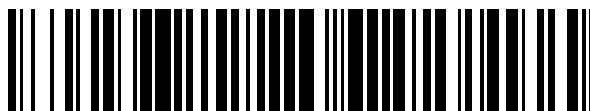


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 753 625**

51 Int. Cl.:

C12Q 1/6886 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **18.09.2015 PCT/EP2015/071524**

87 Fecha y número de publicación internacional: **24.03.2016 WO16042164**

96 Fecha de presentación y número de la solicitud europea: **18.09.2015 E 15778623 (7)**

97 Fecha y número de publicación de la concesión europea: **14.08.2019 EP 3194621**

54 Título: **Un método para predecir el riesgo de recaída del cáncer**

30 Prioridad:

19.09.2014 EP 14185673

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

13.04.2020

73 Titular/es:

**THE PROVOST, FELLOWS, FOUNDATION
SCHOLARS, & THE OTHER MEMBERS OF
BOARD, OF THE COLLEGE OF THE HOLY &
UNDIV. TRINITY OF QUEEN (50.0%)
near Dublin, College Green
Dublin D2, IE y
UNIVERSITY COLLEGE DUBLIN NATIONAL
UNIVERSITY OF IRELAND, DUBLIN (50.0%)**

72 Inventor/es:

**BRACKEN, ADRIAN;
LANIGAN, FIONA y
GALLAGHER, WILLIAM**

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

ES 2 753 625 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

Un método para predecir el riesgo de recaída del cáncer

5 **Campo de la invención**

La invención se refiere a un método para predecir el riesgo de recaída de un tumor en un sujeto. Específicamente, la invención se refiere a un método para predecir el riesgo de cáncer de mama sin afectación ganglionar en fase precoz, cáncer de próstata y otras recaídas de tumores.

10

Antecedentes de la invención

El cáncer de mama es una enfermedad heterogénea que presenta desafíos para los médicos a la hora de predecir la probabilidad de progresión de la enfermedad, en particular en pacientes en los que la enfermedad se detecta en las fases precoces. Para estas mujeres, los parámetros clínico-patológicos convencionales (tamaño del tumor, estado de los ganglios linfáticos, edad del paciente, grado del tumor y expresión de biomarcadores, incluyendo el receptor estrogénico (ER), el receptor de la progesterona (PR), el receptor del factor de crecimiento epidérmico humano 2 (Her2) , Ki67) no son suficientes para caracterizar la complejidad de la enfermedad y predecir con exactitud la probabilidad de recaída del tumor después del tratamiento adyuvante o la extirpación del tumor mediante cirugía. Por lo tanto, debido a una estratificación de riesgo inexacta, muchas de estas pacientes que inherentemente tienen un riesgo bajo de recaída son asignadas a recibir quimioterapia, cuando, de hecho, la mayoría de estas mujeres permanecerían sin cáncer incluso sin este tratamiento tóxico.

15

20

25

30

De hecho, se estima que, para la enfermedad sin afectación ganglionar y positiva para ER, hasta el 85 % de las pacientes estarían sobretatadas si recibieran quimioterapia (Fisher *et al.*, 2004). Además, las pacientes supervivientes tratadas con quimioterapia enfrentan un riesgo más alto de desarrollar un segundo cáncer primario independiente en tejidos no relacionados a lo largo de su vida (Boffetta y Kaldor, 1994). Teniendo en cuenta los efectos secundarios graves, la carga para la salud pública y las futuras implicaciones para la salud de la quimioterapia, el sobretatamiento de pacientes representa un problema importante en el tratamiento clínico del cáncer de mama en fase precoz.

35

40

45

El desafío es desarrollar un método para distinguir con exactitud y de forma reproducible a las pacientes de bajo riesgo de las de alto riesgo de modo que el tratamiento se pueda asignar en consecuencia. Las directrices actuales a menudo dan lugar a opiniones divergentes de los oncólogos de mama en cuanto a si asignar tratamiento neoadyuvante y/o adyuvante, ya que muchos son reacios a renunciar al tratamiento neoadyuvante y/o adyuvante sin una evaluación fiable del riesgo de recaída. La adición de biomarcadores pronósticos y predictivos más exactos y fiables a la evaluación clínica estándar mejoraría enormemente la capacidad tanto los médicos como los pacientes para tomar decisiones sobre el tratamiento mejor fundamentadas. Se está avanzando a este respecto con los ensayos multigénicos Oncotype Dx® Breast Cancer Assay y MammaPrint™, que actualmente se están evaluando en el ensayo de asignación de opciones individualizadas de tratamiento (*Trial Assigning Individualized Options for Treatment* (Rx)) (TAILORx) y los ensayos de micromatrices en la enfermedad sin afectación ganglionar y con afectación de 1 a 3 ganglios linfáticos que pueden evitar la quimioterapia (*Microarray In Node-negative and 1 to 3 positive lymph node Disease may Avoid ChemoTherapy*, MINDACT), respectivamente (Cardoso *et al.*, 2008; Sparano, 2006). MammaPrint™ y Prosigna™ son ejemplos de pruebas pronósticas aprobadas por la Agencia de Alimentos y Medicamentos en este campo.

50

55

El documento WO 2005/039382 describe una serie de conjuntos de genes usados para predecir la probabilidad de recaída del cáncer de mama, también conocida como Oncotype Dx® mencionada anteriormente. La invención está relacionada con un conjunto de genes que comprende 'uno o más' genes de un panel de 50 genes. El documento WO 2104/130825 describe un conjunto de genes que comprende al menos 4 genes de un panel de genes del ciclo celular para detectar el riesgo de cáncer de pulmón. El documento US 7914988 describe una firma de expresión génica para predecir la recidiva en el cáncer de próstata, conocida como puntuación GEX. La invención está relacionada con un conjunto de genes que comprende 'todos o una subcombinación de' genes de un panel de 21 genes.

60

65

El documento WO 2014/066796 describe firmas de expresión génica relacionadas con el pronóstico, el tratamiento y el control del cáncer, específicamente el cáncer de mama. Enumera los genes del sistema inmunitario (ISG) y otros genes de pronóstico del cáncer (OCPG). El grupo de ISG se divide en genes relacionados con linfocitos B, linfocitos T y HLA de clase II, donde una expresión más alta se asocia con un buen pronóstico. El documento WO 2012/030840 describe métodos para determinar la expresión de un panel de 4 CCG seleccionados de una lista de 1299 genes y proporcionar un valor de prueba. Los genes también se pueden usar para el pronóstico del cáncer de próstata, pulmón, vejiga o cerebro; mientras que ningún aumento en la expresión génica se correlaciona con el cáncer de escasa malignidad/no invasor. El documento US 2008/275652 describe métodos para mejorar el pronóstico del cáncer sometiendo a ensayo entre 2 y 90 genes para determinar cuáles están regulados por incremento. Los genes preferentes se limitan a ocho, que incluyen FOXM1 y MYBL2. El documento WO 2013/011479 describe un método para evaluar la(s) vía(s) de señalización celular en una muestra de tejido. D8

describe el ensayo de genes de la vía de Hedgehog, que incluyen FOXM1, en diversas muestras y tipos de cáncer. Ivanesa Pardo *et al.* (Breast Cancer Research, Current Science; vol. 16(2); pp. R26 (2014)) describen el análisis de los niveles de expresión de 17.995 genes en muestras de cáncer de mama. Hui R. *et al.* (Clinical Cancer Research, The American Association for Cancer Research; vol. 6(1); pp. 2777-2787 (2000)) describen que p16INK4A es un marcador de mal pronóstico en el cáncer de mama. Yao Geng *et al.* (Cancer Science; vol. 10492); pp. 194-199 (2013)) describen que los niveles de UHRF1 en plasma son altamente correlativos con el cáncer de mama y su estado y fase.

El uso generalizado de la determinación del perfil de expresión génica ha dado lugar a una rápida expansión en la identificación de firmas de expresión génica que se descubrió que se correlacionan con diferentes aspectos de la progresión tumoral. Estas incluyen las firmas de 'mal pronóstico' (van de Vijver *et al.*, 2002; Wang *et al.*, 2005), 'invasividad' (Liu *et al.*, 2007) y 'grado genómico' (Sotiriou *et al.*, 2006). El documento US 2008/275652 describe cómo esta firma de grado genómico comprende al menos 2 o 4 genes seleccionados de un panel de 97 genes. Sin embargo, a pesar de la capacidad de estas firmas para predecir el pronóstico del cáncer de mama, hay sorprendentemente poco solapamiento entre las firmas. Los solicitantes sugieren que muchos genes en estas firmas pueden ser 'pasajeros', en lugar de 'impulsores' de la progresión tumoral. Los avances recientes en ingeniería inversa hologenómica han permitido identificar con éxito las interacciones reguladoras entre los factores de transcripción y los genes en dirección 3' que eran causales en lugar de correlativas (Carro *et al.*, 2010). Uno de dichos algoritmos, el algoritmo para la reconstrucción de redes celulares exactas (ARACNe) (Margolin *et al.*, 2006), usa redes de interacción génica construidas a partir de conjuntos de datos transcriptómicos para identificar 'nodos', normalmente factores de transcripción, que se predicen para regular directamente múltiples genes en la firma.

Es un objetivo de la presente invención superar al menos uno de los problemas mencionados anteriormente.

Sumario de la invención

La predicción del riesgo de recaída de un tumor y, por tanto, la necesidad de tratamiento adyuvante para pacientes con cáncer de mama sin afectación de ganglios linfáticos (y cáncer de mama con afectación ganglionar y en fase precoz) pueden ser un problema significativo para médicos y pacientes. En el presente documento se ha identificado una 'firma de proliferación central' que está sistemáticamente alta en cultivos primarios en proliferación y está regulada por disminución durante la senescencia celular. Esta firma génica también está altamente expresada en cánceres de mama invasores. Se ha identificado una jerarquía de varios reguladores transcripcionales maestros (MTR - factores de transcripción responsables de la regulación de este conjunto central de genes) en dirección 5' de estos genes de proliferación central. Otro análisis de la expresión de estos factores en conjuntos de datos de cáncer de mama en los niveles de ARNm y proteínas revela una capacidad notable para predecir el riesgo de recaída del cáncer de mama en fase precoz. Sorprendentemente, la combinación de dos de estos factores supera a los biomarcadores clínicos usados actualmente para el riesgo de recaída del cáncer de mama, así como a los ensayos pronósticos multigénicos recientemente desarrollados, tal como Oncotype Dx®. La adición del regulador de la senescencia p16^{INK4A} al panel pronóstico de factores proliferativos permite la identificación de tumores con una vía de senescencia celular alterada, mejorando además el poder pronóstico de la invención. Además, el análisis no sesgado de la supervivencia de varios conjuntos de datos sobre el cáncer de mama ha revelado genes implicados en vías alternativas asociadas al cáncer de mama, tales como resistencia a la apoptosis, invasión y respuesta inmunitaria, que se pueden combinar con el panel de MTR para incrementar aún más el poder pronóstico. Este enfoque ideado por el solicitante ha tenido éxito en la identificación de 'impulsores' de la proliferación del cáncer que, cuando se combina con biomarcadores adicionales, tiene el potencial de convertirse en un ensayo pronóstico superior para el cáncer en fase precoz. Por tanto, mediante la identificación de los 'impulsores' o reguladores en dirección 5' de las firmas clave, se pueden identificar factores de predicción más exactos y fiables del pronóstico del cáncer de mama. El solicitante ha llamado a esta 'firma de proliferación central' OncoMasTR, y este nombre se usará en el presente documento.

De acuerdo con la invención, se proporciona un método para predecir el riesgo de recaída del cáncer en un individuo con cáncer, comprendiendo el método una etapa de someter a ensayo una muestra de cáncer del individuo para determinar la expresión positiva de al menos tres genes (o proteínas codificadas por esos genes) seleccionados del grupo que consiste en FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19, en el que la expresión positiva de al menos tres genes o proteínas codificadas por dichos genes se correlaciona con un riesgo incrementado de recaída de cáncer en comparación con un individuo con cáncer que no presenta una expresión positiva de los al menos tres genes o proteínas codificadas por esos genes; en el que los al menos tres genes, o proteínas codificadas por dichos al menos tres genes, son FOXM1, PTTG1 y ZNF367, incluyendo opcionalmente uno o más de los genes, o proteínas codificadas por dichos genes, UHRF1, E2F1, MYBL2, HMGB2, ATAD2, E2F8 y TCF19.

De acuerdo con la invención, se proporciona un método para predecir el riesgo de recaída del cáncer en un individuo con cáncer después del tratamiento con inhibidores de CDK4/6, comprendiendo el método una etapa de someter a ensayo una muestra de cáncer del individuo para determinar la expresión positiva de al menos tres genes, o proteínas codificadas por dichos genes, seleccionados del grupo que consiste en FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19, en el que la expresión positiva de los al menos tres genes, o

proteínas codificadas por dicho genes, se correlaciona con un riesgo incrementado de recaída del cáncer en un individuo con cáncer después del tratamiento con inhibidores de CDK4/6 en comparación con un individuo con cáncer que no presenta expresión positiva de los al menos tres genes o proteínas codificadas por esos genes; en el que los al menos tres genes, o proteínas codificadas por dichos al menos tres genes, son FOXM1, PTTG1 y ZNF367, incluyendo opcionalmente uno o más de los genes, o proteínas codificadas por dichos genes, UHRF1, E2F1, MYBL2, HMGB2, ATAD2, E2F8 y TCF19.

De acuerdo con la invención, se proporciona un método para determinar una tasa de supervivencia a 5 años o una tasa de supervivencia a 10 años de un individuo diagnosticado de cáncer de mama, comprendiendo el método una etapa de someter a ensayo una muestra de tumor canceroso del individuo para determinar la expresión positiva de al menos tres genes, o proteínas codificadas por esos genes, seleccionados de FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19, en el que la expresión positiva de los al menos tres genes, o proteínas codificadas por esos genes, se correlaciona con una probabilidad disminuida de una tasa de supervivencia a 5 años o una tasa de supervivencia a 10 años en comparación con un individuo con cáncer que no presenta una expresión positiva de los al menos tres genes o proteínas codificadas por esos genes; en el que los al menos tres genes, o proteínas codificadas por dichos al menos tres genes, son FOXM1, PTTG1 y ZNF367, incluyendo opcionalmente uno o más de los genes, o proteínas codificadas por dichos genes, UHRF1, E2F1, MYBL2, HMGB2, ATAD2, E2F8 y TCF19.

En un aspecto, el método comprende además la etapa de someter a ensayo la expresión del gen o proteína p16^{INK4A} además de los al menos tres genes (o proteínas), en el que la expresión regulada incorrectamente de p16^{INK4A} en combinación con la expresión positiva de los al menos tres genes (o proteínas codificadas por esos genes) se correlaciona con un riesgo incrementado de recaída del cáncer o una probabilidad disminuida de una tasa de supervivencia a 5 años o una tasa de supervivencia a 10 años, en comparación con un individuo con cáncer que no presenta una expresión regulada incorrectamente de p16^{INK4A} y expresión positiva de los al menos tres genes (o proteínas codificadas por esos genes). Los pacientes con cáncer de mama con expresión regulada incorrectamente de p16^{INK4A} y expresión positiva de los al menos tres genes (o proteínas codificadas por esos genes) tienen un riesgo incrementado de recaída del cáncer, o una probabilidad disminuida de una tasa de supervivencia a 5 años o una tasa de supervivencia a 10 años, en comparación con pacientes con cáncer que no presentan el patrón de expresión de esta combinación de dichos tres genes (o proteínas codificadas por esos genes).

En un aspecto, los al menos dos genes seleccionados son FOXM1 y UHRF1. En un aspecto, los al menos dos genes seleccionados son FOXM1 y PTTG1. En un aspecto, los al menos dos genes seleccionados son FOXM1 y E2F1. En un aspecto, los al menos dos genes seleccionados son FOXM1 y MYBL2. En un aspecto, los al menos dos genes seleccionados son FOXM1 y HMGB2. En un aspecto, los al menos dos genes seleccionados son UHRF1 y PTTG1. En un aspecto, los al menos dos genes seleccionados son UHRF1 y E2F1. En un aspecto, los al menos dos genes seleccionados son UHRF1 y MYBL2. En un aspecto, los al menos dos genes seleccionados son UHRF1 y HMGB2. En un aspecto, los al menos dos genes seleccionados son PTTG1 y E2F1. En un aspecto, los al menos dos genes seleccionados son PTTG1 y MYBL2. En un aspecto, los al menos dos genes seleccionados son PTTG1 y HMGB2. En un aspecto, los al menos dos genes seleccionados son E2F1 y MYBL2. En un aspecto, los al menos dos genes seleccionados son E2F1 y HMGB2. En un aspecto, los al menos dos genes seleccionados son MYBL2 y HMGB2. En un aspecto, los al menos dos genes seleccionados son FOXM1 y ATAD2. En un aspecto, los al menos dos genes seleccionados son FOXM1 y E2F8. En un aspecto, los al menos dos genes seleccionados son FOXM1 y ZNF367. En un aspecto, los al menos dos genes seleccionados son FOXM1 y TCF19. En un aspecto, los al menos dos genes seleccionados son UHRF1 y ATAD2. En un aspecto, los al menos dos genes seleccionados son UHRF1 y E2F8. En un aspecto, los al menos dos genes seleccionados son UHRF1 y ZNF367. En un aspecto, los al menos dos genes seleccionados son UHRF1 y TCF19. En un aspecto, los al menos dos genes seleccionados son PTTG1 y ATAD2. En un aspecto, los al menos dos genes seleccionados son PTTG1 y E2F8. En un aspecto, los al menos dos genes seleccionados son PTTG1 y ZNF367. En un aspecto, los al menos dos genes seleccionados son PTTG1 y TCF19. En un aspecto, los al menos dos genes seleccionados son E2F1 y ATAD2. En un aspecto, los al menos dos genes seleccionados son E2F1 y E2F8. En un aspecto, los al menos dos genes seleccionados son E2F1 y ZNF367. En un aspecto, los al menos dos genes seleccionados son E2F1 y TCF19. En un aspecto, los al menos dos genes seleccionados son MYBL2 y ATAD2. En un aspecto, los al menos dos genes seleccionados son MYBL2 y E2F8. En un aspecto, los al menos dos genes seleccionados son MYBL2 y ZNF367. En un aspecto, los al menos dos genes seleccionados son MYBL2 y TCF19. En un aspecto, los al menos dos genes seleccionados son HMGB2 y ATAD2. En un aspecto, los al menos dos genes seleccionados son HMGB2 y E2F8. En un aspecto, los al menos dos genes seleccionados son HMGB2 y ZNF367. En un aspecto, los al menos dos genes seleccionados son HMGB2 y TCF19. En un aspecto, los al menos dos genes seleccionados son E2F8 y ATAD2. En un aspecto, los al menos dos genes seleccionados son E2F8 y TCF19. En un aspecto, los al menos dos genes seleccionados son E2F8 y ZNF367. En un aspecto, los al menos dos genes seleccionados son ZNF367 y ATAD2. En un aspecto, los al menos dos genes seleccionados son ZNF367 y TCF19. En un aspecto, los al menos dos genes seleccionados son TCF19 y ATAD2. Preferentemente, los al menos dos genes seleccionados anteriormente se combinan con p16^{INK4A}.

En un aspecto, se seleccionan al menos tres genes y los genes seleccionados son FOXM1, UHRF1 y PTTG1. En un aspecto, los genes seleccionados son FOXM1, UHRF1 y E2F1. En un aspecto, los genes seleccionados son FOXM1, UHRF1 y MYBL2. En un aspecto, los genes seleccionados son FOXM1, UHRF1 y HMGB2. En un aspecto,

5 aspecto, los genes seleccionados son FOXM1, PTTG1, MYBL2 y TCF19. En un aspecto, los genes seleccionados son FOXM1, PTTG1, HMGB2 y ATAD2. En un aspecto, los genes seleccionados son FOXM1, PTTG1, HMGB2 y E2F8. En un aspecto, los genes seleccionados son FOXM1, PTTG1, HMGB2 y ZNF367. En un aspecto, los genes seleccionados son FOXM1, PTTG1, HMGB2 y TCF19. En un aspecto, los genes seleccionados son FOXM1, E2F1, MYBL2 y ATAD2. En un aspecto, los genes seleccionados son FOXM1, E2F1, MYBL2 y E2F8. En un aspecto, los genes seleccionados son FOXM1, E2F1, MYBL2 y ZNF367. En un aspecto, los genes seleccionados son FOXM1, E2F1, HMGB2 y ATAD2. En un aspecto, los genes seleccionados son FOXM1, E2F1, HMGB2 y E2F8. En un aspecto, los genes seleccionados son FOXM1, E2F1, HMGB2 y ZNF367. En un aspecto, los genes seleccionados son FOXM1, E2F1, HMGB2 y TCF19.

10 En un aspecto, los genes seleccionados son FOXM1, MYBL2, HMGB2 y ATAD2. En un aspecto, los genes seleccionados son FOXM1, MYBL2, HMGB2 y E2F8. En un aspecto, los genes seleccionados son FOXM1, MYBL2, HMGB2 y ZNF367. En un aspecto, los genes seleccionados son FOXM1, MYBL2, HMGB2 y TCF19. En un aspecto, los genes seleccionados son UHRF1, PTTG1, E2F1 y ATAD2. En un aspecto, los genes seleccionados son UHRF1, PTTG1, E2F1 y E2F8. En un aspecto, los genes seleccionados son UHRF1, PTTG1, E2F1 y ZNF367. En un aspecto, los genes seleccionados son UHRF1, PTTG1, E2F1 y TCF19. En un aspecto, los genes seleccionados son UHRF1, PTTG1, MYBL2 y ATAD2. En un aspecto, los genes seleccionados son UHRF1, PTTG1, MYBL2 y E2F8. En un aspecto, los genes seleccionados son UHRF1, PTTG1, MYBL2 y ZNF367. En un aspecto, los genes seleccionados son UHRF1, PTTG1, MYBL2 y TCF19. En un aspecto, los genes seleccionados son UHRF1, PTTG1, MYBL2 y ATAD2. En un aspecto, los genes seleccionados son UHRF1, PTTG1, HMGB2 y E2F8. En un aspecto, los genes seleccionados son UHRF1, PTTG1, HMGB2 y ZNF367. En un aspecto, los genes seleccionados son UHRF1, PTTG1, HMGB2 y TCF19. En un aspecto, los genes seleccionados son UHRF1, PTTG1, E2F1, MYBL2 y ATAD2. En un aspecto, los genes seleccionados son PTTG1, E2F1, MYBL2 y E2F8. En un aspecto, los genes seleccionados son PTTG1, E2F1, MYBL2 y ZNF367. En un aspecto, los genes seleccionados son PTTG1, E2F1, MYBL2 y TCF19. En un aspecto, los genes seleccionados son PTTG1, E2F1, HMGB2 y ATAD2. En un aspecto, los genes seleccionados son PTTG1, E2F1, HMGB2 y E2F8. En un aspecto, los genes seleccionados son PTTG1, E2F1, HMGB2 y ZNF367. En un aspecto, los genes seleccionados son PTTG1, E2F1, HMGB2 y TCF19. En un aspecto, los genes seleccionados son E2F1, MYBL2, HMGB2 y ATAD2. En un aspecto, los genes seleccionados son E2F1, MYBL2, HMGB2 y E2F8. En un aspecto, los genes seleccionados son E2F1, MYBL2, HMGB2 y ZNF367. En un aspecto, los genes seleccionados son E2F1, MYBL2, HMGB2 y TCF19. En un aspecto, los genes seleccionados son ATAD2, EDF8, ZNF367 y TCF19. Preferentemente, los al menos cuatro genes seleccionados anteriormente se combinan con p16^{INK4A}.

35 En un aspecto, se seleccionan al menos cinco genes y los genes seleccionados son FOXM1, UHRF1, PTTG1, E2F1 y MYBL2. En un aspecto, los genes seleccionados son FOXM1, UHRF1, PTTG1, E2F1 y HMGB2. En un aspecto, los genes seleccionados son FOXM1, PTTG1, E2F1, MYBL2 y HMGB2. En un aspecto, los genes seleccionados son UHRF1, PTTG1, E2F1, MYBL2 y HMGB2. En un aspecto, los genes seleccionados son FOXM1, UHRF1, PTTG1, E2F1 y ATAD2. En un aspecto, los genes seleccionados son FOXM1, UHRF1, PTTG1, E2F1 y E2F8. En un aspecto, los genes seleccionados son FOXM1, UHRF1, PTTG1, E2F1 y ZNF367. En un aspecto, los genes seleccionados son FOXM1, UHRF1, PTTG1, E2F1 y TCF19. En un aspecto, los genes seleccionados son FOXM1, UHRF1, PTTG1, MYBL2 y ATAD2. En un aspecto, los genes seleccionados son FOXM1, UHRF1, PTTG1, MYBL2 y EFF8. En un aspecto, los genes seleccionados son FOXM1, UHRF1, PTTG1, MYBL2 y ZNF367. En un aspecto, los genes seleccionados son FOXM1, UHRF1, PTTG1, MYBL2 y TCF19. En un aspecto, los genes seleccionados son FOXM1, UHRF1, PTTG1, HMGB2 y ATAD2. En un aspecto, los genes seleccionados son FOXM1, UHRF1, PTTG1, HMGB2 y E2F8. En un aspecto, los genes seleccionados son FOXM1, UHRF1, PTTG1, HMGB2 y ZNF367. En un aspecto, los genes seleccionados son FOXM1, UHRF1, PTTG1, HMGB2 y TCF19. En un aspecto, los genes seleccionados son UHRF1, PTTG1, E2F1, MYBL2 y ATAD2. En un aspecto, los genes seleccionados son UHRF1, PTTG1, E2F1, MYBL2 y E2F8. En un aspecto, los genes seleccionados son UHRF1, PTTG1, E2F1, MYBL2 y ZNF367. En un aspecto, los genes seleccionados son UHRF1, PTTG1, E2F1, MYBL2 y TCF19. En un aspecto, los genes seleccionados son UHRF1, PTTG1, E2F1, HMGB2 y ATAD2. En un aspecto, los genes seleccionados son UHRF1, PTTG1, E2F1, HMGB2 y ZNF367. En un aspecto, los genes seleccionados son UHRF1, PTTG1, E2F1, HMGB2 y TCF19. En un aspecto, los genes seleccionados son PTTG1, E2F1, MYBL2, HMGB2 y ATAD2. En un aspecto, los genes seleccionados son PTTG1, E2F1, MYBL2, HMGB2 y E2F8. En un aspecto, los genes seleccionados son PTTG1, E2F1, MYBL2, HMGB2 y ZNF367. En un aspecto, los genes seleccionados son PTTG1, E2F1, MYBL2, HMGB2 y TCF19. En un aspecto, los genes seleccionados son ATAD2, E2F8, ZNF367, TCF19 y FOXM1. En un aspecto, los genes seleccionados son ATAD2, E2F8, ZNF367, TCF19 y UHRF1. En un aspecto, los genes seleccionados son ATAD2, E2F8, ZNF367, TCF19 y PTTG1. En un aspecto, los genes seleccionados son ATAD2, E2F8, ZNF367, TCF19 y E2F1. En un aspecto, los genes seleccionados son ATAD2, E2F8, ZNF367, TCF19 y MYBL2. En un aspecto, los genes seleccionados son ATAD2, E2F8, ZNF367, TCF19 y HMGB2. Preferentemente, los al menos cinco genes seleccionados anteriormente se combinan con p16^{INK4A}.

65 En un aspecto, los al menos dos genes comprenden FOXM1, y al menos otro gen seleccionado de UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19. Preferentemente, los al menos dos genes se combinan además con p16^{INK4A}.

En un aspecto, los al menos dos genes comprenden UHRF1, y al menos otro gen seleccionado de FOXM1, PTTG1,

- E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19. Preferentemente, los al menos dos genes se combinan además con p16^{INK4A}.
- 5 En un aspecto, los al menos dos genes comprenden PTTG1, y al menos otro gen seleccionado de FOXM1, UHRF1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19. Preferentemente, los al menos dos genes se combinan además con p16^{INK4A}.
- 10 En un aspecto, los al menos dos genes comprenden E2F1, y al menos otro gen seleccionado de FOXM1, PTTG1, UHRF1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19. Preferentemente, los al menos dos genes se combinan además con p16^{INK4A}.
- 15 En un aspecto, los al menos dos genes comprenden MYBL2, y al menos otro gen seleccionado de FOXM1, PTTG1, E2F1, UHRF1, HMGB2, ATAD2, E2F8, ZNF367 y TCF19. Preferentemente, los al menos dos genes se combinan además con p16^{INK4A}.
- 20 En un aspecto, los genes seleccionados son FOXM1, UHRF1, PTTG1, E2F1, MYBL2 y HMGB2. Preferentemente, los genes seleccionados se combinan además con p16^{INK4A}.
- 25 En un aspecto, los genes seleccionados son FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2 y uno o más o todos de ATAD2, E2F8, ZNF367 y TCF19. Preferentemente, los genes seleccionados se combinan además con p16^{INK4A}.
- 30 En un aspecto, los genes seleccionados consisten esencialmente en FOXM1, UHRF1, PTTG1, E2F1, MYBL2 y HMGB2. Preferentemente, los genes se combinan además con p16^{INK4A}. Se debe entender que el término "consiste esencialmente en" significa los seis genes, o cinco genes, o cuatro genes, o tres genes, o dos genes seleccionados de FOXM1, UHRF1, PTTG1, E2F1, MYBL2 y HMGB2.
- 35 En un modo de realización, el cáncer se selecciona del grupo que comprende cáncer de mama sin afectación ganglionar y positivo para ER; cáncer de mama con afectación ganglionar y en fase precoz; mieloma múltiple, cáncer de próstata, glioblastoma, linfoma, fibrosarcoma; mixosarcoma; liposarcoma; condrosarcoma; sarcoma osteogénico; cordoma; angiosarcoma; endoteliosarcoma; linfangiosarcoma; linfangioendoteliosarcoma; sinovioma; mesotelioma; tumor de Ewing; liomiosarcoma; rabdomiosarcoma; carcinoma de colon; cáncer de páncreas; cáncer de mama; cáncer de ovarios; carcinoma de células escamosas; carcinoma basocelular; adenocarcinoma; carcinoma de glándulas sudoríparas; carcinoma de glándulas sebáceas; carcinoma papilar; adenocarcinomas papilares; cistoadenocarcinoma; carcinoma medular; carcinoma broncogénico; carcinoma de células renales; hepatoma; carcinoma de vías biliares; coriocarcinoma; seminoma; carcinoma embrionario; tumor de Wilms; cáncer de cuello
- 40 uterino; cáncer uterino; tumor testicular; carcinoma de pulmón; carcinoma microcítico de pulmón; carcinoma de vejiga; carcinoma epitelial; glioma astrocitoma; meduloblastoma; craneofaringioma; ependimoma; pinealoma; hemangioblastoma; neurinoma del estatoacústico; oligodendroglioma; meningioma; melanoma; retinoblastoma; y leucemias. Adecuadamente, el cáncer es un cáncer epitelial.
- 45 En un modo de realización, el cáncer es preferentemente cáncer de mama o cáncer de próstata. Idealmente, el cáncer de mama es cáncer de mama típicamente sin afectación ganglionar y en fase precoz o cáncer de mama con afectación ganglionar y en fase precoz. Idealmente, el cáncer de mama es cáncer de mama sin afectación ganglionar y en fase precoz o positivo para ER con afectación ganglionar y en fase precoz.
- 50 En un modo de realización, la recaída es el desarrollo de un tumor secundario.
- En un modo de realización, la recaída es desarrollar otro cáncer primario independiente no relacionado con el cáncer del que se obtiene la muestra.
- 55 En un modo de realización de la invención, se proporciona un método para predecir el riesgo de recaída del cáncer de mama en un paciente con cáncer de mama sin afectación ganglionar y en fase precoz, o un paciente con cáncer de mama con afectación ganglionar y en fase precoz, comprendiendo el método una etapa de someter a ensayo una muestra tumoral de cáncer del paciente con cáncer de mama para determinar la expresión positiva de al menos tres genes (o proteínas codificadas por esos genes) seleccionados del grupo que consiste en FOXM1, UHRF1, PTTG1,
- 60 E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19, en el que la expresión positiva de los al menos tres genes (o proteínas codificadas por esos genes) se correlaciona con un riesgo incrementado de recaída del cáncer en comparación con un individuo con cáncer que no muestra presenta positiva de los al menos tres genes (o proteínas codificadas por esos genes); en el que los al menos tres genes, o proteínas codificadas por dichos al menos tres genes, son FOXM1, PTTG1 y ZNF367, incluyendo opcionalmente uno o más de los genes, o proteínas codificadas por dichos genes, UHRF1, E2F1, MYBL2, HMGB2, ATAD2, E2F8 y TCF19.
- 65

En un modo de realización, el método comprende además la etapa de someter a ensayo la expresión del gen p16^{INK4A} (o una proteína codificada por dicho gen) además de los al menos tres genes (o proteínas codificadas por esos genes), en el que la expresión regulada incorrectamente de p16^{INK4A} en combinación con la expresión positiva de una combinación de los al menos tres genes (o proteínas codificadas por esos genes) se correlaciona con un riesgo incrementado de recaída del cáncer en comparación con un individuo con cáncer que no presenta expresión regulada incorrectamente de p16^{INK4A} y expresión positiva de los al menos tres genes (o proteínas codificadas por esos genes). Los pacientes con cáncer de mama con p16^{INK4A} regulado incorrectamente y expresión positiva de los al menos tres genes (o proteínas codificadas por esos genes) tienen un riesgo incrementado de recaída del cáncer en comparación con los individuos con cáncer que no presentan la combinación de expresión positiva de los al menos tres genes y expresión regulada incorrectamente de p16^{INK4A}.

En un modo de realización de la invención, se proporciona un método para identificar a un paciente con cáncer que es adecuado para el tratamiento con un tratamiento para prevenir la recaída o progresión del cáncer, comprendiendo el método una etapa de someter a ensayo una muestra de cáncer del paciente con cáncer para determinar la expresión positiva de al menos tres genes (o proteínas codificadas por esos genes) seleccionados del grupo que consiste en FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19, en el que la expresión positiva de los al menos tres genes (o proteínas codificadas por esos genes) en comparación con un individuo con cáncer que no presenta una expresión positiva de los al menos tres genes (o proteínas codificadas por esos genes) es indicativa de que el paciente con cáncer es adecuado para el tratamiento con un tratamiento para prevenir la recaída o progresión del cáncer; en el que los al menos tres genes, o proteínas codificadas por dichos al menos tres genes, son FOXM1, PTTG1 y ZNF367, incluyendo opcionalmente uno o más de los genes, o proteínas codificadas por dichos genes, UHRF1, E2F1, MYBL2, HMGB2, ATAD2, E2F8 y TCF19.

