(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 938 766**

(51) Int. CI.:

C12Q 1/6886 (2008.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **15.11.2013 PCT/US2013/070373**

(87) Fecha y número de publicación internacional: **22.05.2014 WO14078700**

(96) Fecha de presentación y número de la solicitud europea: **15.11.2013 E 13795976 (3)**

(97) Fecha y número de publicación de la concesión europea: **11.01.2023 EP 2920322**

(54) Título: **Firmas génicas para el pronóstico de cáncer**

(30) Prioridad:

16.11.2012 US 201261727533 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
14.04.2023

(73) Titular/es:

MYRIAD GENETICS, INC. (100.0%)
320 Wakara Way
Salt Lake City, UT 84108, US

(72) Inventor/es:

STONE, STEVEN;
GUTIN, ALEXANDER;
WAGNER, SUSANNE y
REID, JULIA

(74) Agente/Representante:

BERTRÁN VALLS, Silvia

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

Firmas génicas para el pronóstico de cáncer

Referencia cruzada a solicitudes relacionadas

Esta solicitud reivindica la prioridad (por ejemplo, bajo 35 U.S.C. § 119(e)) a la solicitud provisional estadounidense con n.º de serie 61/727.533, presentada el 16 de noviembre de 2012.

Campo de la divulgación

La divulgación se refiere generalmente a una clasificación molecular de enfermedad y particularmente a marcadores moleculares para el pronóstico de cáncer y a métodos de uso de los mismos.

Antecedentes de la divulgación

El cáncer es uno de los principales problemas de salud pública, que representa aproximadamente el 25% de todas las muertes en los Estados Unidos. Aunque se han concebido muchos tratamientos para diversos cánceres, estos tratamientos a menudo varían en cuanto a la gravedad de los efectos secundarios. Resulta útil que los médicos sepan cómo de agresivo es el cáncer de un paciente con el fin de determinar la agresividad con la que tratar el cáncer.

Por ejemplo, a la mayoría de los pacientes con cáncer de próstata asintomático de estadio temprano se les trata con prostatectomía radical o radioterapia y opcionalmente terapia complementaria (por ejemplo, terapia hormonal o quimioterapia), teniendo todas ellas graves efectos secundarios. Sin embargo, para muchos de estos pacientes, estos tratamientos y sus efectos secundarios y costes asociados son innecesarios porque el cáncer en estos pacientes no es agresivo (es decir, crece lentamente y es improbable que provoque mortalidad o morbilidad significativa durante la vida del paciente). En otros pacientes el cáncer es virulento (es decir, es más probable que reaparezca) y es necesario un tratamiento agresivo para salvar la vida del paciente.

Se han concebido algunas herramientas para ayudar a los médicos a decidir qué pacientes necesitan un tratamiento agresivo y cuáles no. De hecho, en la actualidad hay varios parámetros clínicos en uso para este propósito en diversos cánceres diferentes. En el cáncer de próstata, por ejemplo, tales parámetros clínicos incluyen antígeno prostático específico (PSA) en suero, grado de Gleason, estadio patológico y márgenes quirúrgicos. En los últimos años, se ha hecho que los parámetros clínicos sean más útiles mediante su incorporación en nomogramas posoperatorios de múltiples variables continuas que calculan la probabilidad de un paciente de presentar progresión/recidiva de cáncer. Véase, por ejemplo, Kattan *et al*., J. CLIN. ONCOL. (1999) 17:1499-1507; Stephenson *et al*., J. CLIN. ONCOL. (2005) 23:7005-7012. Se describen métodos para evaluar y clasificar el cáncer, por ejemplo, en los documentos WO 2012/030840, WO 2012/006447, WO 2010/080933, US 2012/041274, Cuzick *et al*. The Lancet 12: 245-255, Cuzick *et al*. British Journal of Cancer 106 (6): 1095-99 y Cooperberg *et al*. Journal of Urology 187 (4): E850. Sin embargo, a pesar de estos avances, a muchos pacientes se les administran tratamientos contra el cáncer inadecuados y todavía existe la gran necesidad de herramientas nuevas y mejoradas para predecir la recidiva de cáncer.

