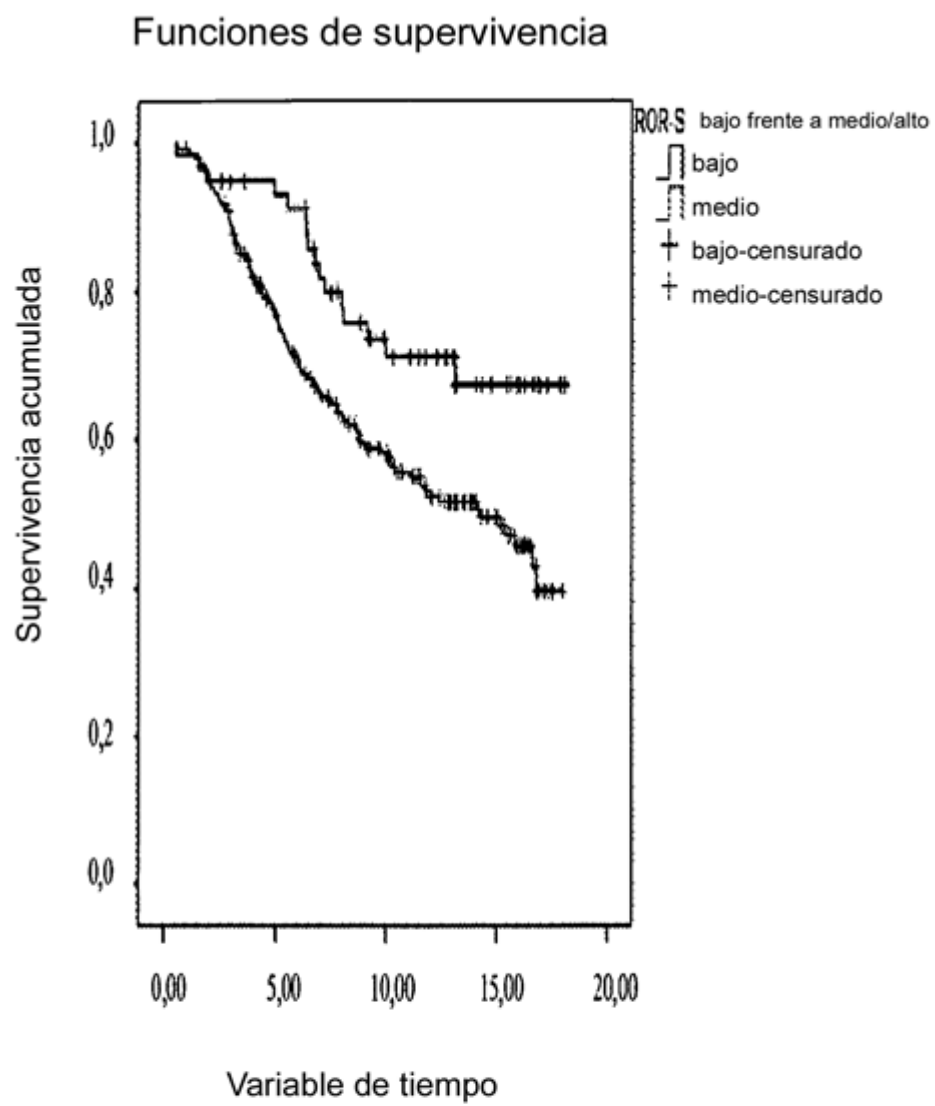


FIG. 10C