En un modo de realización, el tratamiento es un tratamiento neoadyuvante. En la memoria descriptiva, se debe entender que el término "tratamiento neoadyuvante" significa el tratamiento administrado antes del tratamiento primario para incrementar las probabilidades de supervivencia a largo plazo. El tratamiento primario es, en general, cirugía. El tratamiento neoadyuvante se selecciona, en general, de quimioterapia, hormonoterapia, tratamiento dirigido, radioterapia, inmunoterapia o una combinación de los mismos.

En un modo de realización, el tratamiento es un tratamiento adyuvante. En la memoria descriptiva, se debe entender que el término "tratamiento adyuvante" significa cualquier tratamiento administrado después del tratamiento primario para incrementar las probabilidades de supervivencia a largo plazo. El tratamiento primario es, en general, cirugía. El tratamiento adyuvante se selecciona, en general, de quimioterapia, hormonoterapia, tratamiento dirigido, radioterapia, inmunoterapia o una combinación de los mismos.

En un modo de realización, el tratamiento puede ser una combinación de tratamiento neoadyuvante y adyuvante. Se debe entender que, en la memoria descriptiva, los tratamientos "neoadyuvante" y "adyuvante" se puede usar de manera intercambiable.

En un modo de realización, el método comprende además la etapa de someter a ensayo la expresión del gen p16^{INK4A} (o una proteína codificada por dicho gen) además de los al menos tres genes (o proteínas codificadas por esos genes), en el que la expresión regulada incorrectamente de p16^{INK4A} en combinación con la expresión positiva de una combinación de al menos tres de los genes (o proteínas codificadas por esos genes), en comparación con un individuo con cáncer que no presenta una expresión regulada incorrectamente de p16^{INK4A} y una expresión positiva de los al menos tres genes, es indicativa de que el paciente con cáncer es adecuado para el tratamiento con un tratamiento adyuvante para prevenir la recaída o progresión del cáncer. Los pacientes con cáncer de mama con expresión regulada incorrectamente de p16^{INK4A} y expresión positiva de los al menos tres genes (o proteínas codificadas por esos genes) pueden ser adecuados para el tratamiento con un tratamiento adyuvante para prevenir la recaída o progresión del cáncer. En un modo de realización, el paciente con cáncer puede ser adecuado para el tratamiento con un tratamiento neoadyuvante para prevenir la recaída o progresión del cáncer.

En un modo de realización, el cáncer es cáncer de mama sin afectación ganglionar y en fase precoz o cáncer de mama con afectación ganglionar y en fase precoz. Idealmente, el cáncer de mama es cáncer de mama positivo para ER sin afectación ganglionar y en fase precoz o cáncer de mama positivo para ER, con afectación ganglionar y en fase precoz.

En un modo de realización, el tratamiento adyuvante y el tratamiento neoadyuvante es tratamiento quimioterápico. En un modo de realización, el tratamiento adyuvante y el tratamiento neoadyuvante es un tratamiento con inhibidores de CDK4/6 tal como tratamiento con palbociclib (PD 0332991, Pfizer), abemaciclib (LY2835219; Lilly, EE. UU.) o LEE011 (Novartis, Suiza).

En un aspecto de la invención, se proporciona un sistema para obtener datos de al menos una muestra de prueba obtenida de al menos un individuo, comprendiendo el sistema:

un módulo de determinación configurado para recibir al menos una muestra de prueba y realizar al menos un análisis de prueba en la muestra de prueba para someter a ensayo la expresión de al menos dos genes

(o proteínas codificadas por esos genes) seleccionados del grupo que consiste en FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19;

opcionalmente, un sistema de almacenamiento para almacenar datos de expresión generados por el módulo de determinación; y

un módulo de visualización para mostrar un contenido basado en parte en la salida de datos de dicho módulo de determinación, en el que el contenido comprende una señal indicativa de la expresión de los al menos dos genes.

En un aspecto, el módulo de determinación está configurado además para realizar al menos un análisis de prueba en la muestra de prueba para la regulación incorrecta de p16^{INK4A} en combinación con el análisis de prueba en al menos dos genes (o proteínas codificadas por esos genes).

En un aspecto, el sistema comprende un módulo de correlación para correlacionar los datos de expresión de los al menos dos genes (o proteínas codificadas por esos genes) del módulo de determinación con potencial de recaída del cáncer, en el que los datos de expresión de cada gen (o una proteína codificado por el gen) se comparan con un valor de referencia para el gen (o una proteína codificada por el gen) para determinar la expresión positiva del gen (o una proteína codificada por el gen), y en el que la expresión positiva de los al menos dos genes (o proteínas codificadas por esos genes) se correlaciona con un potencial incrementado de recaída en comparación con un individuo con cáncer que no presenta una expresión positiva de los al menos dos genes (o proteínas codificadas por esos genes), y en el que el módulo de visualización muestra un contenido basado en parte en los datos del sistema de correlación, comprendiendo opcionalmente el contenido una señal indicativa del potencial de recaída del cáncer.

En un aspecto, el módulo de correlación correlaciona además los datos de expresión de los al menos dos genes (o proteínas codificadas por esos genes) del módulo de determinación con potencial de recaída del cáncer, conjuntamente con los datos de expresión de p16^{INK4A}, en el que los datos de expresión de cada gen (o una proteína codificada por el gen) y p16^{INK4A} se comparan con un valor de referencia para cada gen (o una proteína codificada por el gen) y p16^{INK4A}, respectivamente, para determinar la expresión positiva del gen (o una proteína codificada por el gen) y la regulación incorrecta de p16^{INK4A}, y en el que la expresión positiva de los al menos dos genes (o proteínas codificadas por esos genes) y la regulación incorrecta de p16^{INK4A} se correlaciona con un potencial incrementado de recaída en comparación con un individuo con cáncer que no presenta una expresión positiva de los al menos dos genes (o proteínas codificadas por esos genes) y regulación incorrecta de p16^{INK4A}, y en el que el módulo de visualización muestra un contenido basado en parte en los datos del sistema de correlación, comprendiendo opcionalmente el contenido una señal indicativa del potencial de recaída del cáncer.

Adecuadamente, el sistema de determinación se puede seleccionar de un aparato de detección inmunohistoquímica, una inmunoelectrotransferencia, una transferencia de tipo Northern, una transferencia de tipo Southern, reacción en cadena de la polimerasa (PCR) cuantitativa, PCR con retrotranscriptasa (RT-PCR), RT-PCR cuantitativa ultrarrápida (qRT-PCR), un ensayo de inmuoadsorción enzimática (ELISA), determinación de proteínas en geles de poliacrilamida y dichos métodos conocidos por los expertos en la técnica. Idealmente, el sistema de determinación comprende un aparato de detección inmunohistoquímica.

En un aspecto de la invención, el contenido basado en el resultado de comparación o el sistema de determinación se muestra en un monitor de ordenador. En un aspecto de la invención, el contenido basado en el resultado de comparación o el sistema de determinación se muestra a través de medios imprimibles. El módulo de visualización puede ser cualquier dispositivo adecuado configurado para recibir de un ordenador y mostrar información legible por ordenador a un usuario. Los ejemplos no limitantes incluyen, por ejemplo, ordenadores de uso general, tales como los basados en un procesador de tipo Intel PENTIUM, procesadores Motorola PowerPC, Sun UltraSPARC, Hewlett-Packard PA-RISC, cualquiera de una variedad de procesadores disponibles de Advanced Micro Devices (AMD) de Sunnyvale, California, o cualquier otro tipo de procesador, dispositivos de visualización visuales tales como pantallas planas, tubos de rayos catódicos y similares, así como impresoras de ordenador de diversos tipos.

En un aspecto, se usa un navegador de World Wide Web para proporcionar una interfaz de usuario para mostrar el contenido basado en el resultado de comparación. Se debe entender que otros módulos de la invención se pueden adaptar para tener una interfaz de navegador web. A través del navegador web, un usuario puede construir solicitudes para recuperar datos del módulo de comparación. Por tanto, el usuario típicamente apuntará y hará clic en elementos de la interfaz de usuario tales como botones, menús desplegados, barras de desplazamiento y similares empleados convencionalmente en interfaces gráficas de usuario.

En un modo de realización de la invención, se proporciona un método para controlar la eficacia del tratamiento del cáncer en un individuo con cáncer, comprendiendo el método una etapa de someter a ensayo una muestra de cáncer del individuo con cáncer para determinar la expresión de al menos tres genes seleccionados del grupo que consiste en FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19, en el que una expresión más alta de al menos tres genes seleccionados del grupo que consiste en FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19 se correlaciona con un tratamiento ineficaz y un mal

resultado en comparación con un individuo con cáncer que tiene una expresión más baja de los al menos tres genes; en el que los al menos tres genes, o proteínas codificadas por dichos al menos tres genes, son FOXM1, PTTG1 y ZNF367, incluyendo opcionalmente uno o más de los genes, o proteínas codificadas por dichos genes, UHRF1, E2F1, MYBL2, HMGB2, ATAD2, E2F8 y TCF19.

5 En un modo de realización, el método comprende además la etapa de someter a ensayo la muestra de cáncer para determinar la expresión del gen p16^{INK4A} (o una proteína codificada por dicho gen) en combinación con someter a ensayo los al menos tres genes (o proteínas codificadas por dichos genes), con lo que la expresión regulada incorrectamente de p16^{INK4A} se correlaciona con un tratamiento ineficaz y un mal resultado en comparación con un individuo con cáncer que tiene una expresión moderada de p16^{INK4A}.

10 En un modo de realización, el tratamiento es un tratamiento neoadyuvante o adyuvante, o una combinación de ambos.

15 En un modo de realización, el tratamiento neoadyuvante y el tratamiento adyuvante es un agente seleccionado de, pero sin limitarse a, trastuzumab (Herceptin®), lapatinib (Tykerb®), neratinib, afatinib (Tovok®), pertuzumab, inhibidores de CDK4/6 (tales como palbociclib (PD 0332991, Pfizer), abemaciclib (LY2835219; Lilly, EE. UU.) y LEE011 (Novartis, Suiza)), ciclofosfamida, metotrexato, 5-fluorouracilo, gemcitabina, adriamicina (doxorubicina), epirubicina, docetaxel (Taxotere®), paclitaxel (Taxol®), capecitabina (Xeloda®) y tamoxifeno.

20 También se describe un método para tratar a un individuo para prevenir o inhibir la recaída del cáncer que comprende una etapa de identificar a un paciente con cáncer en riesgo de recaída utilizando un método de la invención, y, a continuación, tratar al paciente con cáncer con un agente o agentes para prevenir o inhibir la recaída del cáncer. Típicamente, el agente o agentes comprenden tratamiento adyuvante o neoadyuvante, o una combinación de ambos.

25 En un aspecto, se proporciona un método para predecir el riesgo de recaída del cáncer en un individuo con cáncer, comprendiendo el método una etapa de someter a ensayo una muestra de cáncer del individuo para determinar la expresión positiva de al menos cuatro genes, o proteínas codificadas por dichos genes, seleccionados de FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19, en el que la expresión positiva de los al menos cuatro genes, o proteínas codificadas por dichos genes, se correlaciona con un riesgo incrementado de recaída del cáncer en comparación con un individuo con cáncer que no presenta una expresión positiva de los al menos cuatro genes o proteínas codificadas por esos genes.

35 En un aspecto, se proporciona un método para predecir el riesgo de recaída del cáncer en un individuo con cáncer después del tratamiento con inhibidores de CDK4/6, comprendiendo el método una etapa de someter a ensayo una muestra de cáncer del individuo para determinar la expresión positiva de al menos cuatro genes, o proteínas codificadas por dichos genes, seleccionados de FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19, en el que la expresión positiva de los al menos cuatro genes se correlaciona con un riesgo incrementado de recaída del cáncer en un individuo con cáncer después del tratamiento con inhibidores de CDK4/6 en comparación con un individuo con cáncer que no presenta una expresión positiva de los al menos cuatro genes o proteínas codificadas por esos genes.

45 En un aspecto, se proporciona un método para predecir el riesgo de recaída del cáncer de mama en un paciente con cáncer de mama sin afectación ganglionar y en fase precoz, comprendiendo el método una etapa de someter a ensayo una muestra tumoral de cáncer del paciente para determinar la expresión positiva de al menos cuatro genes, o proteínas codificadas por esos genes, seleccionados de FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19, en el que la expresión positiva de los al menos cuatro genes, o proteínas codificadas por esos genes, se correlaciona con un riesgo incrementado de recaída del cáncer en comparación con un paciente con cáncer que no presenta una expresión positiva de los al menos cuatro genes o proteínas codificadas por esos genes.

50 En un aspecto, se proporciona un método para determinar una tasa de supervivencia a 5 años o una tasa de supervivencia a 10 años de un individuo diagnosticado de cáncer de mama, comprendiendo el método una etapa de someter a ensayo una muestra tumoral de cáncer del individuo para determinar la expresión positiva de al menos cuatro genes, o proteínas codificadas por esos genes, seleccionados de FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19, en el que la expresión positiva de los al menos cuatro genes, o proteínas codificadas por esos genes, se correlaciona con una probabilidad disminuida de una tasa de supervivencia a 5 años o una tasa de supervivencia a 10 años en comparación con un individuo con cáncer que no presenta una expresión positiva de los al menos cuatro genes o proteínas codificadas por esos genes.

55 En un aspecto, los métodos comprenden además la etapa de someter a ensayo la expresión del gen p16^{INK4A} o una proteína codificada por dicho gen, en los que la expresión regulada incorrectamente de p16^{INK4A}, en combinación con la expresión positiva de los al menos cuatro genes o proteínas codificadas por esos genes, se correlaciona con un riesgo incrementado de recaída del cáncer o una probabilidad disminuida de una tasa de supervivencia a 5 años o una tasa de supervivencia a 10 años en comparación con un individuo con cáncer que no presenta una expresión

regulada incorrectamente de p16^{INK4A} y una expresión positiva de los al menos cuatro genes o proteínas codificadas por esos genes.

En un aspecto, se proporciona un método para identificar a un paciente con cáncer que es adecuado para el tratamiento con un tratamiento para prevenir la recaída o progresión del cáncer, comprendiendo el método una etapa de someter a ensayo una muestra de cáncer del paciente con cáncer para determinar la expresión positiva de al menos cuatro genes seleccionados de FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19, en el que la expresión positiva de los al menos cuatro genes o proteínas codificadas por esos genes en comparación con un individuo con cáncer que no presenta una expresión positiva de los al menos dos genes o proteínas codificadas por esos genes es indicativa de que el paciente con cáncer es adecuado para el tratamiento con un tratamiento para prevenir la recaída o progresión del cáncer.

En un aspecto, se proporciona un sistema para obtener datos de al menos una muestra de prueba obtenida de al menos un individuo, comprendiendo el sistema un módulo de determinación configurado para recibir al menos una muestra de prueba y realizar al menos un análisis de prueba en la muestra de prueba para someter a ensayo la determinación de la expresión de al menos cuatro genes o proteínas codificadas por esos genes seleccionados de FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19; opcionalmente, un sistema de almacenamiento para almacenar datos de expresión generados por el módulo de determinación; y un módulo de visualización para mostrar un contenido basado en parte en la salida de datos de dicho módulo de determinación, en el que el contenido comprende una señal indicativa de la expresión de al menos dos genes o proteínas codificadas por esos genes.

En un aspecto, se proporciona un método para controlar la eficacia del tratamiento del cáncer en un individuo con cáncer, comprendiendo el método una etapa de someter a ensayo una muestra de cáncer del individuo con cáncer para determinar la expresión de al menos cuatro genes o proteínas codificadas por dicho genes seleccionados de FOXM1, UHRF1, PTTG1, E2F1, MYBL2 y HMGB2, en el que una expresión más alta de al menos cuatro genes seleccionados de FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19 se correlaciona con un tratamiento ineficaz y un mal resultado en comparación con un individuo con cáncer que tiene una expresión más baja de los al menos cuatro genes o proteínas codificadas por esos genes.

En un aspecto, se proporciona un método para predecir el riesgo de recaída o progresión del cáncer de mama en un paciente, y tratar al paciente con un tratamiento para prevenir la recaída del cáncer, comprendiendo el método una etapa de someter a ensayo una muestra de cáncer del paciente para determinar la expresión positiva de al menos cuatro genes seleccionados de FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19, en el que la expresión positiva de los al menos cuatro genes, o proteínas codificadas por esos genes, se correlaciona con un riesgo incrementado de recaída o progresión del cáncer en comparación con un paciente con cáncer que no presenta una expresión positiva de los al menos cuatro genes, o proteínas codificadas por esos genes; y administrar un tratamiento neoadyuvante o adyuvante, o una combinación de ambos, al paciente para prevenir la recaída o progresión del cáncer.

En un aspecto, los al menos cuatro genes, o proteínas codificadas por dichos genes, son FOXM1, PTTG1, UHRF1 y HMGB2.

Breve descripción de los dibujos

La invención se entenderá más claramente a partir de la siguiente descripción de un modo de realización de la misma, dada solo a modo de ejemplo, con referencia a los dibujos que se acompañan, en los que:

La **figura 1** ilustra la identificación de reguladores transcripcionales maestros (MTR) de la proliferación de células mamarias. (A) Análisis de inmunoelectrotransferencia del marcador de proliferación EZH2 y el marcador de senescencia celular p16^{INK4A} en células epiteliales mamarias humanas (HMEC) y fibroblastos embrionarios de ratón (MEF) en crecimiento (pase bajo) y senescentes (pase alto). La β -actina se usó como control de carga. (B) Se alinearon experimentos duplicados de determinación del perfil transcriptómico en cultivos de HMEC y MEF en crecimiento y senescentes para identificar genes expresados a un nivel sistemáticamente alto en células en proliferación. El análisis de la matriz cromática representa todos los genes regulados por incremento o por disminución en más de dos veces en HMEC, y el cambio correspondiente en MEF. (Grupo 1 = 58 genes; Grupo 2 = 193 genes; Grupo 3 = 184 genes; Grupo 4 = 214 genes). El clúster 4 representa una firma de 'proliferación central' que comprende los genes que se regulan por disminución más significativa y sistemáticamente durante pases sucesivos tanto de HMEC como de MEF. (C) Validación por PCR cuantitativa ultrarrápida de los cambios de expresión génica de genes representativos de cada uno de los grupos de genes que se muestran en el panel B. El gen de ARN ribosómico, RPLPO, se usó para la normalización de estos datos. (D) Análisis de ontología génica de grupos de genes individuales. La línea roja indica un valor de p de 0,05. (E) Análisis de enriquecimiento genético de los grupos 1-4 en la firma de MammaPrint y la firma de grado genómico. El factor de multiplicación del solapamiento observado frente a lo que se esperaría por casualidad se representa en el eje de ordenadas. Se muestra el número de genes de 'proliferación central' (número superior) presente en

cada firma de 'mal pronóstico' (número inferior).

La **figura 2** ilustra que E2F1, FOXM1 y MYBL2 se unen a genes proliferativos centrales en HMEC. (A) El análisis de ingeniería inversa usando ARACNe predice 6 reguladores transcripcionales maestros (MTR) en dirección 5' de la firma de 'proliferación central'. Se muestra una red ARACNe representativa de la firma de 'proliferación central' de HMEC/MEF (grupo 4) dentro del conjunto de datos de NKI (van de Vijver *et al.*, 2002). Los MTR se resaltan en rojo y los genes del grupo 4 se resaltan en verde. (B) Validación de la unión de MTR a genes dentro de la firma de 'proliferación central' por ChIP-qRT-PCR. Se analizó ADN precipitado por qRT-PCR usando cebadores dirigidos hacia los promotores de los genes indicados (SEQ ID NO: 1 a 38). Se usó anticuerpo anti-HA como control negativo para ChIP, y los promotores β -ACTB y CHD5 se usaron como controles negativos de promotores para qRT-PCR. Los enriquecimientos por ChIP se presentan como el porcentaje de proteína unida, normalizada a la entrada. Las barras de error indican la desviación estándar de tres muestras técnicas. (C) Análisis de la matriz cromática que muestra datos de ChIP-seq para FOXM1, MYBL2 y E2F1, en células HMEC-Tert. La unión en los promotores de genes de los grupos 1-4 se indica incrementando la señal para cada factor, FOXM1 (rojo), MYBL2 (verde) y E2F1 (azul). Se muestra la región entre -2 y +2 del sitio de inicio de la transcripción (TSS) de estos genes. (D) Pistas de ChIP-seq representativas de los genes indicados, con FOXM1, MYBL2 y E2F1 unidos en sus promotores en células HMEC-Tert. Los datos de RNA-seq de HMEC de pase tanto bajo como alto también se representan para cada gen. El gen KRT2 se incluye como control negativo.

La **figura 3** ilustra que los reguladores transcripcionales maestros predicen la evolución del paciente. (A) Se pronostica que los reguladores transcripcionales maestros estarán en dirección 5' de la firma de mal pronóstico de "grado genómico". Se muestra una red ARACNe representativa de la firma de 'grado genómico' (Sotiriou *et al.*, 2006) dentro del conjunto de datos de Loi (Loi *et al.*, 2007). Los MTR se resaltan en rojo, y los genes de la firma de grado genómico se resaltan en verde. (B) Los análisis de Kaplan-Meier demuestran que la combinación de los 6 MTR (superior) presentan un valor pronóstico superior que Ki67 (inferior) en muestras sin afectación ganglionar sin quimioterapia adyuvante en el conjunto de datos de micromatrices combinadas en términos de supervivencia sin recaída (Loi *et al.*, 2007; Miller *et al.*, 2005; y van de Vijver *et al.*, 2002) (n=457). La puntuación combinada de MTR y los datos de expresión génica de Ki67 se dividieron como grupos de 2 (Lo/Hi) y 3 (Lo/Med/Hi). (C) Ejemplos representativos de tinción inmunohistoquímica para los factores indicados en tumores de bajo y alto riesgo en una micromatriz de tejido de cáncer de mama. Los tumores de bajo riesgo se definieron como aquellos que no recayeron dentro del marco temporal de estudio, mientras que los tumores de alto riesgo sí recayeron. (D) Curvas de supervivencia de Kaplan-Meier para FOXM1, UHRF1, HMGB2 y PTTG1 combinados, en comparación con Ki67 y los criterios de St. Gallen en muestras de TMA (n=408) en términos de supervivencia sin recaída. (E) Matriz cromática que ilustra el poder pronóstico de FOXM1, UHRF1, HMGB2 y PTTG1 y los 4 MTR combinados en los tumores de mama de la cohorte de TMA (n=408) en términos de supervivencia sin recaída. Se incluyeron los resultados de tinción con Ki67 y los criterios de St. Gallen para la comparación. La escala representa $-\log_{10}$ de los valores de p calculados usando la prueba del orden logarítmico.

La **figura 4** ilustra que los niveles de ARNm de CDKN2A y de proteína p16 ausentes y altos predicen un mal pronóstico en el cáncer de mama. (A) Correlaciones de los niveles de expresión de ARNm de CDKN2A con alteraciones del número de copias génicas (CNA) en los locus de genes RB1 y CDKN2A usando la herramienta GISTIC en datos 457 cánceres de mama de TCGA (TCGA, 2012, Nature, 490, 61-70). (B) Curvas de supervivencia de Kaplan-Meier para el ARNm de CDKN2A en cánceres de mama sin afectación ganglionar sin quimioterapia adyuvante en el conjunto de datos de micromatrices combinadas (n=457) en términos de supervivencia sin recaída. Las muestras se estratificaron en 3 grupos basados en los niveles de expresión de ARNm de CDKN2A, cortados en el percentil 33 y 66. Adicionalmente, los grupos de expresión no detectada y alta se combinaron y compararon con el grupo de expresión moderada. Los valores de Chi2 y los valores de p se calcularon utilizando la prueba del orden logarítmico. (C) Ejemplos representativos de tinción inmunohistoquímica para p16 en tumores de bajo y alto riesgo. Los tumores de bajo riesgo se definieron como aquellos que no recayeron dentro del marco temporal del estudio, mientras que los tumores de alto riesgo sí recayeron. (D) Curvas de supervivencia de Kaplan-Meier para niveles de proteína p16 en la cohorte de TMA (n=408) que miden la supervivencia sin recaída. Se estratificó a los pacientes por los niveles de proteína p16 en grupos de expresión negativa, moderada (<50% de células positivas) y alta (>50% de células positivas). Los valores de Chi2 y los valores de p se calcularon utilizando la prueba del orden logarítmico. (E) Curvas de supervivencia de Kaplan-Meier para niveles de proteína p16 en la cohorte de TMA (n = 408) que miden la supervivencia específica del cáncer de mama. Se estratificó a los pacientes como en el panel C.

La **figura 5** ilustra que las mediciones combinadas de los niveles de MTR y p16(INK4A) superan las estimaciones de las estrategias usadas actualmente. (A) Curvas de supervivencia de Kaplan-Meier que comparan el valor pronóstico de la puntuación de ARN de OncoMasTR (combinación de CDKN2A y 6 MTR) con estimaciones de las firmas de Oncotype Dx (21 genes) y MammaPrint (70 genes) en muestras sin afectación ganglionar sin quimioterapia adyuvante en el conjunto de datos de micromatrices combinadas (n=457) en términos de supervivencia sin recaída. Se usaron divisiones de bajo/moderado/alto y bajo/alto

para facilitar la comparación con las firmas pronósticas existentes. (B) Matrices cromáticas que ilustran el valor pronóstico de CDKN2A solo, 6 MTR combinados, puntuación de ARN de OncoMasTR, firma de 70 genes, firma de 21 genes y Ki67 en muestras sin afectación ganglionar sin quimioterapia adyuvante en tres conjuntos de datos de micromatrices de cáncer de mama individuales (Loi *et al.*, 2007; Miller *et al.*, 2005; y van de Vijver *et al.*, 2002) y el conjunto de datos combinados (n=457) en términos de supervivencia sin recaída. Los grupos de riesgo predicho de las firmas de 70 genes y 21 genes se estimaron en base a los datos de expresión génica usando el paquete genefu en R. La escala representa $-\log_{10}$ de los valores de p calculados usando la prueba del orden logarítmico. Se utilizaron divisiones de 2 y 3 grupos para facilitar la comparación con las firmas pronósticas existentes. (C) Curvas de supervivencia de Kaplan-Meier que ilustran la puntuación combinado de 4 MTR (FOX M1, UHRF1, HMGB2, PTTG1) y p16 (puntuación por IHC de OncoMasTR) en todas las muestras (izquierda, n=408) y muestras sin afectación ganglionar (derecha, n=222) de la cohorte de TMA usando datos de supervivencia sin recaída. Los valores pronósticos de los 4 MTR solo, p16 solo, la puntuación por IHC de OncoMasTR, Ki67 y los criterios de St. Gallen se representaron como una matriz cromática basada en el $-\log_{10}$ de los valores de p calculados usando la prueba del orden logarítmico. (D) Supervivencia de Kaplan-Meier como en el panel C, usando datos de supervivencia específicos del cáncer de mama.

La **figura 6** ilustra el rendimiento de la puntuación de ARN de OncoMasTR en pacientes positivos para ER. (A) Curvas de supervivencia de Kaplan-Meier que comparan el valor pronóstico de la puntuación de ARN de OncoMasTR (6 MTR y CDKN2A) con las firmas de 21 genes y 70 genes en pacientes positivos para ER que no recibieron quimioterapia adyuvante, en el conjunto de datos de micromatrices combinadas (n=536) en términos de supervivencia sin recaída. (B) Curvas de supervivencia de Kaplan-Meier como en el panel A en pacientes sin afectación ganglionar y positivos para ER que no recibieron quimioterapia adyuvante, en el conjunto de datos de micromatrices combinadas (n=366).

La **figura 7** ilustra el rendimiento de la puntuación de ARN de OncoMasTR medida por qRT-PCR con Taqman. Curvas de supervivencia de Kaplan-Meier que demuestran el valor pronóstico de la puntuación de ARN de OncoMasTR (4 MTR +/- CDKN2A) como se indica en pacientes positivos para ER, sin afectación ganglionar en el conjunto de datos de NKI que no recibieron quimioterapia adyuvante (n = 151), en términos de supervivencia sin metástasis a distancia. Los pacientes se dividieron en grupos de bajo y alto riesgo, y grupos de bajo, moderado y alto riesgo como se indica. Para hacer esto, se usaron datos de expresión para cada gen MTR para dividir a los pacientes en grupos de bajo/alto en la mediana. La suma de los 6 MTR (+/- CDKN2A) se tomó y se dividió además por la mediana (2 grupos) o por el percentil 33 y 66 (3 grupos). El valor final es la SSMD (censurada a los 10 años).

La **figura 8** ilustra el rendimiento de la puntuación por IHC de OncoMasTR en términos de supervivencia sin metástasis a distancia. Curvas de supervivencia de Kaplan-Meier que demuestran el valor pronóstico de la puntuación por IHC de OncoMasTR (4 MTR +/- CDKN2A) como se indica en pacientes sin afectación ganglionar (LN-) (n = 220), pacientes positivos para ER (ER+) (n = 331) y pacientes LN-ER+ (n = 187), que no recibieron quimioterapia adyuvante, en términos de supervivencia sin metástasis a distancia.

La **figura 9** ilustra el valor pronóstico de MTR adicionales, ATAD2 y TCF19. Curvas de supervivencia de Kaplan-Meier que demuestran el valor pronóstico de ATAD2 y TCF19 en pacientes sin afectación ganglionar, positivos para ER en el conjunto de datos de micromatrices combinadas (n = 375), en términos de supervivencia sin metástasis a distancia, censurada a los 10 años. Los valores de expresión génica para ATAD2 y TCF19 se dividieron en grupos de bajo/alto por la mediana dentro de cada uno de los tres conjuntos de datos. No existe asignación de sondas a E2F8 y ZNF367 en el conjunto de datos de NKI.

La **figura 10** ilustra las curvas de supervivencia de Kaplan-Meier para 6 MTR (FOX M1, UHRF1, MYBL2, HMGB2, E2F1, PTTG1) en el conjunto de datos transcriptómicos del cáncer de próstata de The Cancer Genome Atlas (TCGA) (n=150) en términos de supervivencia sin metástasis.

La **figura 11** ilustra un gráfico de bosque de las 100 combinaciones principales de MTR de la lista de 10 MTR descritos aquí, con al menos 4 MTR en cada combinación.

La **figura 12** ilustra los gráficos de Kaplan-Meier de las 24 combinaciones principales de MTR. En cada caso, la línea negra se refiere a alta expresión de la combinación de marcadores y la gris se refiere a baja expresión de la combinación de marcadores.

La **figura 13** ilustra la puntuación de firma de MTR10 y CDKN2A en líneas celulares humanas tratadas con pablociclib.

Descripción detallada de los dibujos

Definiciones

En esta memoria descriptiva, se debe entender que el término "muestra de cáncer" significa células tumorales, tejido tumoral u otro material biológico derivado de un tumor, por ejemplo, medios acondicionados.

5 En la memoria descriptiva, se debe entender que el término "reguladores transcripcionales maestros (MTR)" significa un conjunto específico de factores de transcripción (TF) que están en dirección 5' de, y se ha demostrado que regulan, genes de proliferación central implicados en la progresión del cáncer y la metástasis. En otras palabras, estos MTR específicos regulan el cáncer y, en particular, la progresión del cáncer de mama.

10 En la memoria descriptiva, se debe entender que el término "expresión positiva" como se aplica a un gen o una proteína codificada por ese gen significa un nivel de expresión del gen o proteína codificada por ese gen que está incrementada por encima de un nivel promedio de expresión del mismo gen o proteína codificada por ese mismo gen que se encuentra en una cohorte de individuos de control emparejados con cáncer (el "grupo de control"). La cohorte de individuos emparejados puede consistir en individuos que no experimentaron una recaída de un cáncer después de la cirugía para extirpar el cáncer. En relación con los controles, la práctica habitual para un experto en la técnica sería usar un control 'estándar', por ejemplo, para inmunohistoquímica (IHC), una línea celular o líneas celulares donde se conoce el nivel de expresión del biomarcador, o para qPCR (reacción en cadena de la polimerasa cuantitativa), se conoce un control estándar similar o un grupo de varias muestras.

20 En la memoria descriptiva, se debe entender que el término "expresión regulada incorrectamente" como se aplica a la expresión de p16^{INK4A} significa un nivel de expresión de p16^{INK4A} que es negativo, está incrementado por encima o disminuido por debajo de un nivel de expresión de p16^{INK4A} que se encuentra en una cohorte de individuos emparejados con cáncer que no recayeron después de la cirugía para extirpar el cáncer.

25 Se debe entender que los términos "expresión normal" o "expresión moderada" como se aplican a un gen o proteína significan un nivel de expresión del gen (o proteína codificada por ese gen) que es equivalente a un nivel de expresión del mismo gen o proteína codificada por ese mismo gen que se encuentra en una cohorte de individuos de control emparejados con cáncer. La cohorte de individuos emparejados puede consistir en individuos que no experimentaron una recaída de un cáncer después de la cirugía para extirpar el cáncer.

30 El método usado para establecer umbrales es diferente para el análisis de micromatrices, análisis por qRT-PCR y expresión de proteínas. Para micromatrices, el umbral es relativo (las muestras se dividieron en tres grupos iguales, por lo que el umbral depende del conjunto de datos), y para la qPCR y la expresión de proteínas se establece en puntos específicos. Para ARN (micromatrices), los niveles de expresión de 'bajo', 'moderado' y 'alto' se refieren a valores de expresión que se encuentran dentro del tercio inferior, medio o superior del intervalo de expresión; o, de forma alternativa, expresión 'baja' y 'alta' se puede referir a valores de expresión que se encuentran dentro de la mitad inferior o superior del intervalo de expresión. Para la qRT-PCR y los niveles de expresión de proteínas, se han establecido umbrales específicos, pero, en general, el término "regulado incorrectamente" se refiere a tumores con valores de expresión que se encuentran por encima o por debajo de valores establecidos en el intervalo de expresión. Para los términos "moderado" y "normal", los términos se refieren a tumores con valores de expresión que se encuentran dentro de valores establecidos en el intervalo de expresión. Por ejemplo, para p16^{INK4A}, los umbrales de qRT-PCR normalizados para la expresión 'moderada' son 0,7 y 1,99. Los umbrales de proteína normalizados (usando IHC) son un 1 % y un 50 % de células positivas. Es decir, una puntuación moderada aquí se refiere a un tumor con >1 % y <50 % de células tumorales positivas para p16^{INK4A}. Estos valores se pueden ajustar en base a cualquier dato nuevo, pero la misma teoría se aplica a los términos "normal", "moderado" y "regulado incorrectamente" con respecto a los niveles de expresión de p16^{INK4A}.

50 En la memoria descriptiva, se debe entender que el término "tratamiento adyuvante" significa cualquier tratamiento administrado después del tratamiento primario para incrementar las probabilidades de supervivencia a largo plazo. En la memoria descriptiva, se debe entender que el término "tratamiento neoadyuvante" significa el tratamiento administrado antes del tratamiento primario para incrementar las probabilidades de supervivencia a largo plazo. El tratamiento primario es, en general, cirugía. El tratamiento adyuvante y el tratamiento neoadyuvante se seleccionan, en general, de quimioterapia, hormonoterapia, tratamiento dirigido, radioterapia, inmunoterapia o una combinación de los mismos.