Sumario de la invención

La presente divulgación se base, en parte, en el sorprendente descubrimiento de que la expresión de esos genes cuya expresión sigue estrechamente el ciclo celular (“genes del ciclo celular” o “CCG” tal como se define adicionalmente a continuación) es particularmente útil en la clasificación de tipos seleccionados de cáncer y la determinación del pronóstico de estos cánceres. La presente invención se define en las reivindicaciones independientes, a las que ahora debe hacerse referencia. Se exponen realizaciones ventajosas en las sub-reivindicaciones.

A menos que se defina lo contrario, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado que entiende habitualmente un experto habitual en la técnica a la que pertenece esta divulgación. Aunque pueden usarse métodos y materiales similares o equivalentes a los descritos en el presente documento en la práctica o pruebas de la presente divulgación, a continuación se describen métodos y materiales adecuados. En caso de conflicto, prevalecerá la presente memoria descriptiva, incluyendo las definiciones. Además, los materiales, métodos y ejemplos son sólo ilustrativos y no se pretende que sean limitativos.

Otras características y ventajas de la divulgación resultarán evidentes a partir de la siguiente descripción detallada, y a partir de las reivindicaciones.

Breve descripción de los dibujos

La figura 1 es una ilustración de la potencia predictiva a lo largo del nomograma para paneles de CCG de diferentes

tamaños.

La figura 2 es una ilustración de CCG que predicen el tiempo hasta la recidiva.

La figura 3 es una ilustración de nomograma que predice el tiempo hasta la recidiva.

La figura 4 es una ilustración de la recidiva no solapante predicha por el nomograma y una firma de CCG.

La figura 5 es una ilustración del tiempo hasta la recidiva para varias poblaciones de pacientes definido por el nomograma y/o estado de CCG.

La figura 6 es una ilustración de un ejemplo de un sistema útil en determinados aspectos y realizaciones de la divulgación.

La figura 7 es un diagrama de flujo que ilustra un ejemplo de un método implementado por ordenador de la divulgación.

La figura 8 es un gráfico de dispersión que compara parámetros clínicos y la puntuación de CCG como predictores de recidiva a partir del ejemplo 5.

La figura 9 ilustra, a partir del ejemplo 5, el umbral de CCG derivado del análisis de la cohorte de entrenamiento para el conjunto de datos de validación, subdividiendo eficazmente la puntuación de firma de CCG a pacientes identificados con bajo riesgo usando parámetros clínicos en pacientes con tasas de recidiva muy bajas y un mayor riesgo de recidiva.

La figura 10 ilustra la tasa de recidiva predicha frente a la puntuación de CCG para pacientes en la cohorte de validación del ejemplo 5.

La figura 11 ilustra la tasa de recidiva predicha frente a la puntuación de CCG para pacientes en la cohorte de validación del ejemplo 5.

La figura 12 ilustra la distribución de puntuación de riesgo clínico en 443 pacientes estudiados en el ejemplo 5. La línea vertical oscura representa el umbral elegido mediante medios de KM para dividir a pacientes con riesgo bajo y alto y usado a lo largo de este estudio.

La figura 13 ilustra la correlación entre la puntuación de CCP y la supervivencia en cáncer de cerebro.

La figura 14 ilustra la correlación entre la puntuación de CCP y la supervivencia en cáncer de vejiga.

La figura 15 ilustra la correlación entre la puntuación de CCP y la supervivencia en cáncer de mama.

La figura 16 ilustra la correlación entre la puntuación de CCP y la supervivencia en cáncer de pulmón.

La figura 17 es una ilustración de la potencia predictiva a lo largo del nomograma para paneles de CCG de diferentes tamaños.

La figura 18 muestra la distribución de casos y controles mediante la puntuación combinada en el ejemplo 7.

La figura 19 muestra la distribución de valores de p observados en comparación con los esperados (al considerar que no existe asociación) en el ejemplo 7.

La figura 20 muestra los perfiles de expresión de ARN subyacentes a los valores de p significativos de seis de los genes destacados en el ejemplo 7.

La figura 21 muestra cómo los niveles de expresión de ARN de *KLK3* predicen el estado de casos y controles independientemente de Gleason en el ejemplo 7.

La figura 22 muestra los perfiles de expresión de ARN subyacentes a los valores de p significativos de seis de los genes destacados en el ejemplo 7.

La figura 23 muestra, gráficamente, pronósticos a modo de ejemplo (por ejemplo, riesgo proporcional de desenlace clínico negativo [recidiva bioquímica o BCR]) correspondientes a diversas puntuaciones combinadas que combinan la puntuación de CCP y variables clínicas según la ecuación:

Puntuación combinada = 0,38*(variable(s) clínica(s)) + 0,57*(puntuación de CPP).

La figura 24 muestra, gráficamente, pronósticos a modo de ejemplo (por ejemplo, riesgo proporcional de desenlace clínico negativo [muerte específica de cáncer de próstata o simplemente “muerte”]) correspondientes a diversas puntuaciones combinadas que combinan la puntuación de CCP y variables clínicas según la ecuación:

Puntuación combinada = 0,39*(variable(s) clínica(s) (por ejemplo, CAPRA)) + 0,57*(puntuación de CCP).

La figura 25 muestra cómo los genes sometidos a prueba en el ejemplo 9 predicen el desenlace tanto independientemente como después de ajustar la puntuación de CCP.

Descripción detallada

I. Determinación de la expresión de genes del ciclo celular

La presente divulgación se basa, en parte, en el descubrimiento de que los genes cuya expresión sigue estrechamente el ciclo celular (“genes del ciclo celular” o “CCG”) son genes particularmente potentes para clasificar cánceres seleccionados incluyendo cáncer de próstata, cáncer de pulmón, cáncer de vejiga, cáncer de cerebro y cáncer de mama, pero no otros tipos de cáncer.

“Gen del ciclo celular” y “CCG” en el presente documento se refieren a un gen cuyo nivel de expresión sigue estrechamente la progresión de la célula a través del ciclo celular. Véase, por ejemplo, Whitfield *et al*., MOL. BIOL. CELL (2002) 13:1977-2000. El término “progresión del ciclo celular” o “CCP” también se usará en esta solicitud y generalmente será intercambiable con CCG (es decir, un gen de CCP es un CCG; una puntuación de CCP es una puntuación de CCG). Más específicamente, los CCG muestran aumentos y disminuciones periódicos en la expresión que coinciden con determinadas fases del ciclo celular, por ejemplo, *STK15* y *PLK* muestran una expresión máxima en G2/M. *Id*. A menudo, los CCG tienen una clara y reconocida función relacionada con el ciclo celular, por ejemplo, en la síntesis o reparación de ADN, en la condensación de cromosomas, en la división celular, etc. Sin embargo, algunos CCG tienen niveles de expresión que siguen el ciclo celular sin tener ningún papel directo obvio en el ciclo celular, por ejemplo, *UBE2S* codifica para una enzima que se conjuga con ubiquitina, aunque su expresión sigue estrechamente el ciclo celular. Por tanto, no es necesario que un CCG según la presente divulgación tenga un papel reconocido en el ciclo celular. Se enumeran CCG a modo de ejemplos en las tablas 1, 2, 7-11, 13, 14 y A. Puede encontrarse una discusión más completa de los CCG en la solicitud internacional n.º PCT/US2010/020397 (n.º de pub. WO/2010/080933) (véase, por ejemplo, la tabla 1 en el documento WO/2010/080933), la solicitud de modelo de utilidad estadounidense con n.º de serie 13/177.887 (n.º de pub. US20120041274), la solicitud internacional n.º PCT/US2011/043228 (n.º de pub. WO/2012/006447) y la solicitud de modelo de utilidad estadounidense con n.º de serie 13/178.380 (n.º de pub. US20120053253).