55 En la memoria descriptiva, se debe entender que el término "muestra" significa células tumorales, tejido tumoral, tejido no tumoral, medios acondicionados, sangre o hemoderivados (suero, plasma, etc.), orina o líquido cefalorraquídeo.

60 La detección de la expresión implica, en general, tinción inmunohistológica de un tejido de biopsia tumoral o un tejido de biopsia de control usando medios adecuados tales como tinción inmunohistoquímica; sin embargo, muchos otros medios para detectar los biomarcadores de la invención serán evidentes para los expertos en la técnica. Por ejemplo, reacción en cadena de la polimerasa (PCR) cuantitativa, PCR con retrotranscriptasa (RT-PCR), RT-PCR cuantitativa ultrarrápida (qRT-PCR), ELISA, inmunoelectrotransferencia, determinación de proteínas en geles de poliacrilamida y similares.

65 En esta memoria descriptiva, se debe entender que el término "cáncer" significa un cáncer que se trata mediante

- 5 pautas quimioterápicas. Un ejemplo de dicho cáncer incluye mieloma múltiple, cáncer de próstata, glioblastoma, linfoma, fibrosarcoma; mixosarcoma; liposarcoma; condrosarcoma; sarcoma osteogénico; cordoma angiosarcoma; endoteliosarcoma; linfangiosarcoma; linfangioendoteliosarcoma; sinovioma; mesotelioma, tumor de Ewing; liomiosarcoma; rabdiosarcoma; carcinoma de colon; cáncer pancreático; cáncer de mama; cáncer de mama sin
- 10 afectación ganglionar positivo para ER; cáncer de mama con afectación ganglionar y en fase precoz; cáncer de mama con afectación ganglionar positivo para ER y en fase precoz; cáncer ovárico; carcinoma de células escamosas; carcinoma basocelular; adenocarcinoma; carcinoma de glándulas sudoríparas; carcinoma de glándulas sebáceas; carcinoma papilar; adenocarcinomas papilares; cistoadenocarcinoma; carcinoma medular; carcinoma broncogénico; carcinoma de células renales; hepatoma; carcinoma de vías biliares; coriocarcinoma; seminoma; carcinoma embrionario; tumor de Wilms; cáncer de cuello uterino; cáncer uterino; tumor testicular; carcinoma de pulmón; carcinoma microcítico de pulmón; carcinoma de vejiga; carcinoma epitelial; glioma; astrocitoma; meduloblastoma; craneofaringioma; ependimoma; pinealoma; hemangioblastoma; neurinoma del estatoacústico; oligodendroglioma; meningioma; melanoma; retinoblastoma; y leucemias.
- 15 En esta memoria descriptiva, se debe entender que el término "fase precoz", como se aplica a un cáncer, especialmente un cáncer de mama, significa tumores que son localmente invasores pero que no se han diseminado a los ganglios linfáticos axilares regionales ni a ninguna otra región del cuerpo fuera del tejido mamario. Es decir, el cáncer no se ha diseminado más allá de la mama o los ganglios linfáticos en la axila en el mismo lado del cuerpo ni a ninguna otra parte del cuerpo.
- 20 En la memoria descriptiva, se debe entender que el término "cáncer de mama con afectación ganglionar y en fase precoz" significa tumores que son localmente invasores y se han diseminado a entre 1-3 ganglios linfáticos axilares regionales, pero no a ninguna otra región del cuerpo fuera del tejido mamario.
- 25 En esta memoria descriptiva, se debe entender que el término "sin afectación ganglionar" como se aplica a un cáncer, especialmente un cáncer de mama, significa tumores que no se han diseminado a los ganglios linfáticos axilares regionales ni a ninguna región fuera del tejido mamario.
- 30 En la memoria descriptiva, los términos "paciente con cáncer de mama" o "paciente" significan un paciente que tiene un tumor de cáncer de mama primario y espera tratamiento para el cáncer o ya se ha sometido o está sometándose a tratamiento para el tumor primario. También se debe entender que el término incluye a un paciente que ha tenido un cáncer de mama primario y está en remisión, por ejemplo, remisión después del tratamiento, incluyendo uno o más de resección tumoral, quimioterapia de primera línea, radioterapia, hormonoterapia, otro tratamiento dirigido o una combinación de lo anterior. Normalmente, el paciente será un paciente con cáncer de mama que se ha
- 35 sometido, o está sometido, a un tratamiento para un tumor primario y que se ha identificado que tiene potencial para desarrollar un fenotipo metastásico. En un modo de realización, el paciente tiene un cáncer de mama sin afectación ganglionar positivo para ER.
- 40 En la memoria descriptiva, se debe entender que el término "recaída" significa la recaída del cáncer del que se están obteniendo muestras en el paciente, en el que el cáncer ha reaparecido en el área de la que se obtienen muestras después del tratamiento, por ejemplo, si se obtienen muestras de cáncer de mama, la recaída del cáncer de mama en el tejido mamario (fuente). También se debe entender que el término significa la recaída de un cáncer primario cuyo sitio es diferente al del cáncer del que inicialmente se obtuvieron muestras, es decir, el cáncer ha reaparecido en un área de la que no se obtuvieron muestras después del tratamiento, tal como las recaídas no locorregionales.
- 45 En esta memoria descriptiva, se debe entender que el término "mal resultado" significa que las probabilidades de supervivencia sin enfermedad son bajas.
- 50 En la memoria descriptiva, se debe entender que el término "tasa de supervivencia" significa el período de tiempo durante el cual un paciente diagnosticado de cáncer, tal como cáncer de mama, probablemente sobrevivirá. La tasa de supervivencia se expresa como una tasa de supervivencia a 5 años, una tasa de supervivencia a 10 años, una tasa de supervivencia a 15 años, una tasa de supervivencia a 20 años, una tasa de supervivencia a 25 años, una tasa de supervivencia a 30 años, una tasa de supervivencia a 35 años, una tasa de supervivencia a 40 años, una tasa de supervivencia a 45 años o una tasa de supervivencia a 50 años. Idealmente, la tasa de supervivencia se expresa como una tasa de supervivencia a 5 años o una tasa de supervivencia a 10 años.
- 55 En esta memoria descriptiva, se debe entender que el término "tratamiento" significa su significado aceptado en general que engloba prohibir, prevenir, restringir y ralentizar, detener o revertir la progresión o gravedad de un fenotipo de cáncer de mama metastásico, en recaída o existente u otro fenotipo de cáncer.
- 60 En esta memoria descriptiva, se debe entender que el término "al menos dos" significa y engloba que al menos dos, al menos tres, al menos cuatro, al menos cinco, al menos seis, al menos siete, al menos ocho, al menos nueve o todos los genes se pueden seleccionar del grupo que consiste en FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367 y TCF19.
- 65 Los medios de almacenamiento legibles por ordenador pueden ser cualquier medio tangible disponible al que se

5 pueda acceder mediante un ordenador. Los medios de almacenamiento legibles por ordenador incluyen medios tangibles volátiles y no volátiles, extraíbles y no extraíbles implementados en cualquier método o tecnología para el almacenamiento de información, tal como instrucciones legibles por ordenador, estructuras de datos, módulos de programa u otros datos. Los medios de almacenamiento legibles por ordenador incluyen, pero no se limitan a, RAM (memoria de acceso aleatorio), ROM (memoria de solo lectura), EPROM (memoria de solo lectura programable borrrable), EEPROM (memoria de solo lectura programable y borrrable eléctricamente), memoria flash u otra tecnología de memoria, CD-ROM (memoria de solo lectura de disco compacto), DVD (discos versátiles digitales) u otros medios de almacenamiento óptico, casetes magnéticos, cinta magnética, almacenamiento en disco magnético u otros medios de almacenamiento magnético, otros tipos de memoria volátil y no volátil, y cualquier otro medio tangible que se pueda usar para almacenar la información deseada y al que se pueda acceder mediante un ordenador incluyendo cualquier combinación adecuada de los anteriores.

15 Los datos legibles por ordenador incorporados en uno o más medios de almacenamiento legibles por ordenador pueden definir instrucciones, por ejemplo, como parte de uno o más programas que, como resultado de su ejecución por un ordenador, dan instrucciones al ordenador para que realice una o más de las funciones descritas en el presente documento, y/o diversos modos de realización, variaciones y combinaciones de las mismas. Dichas instrucciones se pueden escribir en cualquiera de una pluralidad de lenguajes de programación, por ejemplo, Java, J#, Visual Basic, C, C#, C++, Fortran, Pascal, Eiffel, Basic, lenguaje ensamblador COBOL y similares, o cualquiera de una variedad de combinaciones de los mismos. Los medios de almacenamiento legibles por ordenador en los que se incorporan dichas instrucciones pueden residir en uno o más de los componentes de cualquiera de los sistemas, o un medio de almacenamiento legible por ordenador descrito en el presente documento puede distribuirse a través de uno o más de dichos componentes.

25 Los medios de almacenamiento legibles por ordenador pueden ser transportables de modo que las instrucciones almacenadas en los mismos se puedan cargar en cualquier recurso informático para implementar los aspectos de la presente invención analizados en el presente documento. Además, se debe apreciar que las instrucciones almacenadas en el medio legible por ordenador, descrito anteriormente, no se limitan a las instrucciones incorporadas como parte de un programa de aplicación que se ejecuta en un ordenador central. Por el contrario, las instrucciones se pueden incorporar como cualquier tipo de código de ordenador (por ejemplo, programa informático o microcódigo) que se puede emplear para programar un ordenador para implementar aspectos de la presente invención. Las instrucciones ejecutables del ordenador pueden estar escritas en un lenguaje de ordenador adecuado o en una combinación de varios lenguajes. Los expertos en la técnica conocen los métodos básicos de biología computacional y se describen, por ejemplo, en Setubal and Meidanis *et al.*, Introduction to Computational Biology Methods (PWS Publishing Company, Boston, 1997); Salzberg, Searles, Kasif, (Ed.), Computational Methods in Molecular Biology, (Elsevier, Amsterdam, 1998); Rashidi y Buehler, Bioinformatics Basics: Application in Biological Science and Medicine (CRC Press, London, 2000) y Ouelette y Bzevanis Bioinformatics: A Practical Guide for Analysis of Gene and Proteins (Wiley & Sons, Inc., 2.^a ed., 2001).

40 Los módulos funcionales de determinados modos de realización de la invención incluyen como mínimo un sistema de determinación, un dispositivo de almacenamiento, opcionalmente un módulo de comparación, y un módulo de visualización. Los módulos funcionales se pueden ejecutar en uno o varios ordenadores, o mediante el uso de una o varias redes de ordenadores. El sistema de determinación tiene instrucciones ejecutables por ordenador para proporcionar, por ejemplo, niveles de expresión de al menos dos genes (o una proteína codificada por dichos genes) seleccionados del grupo que consiste en FOXM1, UHRF1, PTTG1, E2F1, MYBL2 y HMGB2, y opcionalmente incluye p16^{INK4A}, en forma legible por ordenador.

50 El sistema de determinación puede comprender cualquier sistema para analizar una muestra tumoral de cáncer de mama para la expresión de genes (o proteínas codificadas por dichos genes) seleccionados del grupo que consiste en FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367, TCF19 y p16^{INK4A}. Se pueden emplear métodos estándar, como inmunohistoquímica, inmunoelectrotransferencia, una transferencia de tipo Northern, una transferencia de tipo Southern, reacción en cadena de la polimerasa (PCR) cuantitativa, PCR con retrotranscriptasa (RT-PCR), RT-PCR cuantitativa ultrarrápida (qRT-PCR), un ensayo de inmunoadsorción enzimática (ELISA), determinación de proteínas en geles de poliacrilamida, secuenciación de ARN, micromatrices de ARN y otras técnicas de hibridación o amplificación de ARN, y dichos métodos conocidos por los expertos en la técnica.

60 La información determinada en el sistema de determinación se puede leer mediante el dispositivo de almacenamiento. Como se usa en el presente documento, el "dispositivo de almacenamiento" pretende incluir cualquier aparato informático o de procesamiento adecuado u otro dispositivo configurado o adaptado para almacenar datos o información. Los ejemplos de un aparato electrónico adecuado para su uso con la presente invención incluyen un aparato informático independiente, redes de telecomunicaciones de datos, incluyendo redes de área local (LAN), redes de área amplia (WAN), Internet, Intranet y Extranet, y sistemas de procesamiento informático locales y distribuidos. Los dispositivos de almacenamiento también incluyen, pero no se limitan a: medios de almacenamiento magnético, tales como disquetes, medios de almacenamiento de disco duro, cinta magnética, medios de almacenamiento óptico tales como CD-ROM, DVD, medios de almacenamiento electrónico tales como RAM, ROM, EPROM, EEPROM y similares, discos duros generales e híbridos de estas categorías, tal como medios

de almacenamiento magnéticos/ópticos. El dispositivo de almacenamiento está adaptado o configurado para haber registrado en el mismo información de secuencias de ácidos nucleicos. Dicha información se puede proporcionar en forma digital que se puede transmitir y leer electrónicamente, por ejemplo, por medio de Internet, en disquete, por medio de USB (bus universal en serie) o por medio de cualquier otro modo de comunicación adecuado.

Como se usa en el presente documento, "almacenado" se refiere a un proceso para codificar información en el dispositivo de almacenamiento. Los expertos en la técnica pueden adoptar fácilmente cualquiera de los métodos actualmente conocidos para registrar información en medios conocidos para generar productos que comprenden información relacionada con la expresión de FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367, TCF19 y p16^{INK4A} en una muestra.

En un aspecto, se comparan los datos de referencia almacenados en el dispositivo de almacenamiento que se van a leer por el módulo de comparación.

El "módulo de comparación" puede usar una variedad de programas y formatos de programa informático disponibles para el operativo de comparación para comparar los datos de información de expresión de FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367, TCF19 y p16^{INK4A} determinados en el sistema de determinación para hacer referencia a muestras y/o datos de referencia almacenados. En un aspecto, el módulo de comparación está configurado para usar técnicas de reconocimiento de patrones para comparar información de una o más entradas con uno o más patrones de datos de referencia. El módulo de comparación se puede configurar usando un programa informático existente disponible comercialmente o disponible libremente para comparar patrones, tinción y se puede optimizar para comparaciones de datos particulares que se llevan a cabo. El módulo de comparación proporciona información legible por ordenador relacionada con los niveles de expresión de FOXM1, UHRF1, PTTG1, E2F1, MYBL2, HMGB2, ATAD2, E2F8, ZNF367, TCF19 y p16^{INK4A} de la muestra.

El módulo de comparación, o cualquier otro módulo de la invención, puede incluir un sistema operativo (por ejemplo, UNIX) en el que se ejecuta un sistema de gestión de bases de datos relacionales, una aplicación de la World Wide Web y un servidor de la World Wide Web. La aplicación de la World Wide Web incluye el código ejecutable necesario para generar declaraciones en lenguaje de base de datos (por ejemplo, declaraciones en lenguaje de consulta estructurado (SQL)). En general, los ejecutables incluirán declaraciones SQL incrustadas. Además, la aplicación de la World Wide Web puede incluir un archivo de configuración que contiene punteros y direcciones a las diversas entidades de programa informático que componen el servidor, así como a las diversas bases de datos externas e internas a las que se debe acceder para atender las solicitudes de los usuarios. El archivo de configuración también dirige solicitudes de recursos del servidor al hardware apropiado, ya que puede ser necesario si el servidor se distribuye en dos o más ordenadores separados. En un aspecto, el servidor de la World Wide Web admite un protocolo TCP/IP. Las redes locales como esta a veces se denominan "Intranets". Una ventaja de dichas Intranets es que permiten una fácil comunicación con las bases de datos de dominio público que residen en la World Wide Web (por ejemplo, el sitio de la World Wide Web GenBank o Swiss Pro). Por tanto, en un modo de realización preferente particular de la presente invención, los usuarios pueden acceder directamente a los datos (por medio de enlaces de hipertexto, por ejemplo) que residen en bases de datos de Internet utilizando una interfaz HTML proporcionada por navegadores web y servidores web.

El módulo de comparación proporciona un resultado de comparación legible por ordenador que se puede procesar en forma legible por ordenador mediante criterios predefinidos, o criterios definidos por un usuario, para proporcionar un contenido basado en parte en el resultado de comparación que se puede almacenar y enviar según lo solicite un usuario usando un módulo de visualización.

Por lo tanto, los métodos descritos en el presente documento proporcionan sistemas (y medios legibles por ordenador para producir sistemas informáticos) para realizar métodos como se describe en las declaraciones de invención anteriores, por ejemplo, métodos para diagnosticar el potencial metastásico o el potencial de recaída de un cáncer de mama o un cáncer no de mama en un individuo o métodos para identificar a un paciente con cáncer de mama o un paciente sin cáncer de mama adecuado para tratamiento o prevención de cáncer metastásico o en recaída con un agente terapéutico adyuvante o no adyuvante quimioterápico adecuado.

Los sistemas y los medios legibles por ordenador descritos en el presente documento son meramente modos de realización ilustrativos de la invención para realizar métodos de diagnóstico en un individuo, y no pretenden limitar el alcance de la invención. Son posibles variaciones de los sistemas y los medios legibles por ordenador descritos en el presente documento y se pretenden que estén dentro del alcance de la invención.

Los módulos de la máquina, o los usados en el medio legible por ordenador, pueden asumir numerosas configuraciones. Por ejemplo, se puede proporcionar función en una sola máquina o distribuirse en múltiples máquinas.

Materiales y métodos

Cultivo celular

5 Se cultivaron células HMEC primarias como se describe (Garbe *et al.*, 2009). Se immortalizaron células HMEC-terc usando una construcción pBABE-hTERT-hygro. Se derivaron fibroblastos embrionarios de ratón (MEF) de embriones de ratón C57BL6 el día embrionario 13,5 y se mantuvieron en medio DMEM complementado con FBS al 10 % (v/v) (Hyclone), 100 U/ml penicilina y 100 U/ml de estreptomycin (Gibco).

Secuenciación de ARN

10 Se extrajo ARN total de HMEC en proliferación y senescentes usando el kit RNeasy (Qiagen). Se enriquecieron especies de ARN poliadenilado a partir de 5 µg de ARN total, y se prepararon colecciones de secuenciación a partir de ARN PolyA+ usando el kit TruSeq Sample Prep (Illumina). Las colecciones se usaron directamente para generación de grupos y análisis de secuenciación usando el Genome Analyzer II (Illumina) siguiendo el protocolo del fabricante. Se realizaron la lectura automática de nucleótidos y la asignación al genoma humano (compilación hg19) usando la herramienta de alineación de secuencias BWA. Los factores de multiplicación del ARNm se calcularon en base al número total de lecturas de secuencia asignadas por gen en los dos experimentos.

Análisis de micromatrices de ADN

20 Se extrajo ARN total de MEF en proliferación y senescentes usando el kit RNeasy (Qiagen). Para cada punto de tiempo, se preparó ARN a partir de tres cultivos de MEF independientes y se combinó para reducir la variación experimental. Se preparó ARNc marcado con Cy3, para su uso con una micromatriz de 44k diseñada a medida (Agilent), y se hibridó según las instrucciones del proveedor. Se escanearon micromatrices usando el escáner de micromatrices de ADN de Agilent y se analizaron los datos como se describe previamente (Hokamp *et al.*, 2004). Se llevó a cabo un análisis de ontología génica usando el recurso bioinformático DAVID (<http://david.abcc.ncifcrf.gov/>).

25 Se descargaron conjuntos de datos de micromatrices de cáncer de mama disponibles públicamente de Rosetta Inpharmatics y Gene Expression Omnibus (GSE6532 y GSE3494). Dentro de cada conjunto de datos, los datos de expresión de cada gen se dividieron en la mediana en dos grupos, o en el percentil 33 y 66 en 3 grupos, dependiendo del análisis. Para generar una puntuación de MTR combinada, los valores de expresión génica para cada uno de los 6 genes se dividieron en la mediana, se les dio una puntuación de 1 o 2 en base al nivel de expresión, y la suma de estas puntuaciones se dividió a continuación, como anteriormente, para crear 2 o 3 grupos. La expresión génica de *INK4a* se dividió en 3 grupos (bajo, moderado y alto) en el percentil 33 y 66. Se dio al grupo moderado una puntuación de 1 y los grupos bajo y alto se combinaron y se les dio una puntuación de 2. Para generar la puntuación de ARN de OncoMasTR, la puntuación de MTR combinada y la puntuación de *INK4A* se sumaron y las puntuaciones finales se dividieron en 2 o 3 grupos. Se eliminaron muestras duplicadas en el conjunto de datos de micromatrices combinadas. El paquete *genefu* en R se usó para estimar los grupos de riesgo que se aproximan al ensayo Oncotype Dx® (basado en la firma de 21 genes) y el ensayo MammaPrint (basado en la firma de 70 genes) (Haibe-Kains *et al.*). Para el conjunto de datos de Van de Vijver, se usaron los grupos de riesgo de 70 genes definidos previamente (van de Vijver *et al.*, 2002).

40 PCR cuantitativa ultrarrápida

Se extrajo ARN total de las células usando el kit RNeasy (Qiagen) de acuerdo con el protocolo del fabricante. Se usó 1 µg de ARN para generar ADNc mediante PCR con retrotranscriptasa usando el kit de retrotranscripción con TaqMan (Applied Biosystems). Se determinaron los niveles relativos de expresión de ARNm usando la química de detección SYBR Green I (Applied Biosystems) en el sistema de PCR ultrarrápida ABI Prism 7500 Fast. El constituyente ribosómico RPLPO se usó como gen de control para la normalización (SEQ ID NO: 39 (directa - TTCATTGTGGGAGCAGAC) y SEQ ID NO: 40 (inversa - CAGCAGTTTCTCCAGAC)). Los pares de secuencias de cebadores usados son los siguientes (Div = Cebador directo; Inv = Cebador inverso):

- 50 SEQ ID NO: 1 Div: AGACCGTCCTCAACCAGCTCTTC y SEQ ID NO: 2 Inv: GAAGTGCTTGGAGATCACCGG;
 SEQ ID NO: 3 Div: CAA CAA TAG CCT ATC CAA CAT CCA G y SEQ ID NO: 4 Inv: GGA GCC CAG TCC ATC AGA ACTO C;
- 55 SEQ ID NO: 5 Div: CTGCCTGAAGAGCACCAGATTG y SEQ ID NO: 6 Inv: CAAGGATCATGAGAGGCACTCC;
 SEQ ID NO: 7 Div: CACTGACCAGCAATGCCAGTAC y SEQ ID NO: 8 Inv: CCCCTTGACAAGGTCTGGATTC;
 SEQ ID NO: 9 Div: GCTCCTAAAAGGCCACCATCTG y SEQ ID NO: 10 Inv: TGATCTTTGGGCGATGTTTCAG;
- 60 SEQ ID NO: 11 Div: TGT CAG GAC CTT CGT AGC ATT G y SEQ ID NO: 12 Inv: GGG CTT TGA TCA CCA TAA CCA TC;
- 65 SEQ ID NO: 13 Div: CAA TCT CAA CAA AAC CCT TGG C y SEQ ID NO: 14 Inv: CTC GGC GTA CTT ATT CTC CTC C;

SEQ ID NO: 15 Div: AGAGGATTTGAGGGACAGGGTC y SEQ ID NO: 16 Inv: CCTCTTCTCCTCCGGTGC;

SEQ ID NO: 17 Div: ATGGAGCTGGGTGCTGAGAAC y SEQ ID NO: 18 Inv: CCTTCTCAACTCCATGAGCCC;

5 SEQ ID NO: 19 Div: ACA AAG AAG GAA ATA GAG GGA CCG y SEQ ID NO: 20 Inv: GAT GAG TGG GAG ACT TGG GTT C;

SEQ ID NO: 21 Div: CAGCCCGAGCTTTTGTACAAC y SEQ ID NO: 22 Inv: TTCGCTGCTGACATCTGAGTTC;

10 SEQ ID NO: 23 Div: AAGGTGAGCAAGATGGAAATCC y SEQ ID NO: 24 Inv: CGATCTGCAGGTCCAAGATGTAG

SEQ ID NO: 25 Div: CTCTCTGAGGCCAAGGATCTCC y SEQ ID NO: 26 Inv: CCTTGTTGCAGTATTTGCAGTTG;

15 SEQ ID NO: 27 Div: TGAGCCTGCAGATTTTAAGGTG y SEQ ID NO: 28 Inv: TGGAAAGCTTCTCACGGCATAAC;

SEQ ID NO: 29 Div: AGCTGGCCTGAATCATTAAATACG y SEQ ID NO: 30 Inv: GGTGAAGGTCCATGAGACAAGG;

SEQ ID NO: 31 Div: GGGACAGTAAAAATGTGTCCTGC y SEQ ID NO: 32 Inv: TGCCAGCAATAGATGCTTTTTTG;

20 SEQ ID NO: 33 Div: CAT TCC CGC TCT CCT TCC C y SEQ ID NO: 34 Inv: GCT CGG CTC CCC AGA ATC;

SEQ ID NO: 35 Div: CCTCACTGGAGGAGTGATGCG y SEQ ID NO: 36 Inv: AAGCATCCTAAGCCATTCCATG;

25 SEQ ID NO: 37 Div: CCA TTG AAA ACA AGG ACG ATG C y SEQ ID NO: 38 Inv: CTG TCC CCA ACA ACA TCA AGC.

ChIP y secuenciación por ChIP

30 Se realizaron análisis por ChIP como se describe anteriormente (Bracken *et al.*, 2006). Para ChIP-SEQ, se combinó ADN de 10 experimentos independientes por ChIP y se cuantificó usando un fluorómetro Qubit (Invitrogen). Se generaron colecciones de secuenciación usando 100 ng de ADN inmunoprecipitado usando el kit de preparación de muestras ChIP-SEQ (Illumina). Se purificó ADN de la colección amplificada mediante aislamiento en gel y se verificó la calidad para asegurar la ausencia de contaminación del dímero del adaptador usando el Bioanalyzer 2100 y el ensayo de chip de ADN de alta sensibilidad (Agilent). Se cuantificaron las colecciones de ADN y se diluyeron a 35 10 pM. Las colecciones diluidas se usaron directamente para generación de agrupaciones y análisis de secuenciación usando el Genome Analyzer II (Illumina) siguiendo el protocolo del fabricante. Se realizaron la lectura automática de nucleótidos y la asignación al genoma humano (hg19) de las secuencias de 42 pb usando la herramienta de alineación Bowtie que permite hasta 2 emparejamientos erróneos en cada lectura. Para evitar cualquier sesgo de la PCR, solo se permitieron dos lecturas por posición cromosómica, eliminando por tanto los picos espúreos. Se realizó la detección de picos usando MAC, y se usó ADN de entrada como control para la 40 normalización.

Análisis por ARACNe

45 Se generaron redes transcripcionales de cáncer de mama mediante ARACNe (Margolin *et al.*, 2006), usando conjuntos de datos publicados sobre cáncer de mama (ExPO; Loi *et al.*, 2007; van de Vijver *et al.*, 2002), y se consultaron usando firmas génicas internas o publicadas. Para las redes ExPO y Loi, ARACNe se ejecutó en los conjuntos de datos de expresión completos, mientras que para la red NKI, se aplicó una etapa de filtrado antes de ARACNe para eliminar sondas no informativas. La firma de 70 genes de Mammaprint se derivó a través de una 50 clasificación supervisada de datos de micromatrices de ADN de 78 pacientes sin afectación ganglionar, y predice un corto tiempo hasta la metástasis a distancia (van 't Veer *et al.*, 2002). La firma más grande de 231 genes de la cual se deriva la firma de 70 genes se usó para este análisis. La firma de grado genómico se desarrolló a partir de un conjunto de datos de entrenamiento de 64 tumores de mama positivos para ER, y está compuesta por genes expresados diferencialmente entre grado histológico bajo y alto. La lista más grande de 207 genes de la que se 55 derivó el índice de grado genómico de 97 genes se usó para el análisis por ARACNe (Sotiriou *et al.*, 2006).

Análisis estadístico

60 Se usaron curvas de supervivencia de Kaplan-Meier para el análisis de la supervivencia y se calcularon valores de chi cuadrado y de p usando la prueba del orden logarítmico. Se usó un análisis multivariado de riesgos proporcionales de Cox para evaluar el valor pronóstico sumado de genes individuales y puntuaciones combinadas, además de un modelo clínico estándar que incluye la edad (<50, >=50 años), el estado ganglionar (con o sin afectación), el tamaño del tumor (<2 cm, >=2 cm), el grado tumoral (1 frente a 2 y 3), el estado del tratamiento, y el estado de ER y HER2. También se llevó a cabo un análisis multivariado usando el modelo clínico estándar anterior, 65 más el grupo de riesgo predicho con firma de 21 genes. La contribución de cada marcador se evaluó mediante el cambio en el cociente de verosimilitud (LR-Chi, $df=1$) y se calcularon los valores de p. Un valor de p de menos de

0,05 se consideró significativo. El criterio principal de valoración clínica usado para el análisis de los datos de micromatrices y TMA fue la supervivencia sin recaída (SSR). Todos los análisis estadísticos se llevaron a cabo usando el lenguaje de programación R (versión 2.15.0). Se crearon matrices cromáticas usando una herramienta en línea (<http://chibi.ubc.ca/matrix2png>). Se llevó a cabo un análisis de enriquecimiento calculando el número de genes únicos de "mal pronóstico" presentes en la firma de "proliferación central", en comparación con lo que se esperaría en todo el genoma (Observado/Esperado). Los genes únicos en las firmas de "mal pronóstico" fueron n=61 para la firma de MammaPrint y n=207 para la firma de grado genómico, y el análisis se normalizó en base a la plataforma experimental usada para derivar la firma.

10 Cohorte de TMA

La micromatriz tisular (TMA) usada en este estudio se derivó de una cohorte de referencia de 512 casos consecutivos de cáncer de mama invasor diagnosticados en el Departamento de Anatomopatología, Hospital Universitario de Malmo, Malmo, Suecia, entre 1988 y 1992, y se ha descrito previamente (Svensson *et al.*, 2005). En resumen, la mediana de edad fue de 65 años (intervalo 27-96) y la mediana del tiempo de seguimiento con respecto a la supervivencia específica de la enfermedad y global fue de 11 años (intervalo 0-17). Se excluyeron pacientes con enfermedad en recaída y tratamientos sistémicos previos, así como varios casos de carcimona ductal *in situ* (CDIS) ductal clasificados erróneamente. Doscientos sesenta y tres pacientes habían muerto en el último seguimiento (diciembre de 2004), 90 de los cuales se clasificaron como muertes específicas por cáncer de mama. Se extrajeron núcleos de tejido (1 mm) de áreas representativas de cáncer invasor de bloques de donantes y se presentaron por duplicado. Este estudio ha sido aprobado por el Comité de Ética de la Universidad de Lund y el Hospital Universitario de Malmo.

25 Inmunohistoquímica

Se desparafinaron portaobjetos de TMA en xileno y se rehidrataron en alcoholes de gradiente descendente. Se realizó la recuperación de antígeno mediada por calor usando tampón de citrato de sodio 10 mM (pH 6,0) en un módulo PT (LabVision, Reino Unido) durante 15 min a 95 °C. Se usó el kit de IHC de LabVision (LabVision, Reino Unido) para la tinción. Se bloqueó la actividad de peroxidasa endógena por incubación con peróxido de hidrógeno al 3 % durante 10 min. Se bloquearon secciones durante 10 min en un agente de bloqueo UV y el anticuerpo primario relevante se incubó durante 1 h. Se lavaron las secciones en solución salina tamponada con fosfato con Tween 20 al 0,1 % (PBS-T), después de lo cual se aplicó potenciador de anticuerpo primario durante 20 min, y se lavaron las secciones en PBS-T. A continuación, se incubaron las secciones con polímero HRP durante 15 min, se lavaron en PBS-T y a continuación se desarrollaron durante 10 min usando una solución de diaminobencidina (DAB) (LabVision, Reino Unido). Todas las incubaciones y etapas de lavado se llevaron a cabo a temperatura ambiente. Las secciones se contratificaron en hematoxilina, se deshidrataron en alcohol y xileno, y se montaron usando medio de montaje DPX. Como control negativo, el anticuerpo primario se sustituyó con PBS-T.

Los anticuerpos primarios usados fueron HMGB2 (Abcam; 1:1500), UHRF1 (BD Biosciences; 1:1000), PTTG1 (Invitrogen; 1:500), FOXM1 (Santa Cruz, C20; 1:300) y p16 (Clon JC8; 1:5000). Se habían teñido previamente secciones de TMA en el Ventana Benchmark (Ventana Medical Systems Inc, EE. UU.) usando anticuerpos prediluidos para ER (clon 6F11, Ventana), PR (clon 16, Ventana) y Her2 (Pathway CB-USA 760-2694), o en el Dako Techmate 500 (Dako, Dinamarca) para Ki-67 (1:200, M7240, Dako).

45 Análisis de TMA

Se escanearon portaobjetos a 20× aumentos utilizando un escáner de portaobjetos ScanScope XT (Aperio Technologies, CA). Para la puntuación manual, se evaluó la tinción de las células tumorales por un anatomopatólogo en base a la intensidad, en una escala de negativa (0), débil (1), moderada (2) y fuerte (3); y el porcentaje, en una escala de 0-6 (0 = 0-1 %; 1 = 1-10 %; 2 = 10-25 %; 3 = 25-50 %; 4 = 50-75 %; 5 = 75-90 %; 6 = 90-100 %). La tinción para los factores HMGB2 y UHRF1 fue predominantemente nuclear, mientras que PTTG1, FOXM1 y p16^{INK4A} tiñeron los compartimientos tanto nuclear como citoplasmático y se puntuaron en consecuencia. Para UHRF1, PTTG1 y p16^{INK4A}, el porcentaje de núcleos tumorales positivos fue la variable más significativa en relación con el resultado y se utilizó en todos los demás análisis. Para HMGB2, se usó una puntuación de Allred modificada (intensidad más porcentaje) y, para FOXM1, el porcentaje de positividad citoplasmática dentro de las células tumorales fue la variable más significativa. Para el análisis de los cuatro MTR, se aplicó un umbral de positividad independientemente para cada variable, para crear una puntuación binaria con expresión baja (0) y alta (1). Para p16^{INK4A}, los grupos de expresión 'negativo' (0 % de células positivas) y 'alto' (>50 % de células positivas) se combinaron y se les dio una puntuación de 1, y se compararon con el grupo 'moderado' con una puntuación de 0. Para generar una puntuación de MTR combinada al nivel de proteína, se generó la suma de las puntuaciones binarios para los cuatro MTR. Los tumores con expresión alta de >1 MTR se clasificaron como que tenían una puntuación de MTR alta. Para generar la puntuación de 4MTR + p16^{INK4A} combinada (puntuación por IHC de OncoMasTR), la puntuación de 4MTR binaria se combinó con la puntuación de p16^{INK4A} binaria, y se dividió en dos grupos con un umbral de >2.