Si un gen particular es un CCG puede determinarse mediante cualquier técnica conocida en la técnica, incluyendo aquellas enseñadas en Whitfield *et al*., MOL. BIOL. CELL (2002) 13:1977-2000; Whitfield *et al*., MOL. CELL. BIOL. (2000) 20:4188-4198; el documento WO/2010/080933 (¶ [0039]). Todos los CCG en la tabla 2 a continuación forman un panel de CCG (“panel A”) útil en la divulgación. Tal como se mostrará en detalle a lo largo de este documento, en la divulgación también pueden usarse CCG individuales (por ejemplo, CCG en la tabla 2) y subconjuntos de estos genes.

Tabla 2

Símbolo de gen	ID de gen de Entrez	ID de ensayo de ABI	N.os de registro de RefSeq
*APOBEC3B**	9582	Hs00358981_m1	NM_004900.3
*ASF1B**	55723	Hs00216780_m1	NM_018154.2
*ASPM**	259266	Hs00411505_m1	NM_018136.4
*ATAD2**	29028	Hs00204205_m1	NM_014109.3
*BIRC5**	332	Hs00153353_m1; Hs03043576_m1	NM_001012271.1; NM_001012270.1; NM_001168.2
*BLM**	641	Hs00172060_m1	NM_000057.2
BUB1	699	Hs00177821_m1	NM_004336.3

*BUB1B**	701	Hs01084828_m1	NM_001211.5
*C12orf48**	55010	Hs00215575_m1	NM_017915.2
*C18orf24**	220134	Hs00536843_m1	NM_145060.3; NM_001039535.2
*C1orf135**	79000	Hs00225211_m1	NM_024037.1
*C21orf45**	54069	Hs00219050_m1	NM_018944.2
*CCDC99**	54908	Hs00215019_m1	NM_017785.4
*CCNA2**	890	Hs00153138_m1	NM_001237.3
*CCNB1**	891	Hs00259126_m1	NM_031966.2
*CCNB2**	9133	Hs00270424_m1	NM_004701.2
*CCNE1**	898	Hs01026536_m1	NM_001238.1; NM_057182.1
*CDC2**	983	Hs00364293_m1	NM_033379.3; NM_001130829.1; NM_001786.3
*CDC20**	991	Hs03004916_g1	NM_001255.2
*CDC45L**	8318	Hs00185895_m1	NM_003504.3
*CDC6**	990	Hs00154374_m1	NM_001254.3
*CDCA3**	83461	Hs00229905_m1	NM_031299.4
*CDCA8**	55143	Hs00983655_m1	NM_018101.2
*CDKN3**	1033	Hs00193192_m1	NM_001130851.1; NM_005192.3
*CDT1**	81620	Hs00368864_m1	NM_030928.3
CENPA	1058	Hs00156455_m1	NM_001042426.1; NM_001809.3
*CENPE**	1062	Hs00156507_m1	NM_001813.2
*CENPF**	1063	Hs00193201_m1	NM_016343.3
*CENPI**	2491	Hs00198791_m1	NM_006733.2
*CENPM**	79019	Hs00608780_m1	NM_024053.3
*CENPN**	55839	Hs00218401_m1	NM_018455.4; NM_001100624.1; NM_001100625.1
*CEP55**	55165	Hs00216688_m1	NM_018131.4; NM_001127182.1
*CHEK1**	1111	Hs00967506_m1	NM_001114121.1; NM_001114122.1; NM_001274.4