65 Resultados

Identificación de una firma de expresión génica de 'proliferación central'

El solicitante se propuso identificar un conjunto de genes de 'proliferación central' que se expresan de manera consecuente en células en crecimiento activo de una manera independiente del linaje. Para hacer esto, el solicitante aisló células epiteliales mamarias humanas (HMEC) y fibroblastos embrionarios de ratón (MEF) y los sometió a pases hacia la senescencia celular, como se caracteriza por un incremento en los niveles de p16^{INK4A} (Zindy *et al.*, 1997), y una disminución en los niveles del gen diana de E2F, EZH2 (Bracken *et al.*, 2003) (figura 1A). Luego, el solicitante realizó un análisis de expresión de ARNm de todo el genoma en cultivos de HMEC y MEF en proliferación y senescentes e identificó cuatro grupos de genes expresados diferencialmente (figura 1B). Los cambios de expresión de genes representativos de cada grupo se validaron por RT-PCR cuantitativa (figura 1C). Los genes del grupo 3, que estaban regulados por disminución durante pases sucesivos de las células HMEC, incluían varios genes implicados en procesos específicos de células epiteliales mamarias, tal como la citoqueratina luminal KRT19 y la proteína de unión hermética CLDN3. Esto es consecuente con el hecho de que la proporción de células luminales y mioepiteliales varía durante pases sucesivos de células HMEC (Garbe *et al.*, 2009). Por lo tanto, el solicitante razonó que muchos de los genes dentro del grupo 3 estaban regulados por disminución independientemente de la disminución progresiva de la tasa de proliferación. Consecuentemente con esto, un análisis de ontología génica para cada uno de los cuatro grupos de genes reveló un mayor enriquecimiento de las categorías funcionales vinculadas al ciclo celular y la proliferación en el grupo 4, en comparación con el grupo 3 (figura 1D). Por lo tanto, la estrategia para combinar los cambios de expresión de MEF y HMEC sometidas a pases sucesivos permitió la identificación de genes de 'proliferación central' en células epiteliales mamarias.

Luego, el solicitante deseaba determinar hasta qué punto estaban enriquecidos los genes del grupo 4 de 'proliferación central' en dos de las firmas más conocidas de 'mal pronóstico de cáncer de mama', la firma de 70 genes de MammaPrint y la firma de 'grado genómico' (Sotiriou *et al.*, 2006; van 't Veer *et al.*, 2002). Esto reveló un enriquecimiento significativo de los genes del grupo 4, pero no de los genes de los grupos 1-3, en ambas firmas de mal pronóstico (figura 1E), lo que respalda la opinión, tal vez no sorprendente, de que un factor contribuyente importante al poder pronóstico de estas dos firmas es su capacidad de medir simplemente la proliferación de células tumorales (Mosley y Keri, 2008; Wirapati *et al.*, 2008).

Identificación de reguladores transcripcionales maestros (MTR) en dirección 5' de la firma de 'proliferación central'

De forma interesante, a pesar de la capacidad de varias firmas de mal pronóstico establecidas para predecir el resultado del cáncer de mama, hay sorprendentemente poco solapamiento entre las propias firmas (Fan *et al.*, 2006; Haibe-Kains *et al.*, 2008). El solicitante razonó que los genes proliferativos dentro de estas firmas, varios de los cuales son genes de 'proliferación central' en el análisis presentado en el presente documento (figura 1E), pueden ser de hecho solo pasajeros, en lugar de impulsores de la proliferación de células tumorales. Por lo tanto, el solicitante planteó la hipótesis de que los reguladores transcripcionales en dirección 5' de los genes de 'proliferación central' serían predictores más fiables del pronóstico del cáncer de mama.

Teniendo en cuenta la naturaleza jerárquica de la regulación de la expresión génica, el solicitante deseaba identificar los reguladores transcripcionales clave en dirección 5' de la firma de proliferación central. Para identificar los reguladores transcripcionales maestros (MTR) en dirección 5' de los genes de 'proliferación central', se usó un enfoque bioinformático llamado ARACNe (Carro *et al.*, 2010; Margolin *et al.*, 2006). Este enfoque utiliza redes de interacción construidas a partir de conjuntos de datos de expresión génica para inferir interacciones transcripcionales directas. ARACNe se aplicó a tres conjuntos de datos de expresión génica de cáncer de mama disponibles públicamente (ExPO; Loi *et al.*, 2007; van de Vijver *et al.*, 2002) y predijo varios MTR en dirección 5' de los genes de 'proliferación central' en el cáncer de mama (figura 2A y tabla 1). Entre los MTR de mayor puntuación se encontraban Forkhead Box M1 (FOXM1), PHD similar a ubiquitina y dedo RING 1 (UHRF1), Securina o gen transformador de tumor hipofisario 1 (PTTG1), Factor de transformación E2F 1 (E2F1), homólogo del oncogén vírico de mieloblastosis v-myb (aviar) de tipo 2 (MYBL2) y caja del grupo de alta movilidad 2 (HMGB2), que fueron relativamente consecuentes en los tres conjuntos de datos de cáncer de mama independientes, lo que respalda la idea de que los MTR demostrarían ser indicadores más fiables de la proliferación de células tumorales que sus genes diana en dirección 3'. También se identificaron cuatro genes adicionales consecuentemente en los conjuntos de datos como que estaban en dirección 5' de los genes de 'proliferación central'. Estos son ATAD2, E2F8, ZNF367 y TCF19.

Tabla 1: Reguladores transcripcionales maestros principales de las firmas de expresión indicadas según lo predicho por ARACNe

Rango	Firma de proliferación central	Firma de mal pronóstico	Firma de grado genómico
1	FOXM1	PTTG1	PTTG1
2	PTTG1	FOXM1	FOXM1
3	UHRF1	UHRF1	UHRF1
4	MYBL2	ATAD2	MYBL2

5	HMGB2	MYBL2	ATAD2
6	ATAD2	ZNF367	HMGB2
7	E2F1	HMGB2	ZBTB20
8	E2F8	TCF19	E2F1
9	ZNF367	E2F8	E2F8
10	TCF19	E2F1	ZNF367

Luego, el solicitante deseaba determinar si algunos de los MTR se unían directamente a los promotores del grupo 4, genes de 'proliferación central', como se predice. Las inmunoprecipitaciones de cromatina (ChIP) seguidas de PCR cuantitativa ultrarrápida (qPCR) confirmaron la unión directa de cuatro de los MTR (FOX M1, MYBL2, E2F1 y HMGB2) a los promotores de los genes de 'proliferación central' en células HMEC-Terc (figura 2B) . Para obtener una visión más amplia sobre la unión de MTR en todo el genoma, se realizó ChIP seguido de secuenciación de alto rendimiento (ChIP-seq) en células HMEC-Terc para E2F1, MYBL2 y FOX M1. Esto reveló que los tres MTR se asocian principalmente con los promotores del grupo 4, los genes de 'proliferación central' y, en menor medida, algunos genes del grupo 3 (figura 2C). Las pistas de ChIP-seq de tres genes representativos muestran picos que representan la unión de E2F1, MYBL2 y FOX M1 en los promotores de los genes *CCNB1*, *UBE2C* y *CENPA* (figura 2D), pero no en el promotor de un gen no expresado en HMEC, *KRT2*. El solicitante no pudo investigar los patrones de unión en todo el genoma de PTTG1 o UHRF1 debido a la falta de anticuerpos adecuados de grado ChIP de alta calidad. Sin embargo, el hecho de que se haya informado de que PTTG1 tiene un papel en la activación transcripcional de genes del ciclo celular, respalda las predicciones de ARACNe (Tong y Eigler, 2009; Tong *et al.*, 2007). Por otro lado, UHRF1 se considera, en general, un represor transcripcional, que se requiere para el mantenimiento de la metilación del ADN durante la división celular (Bostick *et al.*, 2007). Por lo tanto, es poco probable que UHRF1 regule directamente los genes de proliferación central, y es más probable que sea un gen proliferativo corregulado. Respalda esta posibilidad, E2F1, MYBL2 y FOX M1 también se unen al promotor del gen *UHRF1* en células HMEC.

Paralelamente a la identificación de estos MTR, el solicitante también llevó a cabo un análisis no sesgado de la supervivencia de 565 pacientes sin afectación ganglionar de cuatro conjuntos de datos independientes de expresión génica del cáncer de mama (Buffa *et al.*, 2011; Ivshina *et al.*, 2006; Loi *et al.*, 2007; van de Vijver *et al.*, 2002), para identificar los genes asociados con la supervivencia del paciente en orden de clasificación (tabla 2). Sorprendentemente, este análisis identificó varios de los MTR de proliferación entre los 20 genes principales asociados con el resultado del cáncer de mama en estos pacientes sin afectación ganglionar, con varios de estos MTR de proliferación con puntuaciones más altas que los biomarcadores clínicos convencionales (ER, PR, Ki67) o genes incorporados en el ensayo Oncotype Dx® (BIRC5, CCNB1, BCL2, CTSL2). Este resultado ilustra el poder de estos MTR como biomarcadores pronósticos, y nos inspiró a investigarlos más a fondo.

Tabla 2: Análisis no sesgado de la supervivencia de todos los genes en cuatro conjuntos de datos de cáncer de mama (Van de Vijver *et al.*, Loi *et al.*, Ivshina *et al.*, Buffa *et al.*). Los valores de expresión génica se dividieron en la mediana, se analizaron en relación con la supervivencia global usando la prueba del orden logarítmico, y se clasificaron en orden de poder pronóstico (cursiva = MTR; negrita = biomarcadores clínicos convencionales; sombreado = genes de la vía de OncoMasTR)

Rango	Gen	Función
1	PRC1	Ciclo celular
2	<i>UHRF1</i>	<i>MTR de proliferación</i>
3	ZWINT	Ciclo celular
4	IGBP1	Transducción de señales
5	RPL29	Proteína ribosómica
6	CCNB2	Proliferación
7	TRIP13	Reparación de ADN
8	CDC45L	Ciclo celular
9	TROPA	Adhesión celular
10	TACC3	Proliferación
11	LRP2	Lipoproteína/Señalización hormonal/Respuesta al estrés
12	MAD2L1	Ciclo celular
13	BLM	Replicación y reparación de ADN

ES 2 753 625 T3

14	CDKN3	Ciclo celular
15	SEC14L2	Biosíntesis de colesterol
16	MYBL2	<i>MTR de proliferación (y Oncotype Dx)</i>
17	BIRC5	Oncotype Dx® (antiapoptosis)
18	PTTG1	<i>MTR de proliferación</i>
19	H2AFZ	Remodelación de la cromatina
20	TK1	Replicación de ADN
21	FBXO5	Vía de la ubiquitina
22	EIF2C2	Vía de ARNi
23	EBP	Biosíntesis de colesterol
24	PLP2	Proteína del retículo endoplásmico
25	EZH2	Proliferación/proteína Polycomb
26	FOXM1	<i>MTR de proliferación</i>
27	PDZK1	Proteína de andamiaje/metabolismo del colesterol
28	FEN1	Reparación de ADN
29	TXNRD1	Estrés oxidativo
30	COL4A1	Componente de membrana basal
31	STC2	Homeostasis del calcio/Señalización de estrógenos
32	GPR56	Señalización celular
33	SQLE	Biosíntesis de esteroides
34	EXO1	Reparación de ADN
35	YWHAZ	Antiapoptosis
36	GATA3	Respuesta hormonal
37	KIF4A	Ciclo celular
38	ADM	Señalización hormonal
39	CREBL2	Ciclo celular
40	TTK	Proliferación
41	BUB1	Ciclo celular/apoptosis
42	CTPS	Síntesis de ADN
43	CHST3	Migración celular/Respuesta a herida
44	CAMLG	Apoptosis/homeostasis del calcio
45	PSMD1	Componente del proteasoma
46	KIF13B	Vía de daño del ADN
47	NRM	Proteína de membrana nuclear
48	STXBP2	Transporte de vesículas
49	GALT	Metabolismo de glucoproteínas
50	GPI	Metabolismo del glucógeno/angiogénesis
51	POLD1	Replicación de ADN
52	RRM2	Replicación de ADN
53	MYB	Proliferación/Diferenciación
54	CDC20	Ciclo celular
55	SERPINH1	Respuesta inflamatoria/Protólisis

ES 2 753 625 T3

56	SERPINA3	Proteólisis
57	HMMR	Motilidad celular
58	PDCD4	Invasión/Apoptosis
59	PGK1	Metabolismo de la glucosa
60	RQCD1	Diferenciación celular
61	NDRG1	Respuesta al estrés/apoptosis
62	SLU7	Ajuste de ARNm
63	ESR1	Oncotype Dx® (respuesta hormonal)
64	SPARCL1	Migración celular/Invasión
65	NME5	Antiapoptosis
66	BTG2	Antiproliferativo
67	WDR5	Modificación de histonas
68	HMGCL	Cetogénesis
69	SERPINE1	Invasión de migración celular
70	BTN2A1	Metabolismo lipídico
71	CELSR2	Adhesión/señalización célula a célula
72	PKM2	Metabolismo de la glucosa
73	ORC1L	Replicación de ADN
74	FANCA	Reparación de ADN
75	FLT3	Angiogénesis
76	TYMS	
77	SIRT1	
78	GARS	
79	XPOT	
80	FUT8	Glucosilación de proteínas
81	BTD	
82	LZTFL1	
83	STIP1	
84	ME1	
85	UCP2	
86	RPL14	
87	NP	
88	CIRBP	
89	ORC6L	
90	PSMD7	
91	CCNE2	
92	CENPA	
93	CDC25B	
94	<i>E2F1</i>	<i>MTR de proliferación</i>
95	CCNB1	Proliferación (Oncotype Dx®)
96	H2AFX	
97	RAD54L	

ES 2 753 625 T3

98	ADAMTS7	
99	LEPR	
100	KIAA1609	
101	KIAA1407	
102	CCNA2	Ciclo celular
103	PFKL	
104	KIAA0999	
105	SLC23A2	
106	FUCA1	
107	RFC2	
108	CCNI	
109	NEK2	
110	HS3ST1	
111	DYSF	
112	AGTR1	
113	VAV3	
114	PDE6B	
115	POLA2	
116	ATP5G3	
117	KIAA0831	
118	PTMA	
119	GSTM3	
120	PHB	
121	MAP4K4	
122	PGR	Oncotype Dx® (respuesta hormonal)
123	BCL2	Oncotype Dx® (antiapoptosis)
124	IGFBP4	
125	CENPE	
126	CYC1	
127	CD01	
128	MYCBP	
129	SKP2	
130	RAB3D	
131	DHCR7	
132	KIAA1324	
133	ATP11A	
134	BECN1	
135	HDGF	
136	PCYT1A	
137	TNNC1	
138	CENPF	
139	ADCY1	Metabolismo de ATP

ES 2 753 625 T3

140	MKI67	Oncotype Dx® (proliferación)
141	KIAA0101	
142	KCNN3	
143	SLC19A1	
144	EPHA4	Adhesión celular/señalización/migración/invasión
145	CDC25C	
146	NFATC1	
147	PDE5A	
148	ABCF1	
149	CKS2	
150	PRRG2	Señalización de calcio y vitamina K
151	CLDN4	Adhesión celular
152	GTSE1	
153	RAI2	
154	PRLR	Señalización hormonal
155	SEMA7A	
156	CPT1A	
157	PDHA1	
158	RAB27B	
159	MCM2	
160	FLNB	
161	SLC2A3	Transporte/metabolismo de glucosa
162	IMPDH2	
163	<i>HMGB2</i>	<i>MTR de proliferación</i>
164	HOXB13	Proteína Homeobox
165	NFRKB	
166	RPS6KA5	
167	CRIP2	
168	BTF3	
169	MAGED1	
170	NAPG	
171	ASNS	
172	PTTG2	
173	TPST1	Respuesta a herida
174	RPLP1	
175	GLTSCR2	
176	PLA2R1	
177	POLQ	
178	CSTB	
179	CALU	Señalización dependiente del calcio
180	PPARD	
181	TXN	

182	NAT1	
183	MYO7A	
184	EIF4G1	
185	SHMT2	
186	PTDSS1	
187	LHX2	
188	PLA2G10	
189	ANLN	
190	ATP5J	
191	POLR2D	
192	SERF1A	
193	EPHB4	
194	CDC23	
195	PTPN14	
196	PEX12	
197	PPP1R11	
198	CSPG5	
199	DONSON	
200	CTSL2	Oncotype Dx® (Invasión)

Los MTR proliferativos son excelentes predictores del pronóstico del cáncer de mama en los niveles de ARN y proteínas

5 Luego, se exploró la posible importancia clínica de los MTR como marcadores pronósticos en el cáncer de mama. El solicitante comenzó realizando un análisis no sesgado por ARACNe de las firmas de MammaPrint y grado genómico, que han demostrado que predicen el resultado clínico en pacientes con cáncer de mama (Sotiriou *et al.*, 2006; van 't Veer *et al.*, 2002). Sorprendentemente, en los tres conjuntos de datos independientes analizados (ExPO; 10 *Loi et al.*, 2007; van de Vijver *et al.*, 2002), se predijo que FOXM1, E2F1, MYBL2, UHRF1, PTTG1, HMGB2, ATAD2, E2F8, ZNF367 y TCF19 estaban entre los principales reguladores en dirección 5' de ambas firmas de "mal pronóstico" (figura 3A y tabla 1). Esto sugiere que estos MTR regulan directamente la expresión de muchos genes dentro de las firmas de pronóstico tanto MammaPrint y como grado genómico.

15 A continuación, el solicitante deseaba explorar la posibilidad de que los MTR puedan ser predictores fiables del mal pronóstico. La asociación de cada MTR individual con la supervivencia del paciente se examinó en un conjunto de datos combinados de tres estudios de micromatrices publicados que representan la expresión del ARNm del genoma completo 457 tumores de mama sin afectación ganglionar no tratados mediante quimioterapia (*Loi et al.*, 2007; *Miller et al.*, 2005; van de Vijver *et al.*, 2002). Esto reveló que los altos niveles de expresión de ARNm de cualquiera de FOXM1, E2F1, MYBL2, UHRF1, PTTG1, HMGB2 en tumores de mama se asociaron 20 significativamente con un tiempo de supervivencia sin recaída reducido y una combinación de los seis MTR fue más poderosa para estratificar a los pacientes en comparación con cualquier MTR solo (figura 3B). Significativamente, usando una estrategia de clasificación en bajo/alto o bajo/moderado/alto, la combinación de seis MTR fue mejor para predecir la supervivencia sin recaída que el marcador de proliferación establecido Ki67 (figura 3B). Estos seis MTR ahora forman el panel 'central' del método o ensayo de la presente invención, también llamado ensayo OncoMasTR. 25 Los altos niveles de expresión de ARNm de ATAD2 y TCF19 en tumores de mama también se asociaron significativamente con un tiempo de supervivencia sin recaída reducido en esta cohorte (figura 9). La información de expresión no estaba disponible en esta cohorte para E2F8 y ZNF367.

30 Luego, se examinaron los niveles de proteína de los MTR en una cohorte de pacientes de cáncer de mama independiente por medio de inmunohistoquímica (IHC). Se cribaron anticuerpos para los 6 MTR y se identificaron cuatro que reconocían específicamente FOXM1, HMGB2, PTTG1 y UHRF1. Se evaluaron micromatrices tisulares (TMA) que representaban 512 tumores de mama invasores para los niveles de proteína de cada uno de estos MTR (figura 3C). Los TMA teñidas se puntuaron manualmente y los resultados se analizaron en relación con la supervivencia sin recaída para los 430 tumores con información sobre los cuatro MTR (figura 3D). Cada MTR se 35 asoció individualmente con un mal pronóstico, y la combinación de los cuatro MTR fue más poderosa para estratificar a los pacientes en relación con la supervivencia, en comparación con los indicadores pronósticos

existentes tales como Ki67 o los criterios de St. Gallen, un índice pronóstico basado en la edad, el estado ganglionar, el tamaño tumoral, el estado de ER/PR y el grado tumoral (Goldhirsch *et al.*, 2001) (figura 3D). Los resultados de este análisis de Kaplan-Meier también se representaron en un formato de matriz cromática para indicar la fuerza de la asociación con la supervivencia sin recaída (figura 3E). Según el conocimiento de los autores de la invención, esta disposición de la matriz cromática no se ha usado previamente para presentar el análisis de la supervivencia a gran escala, y proporciona una manera intuitiva de determinar la mejor combinación de pronóstico en cualquier conjunto de datos en particular.

Para refinar aún más el método de predicción de la invención reivindicada y complementar el enfoque adoptado por el solicitante, se tuvieron en cuenta las otras vías cruciales, además del control de la proliferación, implicadas en la progresión del cáncer de mama. Se cribaron genes adicionales del análisis no sesgado de cuatro conjuntos de datos de cáncer de mama independientes (descritos anteriormente y en la tabla 2), que se correlacionan fuertemente con la supervivencia y representan otros aspectos de la progresión tumoral como distintos de la proliferación, tal como la migración/invasión, la apoptosis y las vías de señalización hormonal (tabla 3). Cuando se combinan con los MTR de proliferación, estos genes añaden otra capa de información e incrementan aún más el poder predictivo de la combinación de genes. Estos genes forman la base del panel de vías de OncoMasTR que, cuando se combina con los genes centrales de OncoMasTR, mejora aún más el poder pronóstico del método.

Tabla 3: Resumen de los paneles de genes central y de vías de OncoMasTR

Panel central de OncoMasTR	Panel de vías de OncoMasTR			
	Migración/ Invasión	Apoptosis	Señalización hormonal/de factores de crecimiento	Otra función
UHRF1 FOXM1 MYBL2 PTTG1 E2F1 HMGB2	EPHA4 HOXB13 CLDN4 SERPINE1	BIRC5 BCL2 TXNRD1 NDRG1	CAMLG PRLR ADM PRRG2	IGBP1 FUT8 CALU ADCY1
	Señalización de estrógenos	Inflamación/respuesta a herida	Angiogénesis	Vías metabólicas
	GATA3 PDZK1	TPST1 SERPINA3	FLT3	SLC2A3 LRP2

Se puede inferir una interrupción de las vías de senescencia celular usando una combinación de MTR y niveles de p16^{INK4A} y es un fuerte predictor de malos resultados en el cáncer de mama

Luego, el solicitante deseaba examinar si los niveles de p16^{INK4A}, un indicador indirecto potencial para eludir el punto de control de la senescencia celular en el cáncer, podrían contribuir al poder pronóstico de los MTR. En primer lugar, para confirmar que los niveles de ARNm de *CDKN2A* regulados incorrectamente se correlacionaban con la perturbación genética del puesto de control de la senescencia celular, se analizó el conjunto de datos del cáncer de mama del Atlas del Genoma del Cáncer (TCGA) (Cancer Genoma Atlas, 2012), y se descubrió que unos niveles altos de ARNm de *CDKN2A* se correlacionaban con delección de *RBI*, como se informó previamente en otros estudios (Hara *et al.*, 1996; Kotake *et al.*, 2007; Li *et al.*, 1994; Tam *et al.*, 1994), mientras que la delección de *CDKN2A* se correlacionaba con niveles de ARNm disminuidos (figura 4A). Sorprendentemente, se descubrió que unos niveles moderados de ARNm de *INK4A* se correlacionan con una mejor supervivencia sin recaída en 457 pacientes con cáncer de mama sin afectación ganglionar, mientras que niveles muy bajos o muy altos se correlacionan con una supervivencia sin recaída más corta (Loi *et al.*, 2007; Miller *et al.*, 2005; van de Vijver *et al.*, 2002) (figura 4B). Luego, el solicitante realizó IHC para la proteína p16^{INK4A} en los mismos TMA de cáncer de mama usados previamente (figura 4C). Esto confirmó que los niveles de proteína p16^{INK4A} muy altos o muy bajos también se correlacionaban tanto con una supervivencia sin recaída como con una supervivencia específica del cáncer de mama más cortas, mientras que los niveles moderados se correlacionaban con una supervivencia prolongada (figura 4D-E).

En base a estas observaciones, el solicitante razonó que los cánceres de mama con niveles de proteína p16^{INK4A} muy altos o muy bajos habían eludido el punto de control de la senescencia celular, y esto podría explicar potencialmente su mal pronóstico. Los cánceres de mama con niveles bajos de proteína p16^{INK4A} tenían más probabilidades de tener una delección en el locus del gen *INK4A*, mientras que aquellos con niveles aberrantemente altos probablemente tenían mutaciones en el gen *INK4A* o regulación incorrecta de los miembros de la vía E2F-pRB en dirección 3' tal como ciclina D1 o pRB. Por el contrario, los tumores con expresión moderada de *INK4A* estaban más probablemente enriquecidos en células que no habían eludido el punto de control de la senescencia celular y, por lo tanto, tenían un pronóstico más favorable.

Estudios previos de la expresión de p16^{INK4A} en relación con el pronóstico del cáncer de mama han informado de resultados contradictorios; mientras que se descubrió que p16^{INK4A} está asociado con un mal pronóstico en algunas cohortes (Hui *et al.*, 2000; Milde-Langosch *et al.*, 2001), otros estudios mostraron una asociación con resultados mejorados (Peurala *et al.*, 2013). Estos estudios han dividido, en general, los valores de expresión en dos grupos, bajo/negativo y alto, para el análisis. Sin embargo, en base a lo que se conoce de la biología de p16^{INK4A} y la vía p16-Rb en el cáncer, el solicitante propone que el mejor enfoque puede ser examinar la expresión de p16^{INK4A} en tres grupos, expresión baja/negativa, moderada y alta. Esto puede separar los tumores que probablemente hayan eliminado o inactivado p16^{INK4A} (expresadores bajos) y aquellos que tienen niveles aberrantemente altos de p16^{INK4A} y es probable que tengan una vía p16-Rb regulada incorrectamente (expresadores altos) de los tumores con una respuesta de senescencia funcional (expresadores moderados).

Una combinación de medición de MTR proliferativos y niveles de p16^{INK4A} (puntuación de OncoMasTR) supera los enfoques usados actualmente para predecir el pronóstico del cáncer de mama

La capacidad de pronóstico de una combinación de p16^{INK4A} y los MTR proliferativos se evaluaron luego. Para hacer esto, se desarrolló una puntuación que englobaba tanto los MTR proliferativos como la expresión de p16^{INK4A}, denominada 'puntuación de ARN de OncoMasTR', y se comparó con las estimaciones de otros ensayos pronósticos multigénicos principales (figura 5A). Esto reveló que la puntuación de ARN de OncoMasTR se comparaba favorablemente con las estimaciones sustitutivas de las firmas de MammaPrint™ y OncotypeDx®, usando categorías de bajo/alto para la comparación con MammaPrint™, y categorías de bajo/moderado/alto para la comparación con Oncotype Dx®. Para demostrar además la capacidad de pronóstico de la puntuación de ARN de OncoMasTR, el solicitante analizó cada conjunto de datos individual y el conjunto de datos combinado, y representó los resultados en un formato de matriz cromática (figura 5B). Este análisis extendido reveló que, mientras que la firma de 70 genes de MammaPrint™ funcionó mejor en el conjunto de datos que incluía muestras usadas en su derivación (van 't Veer *et al.*, 2002; van de Vijver *et al.*, 2002), la puntuación de ARN de OncoMasTR superó las estimaciones de los ensayos MammaPrint™ y Oncotype Dx® en general cuando se combinaron los tres conjuntos de datos.

Luego, para validar estas observaciones al nivel de proteína, el solicitante combinó la proteína p16^{INK4A} y el panel de 4-MTR basado en IHC, llamado 'puntuación por IHC de OncoMasTR', y sometió a prueba esta combinación en todos los pacientes y en pacientes sin afectación ganglionar, en relación con la supervivencia sin recaída (figura 5C) y la supervivencia específica del cáncer de mama (figura 5D). Esto reveló que, cuando se añade p16^{INK4A} al panel de MTR basado en IHC, la combinación de niveles altos de proteínas de MTR proliferativos y proteína p16^{INK4A} baja o aberrantemente alta se asoció fuertemente con un mal pronóstico, y hubo una mejora notable en la capacidad para predecir la supervivencia del paciente en comparación con los cuatro MTR sin p16^{INK4A}, ya sea en todos los pacientes (figura 5C) o en una subcohorta sin afectación ganglionar (figura 5D).

La puntuación de ARN de OncoMasTR supera las estimaciones sustitutivas de MammaPrint™ y Oncotype Dx® en pacientes positivos para ER

Para evaluar además la utilidad clínica potencial de la puntuación de ARN de OncoMasTR, se examinó su poder pronóstico en 366 pacientes sin afectación ganglionar y positivos para ER, lo que refleja los criterios de inclusión para el ensayo Oncotype Dx®. La puntuación de ARN de OncoMasTR superó las estimaciones sustitutivas de los ensayos MammaPrint™ (grupos de bajo/alto) y Oncotype Dx® (grupos de bajo/mod/alto) tanto en la cohorte completa (figura 6A) como en la cohorte de pacientes con sin afectación ganglionar (figura 6B). La puntuación de ARN de OncoMasTR también se evaluó usando un enfoque de qRT-PCR con Taqman® en 151 pacientes sin afectación ganglionar y positivos para ER, usando la SSMD como criterio de valoración, emparejado con la cohorte usada para la validación por IHC (figura 7). Esto demostró que la puntuación de ARN de OncoMasTR, cuando se midió mediante el análisis por qRT-PCR con Taqman®, mostró un rendimiento análogo al análisis basado en micromatrices. Además, la puntuación por IHC de OncoMasTR también demostró utilidad en este grupo de pacientes, usando bien la supervivencia sin recaída o la supervivencia sin metástasis a distancia (figura 8) como criterio de valoración.

La puntuación de OncoMasTR tiene un valor pronóstico independiente en todos los pacientes y en pacientes sin afectación ganglionar

Luego, para determinar si la combinación de MTR e *INK4A*/p16^{INK4A} puede proporcionar información de pronóstico adicional independiente de las variables clínico-patológicas estándar, el solicitante realizó un análisis multivariado usando modelos de riesgos proporcionales de Cox. Se descubrió que la puntuación de OncoMasTR contribuía con información pronóstica añadida a un modelo de variable clínico-patológica estándar, en términos de supervivencia sin recaída, tanto a nivel de ARNm (tabla 4) como de proteína (tabla 5). Esto también se observó en la cohorte de pacientes sin afectación ganglionar. El valor pronóstico añadido de la puntuación de OncoMasTR en la parte superior del modelo clínico estándar es superior a todos los demás indicadores pronósticos, incluyendo Ki67, la firma de 70 genes (MammaPrint™) y la firma de 21 genes (Oncotype Dx®). Además, se descubrió que la puntuación de ARN de OncoMasTR proporciona información de pronóstico adicional significativa a un modelo que comprende las

variables clínicas estándar conjuntamente con la estimación sustitutiva de Oncotype Dx®.

Tabla 4: Análisis multivariado de regresión de Cox usando un modelo de variable clínica estándar* en los conjuntos de datos de micromatrices combinadas

Variable	Todos los pacientes (n = 567)		Pacientes sin afectación ganglionar (n=410)		
	Chi2**	Valor de p	Chi2	Valor de p	
Lo/Med/Hi	FOXM1	24,14	<0,001	26,59	<0,001
	E2F1	25,28	<0,001	15,56	<0,001
	HMGB2	10,89	<0,001	7,47	0,006
	MYBL2	25,43	<0,001	15,91	<0,001
	PTTG1	12,37	<0,001	10,16	0,001
	UHRF1	22,71	<0,001	17,61	<0,001
	CDKN2A	2,23	0,135	13,82	<0,001
	6MTR	33,80	<0,001	20,27	<0,001
	Puntuación ARN OncoMasTR	43,87	<0,001	44,04	<0,001
	21 genes	29,02	<0,001	38,03	<0,001
	Ki67	8,30	0,004	7,45	0,006
Lo/Hi	6MTR	23,82	<0,001	29,32	<0,001
	Puntuación ARN OncoMasTR	29,62	<0,001	32,20	<0,001
	70 genes	30,20	<0,001	28,88	<0,001
	Ki67	5,52	0,018	8,99	0,003

*Variables clínicas usadas: Edad (≥ 50 años), estado ganglionar, tamaño del tumor (≥ 2 cm), grado del tumor (>1), tratamiento (tratamiento endocrino) y estado de ER.

5 **Valor pronóstico añadido de cada variable, representado por el cambio en el valor de Chi2 del modelo de solo variables clínicas al modelo de variable clínica + marcador en los tres conjuntos de datos de micromatrices combinadas. Se usó la supervivencia sin recaída como criterio de valoración para este análisis.

Tabla 5: Análisis multivariado de regresión de Cox usando un modelo de variable clínica estándar* en micromatrices tisulares

Variable	Todos los pacientes (n = 272)		Pacientes sin afectación ganglionar (n=171)		
	Chi2**	Valor de p	Chi2	Valor de p	
Lo/Hi	FOXM1	1,60	0,207	0,49	0,485
	HMGB2	0,05	0,819	2,53	0,112
	PTTG1	4,03	0,044	0,17	0,677
	UHRF1	4,53	0,033	0,77	0,379
	p16	6,73	0,009	7,23	0,007
	4 MTR	12,24	<0,001	0,28	0,597
	Puntuación IHC OncoMasTR	24,86	<0,001	7,28	0,007
	Ki67	5,23	0,022	3,42	0,064

10 *Variables clínicas usadas: Edad (≥ 50 años), estado ganglionar, tamaño del tumor (≥ 2 cm), grado del tumor (>1), tratamiento (quimioterapia, tratamiento endocrino, radioterapia), estado de ER y HER2. **Valor pronóstico añadido de cada variable, representado por el cambio en el valor de Chi2 del modelo de solo variables clínicas al modelo de variable clínica + marcador en los tres conjuntos de datos de micromatrices tisulares. Se usó la supervivencia sin recaída como criterio de valoración para este análisis.

15 Poder pronóstico en una cohorte de cáncer de próstata

Este proyecto actual describe la validación del panel de OncoMasTR como un pronóstico de cáncer de mama en cohortes independientes; sin embargo, el panel también se puede usar para otros tipos de cáncer tales como los enumerados anteriormente. Por ejemplo, se analizó un conjunto de datos transcriptómicos de cáncer de próstata disponible públicamente (Taylor *et al.*, 2010), que reveló que el panel de OncoMasTR mostró capacidad de pronóstico en términos de supervivencia sin metástasis en este tipo de cáncer (véase la figura 10). Se descubrió que los pacientes con cáncer de próstata con expresión alta del panel de 6 MTR (FOX1, E2F1, MYBL2, UHRF1, PTTG1, HMGB2) tiene un mal resultado en comparación con los pacientes con expresión baja de estos genes.

Un método de predicción basado en la expresión de estos MTR y p16^{INK4A} podrá abordar la necesidad no cubierta de los pacientes con cáncer de mama en fase precoz y proporcionarles las herramientas necesarias para tomar decisiones de tratamiento mejor fundamentadas. La adición de genes de vías adicionales, o novedosos MTR como ATAD2, E2F8, ZNF367 y TCF19, algunos de los cuales han demostrado predecir un mal pronóstico en pacientes con cáncer de mama (figura 9), también puede mejorar aún más la capacidad de pronóstico de este ensayo. Dicha prueba mejorará lo que está disponible actualmente en base al hecho de que cada uno de estos MTR está en dirección 5' de muchos genes implicados en la proliferación del cáncer de mama y, por tanto, midiendo estos MTR, se está midiendo eficazmente el estado de una 'firma de proliferación' mucho más grande. El poder predictivo de este panel de MTR de proliferación se ha aumentado mediante la adición del regulador de senescencia p16^{INK4A}. Combinando estos genes "centrales" con genes seleccionados de 'vías', se pueden diseccionar completamente las complejidades moleculares del cáncer de mama, y determinar con exactitud la probabilidad de recaída.

El potencial pronóstico de estos 10 MTR, en combinación con p16^{INK4A}, se analizó posteriormente individualmente usando BreastMark (Madden, S.F. *et al.* BreastMark: an integrated approach to mining publicly available transcriptomic datasets relating to breast cancer outcome. *Breast Cancer Res* 15, R52, doi:10.1186/bcr3444 (2013)), un enfoque integrado para realizar análisis de supervivencia de conjuntos de datos cruzados en el cáncer de mama (tabla 6). Este algoritmo integra datos de expresión génica y supervivencia de 26 conjuntos de datos en 12 plataformas de micromatrices diferentes correspondientes a aproximadamente 17.000 genes en hasta 4738 muestras. El desglose de la información clínica individual disponible con cada conjunto de datos se describe en detalle en el artículo original, junto con los métodos usados para analizar/normalizar los datos de expresión génica. Se facilita por gen el análisis de supervivencia de conjuntos de datos cruzados en múltiples plataformas de micromatrices dispares, centrando los datos para eliminar la información específica de sonda y dicotomizar las muestras dentro de cada conjunto de datos antes de combinarlas para realizar un análisis de supervivencia global agrupado. En el análisis presentado en el presente documento, se eligió la supervivencia sin enfermedad (SSE) como criterio de valoración de la supervivencia y se usó la mediana de la expresión génica para dicotomizar los datos.

Existen más de 1000 combinaciones de MTR con cuatro o más genes que se pueden elegir de la lista de 10 MTR descritos en el presente documento, cada una de las cuales se puede combinar y evaluar para su potencial de pronóstico. Para identificar la combinación óptima de estos MTR, BreastMark se adaptó de la siguiente manera. Para cada combinación de MTR, se tomaron los conjuntos de datos procesados de BreastMark y, dentro de cada conjunto de datos, los datos de expresión de cada MTR se dividieron en la mediana en dos grupos. Una vez que las muestras se han dicotomizado, los datos de expresión génica ya no se usan, lo que permite comparaciones entre diferentes conjuntos de datos/plataformas. Para generar una puntuación combinada del regulador transcripcional maestro (MTR), los valores de expresión génica para cada una de los MTR en una combinación en particular se dividieron en la mediana, dando una puntuación de 1 o 2 en base al nivel de expresión. Esto da como resultado que cada muestra en un conjunto de datos en particular obtenga una puntuación de MTR basada en la suma de sus puntuaciones de MTR individuales. Por ejemplo, si una combinación de MTR en particular contenía 6 genes, y cada gen en una muestra en particular se expresó a un nivel por debajo de la mediana de expresión de ese gen en ese conjunto de datos, la puntuación de MTR sería 6, la suma de la puntuación de 1 para cada uno de los 6 MTR. Esto da como resultado un intervalo de puntuaciones de MTR entre 6 (todos los MTR se expresan de manera baja) y 12 (todos los MTR se expresan de manera alta), que a continuación se puede dicotomizar en base a la mediana de la puntuación de MTR para ese conjunto de datos y combinar con la información de SSE para identificar si esta combinación de MTR es pronóstica (un valor de p significativo) y cuál es su pronóstico (la proporción de riesgos).

Las 100 mejores combinaciones de MTR se pueden ver en el diagrama de bosque en la figura 11, y las curvas de Kaplan-Meier individuales para las 24 mejores combinaciones se pueden ver en la figura 12. Las muestras se clasificaron en base al tamaño de la proporción de riesgos una vez que se había establecido la significación (con ajuste para múltiples pruebas usando el método de Benjamini y Hochberg (Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N. y Golani, I. Controlling the false discovery rate in behavior genetics research. *Behavioural brain research* 125, 279-284 (2001)). Cabe señalar que los tamaños de muestra varían dependiendo de la combinación de MTR usados, ya que no todos los MTR están presentes en los 26 conjuntos de datos de BreastMark, por ejemplo, ZNF367 solo está presente en cuatro conjuntos de datos con un total de 295 muestras.

Tabla 6. Análisis individual de supervivencia del cáncer de mama de los diez principales reguladores transcripcionales maestros identificados por ARACNe, usando el algoritmo BreastMark.

Factor de transcripción	ID del gen Entrez	Proporción de riesgos	Valor de p	Numero de muestra
ATAD2	29028	1,378 (1,224 - 1,552)	1,03E-07	2576
E2F1	1869	1,301 (1,15 - 1,472)	2,92E-05	2357
E2F8	79733	1,375 (1,214 - 1,558)	4,74E-07	2281
FOXM1	2305	1,578 (1,392 - 1,788)	5,45E-13	2357
HMGB2	3148	1,271 (1,122 - 1,439)	0,0001493	2357
MYBL2	4605	1,506 (1,339 - 1,694)	7,08E-12	2652
PTTG1	9232	1,586 (1,402 - 1,794)	1,25E-13	2437
TCF19	6941	1,27 (1,097 - 1,471)	0,00136	1378
UHRF1	29128	1,318 (1,144 - 1,52)	0,0001328	1533
ZNF367	195828	1,08 (0,8274 - 1,41)	0,571	295

Basándose en los datos mecanicistas que sustentan el panel de OncoMasTR, los solicitantes también creen que el poder predictivo del panel tendrá la capacidad de predecir la respuesta a los inhibidores de CDK4/6 como pablociclib. Pablociclib es un inhibidor altamente selectivo y activo por vía oral de las cinasas dependientes de ciclina CDK4/6, que se evaluó inicialmente como una politerapia con letrozol en el cáncer de mama avanzado ER+ Her2+, en el ensayo PALOMA-1 (Richard S. Finn, 2014) . Los resultados de este ensayo han demostrado que la adición de pablociclib a una pauta estándar prolonga la supervivencia en 10 meses, lo que es un resultado muy prometedor en estos pacientes en fase tardía. Basándose en los datos mecanicistas que sustentan OncoMasTR, el solicitante cree que es probable que tenga una utilidad predictiva en términos de respuesta a este novedoso tratamiento.

Cálculo de la puntuación de las firmas de MTR10 + CDKN2A en líneas celulares tratadas con pablociclib

Pablociclib es un inhibidor de las cinasas dependientes de ciclinas D y sus efectos sobre las líneas celulares de cáncer de mama humano se examinaron previamente por Finn *et al.* En resumen, 47 líneas celulares humanas, que representan los subtipos moleculares del cáncer de mama, se trataron con pablociclib y se calcularon sus perfiles de expresión génica, junto con sus valores de CI50. Los datos de expresión génica se descargaron del Gene Expression Omnibus para las 47 líneas celulares, junto con los datos IC50 que los acompañaban (número de acceso GSE18496). Los datos de expresión génica para los 10 MTR descritos aquí se dividieron gen por gen usando la mediana de la expresión en todas las líneas celulares como punto de corte. A esas líneas celulares con una expresión mayor o más baja que la mediana de un gen se les dio un valor de 2 o 1 para ese gen, respectivamente. Esto se repitió para cada uno de los diez genes. La expresión de CDKN2A en las líneas celulares se dividió por igual en tres, a aquellas líneas celulares con expresión alta o baja se les dio un valor de 2 y a aquellas con un nivel de expresión intermedio se les dio un valor de 1. A continuación, se calculó una puntuación para cada línea celular sumando las puntuaciones de genes individuales. La figura 13 muestra una curva de los valores de CI50 frente a la puntuación de la firma (coeficiente de correlación=0,319, valor de p=0,03). El valor de p significativo de los datos *in vitro* sugiere que los MTR pueden proporcionar un valor predictivo con respecto a los pacientes que reciben inhibidores de CDK4/6 para tratar el cáncer.

En la memoria descriptiva, los términos "comprenden, comprende, comprendido y que comprenden" o cualquier variación de los mismos y los términos "incluyen, incluye, incluido y que incluyen" o cualquier variación de los mismos se consideran totalmente intercambiables y se debe dar a todos la interpretación más amplia posible y viceversa.

Referencias

Boffetta, P. y Kaldor, JM (1994). Secondary malignancies following cancer chemotherapy. *Acta Oncol* 33, 591-598.

Boonyaratanakornkit, V., Melvin, V., Prendergast, P., Altmann, M., Ronfani, L., Bianchi, M. E., Taraseviciene, L., Nordeen, S. K., Allegretto, E. A., y Edwards, D. P. (1998). High-mobility group chromatin proteins 1 and 2 functionally interact with steroid hormone receptors to enhance their DNA binding in vitro and transcriptional activity in mammalian cells. *Mol Cell Biol* 18, 4471-4487.

Bostick, M., Kim, J. K., Esteve, P. O., Clark, A., Pradhan, S., y Jacobsen, S. E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* 317, 1760-1764.

Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H., y Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev* 20, 1123-1136.

Bracken, A. P., Pasini, D., Capra, M., Prosperini, E., Colli, E., y Helin, K. (2003). EZH2 is downstream of the pRB-

- E2F pathway, essential for proliferation and amplified in cancer. *The EMBO journal* 22, 5323-5335.
- 5 Buffa, F. M., Camps, C., Winchester, L., Snell, C. E., Gee, H. E., Sheldon, H., Taylor, M., Harris, A. L., y Ragoussis, J. (2011). microRNA-associated progression pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression profiling in breast cancer. *Cancer Res* 71, 5635-5645.
- Cancer Genome Atlas, N. (2012). Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61-70.
- 10 Cardoso, F., Van't Veer, L., Rutgers, E., Loi, S., Mook, S., y Piccart-Gebhart, M. J. (2008). Clinical application of the 70-gene profile: the MINDACT trial. *J Clin Oncol* 26, 729-735.
- Carro, M. S., Lim, W. K., Alvarez, M. J., Bollo, R. J., Zhao, X., Snyder, E. Y., Sulman, E. P., Anne, S. L., Doetsch, F., Colman, H., *et al.* (2010). The transcriptional network for mesenchymal transformation of brain tumours. *Nature* 463, 318-325.
- 15 Curtis, C., Shah, S. P., Chin, S. F., Turashvili, G., Rueda, O. M., Dunning, M. J., Speed, D., Lynch, A. G., Samarajiwa, S., Yuan, Y., *et al.* (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. ExPO <http://www.intgen.org/expo/>.
- 20 Fan, C., Oh, D. S., Wessels, L., Weigelt, B., Nuyten, D. S., Nobel, A. B., van't Veer, L. J., y Perou, C. M. (2006). Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 355, 560-569.
- Finn *et al*, *Breast Cancer Res.* 2009;11(5):R77. doi: 10.1186/bcr2419
- 25 Fisher, B., Jeong, J. H., Bryant, J., Anderson, S., Dignam, J., Fisher, E. R., y Wolmark, N. (2004). Treatment of lymph-node-negative, oestrogen-receptor-positive breast cancer: long-term findings from National Surgical Adjuvant Breast and Bowel Project randomised clinical trials. *Lancet* 364, 858-868.
- 30 Garbe, J. C., Bhattacharya, S., Merchant, B., Bassett, E., Swisshelm, K., Feiler, H. S., Wyrobek, A. J., y Stampfer, M. R. (2009). Molecular distinctions between stasis and telomere attrition senescence barriers shown by long-term culture of normal human mammary epithelial cells. *Cancer Res* 69, 7557-7568.
- 35 Goldhirsch, A., Glick, J. H., Gelber, R. D., Coates, A. S., y Senn, H. J. (2001). Meeting highlights: International Consensus Panel on the Treatment of Primary Breast Cancer. Seventh International Conference on Adjuvant Therapy of Primary Breast Cancer. *J Clin Oncol* 19, 3817-3827.
- Haibe-Kains, B., Desmedt, C., Piette, F., Buyse, M., Cardoso, F., Van't Veer, L., Piccart, M., Bontempi, G., y Sotiriou, C. (2008). Comparison of prognostic gene expression signatures for breast cancer. *BMC Genomics* 9, 394.
- 40 Hara, E., Smith, R., Parry, D., Tahara, H., Stone, S., y Peters, G. (1996). Regulation of p16CDKN2 expression and its implications for cell immortalization and senescence. *Mol Cell Biol* 16, 859-867.
- 45 Hokamp, K., Roche, F. M., Acab, M., Rousseau, M. E., Kuo, B., Goode, D., Aeschliman, D., Bryan, J., Babiuk, L. A., Hancock, R. E., y Brinkman, F. S. (2004). ArrayPipe: a flexible processing pipeline for microarray data. *Nucleic Acids Res* 32, W457-459.
- Hui, R., Macmillan, R. D., Kenny, F. S., Musgrove, E. A., Blarney, R. W., Nicholson, R. I., Robertson, J. F., y Sutherland, R. L. (2000). INK4a gene expression and methylation in primary breast cancer: overexpression of p16INK4a messenger RNA is a marker of poor prognosis. *Clin Cancer Res* 6, 2777-2787.
- 50 Ivshina, A. V., George, J., Senko, O., Mow, B., Putti, T. C., Smeds, J., Lindahl, T., Pawitan, Y., Hall, P., Nordgren, H., *et al.* (2006). Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res* 66, 10292-10301.
- 55 Kotake, Y., Cao, R., Viatour, P., Sage, J., Zhang, Y., y Xiong, Y. (2007). pRB family proteins are required for H3K27 trimethylation and Polycomb repression complexes binding to and silencing p16INK4alpha tumor suppressor gene. *Genes Dev* 21, 49-54.
- 60 Laoukili, J., Kooistra, M. R., Bras, A., Kauw, J., Kerkhoven, R. M., Morrison, A., Clevers, H., y Medema, R. H. (2005). FoxM1 is required for execution of the mitotic programme and chromosome stability. *Nat Cell Biol* 7, 126-136.
- Li, Y., Nichols, M. A., Shay, J. W., y Xiong, Y. (1994). Transcriptional repression of the D-type cyclin-dependent kinase inhibitor p16 by the retinoblastoma susceptibility gene product pRb. *Cancer Res* 54, 6078-6082.
- 65 Liu, R., Wang, X., Chen, G. Y., Dalerba, P., Gurney, A., Hoey, T., Sherlock, G., Lewicki, J., Shedden, K., y Clarke, M. F. (2007). The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med* 356, 217-226.

- Loi, S., Haibe-Kains, B., Desmedt, C., Lallemand, F., Tutt, A. M., Gillet, C., Ellis, P., Harris, A., Bergh, J., Foekens, J. A., *et al.* (2007). Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol* 25, 1239-1246.
- 5 Margolin, A. A., Wang, K., Lim, W. K., Kustagi, M., Nemenman, I., y Califano, A. (2006). Reverse engineering cellular networks. *Nat Protoc* 1, 662-671.
- Milde-Langosch, K., Bamberger, A. M., Rieck, G., Kelp, B., y Loning, T. (2001). Overexpression of the p16 cell cycle inhibitor in breast cancer is associated with a more malignant phenotype. *Breast Cancer Res Treat* 67, 61-70.
- 10 Mosley, J. D., y Keri, R. A. (2008). Cell cycle correlated genes dictate the prognostic power of breast cancer gene lists. *BMC Med Genomics* 1, 11.
- 15 Oh, I. H., y Reddy, E. P. (1998). The C-terminal domain of B-Myb acts as a positive regulator of transcription and modulates its biological functions. *Mol Cell Biol* 18, 499-511.
- Peurala, E., Koivunen, P., Haapasaaari, K. M., Bloigu, R., y Jukkola-Vuorinen, A. (2013). The prognostic significance and value of cyclin D1, CDK4 and p16 in human breast cancer. *Breast Cancer Res* 15, R5.
- 20 Richard S. Finn, J. P. C., Istvan Lang, Katalin Boer, Igor M. Bondarenko, Sergey O. Kulyk, Johannes Ettl, Ravindranath Patel, Tamas Pinter, Marcus Schmidt, Yaroslav V. Shparyk, Anu R. Thummala, Nataliya L. Voytko, Xin Huang, Sindy T. Kim, Sophia S. Randolph, Dennis J. Slamon (2014). Final results of a randomized Phase II study of PD 0332991, a cyclin-dependent kinase (CDK)-4/6 inhibitor, in combination with letrozole vs letrozole alone for first-line treatment of ER+/HER2- advanced breast cancer (PALOMA-1; TRIO-18). Proceedings of the 105th Annual Meeting of the American Association for Cancer Research.
- 25 Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., Nordgren, H., Farmer, P., Praz, V., Haibe-Kains, B., *et al.* (2006). Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 98, 262-272.
- 30 Sparano, J. A. (2006). TAILORx: trial assigning individualized options for treatment (Rx). *Clin Breast Cancer* 7, 347-350.
- 35 Svensson, S., Jirstrom, K., Ryden, L., Roos, G., Emdin, S., Ostrowski, M. C., y Landberg, G. (2005). ERK phosphorylation is linked to VEGFR2 expression and Ets-2 phosphorylation in breast cancer and is associated with tamoxifen treatment resistance and small tumours with good prognosis. *Oncogene* 24, 4370-4379.
- Tam, S. W., Shay, J. W., y Pagano, M. (1994). Differential expression and cell cycle regulation of the cyclin-dependent kinase 4 inhibitor p16Ink4. *Cancer Res* 54, 5816-5820.
- 40 Tong, Y., y Eigler, T. (2009). Transcriptional targets for pituitary tumor-transforming gene-1. *J Mol Endocrinol* 43, 179-185.
- 45 Taylor, B. S., Schultz, N., Hieronymus, H., Gopalan, A., Xiao, Y., Carver, B. S., Arora, V. K., Kaushik, P., Cerami, E., Reva, B., *et al.* (2010). Integrative genomic profiling of human prostate cancer. *Cancer Cell* 18, 11-22.
- Tong, Y., Tan, Y., Zhou, C., y Melmed, S. (2007). Pituitary tumor transforming gene interacts with Spl to modulate G1/S cell phase transition. *Oncogene* 26, 5596-5605.
- 50 van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., *et al.* (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530-536.
- 55 Wang, Y., Klijn, J. G., Zhang, Y., Sieuwerts, A. M., Look, M. P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M. E., Yu, J., *et al.* (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365, 671-679.
- 60 Wirapati, P., Sotiriou, C., Kunkel, S., Farmer, P., Pradervand, S., Haibe-Kains, B., Desmedt, C., Ignatiadis, M., Sengstag, T., Schutz, F., *et al.* (2008). Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res* 10, R65.
- Wu, L., Timmers, C., Maiti, B., Saavedra, H. I., Sang, L., Chong, G. T., Nuckolls, F., Giangrande, P., Wright, F. A., Field, S. J., *et al.* (2001). The E2F1-3 transcription factors are essential for cellular proliferation. *Nature* 414, 457-462.
- 65 Zindy, F., Quelle, D. E., Roussel, M. F., y Sherr, C. J. (1997). Expression of the p16INK4a tumor suppressor versus

other INK4 family members during mouse development and aging. Oncogene 15, 203-211.

LISTADO DE SECUENCIAS

- 5 <110> The Provost, Fellows, Foundation Scholars y los otros miembros de la Junta, del Colegio de la Santísima e Indivisa Trinidad de la Reina Isabel, cerca del Dublin University College de Dublín, la Universidad Nacional de Irlanda, Dublín
- <120> Un método para predecir el riesgo de recaída del cáncer
- 10 <130> P11144PC00
- <150> EP 14185673.2
- <151> 2014-09-19
- 15 <160> 40
- <170> PatentIn versión 3.5
- 20 <210> 1
- <211> 23
- <212> ADN
- <213> Secuencia artificial
- 25 <220>
- <223> Cebador directo para UHRF1
- <400> 1
- 30 agaccgtcct caaccagctc ttc 23
- <210> 2
- <211> 21
- <212> ADN
- <213> Secuencia artificial
- 35 <220>
- <223> Cebador inverso para UHRF1
- <400> 2
- 40 gaagtgcctg gagatcaccg g 21
- <210> 3
- <211> 25
- <212> ADN
- 45 <213> Secuencia artificial
- <220>
- <223> Cebador directo para FOXM1
- 50 <400> 3
- caacaatagc ctatccaaca tccag 25
- <210> 4
- <211> 22
- 55 <212> ADN
- <213> Secuencia artificial
- <220>
- <223> Cebador inverso para FOXM1
- 60 <400> 4
- ggagcccagt ccatcagaac tc 22
- 65 <210> 5
- <211> 22

<212> ADN
 <213> Secuencia artificial

 <220>
 5 <223> Cebador directo para PTTG1

 <400> 5
 ctgcctgaag agcaccagat tg 22

 10 <210> 6
 <211> 22
 <212> ADN
 <213> Secuencia artificial

 15 <220>
 <223> Cebador inverso para PTTG1

 <400> 6
 caaggatcat gagaggcact cc 22
 20
 <210> 7
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 25
 <220>
 <223> Cebador directo para MYBL2

 <400> 7
 30 cactgaccag caatgccagt ac 22

 <210> 8
 <211> 22
 <212> ADN
 35 <213> Secuencia artificial

 <220>
 <223> Cebador inverso para MYBL2

 <400> 8
 40 ccccttgaca aggtctggat tc 22

 <210> 9
 <211> 22
 45 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Cebador directo para HMGB2
 50
 <400> 9
 gctcctaaaa ggccaccatc tg 22

 <210> 10
 55 <211> 21
 <212> ADN
 <213> Secuencia artificial

 <220>
 60 <223> Cebador inverso para HMGB2

 <400> 10
 tgatcttgg gcgatgtca g 21

 65 <210> 11
 <211> 22

<212> ADN
 <213> Secuencia artificial

 <220>
 5 <223> Cebador directo para E2F1

 <400> 11
 tgtcaggacc ttcgtagcat tg 22

 10 <210> 12
 <211> 23
 <212> ADN
 <213> Secuencia artificial

 15 <220>
 <223> Cebador inverso para E2F1

 <400> 12
 20 gggcttgat caccataacc atc 23

 <210> 13
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 25
 <220>

 <223> Cebador directo para E2F8
 <400> 13
 30 caatctcaac aaaacccttg gc 22

 <210> 14
 <211> 22
 <212> ADN
 35 <213> Secuencia artificial

 <220>
 <223> Cebador inverso para E2F8

 <400> 14
 40 ctcgcggtac ttattctct cc 22

 <210> 15

 <211> 22
 <212> ADN
 <213> Secuencia artificial

 <220>
 50 <223> Cebador directo para INK4A

 <400> 15
 agaggatttg agggacaggg tc 22

 <210> 16
 <211> 20
 <212> ADN
 <213> Secuencia artificial

 60 <220>
 <223> Cebador inverso para INK4A

 <400> 16
 65 cctcttctt cctccggtgc 20

 <210> 17

<211> 21
 <212> ADN
 <213> Secuencia artificial
 5 <220>
 <223> Cebador directo para PITX
 <400> 17
 atggagctgg gtgctgagaa c 21
 10 <210> 18
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 15 <220>
 <223> Cebador inverso para PITX
 <400> 18
 20 ccttctcaa ctccatgagc cc 22
 <210> 19
 <211> 24
 <212> ADN
 25 <213> Secuencia artificial
 <220>
 <223> Cebador directo para HSF4
 30 <400> 19
 acaaagaagg aaatagaggg accg 24
 <210> 20
 <211> 22
 35 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador inverso para HSF4
 40 <400> 20
 gatgagtggg agacttgggt tc 22
 <210> 21
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 45 <220>
 <223> Cebador directo para KLF6
 <400> 21
 cagcccgagc tttgttaca ac 22
 55 <210> 22
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 60 <220>
 <223> Cebador inverso para KLF6
 <400> 22
 65 ttcgctgctg acatctgagt tc 22
 <210> 23

<211> 22
 <212> ADN
 <213> Secuencia artificial
 5 <220>
 <223> Cebador directo para ID2
 <400> 23
 aaggtagca agatggaaat cc 22
 10 <210> 24
 <211> 23
 <212> ADN
 <213> Secuencia artificial
 15 <220>
 <223> Cebador inverso para ID2
 <400> 24
 20 cgatctgcag gtccaagatg tag 23
 <210> 25
 <211> 22
 <212> ADN
 25 <213> Secuencia artificial
 <220>
 <223> Cebador directo para SNAI1
 30 <400> 25
 ctctctgagg ccaagatct cc 22
 <210> 26
 <211> 23
 35 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador inverso para SNAI1
 40 <400> 26
 cctgttgca gtattgcag ttg 23
 <210> 27
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 45 <220>
 <223> Cebador directo para CDH2
 <400> 27
 tgagcctgca gatttaagg tg 22
 55 <210> 28
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 60 <220>
 <223> Cebador inverso para CDH2
 <400> 28
 65 tggaaagctt ctcacggcat ac 22
 <210> 29

<211> 23
 <212> ADN
 <213> Secuencia artificial

5 <220>
 <223> Cebador directo para CCNA2

<400> 29
 agctggcctg aatcattaat acg 23

10 <210> 30
 <211> 22
 <212> ADN
 <213> Secuencia artificial

15 <220>
 <223> Cebador inverso para CCNA2

<400> 30
 ggtgaaggtc catgagacaa gg 22

<210> 31
 <211> 23
 <212> ADN
 25 <213> Secuencia artificial

<220>
 <223> Cebador directo para EZH2

30 <400> 31
 gggacagtaa aaatgtgtcc tgc 23

<210> 32
 <211> 22
 35 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Cebador inverso para EZH2

40 <400> 32
 tgccagcaat agatgctttt tg 22

<210> 33
 <211> 19
 45 <212> ADN
 <213> Secuencia artificial

<220>
 50 <223> Cebador directo para DEK

<400> 33
 cattcccgt ctcttccc 19

<210> 34
 <211> 18
 55 <212> ADN
 <213> Secuencia artificial

60 <220>
 <223> Cebador inverso para DEK

<400> 34
 gctcggctcc ccagaatc 18

65 <210> 35

<211> 21
 <212> ADN
 <213> Secuencia artificial

5 <220>
 <223> Cebador directo para HELLS

<400> 35
 cctcactgga ggagtgatgc g 21

10 <210> 36
 <211> 22
 <212> ADN
 <213> Secuencia artificial

15 <220>
 <223> Cebador inverso para HELLS

<400> 36
 aagcatccta agccattcca tg 22

<210> 37
 <211> 22
 <212> ADN
 25 <213> Secuencia artificial

<220>
 <223> Cebador directo para E2F7

30 <400> 37
 ccattgaaaa caaggacgat gc 22

<210> 38
 <211> 21
 35 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Cebador inverso para E2F7

40 <400> 38
 ctgtcccaaa caacatcaag c 21

<210> 39
 45 <211> 18
 <212> ADN
 <213> Secuencia artificial

<220>
 50 <223> Cebador directo para RPLPO (Normalizador)

<400> 39
 ttcattgtgg gagcagac 18

55 <210> 40
 <211> 18
 <212> ADN
 <213> Secuencia artificial

60 <220>
 <223> Cebador inverso para RPLPO (Normalizador)

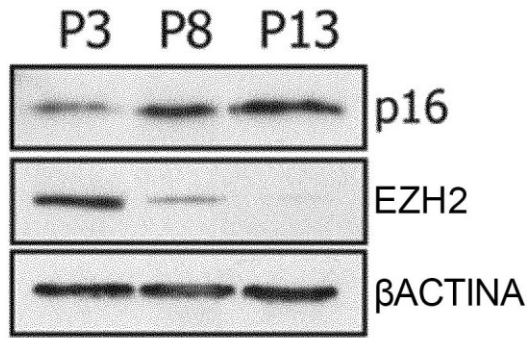
<400> 40
 65 cagcagtttc tccagagc 18

REIVINDICACIONES

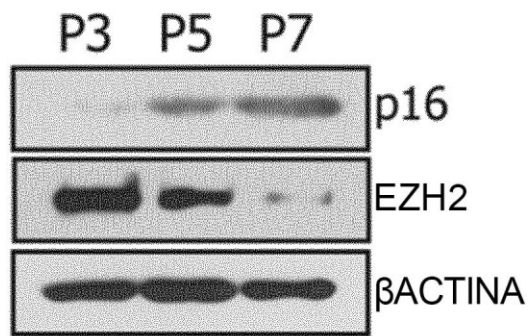
1. Un método para predecir el riesgo de recaída del cáncer en un individuo con cáncer, comprendiendo el método una etapa de someter a ensayo una muestra de cáncer del individuo para determinar la expresión positiva de al menos tres genes, o proteínas codificadas por dichos genes, seleccionados de FOXM1, PTTG1, ZNF367, UHRF1, E2F1, MYBL2, HMGB2, ATAD2, E2F8 y TCF19, en el que la expresión positiva de los al menos tres genes, o proteínas codificadas por dichos genes, se correlaciona con un riesgo incrementado de recaída del cáncer en comparación con un individuo con cáncer que no presenta una expresión positiva de los al menos tres genes o proteínas codificadas por esos genes; en el que los al menos tres genes, o proteínas codificadas por dichos al menos tres genes, son FOXM1, PTTG1 y ZNF367, incluyendo opcionalmente uno o más de los genes, o proteínas codificadas por dichos genes, UHRF1, E2F1, MYBL2, HMGB2, ATAD2, E2F8 y TCF19.
2. Un método para predecir el riesgo de recaída del cáncer en un individuo con cáncer después del tratamiento con inhibidores de CDK4/6, comprendiendo el método una etapa de someter a ensayo una muestra de cáncer del individuo para determinar la expresión positiva de al menos tres genes, o proteínas codificadas por dichos genes, seleccionados de FOXM1, PTTG1, ZNF367, UHRF1, E2F1, MYBL2, HMGB2, ATAD2, E2F8 y TCF19, en el que la expresión positiva de los al menos tres genes, o proteínas codificadas por dicho genes, se correlaciona con un riesgo incrementado de recaída del cáncer en un individuo con cáncer después del tratamiento con inhibidores de CDK4/6 en comparación con un individuo con cáncer que no presenta expresión positiva de los al menos tres genes o proteínas codificadas por esos genes; en el que los al menos tres genes, o proteínas codificadas por dichos al menos tres genes, son FOXM1, PTTG1 y ZNF367, incluyendo opcionalmente uno o más de los genes, o proteínas codificadas por dichos genes, UHRF1, E2F1, MYBL2, HMGB2, ATAD2, E2F8 y TCF19.
3. Un método de acuerdo con una cualquiera de las reivindicaciones 1 a 2, en el que el cáncer se selecciona del grupo que comprende cáncer de mama sin afectación ganglionar y positivo para ER; cáncer de mama con afectación ganglionar y en fase precoz; mieloma múltiple, cáncer de próstata, glioblastoma, linfoma, fibrosarcoma; mixosarcoma; liposarcoma; condrosarcoma; sarcoma osteogénico; cordoma; angiosarcoma; endoteliosarcoma; linfangiosarcoma; linfangioendoteliosarcoma; sinovioma; mesotelioma; tumor de Ewing; liomiosarcoma; rabdmiosarcoma; carcinoma de colon; cáncer de páncreas; cáncer de mama; cáncer de ovarios; carcinoma de células escamosas; carcinoma basocelular; adenocarcinoma; carcinoma de glándulas sudoríparas; carcinoma de glándulas sebáceas; carcinoma papilar; adenocarcinomas papilares; cistoadenocarcinoma; carcinoma medular; carcinoma broncogénico; carcinoma de células renales; hepatoma; carcinoma de vías biliares; coriocarcinoma; seminoma; carcinoma embrionario; tumor de Wilms; cáncer de cuello uterino; cáncer uterino; tumor testicular; carcinoma de pulmón; carcinoma microcítico de pulmón; carcinoma de vejiga; carcinoma epitelial; glioma astrocitoma; meduloblastoma; craneofaringioma; ependimoma; pinealoma; hemangioblastoma; neurinoma del estatoacústico; oligodendroglioma; meningioma; melanoma; retinoblastoma; y leucemias.
4. El método de acuerdo con la reivindicación 3, en el que el cáncer es cáncer de mama.
5. Un método de acuerdo con la reivindicación 4, en el que el cáncer de mama es cáncer de mama sin afectación ganglionar y en fase precoz, y opcionalmente en el que el cáncer de mama es cáncer de mama sin afectación ganglionar y en fase precoz o sin afectación ganglionar, positivo para ER y en fase precoz.
6. Un método para predecir el riesgo de recaída del cáncer en un paciente con cáncer de mama sin afectación ganglionar y en fase precoz, comprendiendo el método una etapa de someter a ensayo una muestra tumoral de cáncer del paciente para determinar la expresión positiva de al menos tres genes, o proteínas codificadas por esos genes, seleccionados de FOXM1, PTTG1, ZNF367, UHRF1, E2F1, MYBL2, HMGB2, ATAD2, E2F8 y TCF19, en el que la expresión positiva de los al menos tres genes, o proteínas codificadas por esos genes, se correlaciona con un riesgo incrementado de recaída del cáncer en comparación con un paciente con cáncer que no presenta una expresión positiva de los al menos tres genes o proteínas codificadas por esos genes; en el que los al menos tres genes, o proteínas codificadas por dichos al menos tres genes, son FOXM1, PTTG1 y ZNF367, incluyendo opcionalmente uno o más de los genes, o proteínas codificadas por dichos genes, UHRF1, E2F1, MYBL2, HMGB2, ATAD2, E2F8 y TCF19.
7. Un método para determinar una tasa de supervivencia a 5 años o una tasa de supervivencia a 10 años de un individuo diagnosticado de cáncer de mama, comprendiendo el método una etapa de someter a ensayo una muestra tumoral de cáncer del individuo para determinar la expresión positiva de al menos tres genes, o proteínas codificadas por esos genes, seleccionados de FOXM1, PTTG1, ZNF367, UHRF1, E2F1, MYBL2, HMGB2, ATAD2, E2F8 y TCF19, en el que la expresión positiva de los al menos tres genes, o proteínas codificadas por esos genes, se correlaciona con una probabilidad disminuida de una tasa de supervivencia a 5 años o una tasa de supervivencia a 10 años en comparación con un individuo con cáncer que no presenta una expresión positiva de los al menos tres genes o proteínas codificadas por esos genes; en el que los al menos tres genes, o proteínas codificadas por dichos al menos tres genes, son FOXM1, PTTG1 y ZNF367, incluyendo opcionalmente uno o más de los genes, o proteínas codificadas por dichos genes, UHRF1, E2F1, MYBL2, HMGB2, ATAD2, E2F8 y TCF19.
8. Un método de acuerdo con la reivindicación 1, la reivindicación 2, la reivindicación 6 o la reivindicación 7,