*CKAP2**	26586	Hs00217068_m1	NM_018204.3; NM_001098525.1
*CKS1B**	1163	Hs01029137_g1	NM_001826.2
*CKS2**	1164	Hs01048812_g1	NM_001827.1
*CTPS**	1503	Hs01041851_m1	NM_001905.2
*CTSL2**	1515	Hs00952036_m1	NM_001333.2
*DBF4**	10926	Hs00272696_m1	NM_006716.3
*DDX39**	10212	Hs00271794_m1	NM_005804.2
*DLGAP5/DLG7**	9787	Hs00207323_m1	NM_014750.3
*DONSON**	29980	Hs00375083_m1	NM_017613.2
*DSN1**	79980	Hs00227760_m1	NM_024918.2
*DTL**	51514	Hs00978565_m1	NM_016448.2
*E2F8**	79733	Hs00226635_m1	NM_024680.2
*ECT2**	1894	Hs00216455_m1	NM_018098.4
*ESPL1**	9700	Hs00202246_m1	NM_012291.4
*EXO1**	9156	Hs00243513_m1	NM_130398.2; NM_003686.3; NM_006027.3
*EZH2**	2146	Hs00544830_m1	NM_152998.1; NM_004456.3
*FANCI**	55215	Hs00289551_m1	NM_018193.2; NM_001113378.1
*FBXO5**	26271	Hs03070834_m1	NM_001142522.1; NM_012177.3
*FOXM1**	2305	Hs01073586_m1	NM_202003.1; NM_202002.1; NM_021953.2
*GINS1**	9837	Hs00221421_m1	NM_021067.3
*GMPS**	8833	Hs00269500_m1	NM_003875.2
*GPSM2**	29899	Hs00203271_m1	NM_013296.4
*GTSE1**	51512	Hs00212681_m1	NM_016426.5
*H2AFX**	3014	Hs00266783_s1	NM_002105.2
*HMMR**	3161	Hs00234864_m1	NM_001142556.1; NM_001142557.1; NM_012484.2; NM_012485.2

*HN1**	51155	Hs00602957_m1	NM_001002033.1; NM_001002032.1; NM_016185.2
*KIAA0101**	9768	Hs00207134_m1	NM_014736.4
*KIF11**	3832	Hs00189698_m1	NM_004523.3
*KIF15**	56992	Hs00173349_m1	NM_020242.2
*KIF18A**	81930	Hs01015428_m1	NM_031217.3
*KIF20A**	10112	Hs00993573_m1	NM_005733.2
*KIF20B/MPHOSPH1**	9585	Hs01027505_m1	NM_016195.2
*KIF23**	9493	Hs00370852_m1	NM_138555.1; NM_004856.4
*KIF2C**	11004	Hs00199232_m1	NM_006845.3
*KIF4A**	24137	Hs01020169_m1	NM_012310.3
*KIFC1**	3833	Hs00954801_m1	NM_002263.3
KPNA2	3838	Hs00818252_g1	NM_002266.2
*LMNB2**	84823	Hs00383326_m1	NM_032737.2
MAD2L1	4085	Hs01554513_g1	NM_002358.3
*MCAM**	4162	Hs00174838_m1	NM_006500.2
*MCM10**	55388	Hs00960349_m1	NM_018518.3; NM_182751.1
*MCM2**	4171	Hs00170472_m1	NM_004526.2
*MCM4**	4173	Hs00381539_m1	NM_005914.2; NM_182746.1
*MCM6**	4175	Hs00195504_m1	NM_005915.4
*MCM7**	4176	Hs01097212_m1	NM_005916.3; NM_182776.1
MELK	9833	Hs00207681_m1	NM_014791.2
*MKI67**	4288	Hs00606991_m1	NM_002417.3
*MYBL2**	4605	Hs00231158_m1	NM_002466.2
*NCAPD2**	9918	Hs00274505_m1	NM_014865.3
*NCAPG**	64151	Hs00254617_m1	NM_022346.3
*NCAPG2**	54892	Hs00375141_m1	NM_017760.5
*NCAPH**	23397	Hs01010752_m1	NM_015341.3
*NDC80**	10403	Hs00196101_m1	NM_006101.2
*NEK2**	4751	Hs00601227_mH	NM_002497.2