- comprendiendo además el método la etapa de someter a ensayo la expresión del gen p16^{INK4A} o una proteína codificada por dicho gen, en los que la expresión regulada incorrectamente de p16^{INK4A}, en combinación con la expresión positiva de los al menos tres genes o proteínas codificadas por esos genes, se correlaciona con un riesgo incrementado de recaída del cáncer o una probabilidad disminuida de una tasa de supervivencia a 5 años o una tasa de supervivencia a 10 años en comparación con un individuo con cáncer que no presenta una expresión regulada incorrectamente de p16^{INK4A} y una expresión positiva de los al menos tres genes o proteínas codificadas por esos genes.
- 5
9. Un método para identificar a un paciente con cáncer que es adecuado para el tratamiento con un tratamiento para prevenir la recaída o progresión del cáncer, comprendiendo el método una etapa de someter a ensayo una muestra de cáncer del paciente con cáncer para determinar la expresión positiva de al menos tres genes seleccionados de FOXM1, PTTG1, ZNF367, UHRF1, E2F1, MYBL2, HMGB2, ATAD2, E2F8 y TCF19, en el que la expresión positiva de los al menos tres genes o proteínas codificadas por esos genes en comparación con un individuo con cáncer que no presenta una expresión positiva de los al menos tres genes o proteínas codificadas por esos genes es indicativa de que el paciente con cáncer es adecuado para el tratamiento con un tratamiento para prevenir la recaída o progresión del cáncer; en el que los al menos tres genes, o proteínas codificadas por dichos al menos tres genes, son FOXM1, PTTG1 y ZNF367, incluyendo opcionalmente uno o más de los genes, o proteínas codificadas por dichos genes, UHRF1, E2F1, MYBL2, HMGB2, ATAD2, E2F8 y TCF19.
- 10
10. Un método de acuerdo con la reivindicación 9, comprendiendo además el método la etapa de someter a ensayo la expresión del gen p16^{INK4A} o una proteína codificada por dicho gen, en el que la expresión regulada incorrectamente de p16^{INK4A} en combinación con la expresión positiva de los al menos tres genes o proteínas codificadas por esos genes, en comparación con un individuo con cáncer que no presenta una expresión regulada incorrectamente de p16^{INK4A} y una expresión positiva de los al menos tres genes o proteínas codificadas por esos genes, es indicativa de que el paciente con cáncer es adecuado para el tratamiento con un tratamiento adyuvante para prevenir la recaída o progresión del cáncer.
- 15
11. Un método de acuerdo con la reivindicación 9 o la reivindicación 10, en el que el tratamiento es un tratamiento neoadyuvante o adyuvante, o una combinación de ambos.
- 20
12. Un método de acuerdo con la reivindicación 11, en el que el tratamiento neoadyuvante y el tratamiento adyuvante es un agente seleccionado de trastuzumab, lapatinib, neratinib, afatinib, pertuzumab, inhibidores de CDK4/6, ciclofosfamida, metotrexato, 5-fluorouracilo, gemcitabina, adriamicina (doxorubicina), epirubicina, docetaxel, paclitaxel, capecitabina y tamoxifeno.
- 25
13. Un método para controlar la eficacia del tratamiento del cáncer en un individuo con cáncer, comprendiendo el método una etapa de someter a ensayo una muestra de cáncer del individuo con cáncer para determinar la expresión positiva de al menos tres genes o proteínas codificadas por dichos genes seleccionados de FOXM1, PTTG1, ZNF367, UHRF1, E2F1, MYBL2, HMGB2, ATAD2, E2F8 y TCF19, en el que la expresión más alta de los al menos tres genes se correlaciona con un tratamiento ineficaz y un mal resultado en comparación con un individuo con cáncer que tiene una expresión más baja de los al menos tres genes o proteínas codificadas por esos genes; en el que los al menos tres genes, o proteínas codificadas por dichos al menos tres genes, son FOXM1, PTTG1 y ZNF367, incluyendo opcionalmente uno o más de los genes, o proteínas codificadas por dichos genes, UHRF1, E2F1, MYBL2, HMGB2, ATAD2, E2F8 y TCF19.
- 30
14. El método de acuerdo con la reivindicación 13, que comprende además la etapa de someter a ensayo la muestra de cáncer para determinar la expresión del gen p16^{INK4A} o una proteína codificada por dicho gen en combinación con someter a ensayo los al menos tres genes o proteínas codificadas por dichos genes, con lo que la expresión regulada incorrectamente de p16^{INK4A} se correlaciona con un tratamiento ineficaz y un mal resultado en comparación con un individuo con cáncer que tiene una expresión moderada de p16^{INK4A}.
- 35
15. Un método de acuerdo con las reivindicaciones 13 a 14, en el que el tratamiento es un tratamiento neoadyuvante o un tratamiento adyuvante, o una combinación de ambas, y en el que el tratamiento neoadyuvante y el tratamiento adyuvante es opcionalmente un agente seleccionado de trastuzumab, lapatinib, neratinib, afatinib, pertuzumab, inhibidores de CDK4/6, ciclofosfamida, metotrexato, 5-fluorouracilo, gemcitabina, adriamicina (doxorubicina), epirubicina, docetaxel, paclitaxel, capecitabina y tamoxifeno.
- 40
- 45
- 50
- 55

Figura 1A



HMEC



MEF

Figura 1B

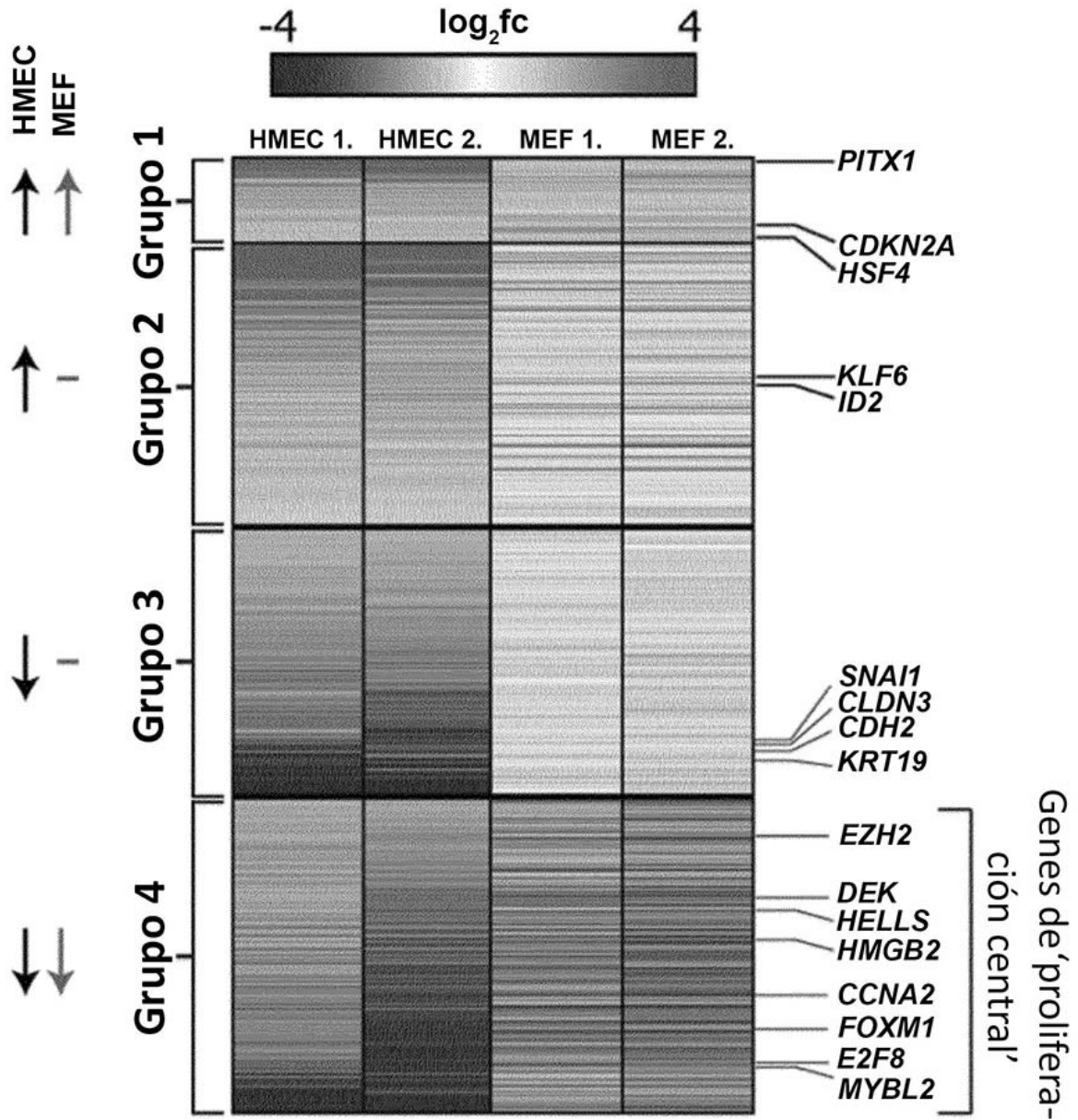


Figura 1C

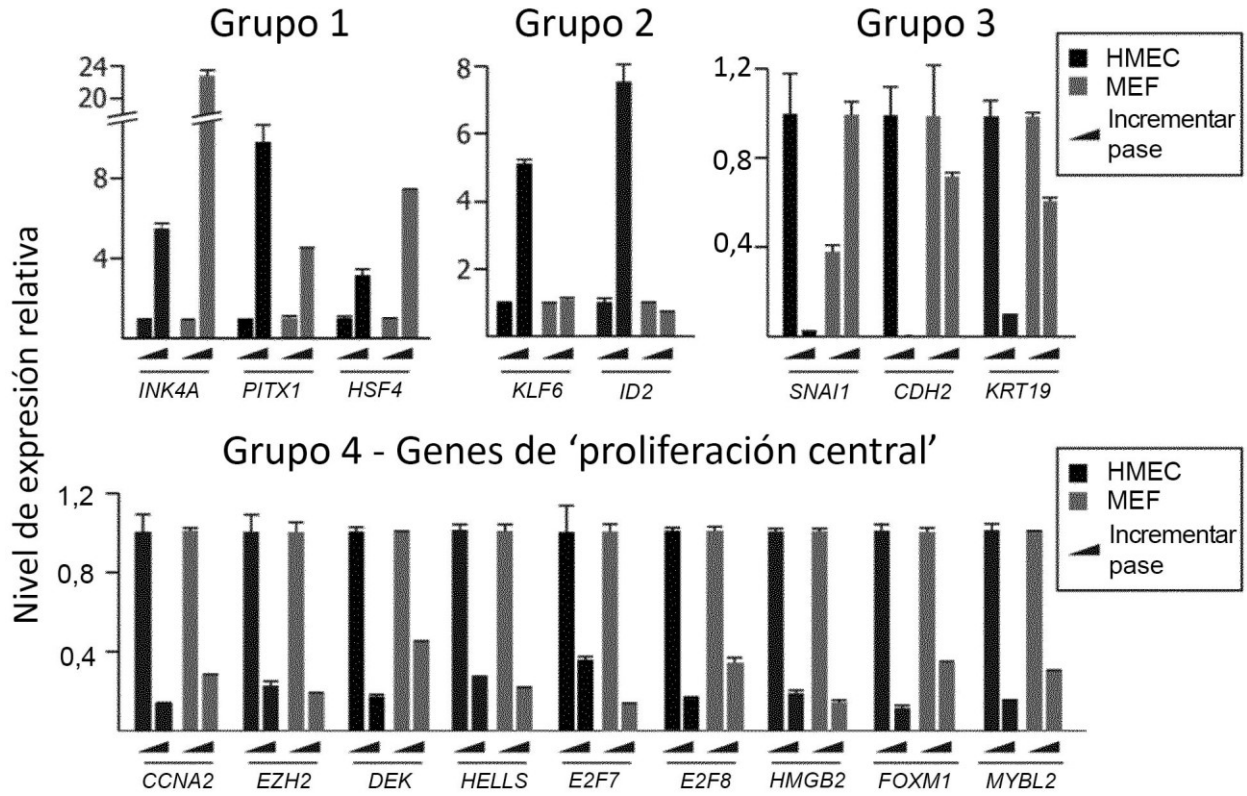


Figura 1D

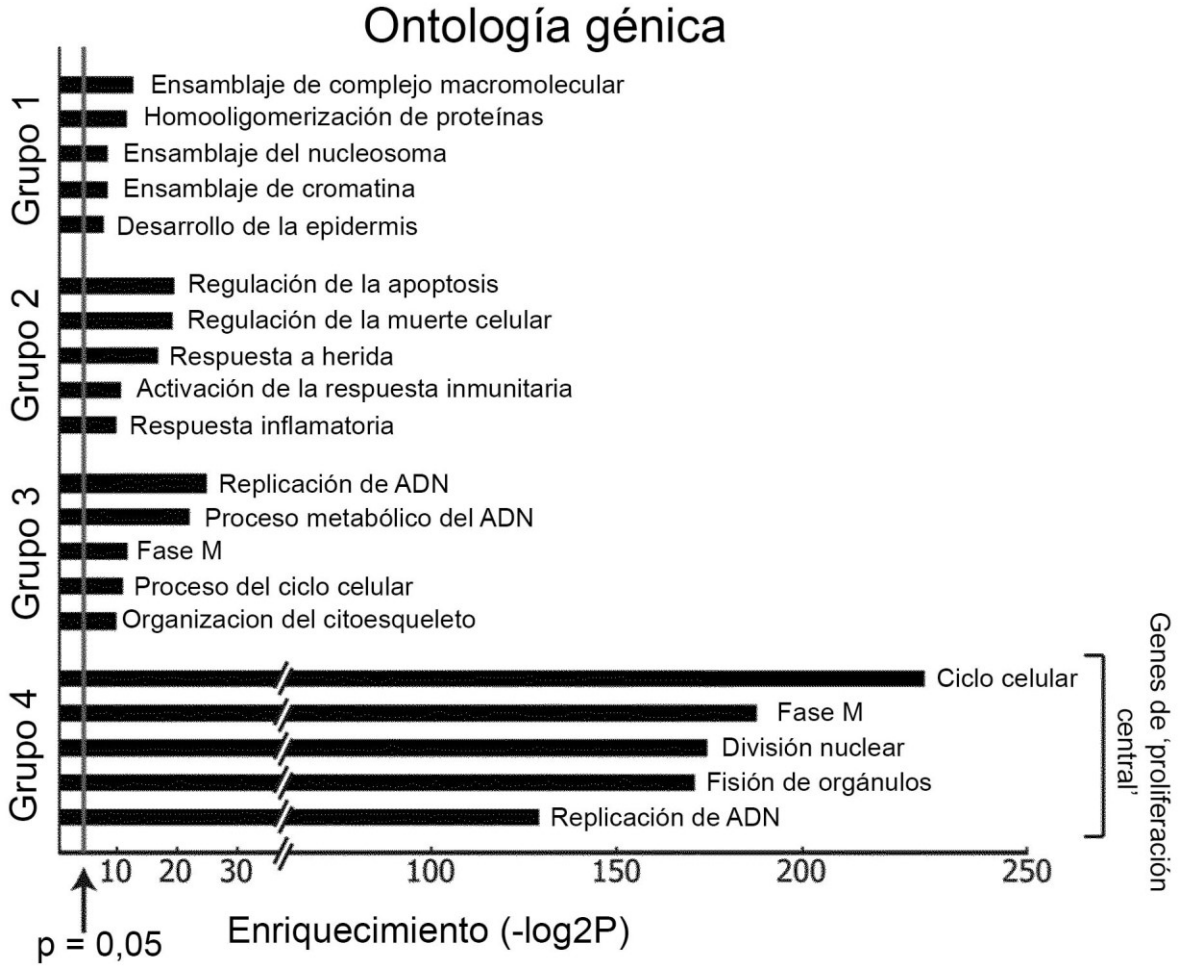


Figura 1E

Enriquecimiento de genes de 'proliferación central' dentro de las firmas de 'mal pronóstico' del cáncer de mama

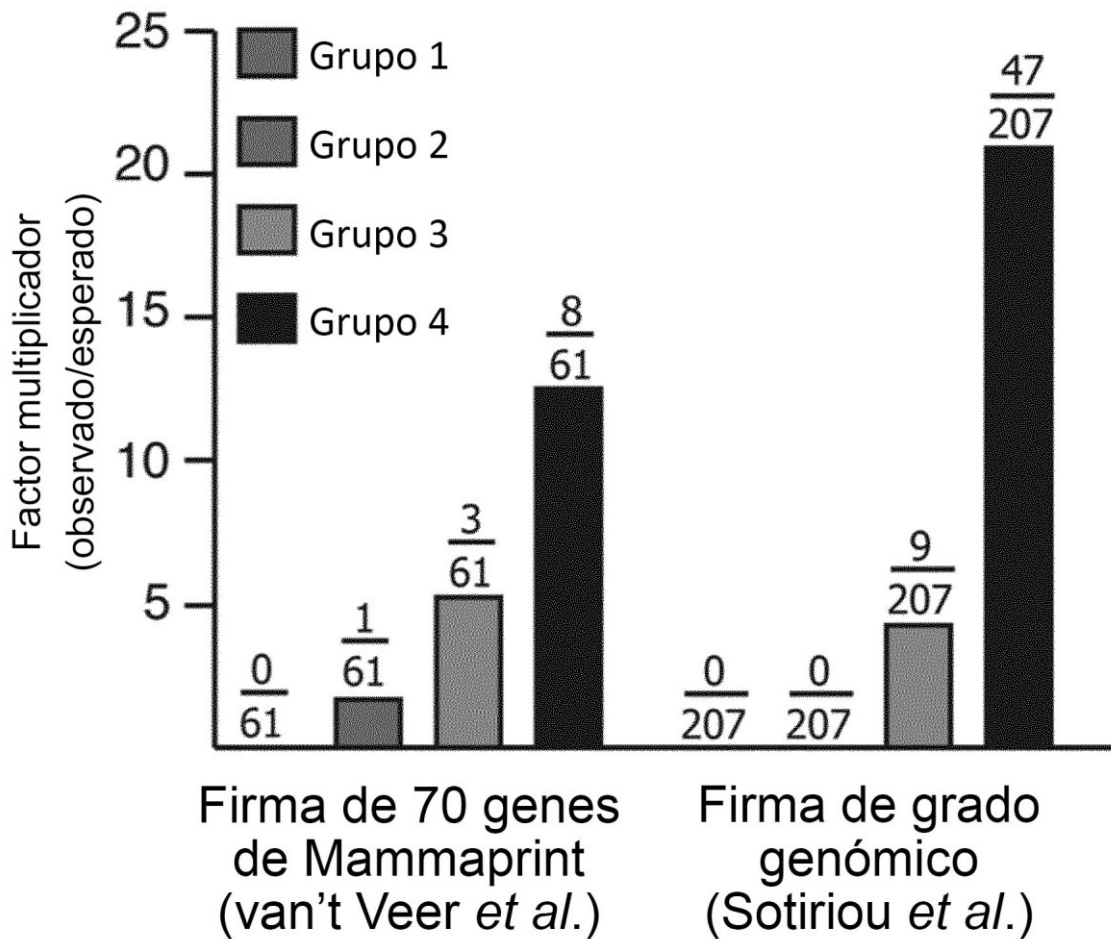


Figura 2B

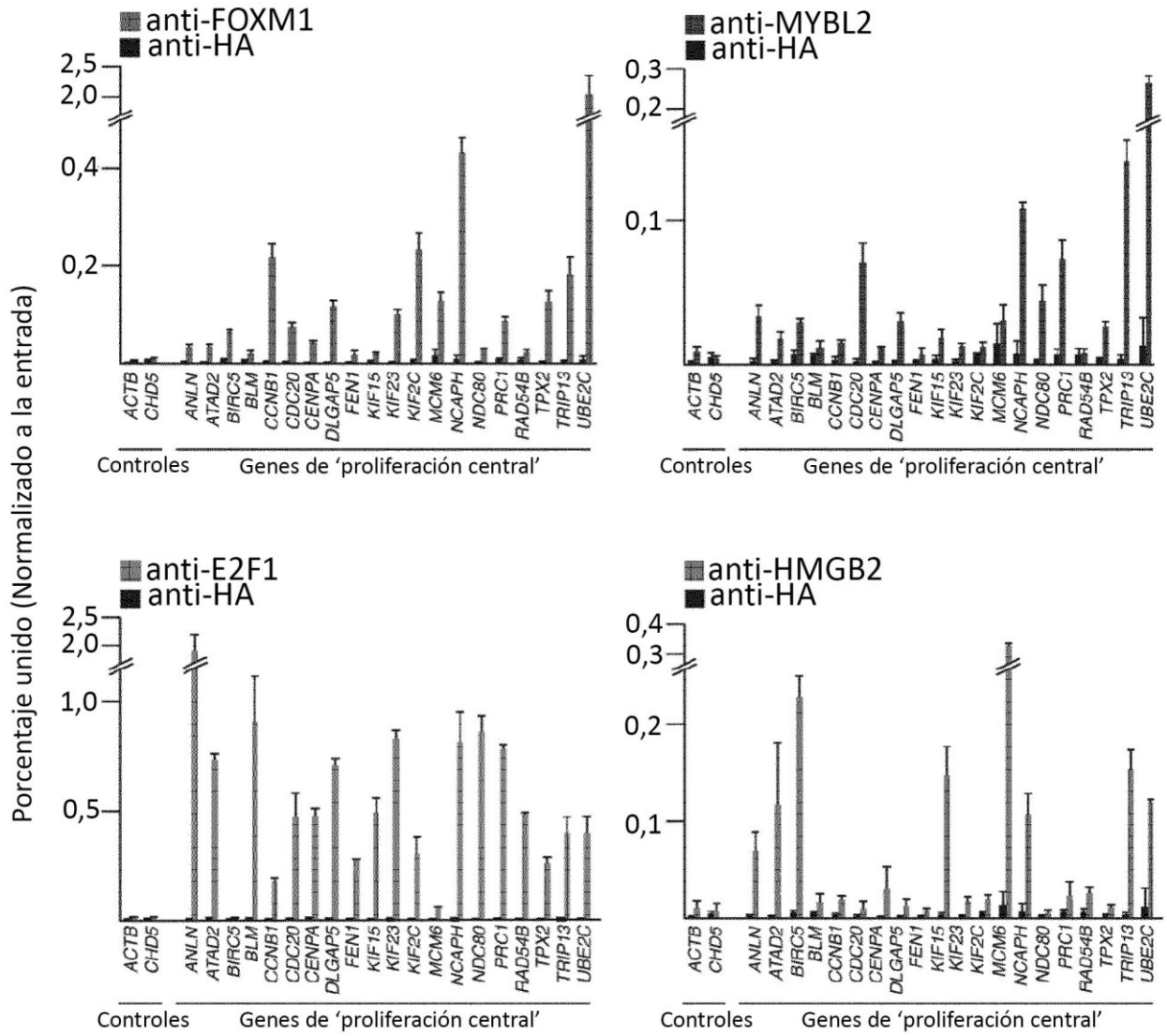


Figura 2C

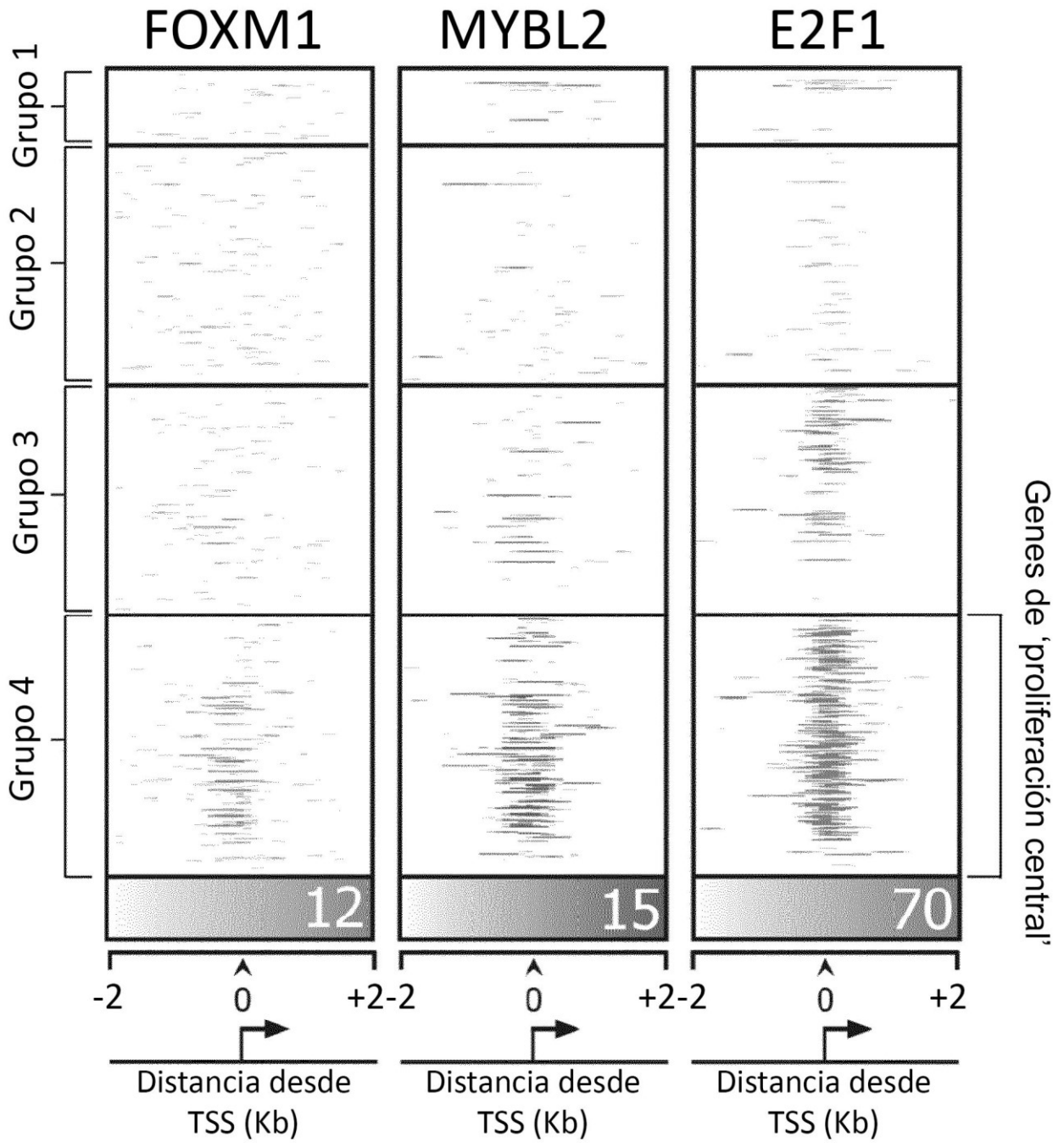


Figura 2D

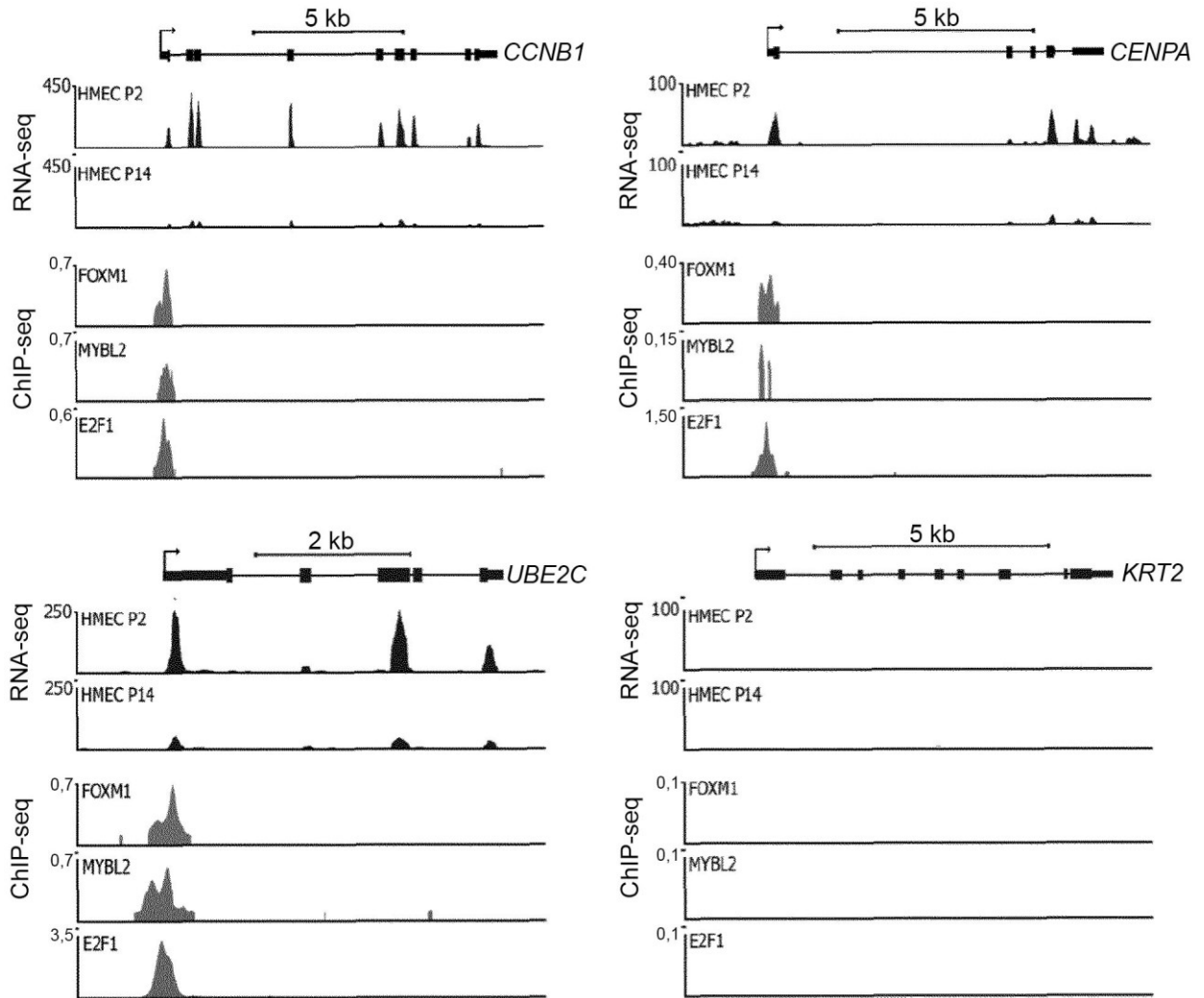


Figura 3A

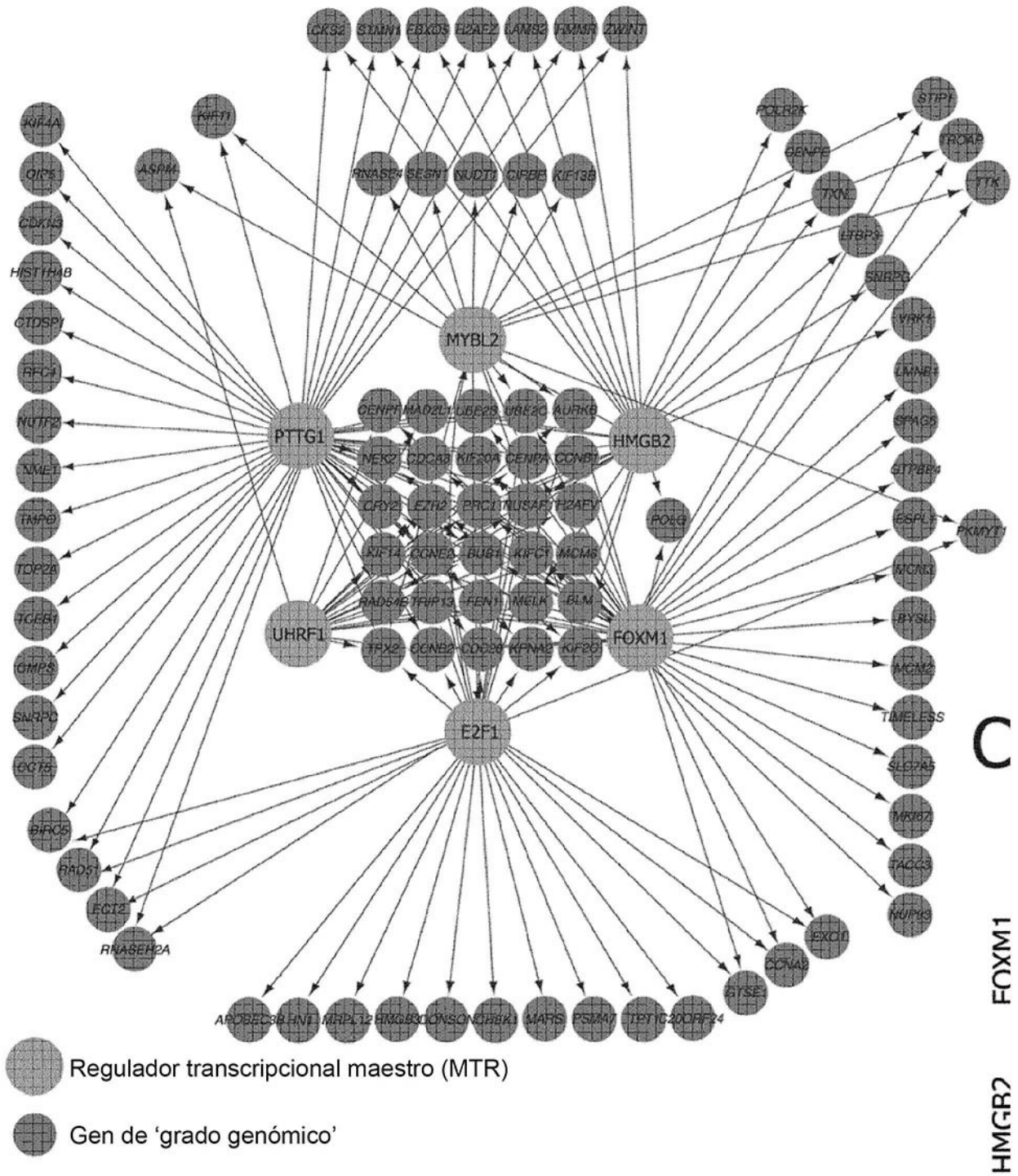


Figura 3B

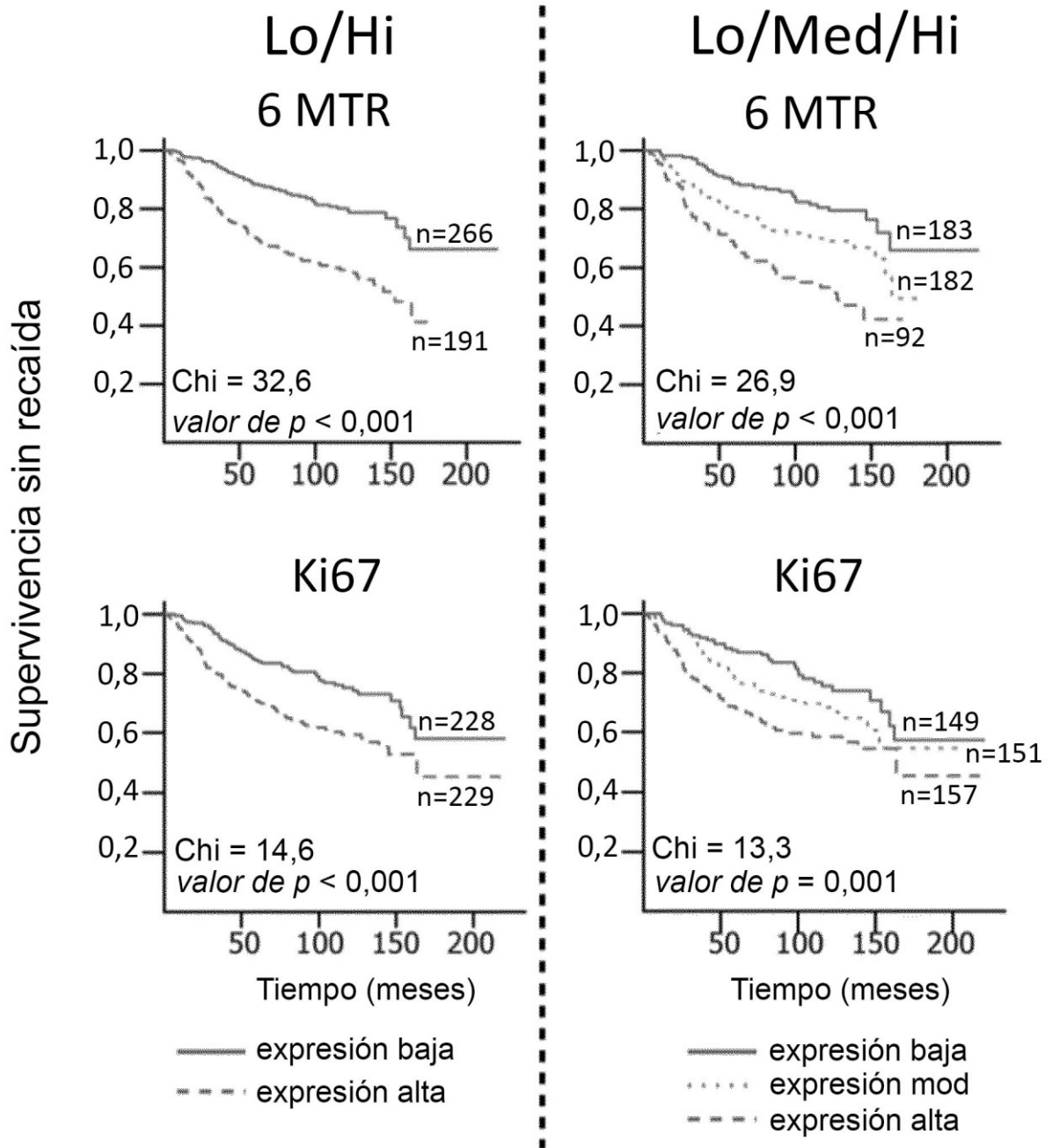


Figura 3C

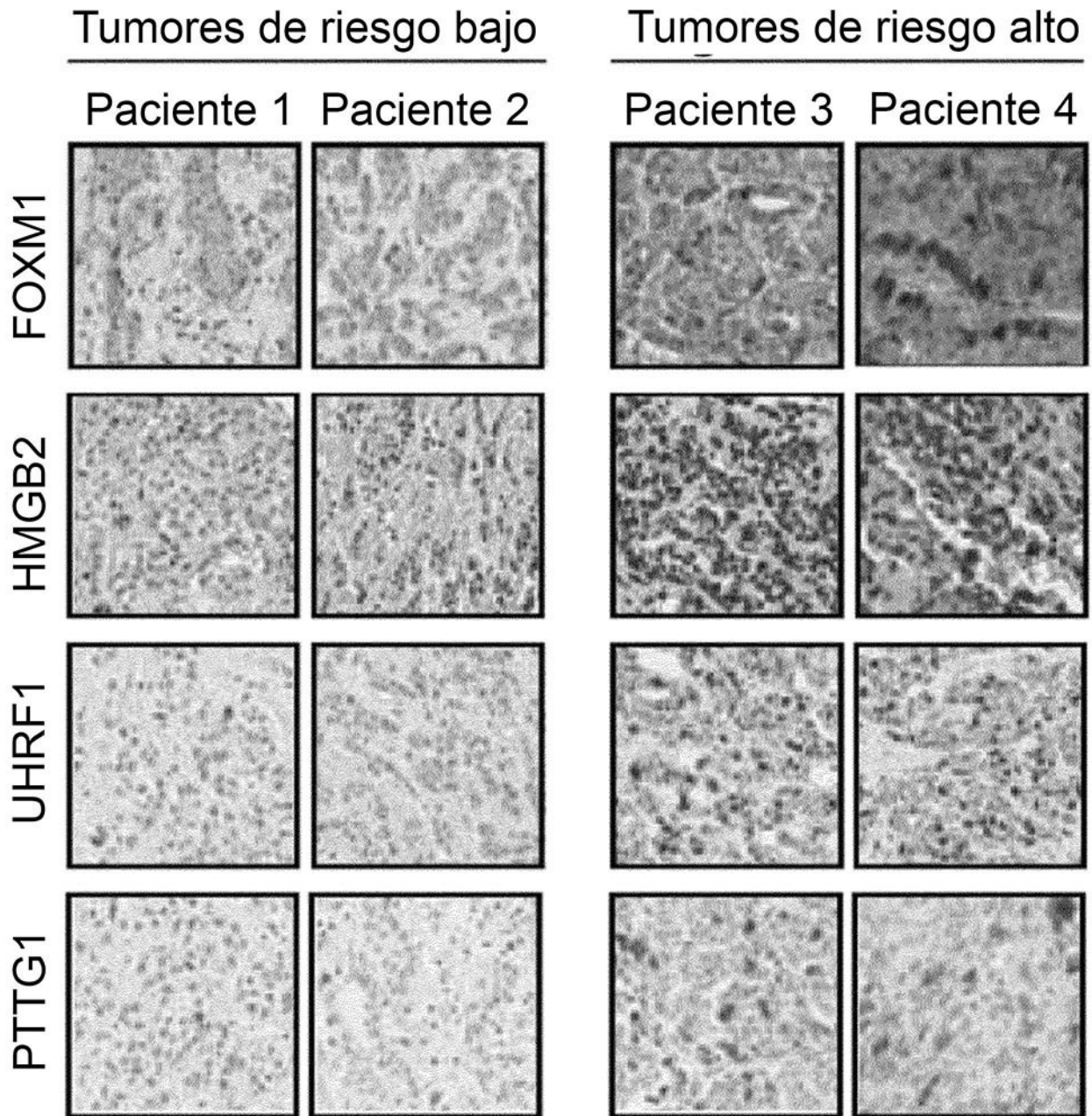


Figura 3D

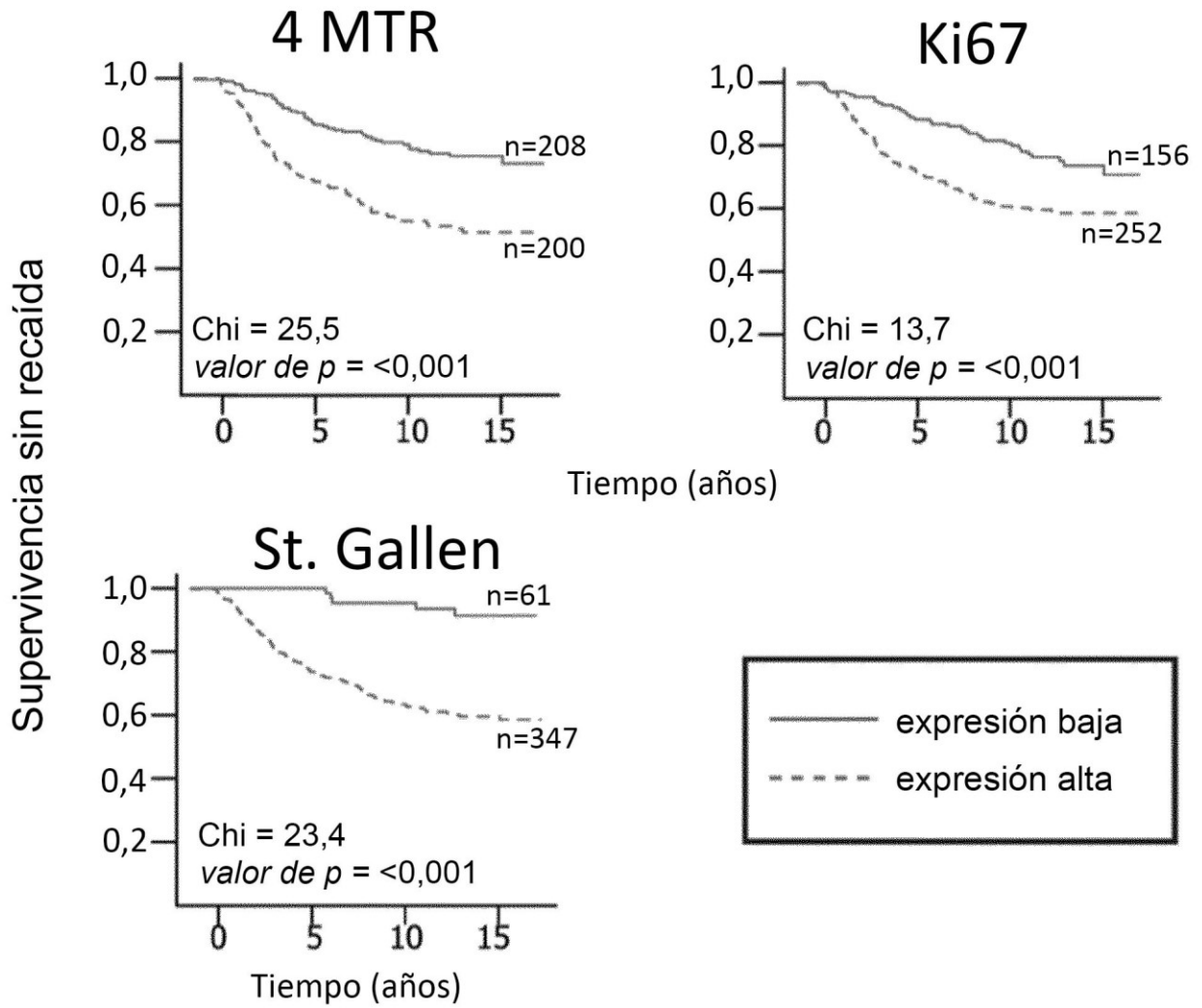


Figura 3E

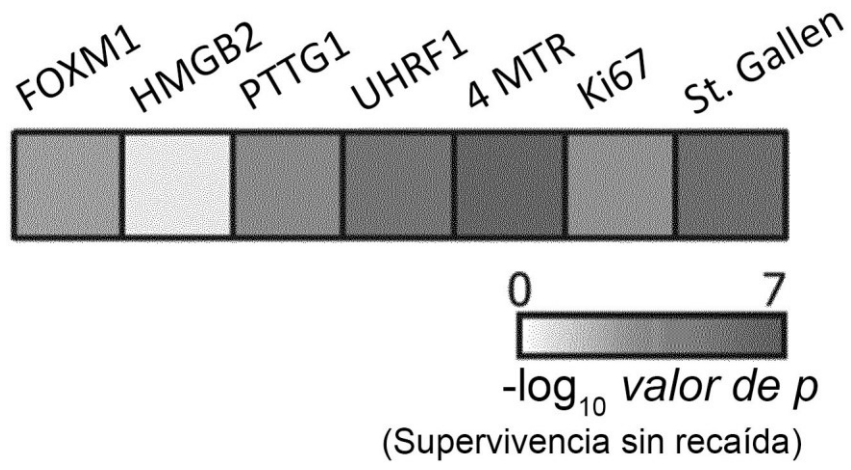


Figura 4A

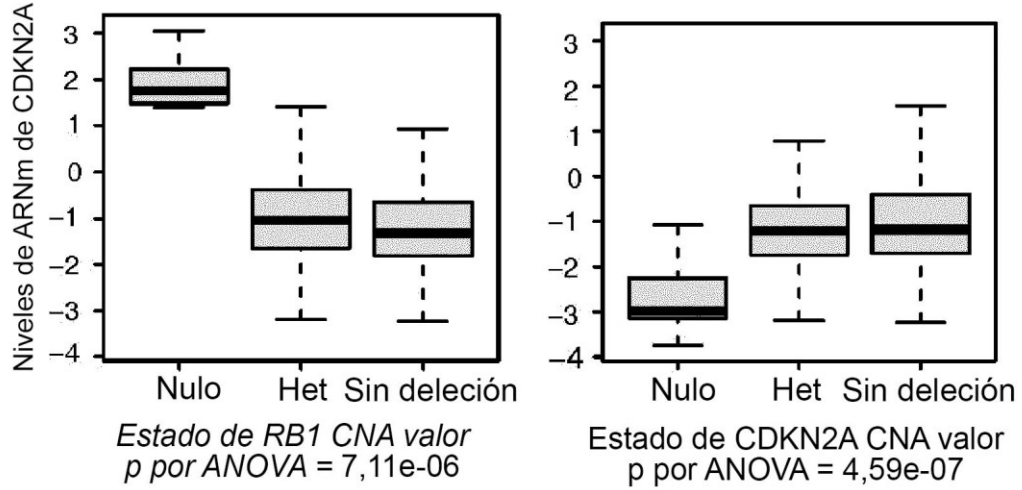


Figura 4B

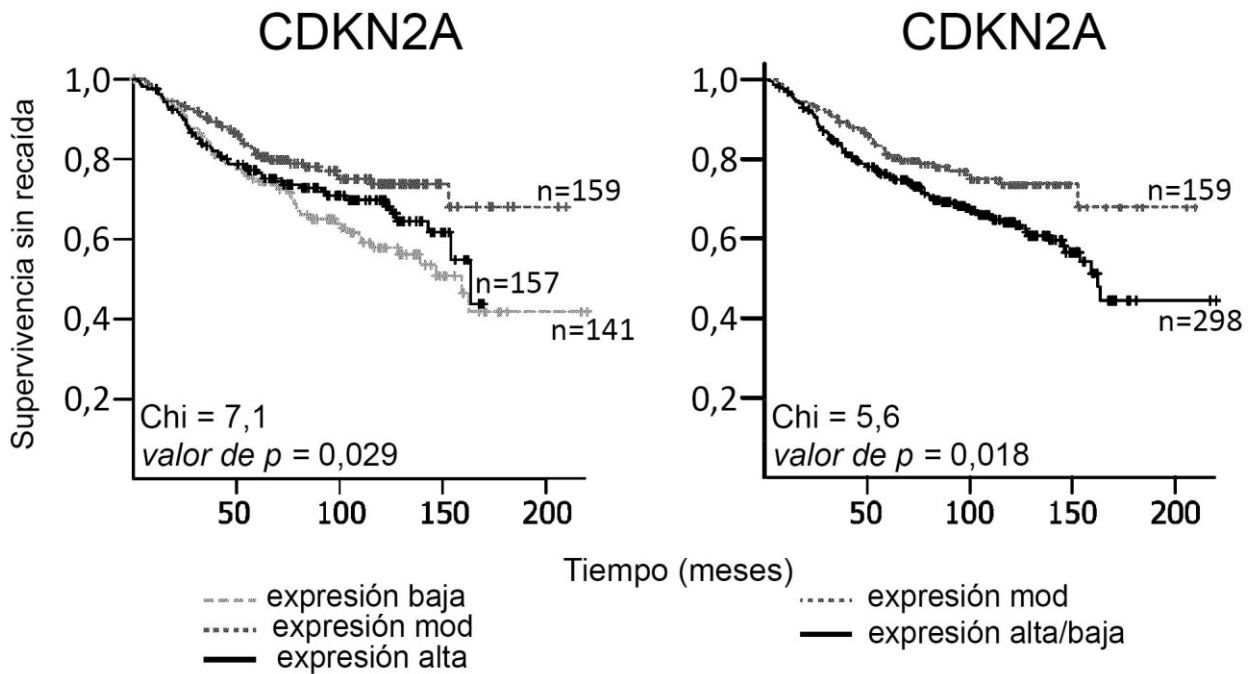
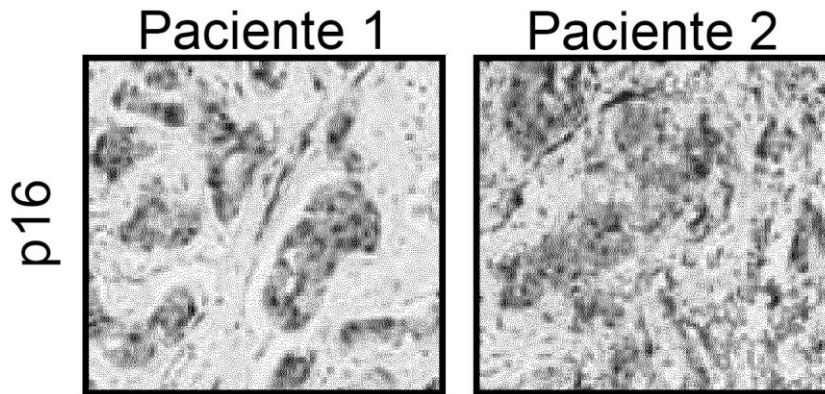
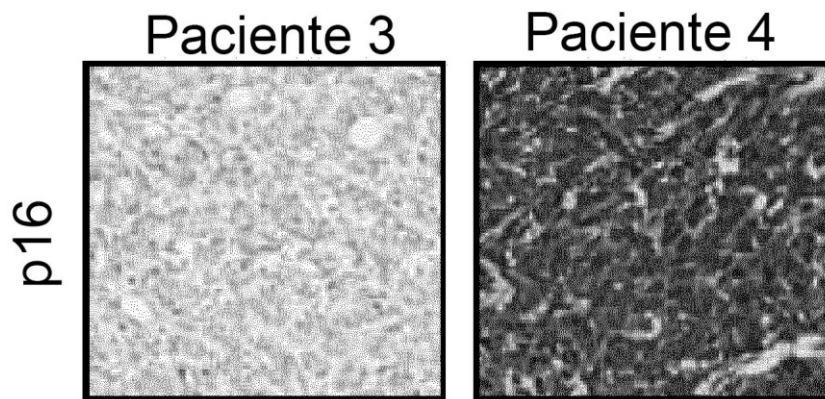


Figura 4C

Tumores de riesgo bajo representativos
(niveles de p16 moderados)



Tumores de riesgo alto representativos
(niveles de p16 muy bajos o altos)