*NUSAP1**	51203	Hs01006195_m1	NM_018454.6; NM_001129897.1; NM_016359.3
*OIP5**	11339	Hs00299079_m1	NM_007280.1
*ORC6L**	23594	Hs00204876_m1	NM_014321.2
*PAICS**	10606	Hs00272390_m1	NM_001079524.1; NM_001079525.1; NM_006452.3
*PBK**	55872	Hs00218544_m1	NM_018492.2
*PCNA**	5111	Hs00427214_g1	NM_182649.1; NM_002592.2
*PDSS1**	23590	Hs00372008_m1	NM_014317.3
*PLK1**	5347	Hs00153444_m1	NM_005030.3
*PLK4**	10733	Hs00179514_m1	NM_014264.3
*POLE2**	5427	Hs00160277_m1	NM_002692.2
*PRC1**	9055	Hs00187740_m1	NM_199413.1; NM_199414.1; NM_003981.2
*PSMA7**	5688	Hs00895424_m1	NM_002792.2
*PSRC1**	84722	Hs00364137_m1	NM_032636.6; NM_001005290.2; NM_001032290.1; NM_001032291.1
*PTTG1**	9232	Hs00851754_u1	NM_004219.2
*RACGAP1**	29127	Hs00374747_m1	NM_013277.3
*RAD51**	5888	Hs00153418_m1	NM_133487.2; NM_002875.3
*RAD51AP1**	10635	Hs01548891_m1	NM_001130862.1; NM_006479.4
*RAD54B**	25788	Hs00610716_m1	NM_012415.2
*RAD54L**	8438	Hs00269177_m1	NM_001142548.1; NM_003579.3

*RFC2**	5982	Hs00945948_m1	NM_181471.1; NM_002914.3
*RFC4**	5984	Hs00427469_m1	NM_181573.2; NM_002916.3
*RFC5**	5985	Hs00738859_m1	NM_181578.2; NM_001130112.1; NM_001130113.1; NM_007370.4
*RNASEH2A**	10535	Hs00197370_m1	NM_006397.2
*RRM2**	6241	Hs00357247_g1	NM_001034.2
*SHCBP1**	79801	Hs00226915_m1	NM_024745.4
*SMC2**	10592	Hs00197593_m1	NM_001042550.1; NM_001042551.1; NM_006444.2
*SPAG5**	10615	Hs00197708_m1	NM_006461.3
*SPC25**	57405	Hs00221100_m1	NM_020675.3
*STIL**	6491	Hs00161700_m1	NM_001048166.1; NM_003035.2
*STMN1**	3925	Hs00606370_m1; Hs01033129_m1	NM_005563.3; NM_203399.1
*TACC3**	10460	Hs00170751_m1	NM_006342.1
*TIMELESS**	8914	Hs01086966_m1	NM_003920.2
*TK1**	7083	Hs01062125_m1	NM_003258.4
*TOP2A**	7153	Hs00172214_m1	NM_001067.2
*TPX2**	22974	Hs00201616_m1	NM_012112.4
*TRIP13**	9319	Hs01020073_m1	NM_004237.2
*TTK**	7272	Hs00177412_m1	NM_003318.3
*TUBA1C**	84790	Hs00733770_m1	NM_032704.3
*TYMS**	7298	Hs00426591_m1	NM_001071.2
UBE2C	11065	Hs00964100_g1	NM_181799.1; NM_181800.1; NM_181801.1; NM_181802.1; NM_181803.1; NM_007019.2

UBE2S	27338	Hs00819350_m1	NM_014501.2
*VRK1**	7443	Hs00177470_m1	NM_003384.2
*ZWILCH**	55055	Hs01555249_m1	NM_017975.3; NR_003105.1
*ZWINT**	11130	Hs00199952_m1	NM_032997.2; NM_001005413.1; NM_007057.3

* Subconjunto de 124 genes de CCG útiles en la divulgación (“panel B”). ID de ensayo ABI significa el número de ID de catálogo para el ensayo de expresión de genes disponible comercialmente de Applied Biosystems Inc. (Foster City, CA) para el gen particular.

Se proporciona, pero no forma parte de la invención, un método para determinar la expresión de genes en una muestra tumoral de un paciente (por ejemplo, uno que se ha identificado que padece cáncer de próstata, cáncer de pulmón, cáncer de vejiga o cáncer de cerebro). Generalmente, el método incluye al menos las siguientes etapas: (1) obtener una muestra tumoral de un paciente (por ejemplo, uno que se ha identificado que padece cáncer de próstata, cáncer de pulmón, cáncer de vejiga o cáncer de cerebro); (2) determinar la expresión de un panel de genes en la muestra tumoral que incluye al menos 2, 4, 6, 8 ó 10 genes del ciclo celular; y (3) proporcionar un valor de prueba mediante (a) ponderar la expresión determinada de cada uno de una pluralidad de genes de prueba seleccionados de dicho panel de genes con un coeficiente predefinido, y (b) combinar la expresión ponderada para proporcionar dicho valor de prueba, en el que al menos el 20%, el 50%, al menos el 75% o al menos el 90% de dicha pluralidad de genes de prueba son genes del ciclo celular. En algunas realizaciones, los genes de prueba se ponderan de tal manera que los genes del ciclo celular se ponderan para contribuir al menos el 50%, al menos el 55%, al menos el 60%, al menos el 65%, al menos el 75%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, al menos el 99% o el 100% del valor de prueba. En algunas realizaciones, el 20%, el 25%, el 30%, el 35%, el 40%, el 45%, el 50%, el 55%, el 60%, el 65%, el 75%, el 80%, el 85%, el 90%, el 95% o al menos el 99% o el 100% de la pluralidad de genes de prueba son genes del ciclo celular.

La expresión de genes puede determinarse o bien a nivel de ARN (es decir, ARNm o ARN no codificante (ARNnc)) (por ejemplo, miARN, ARNt, ARNr, ARNpn, ARNip y ARNpi) o bien a nivel de proteína. Medir la expresión de genes a nivel de ARNm incluye medir los niveles de ADNc correspondiente al ARNm. Los niveles de proteínas en una muestra tumoral pueden determinarse mediante cualquier técnica conocida en la técnica, por ejemplo, HPLC, espectrometría de masas o usando anticuerpos específicos de proteínas seleccionadas (por ejemplo, IHC, ELISA, etc.).

En una realización preferida, la cantidad de ARN transcrito a partir del panel de genes que incluye genes de prueba se mide en la muestra tumoral. Además, también se mide la cantidad de ARN de uno o más genes de mantenimiento en la muestra tumoral, y se usa para normalizar o calibrar la expresión de los genes de prueba. Los términos “genes de normalización” y “genes de mantenimiento” se definen a continuación en el presente documento.

Se proporciona, pero no forma parte de la invención, una “pluralidad de genes de prueba”, la pluralidad de genes de prueba puede incluir al menos 2, 3 ó 4 genes del ciclo celular, que constituyen al menos el 50%, el 75% o el 80% de la pluralidad de genes de prueba, y preferiblemente el 100% de la pluralidad de genes de prueba.

La pluralidad de genes de prueba incluye al menos 5, 6, 7 o al menos 8 genes del ciclo celular, que constituyen al menos el 20%, el 25%, el 30%, el 40%, el 50%, el 60%, el 70%, el 75%, el 80% o el 90% de la pluralidad de genes de prueba, y preferiblemente el 100% de la pluralidad de genes de prueba. Tal como resultará claro a partir del contexto de este documento, un panel de genes es una pluralidad de genes. Normalmente estos genes se someten a ensayo juntos en una o más muestras de un paciente.

La pluralidad de genes de prueba incluye al menos 8, 10, 12, 15, 20, 25 ó 30 genes del ciclo celular, que constituyen al menos el 20%, el 25%, el 30%, el 40%, el 50%, el 60%, el 70%, el 75%, el 80% o el 90% de la pluralidad de genes de prueba, y preferiblemente el 100% de la pluralidad de genes de prueba.

Tal como resultará evidente para un experto en la técnica al tanto de la presente divulgación y la divulgación en el

presente documento, “muestra tumoral” significa cualquier muestra biológica que contiene una o más células tumorales, o uno o más ARN o proteínas derivados de tumor, y obtenida de un paciente con cáncer. Por ejemplo, una muestra tisular obtenida de un tejido tumoral de un paciente con cáncer es una muestra tumoral útil en la presente divulgación. La muestra tisular puede ser una muestra FFPE, o muestra congelada recién preparada, y preferiblemente contener en gran medida células tumorales. Una única célula maligna de un tumor de paciente con cáncer también es una muestra tumoral útil. Una célula maligna de este tipo puede obtenerse directamente del tumor del paciente, o purificarse a partir del líquido corporal del paciente (por ejemplo, sangre, orina). Por tanto, un líquido corporal tal como sangre, orina, esputo y saliva que contiene una o células tumorales, o ARN o proteínas derivados de tumor, también pueden ser útiles como muestra tumoral para los propósitos de poner en práctica la presente divulgación.