D

Figura 4D

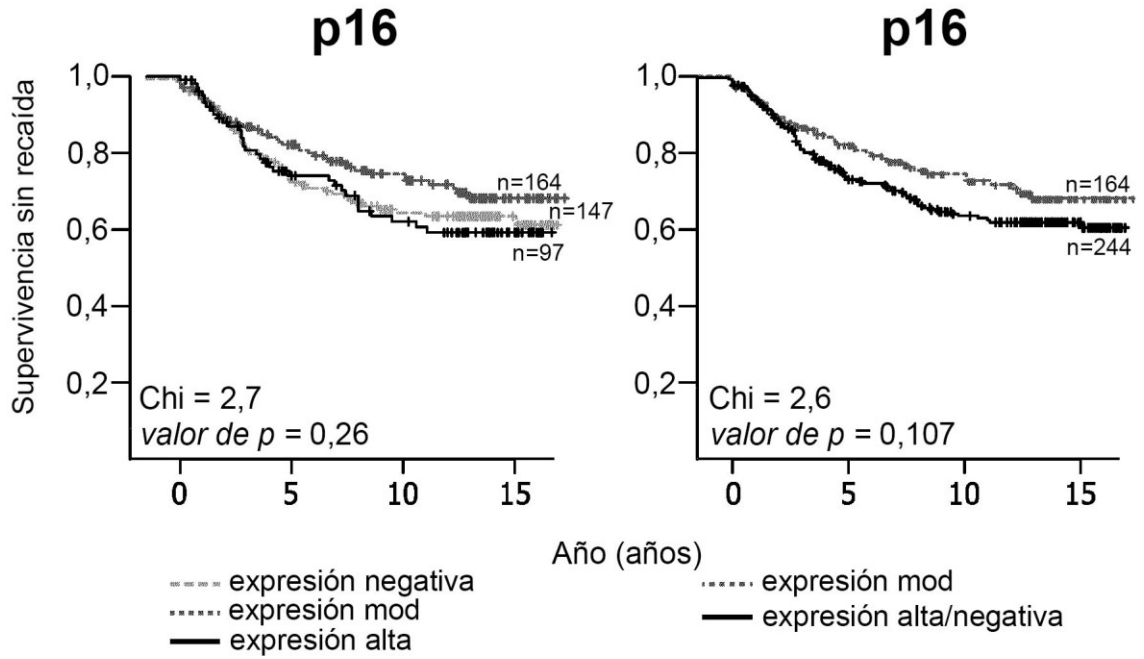


Figura 4E

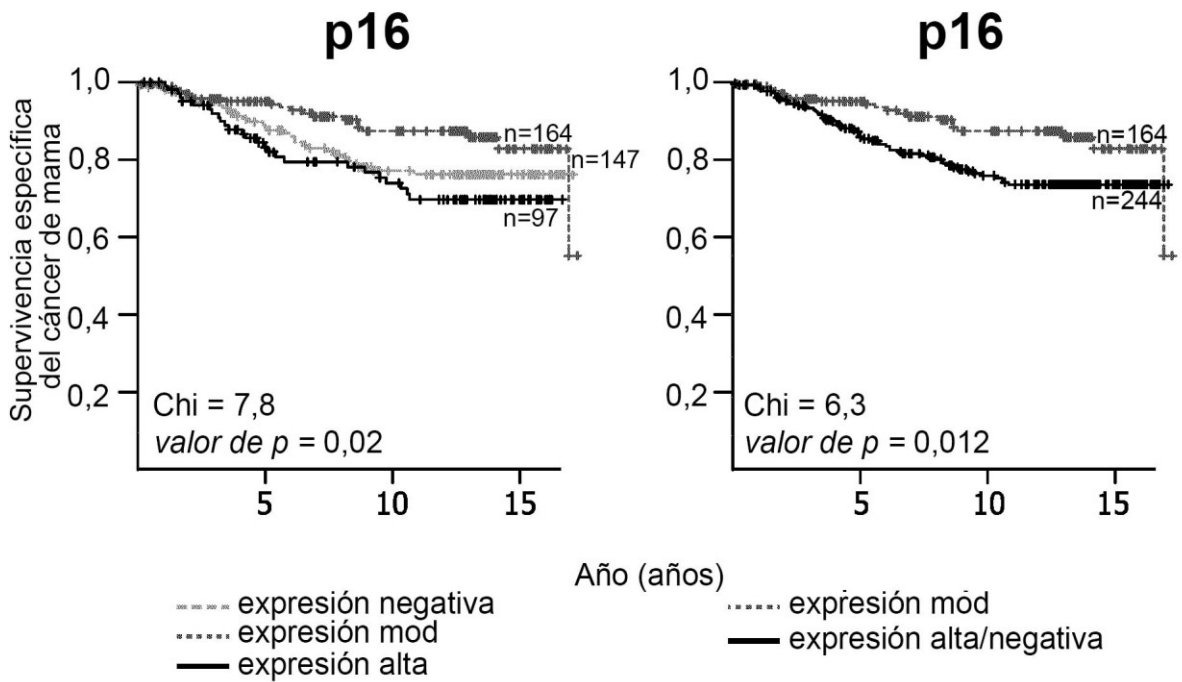


Figura 5A

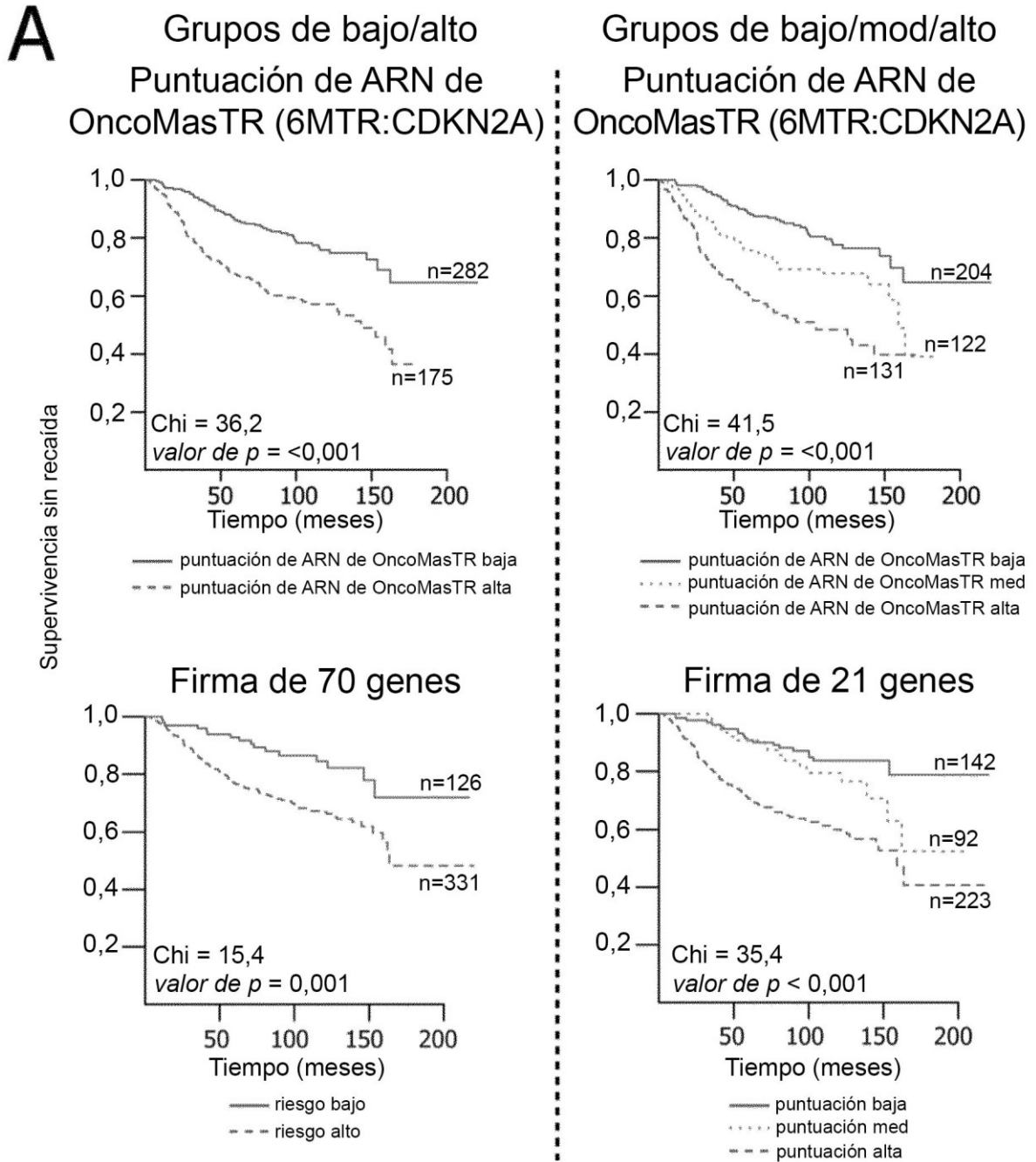


Figura 5B

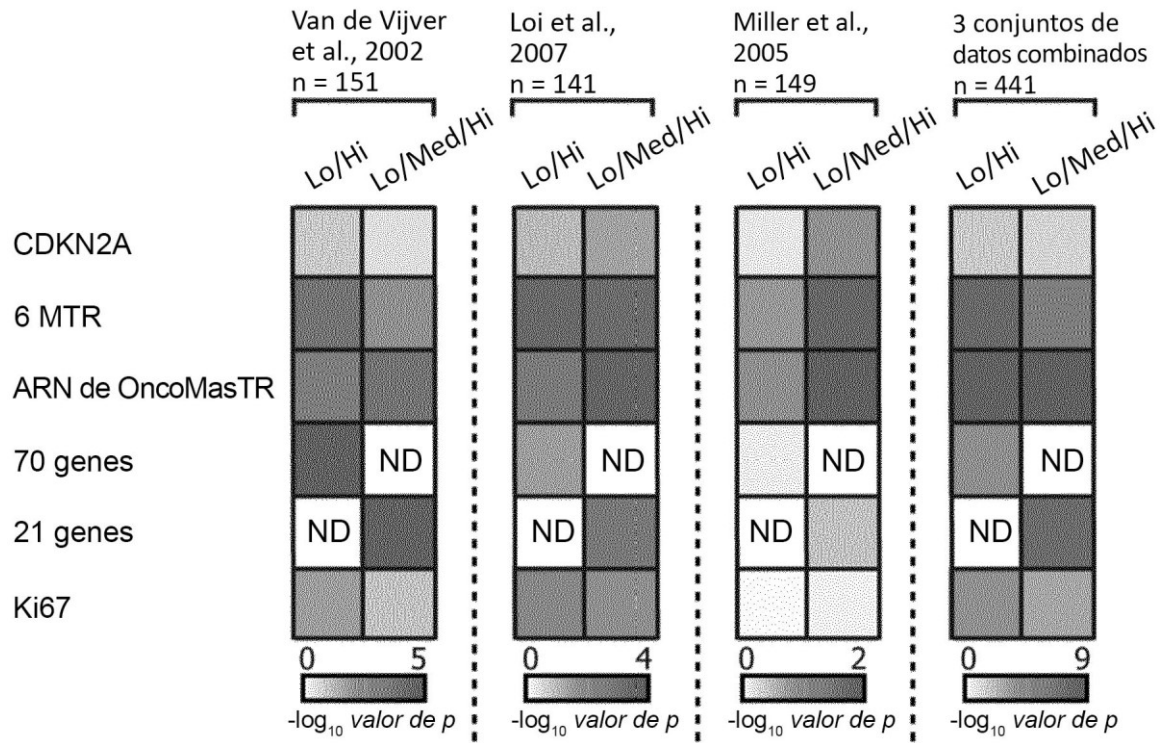


Figura 5C

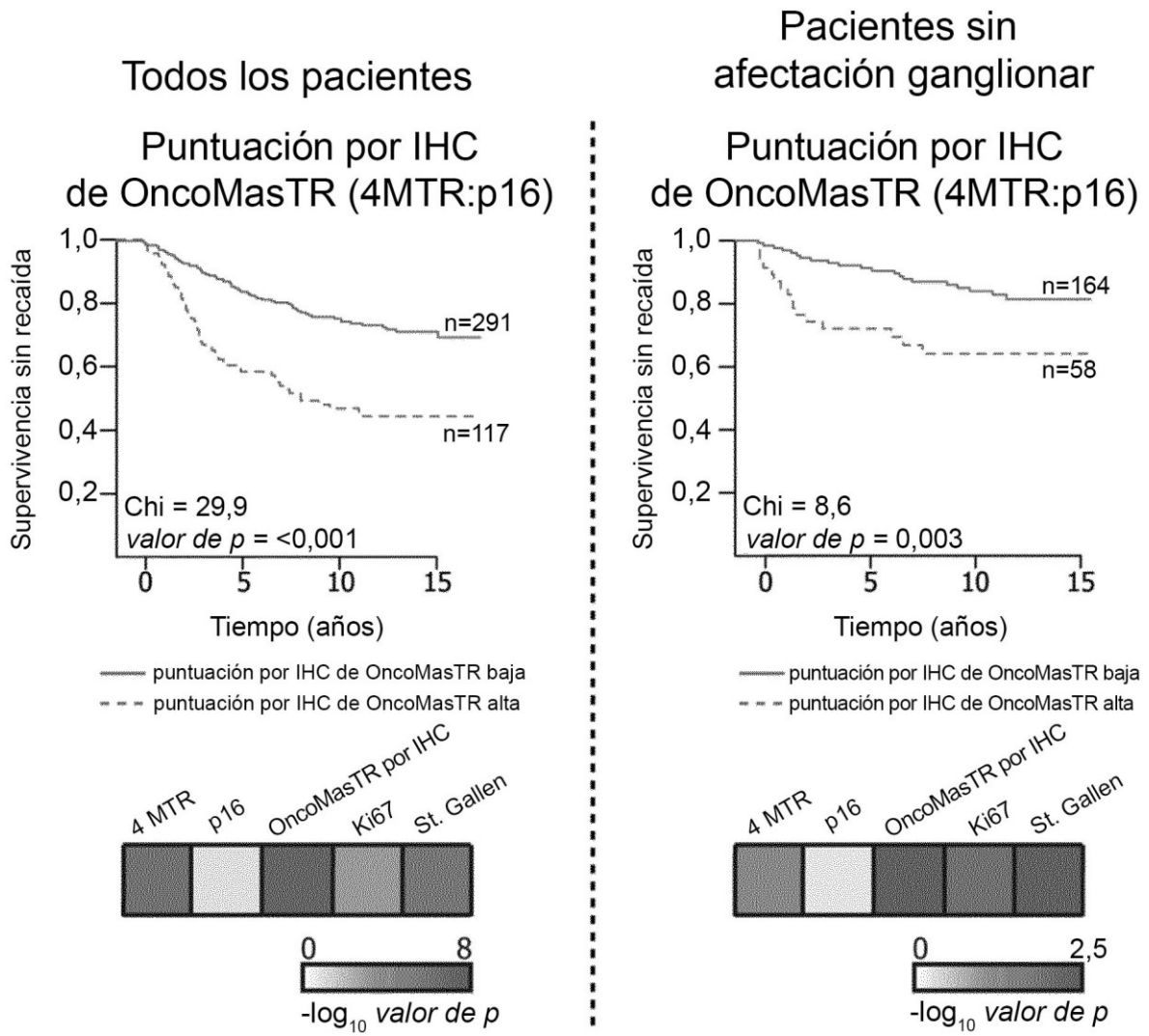


Figura 5D

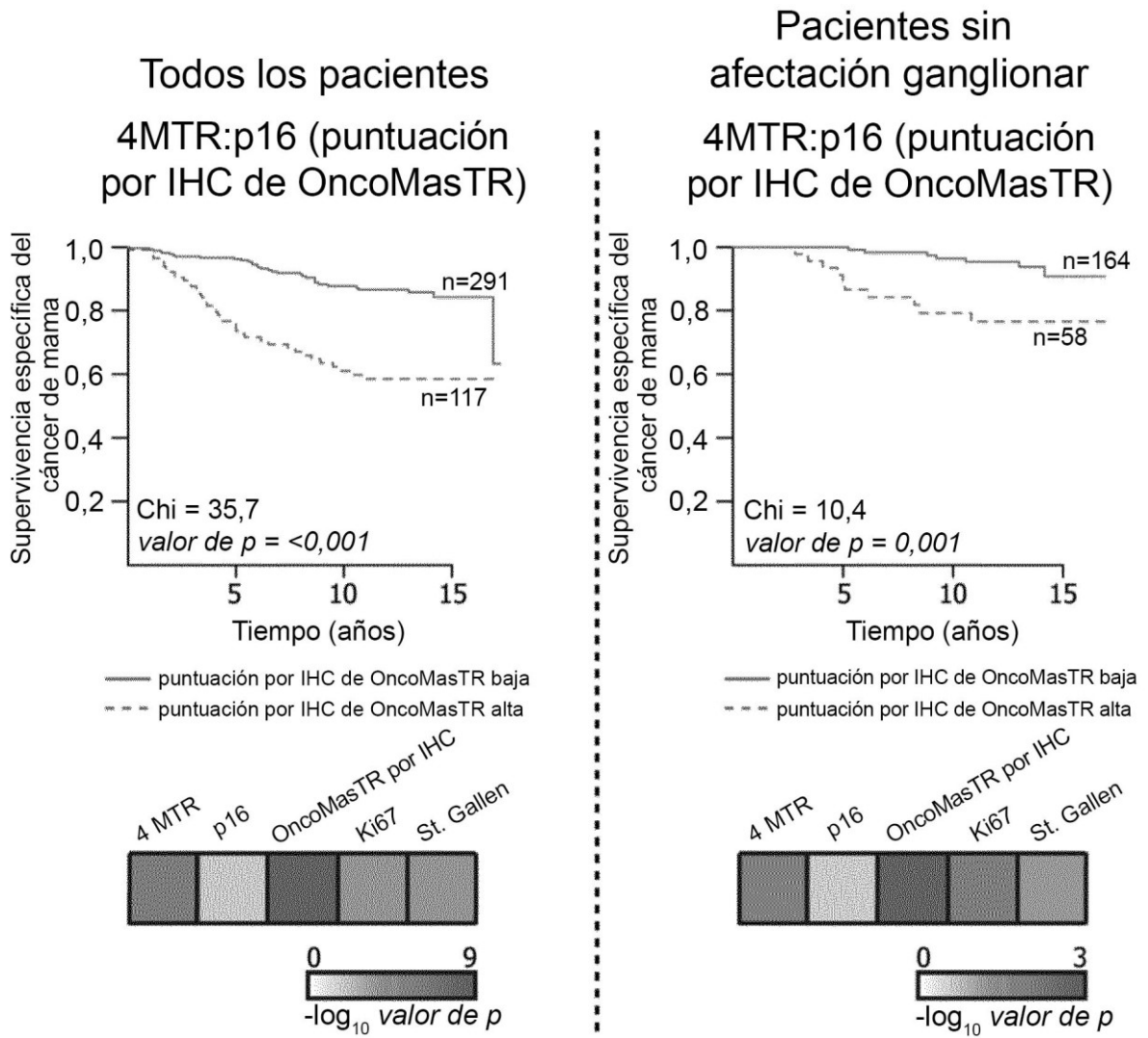


Figura 6A

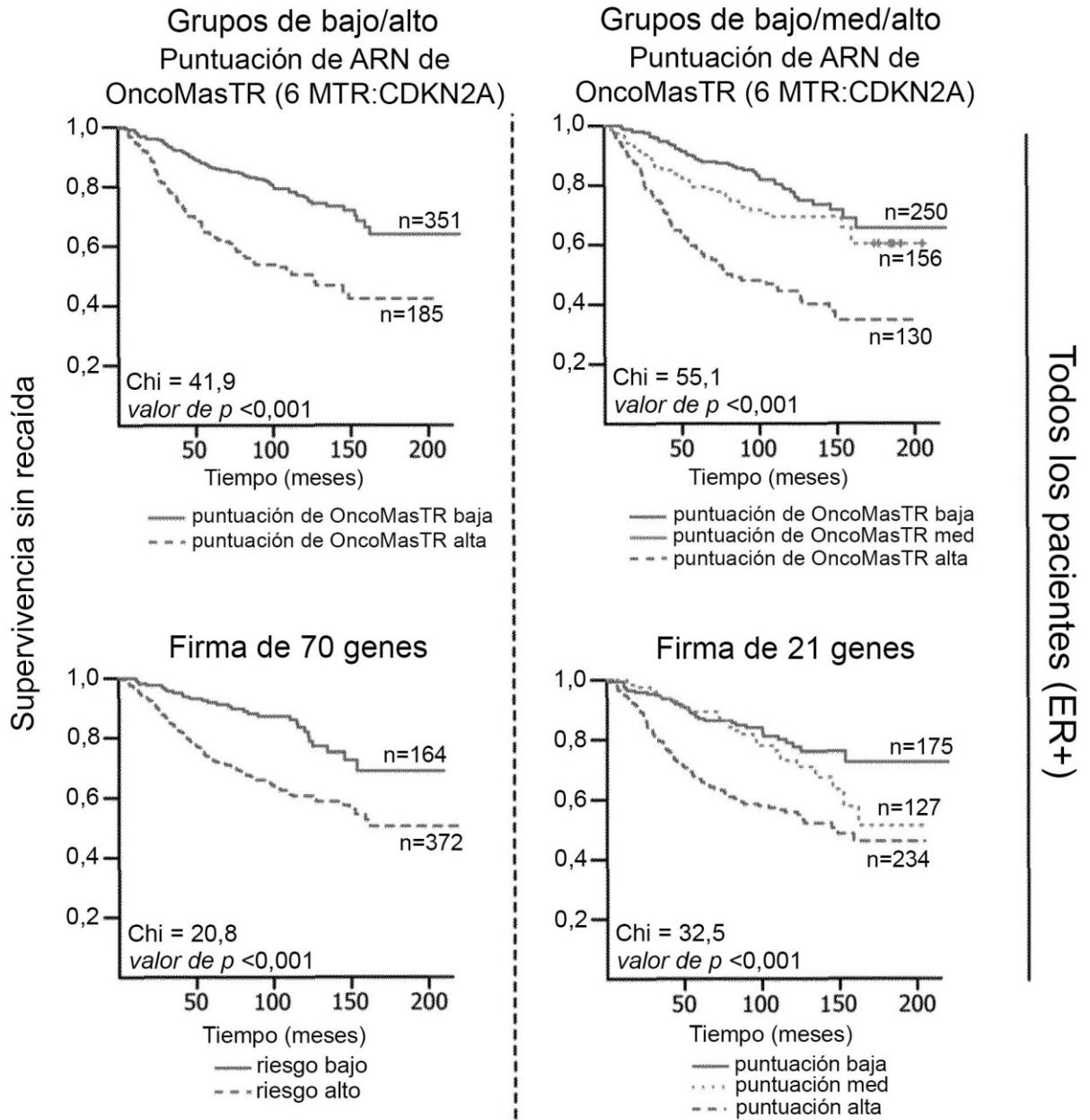
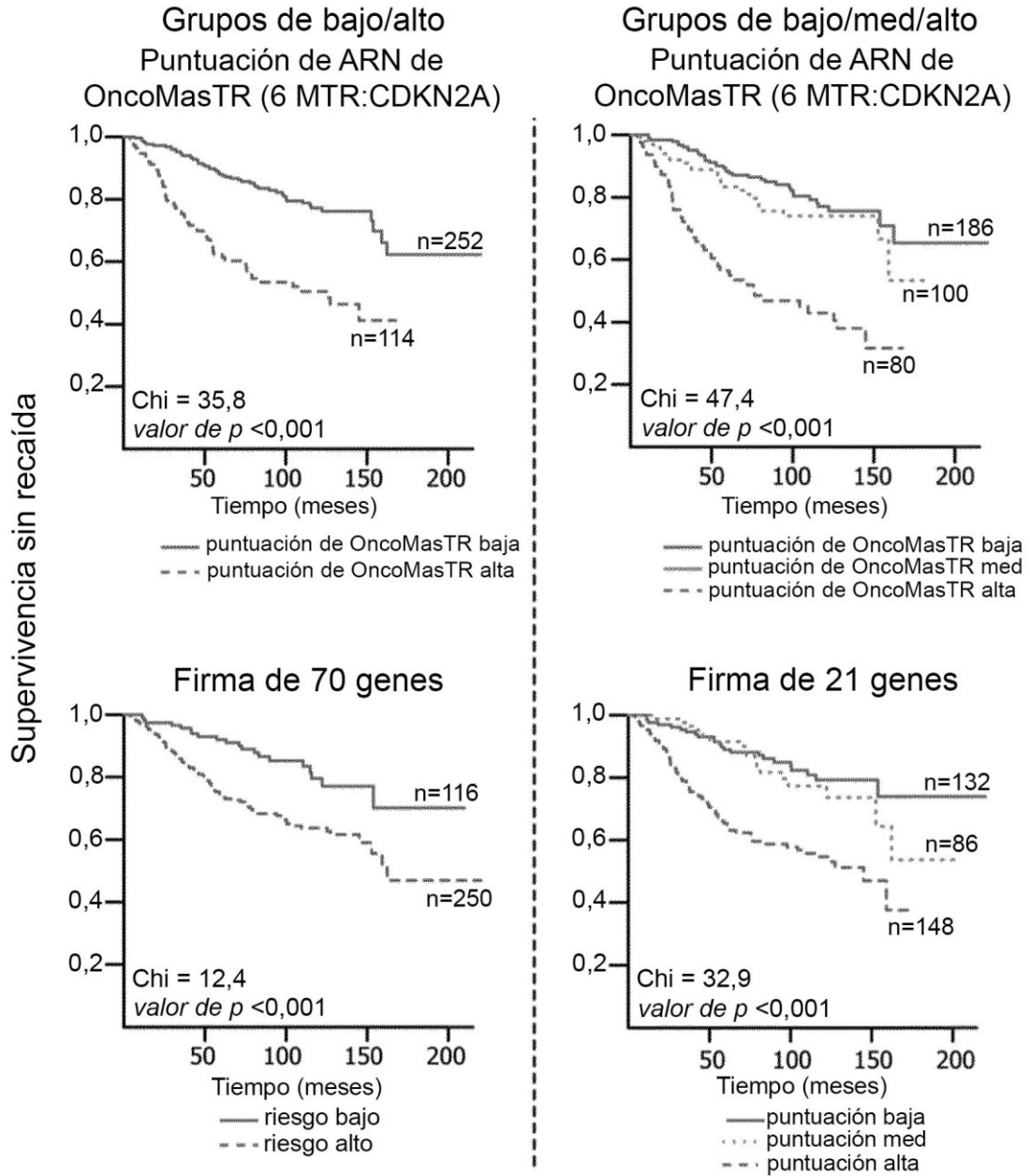


Figura 6B



Grupos de bajo/alto

Grupos de bajo/mod/alto

6 MTR

6 MTR + p16

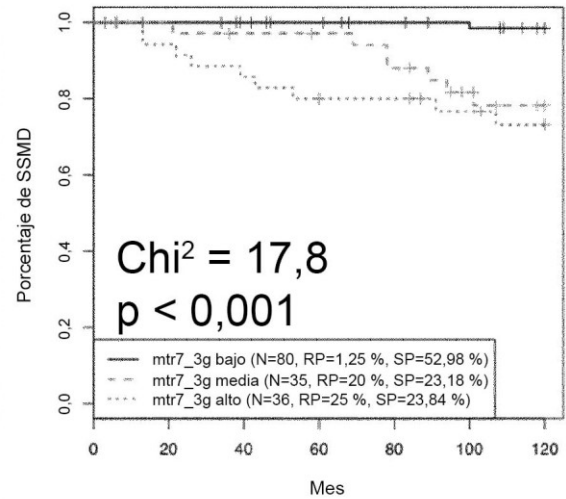
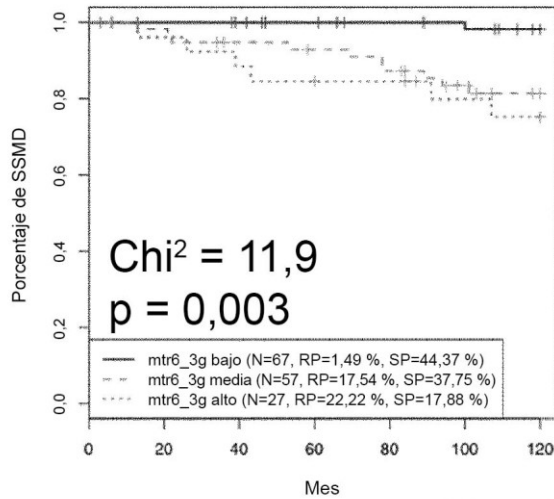
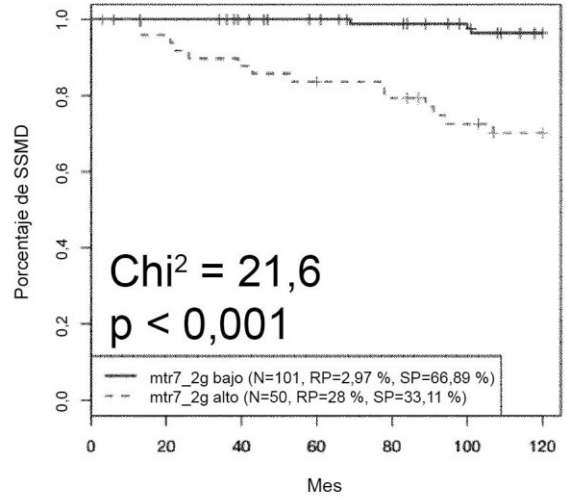
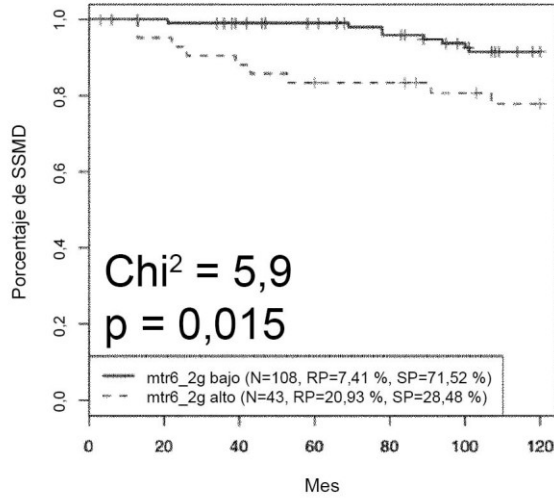
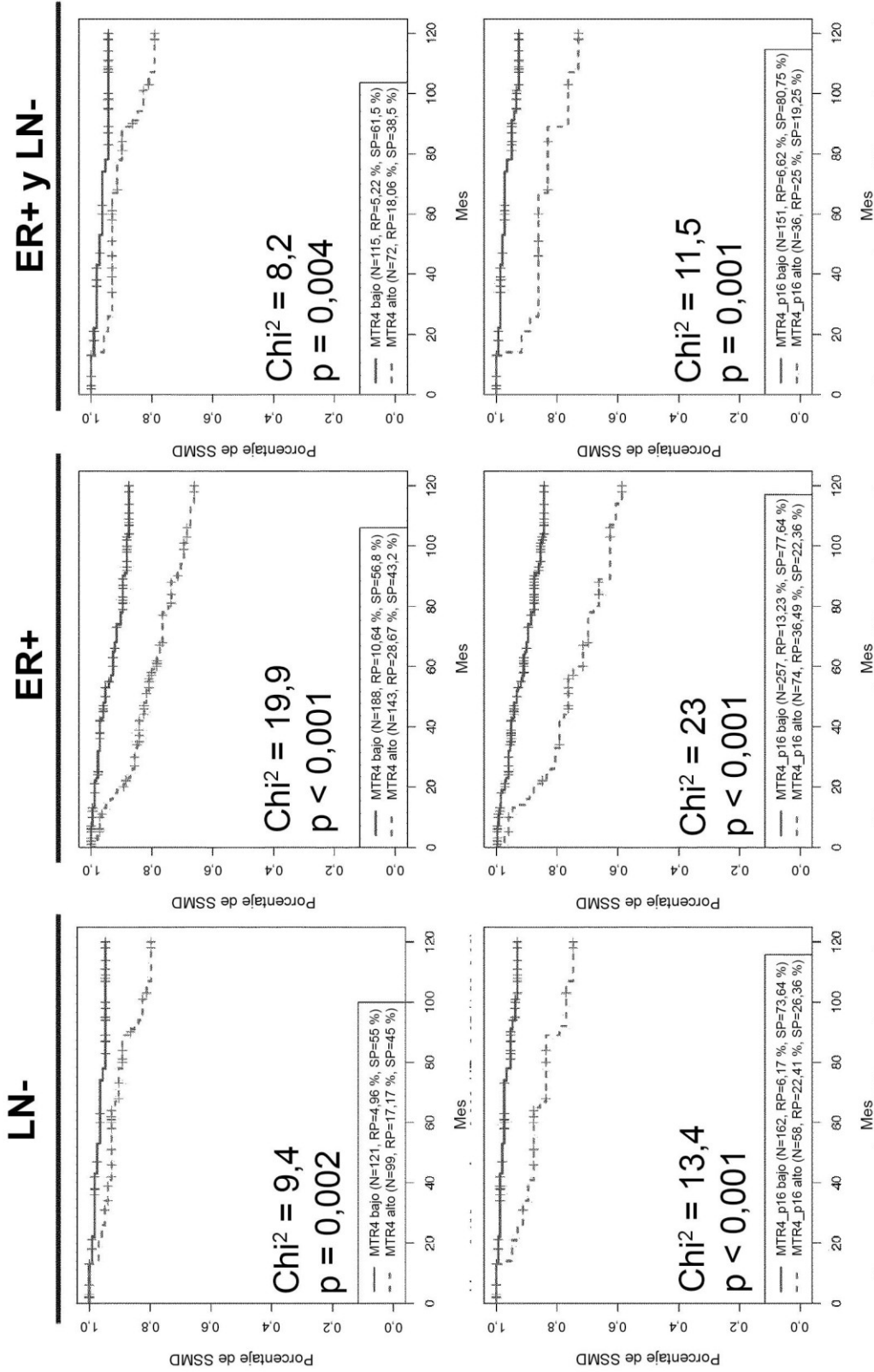


Figura 7

Figura 8



MTR4 + p (9) SUBSTITUTE SHEET (RULE 26)

Figura 9

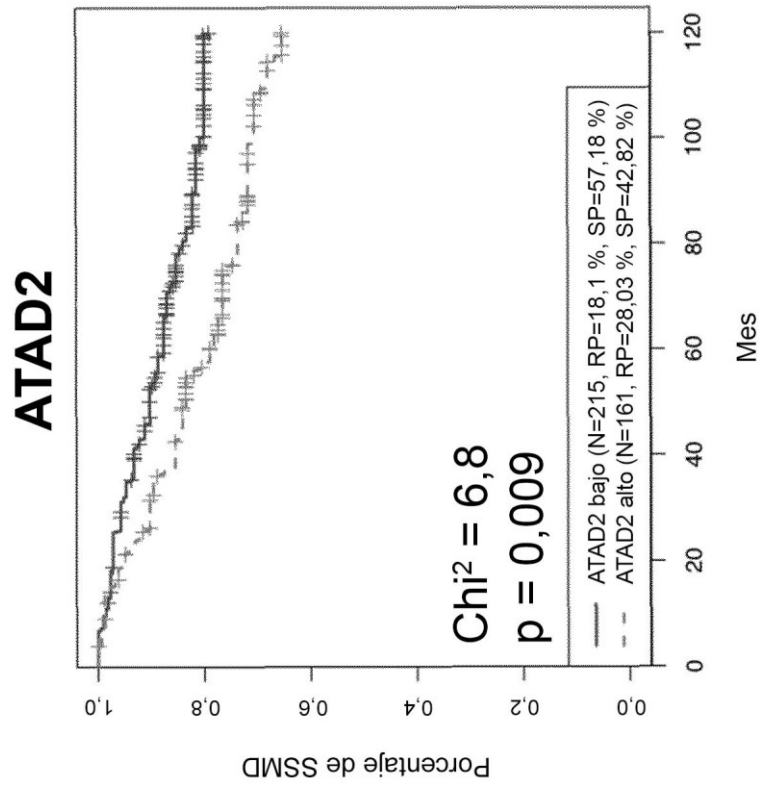
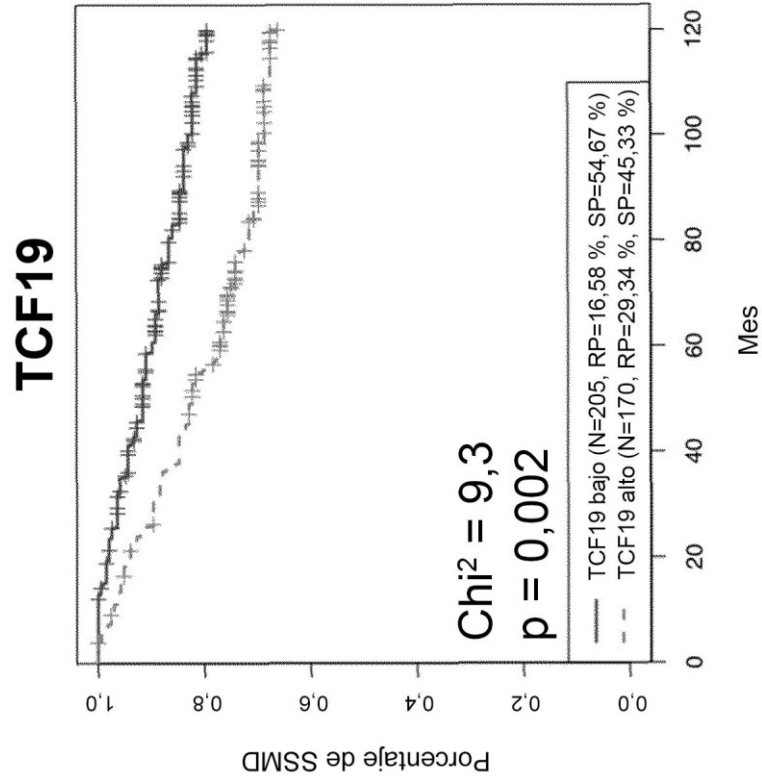
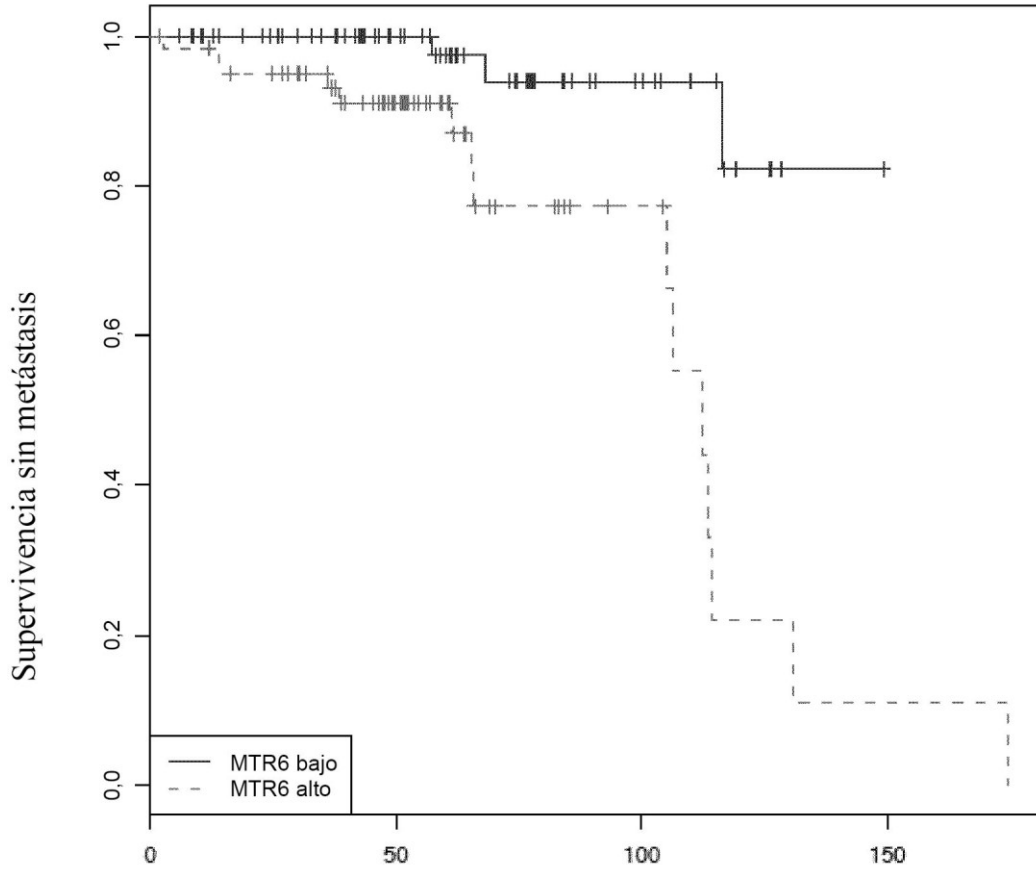


Figura 10



Chisq=16,601, valor de p=0, HR=7,69 (2,17-27,18)

Figura 11

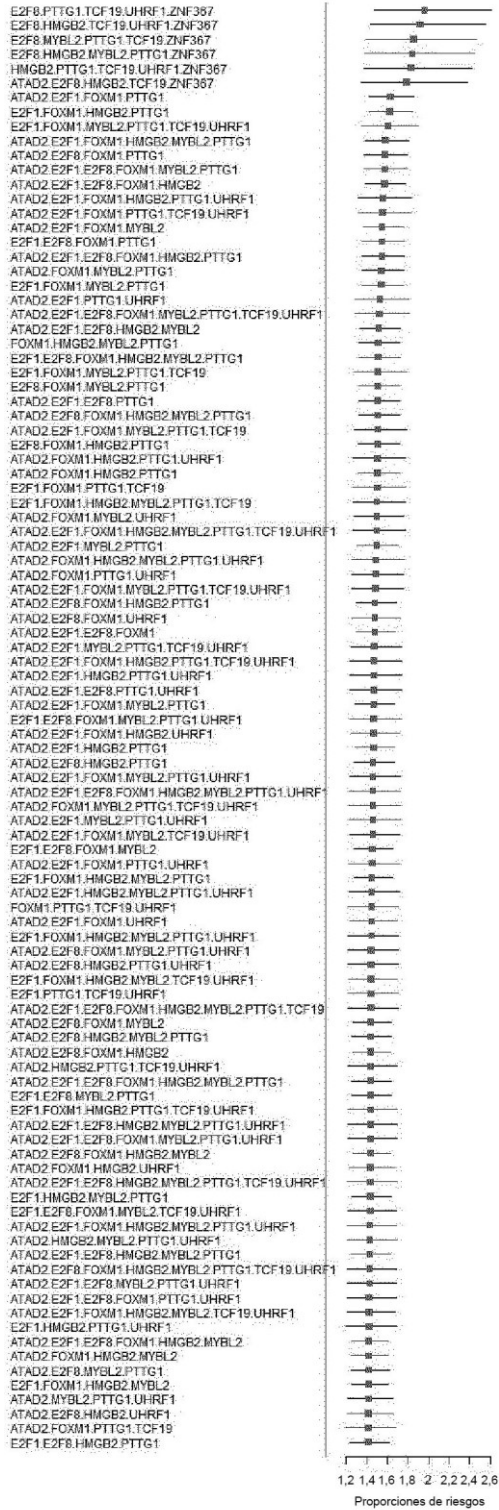


Figura 12

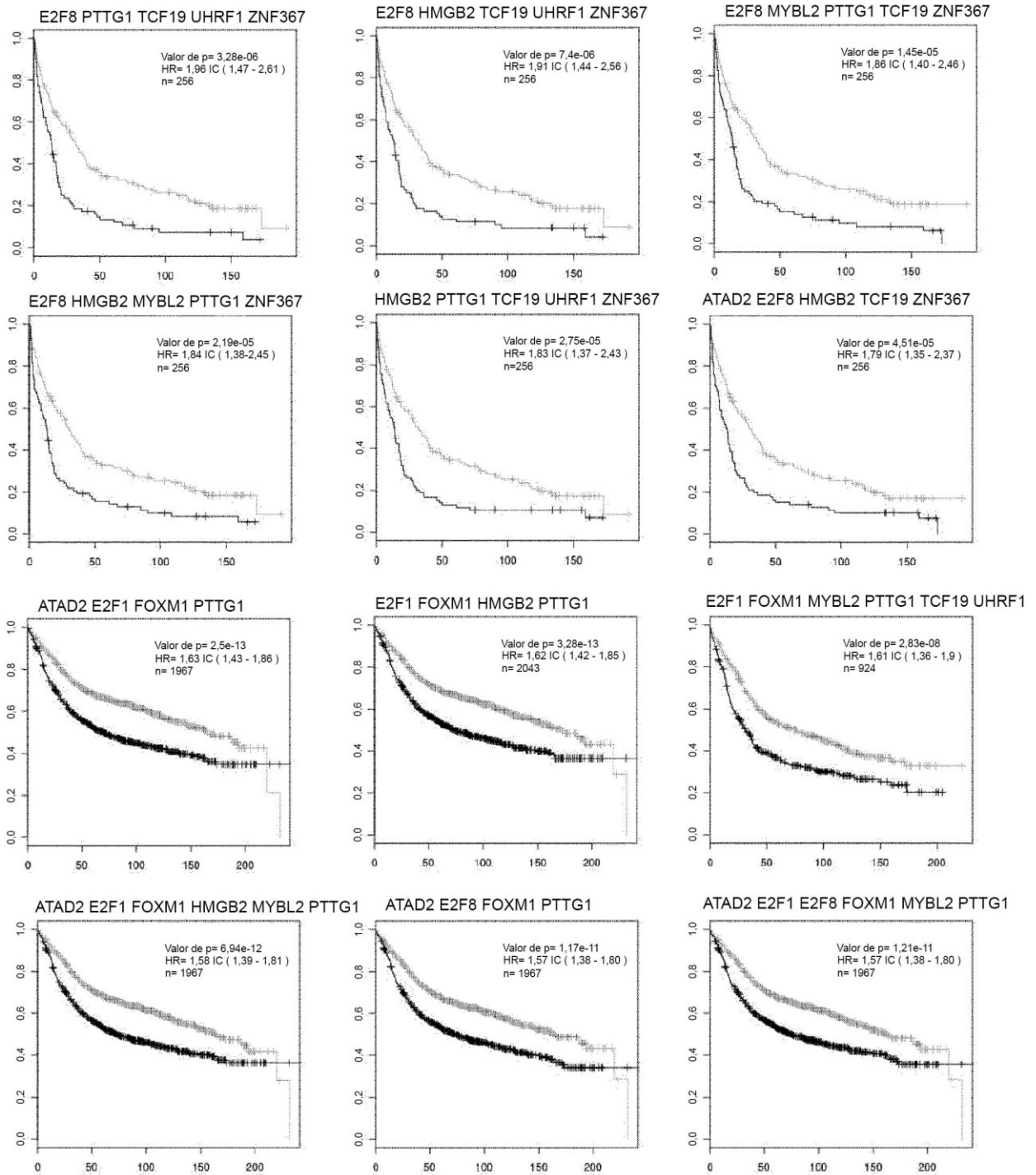


Figura 12 (cont.)

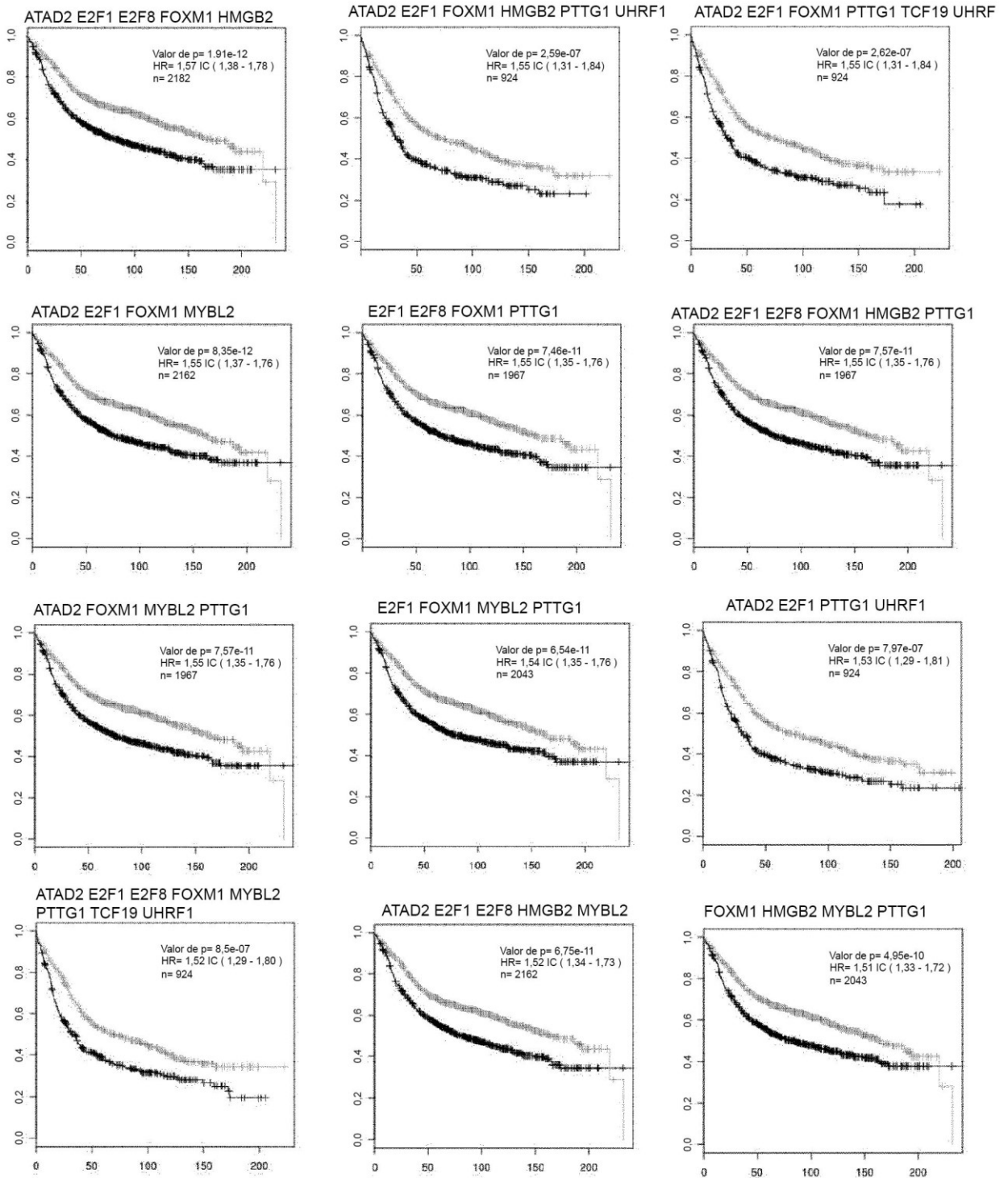


Figura 13