Los expertos en la técnica están familiarizados con diversas técnicas para determinar el estado de un gen o una proteína en una muestra tisular o celular, incluyendo, pero sin limitarse a, análisis por microalineamiento (por ejemplo, para someter a ensayo la expresión de ARNm o microARN, el número de copias, etc.), PCR™ cuantitativa en tiempo real (“qRT-PCR™”, por ejemplo, TaqMan™), inmunoanálisis (por ejemplo, ELISA, inmunohistoquímica), etc. Puede usarse el nivel de actividad de un polipéptido codificado por un gen de la misma manera que el nivel de expresión del gen o polipéptido. A menudo, mayores niveles de actividad indican mayores niveles de expresión, mientras que menores niveles de actividad indican menores niveles de expresión. Por tanto, en algunas realizaciones, la divulgación proporciona cualquiera de los métodos comentados anteriormente, en los que se determina el nivel de actividad de un polipéptido codificado por el CCG en lugar del, o además del, nivel de expresión del CCG. Los expertos en la técnica están familiarizados con técnicas para medir la actividad de tales diversas proteínas, incluyendo aquellas codificadas por los genes enumerados en las tablas 1 y 2. Los métodos de la divulgación pueden ponerse en práctica independientemente de la técnica particular usada.

En realizaciones preferidas, también se obtiene la expresión de uno o más genes de normalización (a menudo denominados “de mantenimiento”) para su uso en normalizar la expresión de genes de prueba. Tal como se usa en el presente documento, “genes de normalización” se refiere a los genes cuya expresión se usa para calibrar o normalizar la expresión medida del gen de interés (por ejemplo, genes de prueba). De manera importante, la expresión de genes de normalización debe ser independiente del desenlace/pronóstico del cáncer, y la expresión de los genes de normalización es muy similar entre todas las muestras tumorales. La normalización garantiza una comparación precisa de la expresión de un gen de prueba entre diferentes muestras. Para este propósito, pueden usarse genes de mantenimiento conocidos en la técnica. Los genes de mantenimiento se conocen bien en la técnica, incluyendo los ejemplos, pero no se limitan a, GUSB (glucuronidasa, beta), HMBS (hidroximetilbilano sintasa), SDHA (complejo de succinato deshidrogenasa, subunidad A, flavoproteína), UBC (ubiquitina C) y YWHAZ (proteína de activación de tirosina 3-monooxigenasa/triptófano 5-monooxigenasa, polipéptido zeta). Pueden usarse uno o más genes de mantenimiento. Preferiblemente, se usan al menos 2, 5, 10 ó 15 genes de mantenimiento para proporcionar un conjunto de genes de normalización combinados. La cantidad de expresión de genes de tales genes de normalización puede promediarse, combinarse conjuntamente mediante adiciones directas o mediante un algoritmo definido. Algunos ejemplos de genes de mantenimiento particularmente útiles para su uso en los métodos y las composiciones de la divulgación incluyen aquellos enumerados en la tabla 3 a continuación.

Tabla 3

Símbolo de gen	ID de gen de Entrez	ID de ensayo de Applied Biosystems	N.os de registro de RefSeq
*CLTC**	1213	Hs00191535_m1	NM_004859.3
GUSB	2990	Hs99999908_m1	NM_000181.2
HMBS	3145	Hs00609297_m1	NM_000190.3
*MMADHC**	27249	Hs00739517_g1	NM_015702.2
*MRFAP1**	93621	Hs00738144_g1	NM_033296.1
*PPP2CA**	5515	Hs00427259_m1	NM_002715.2

*PSMA1**	5682	Hs00267631_m1	
*PSMC1**	5700	Hs02386942_g1	NM_002802.2
*RPL13A**	23521	Hs03043885_g1	NM_012423.2
*RPL37**	6167	Hs02340038_g1	NM_000997.4
*RPL38**	6169	Hs00605263_g1	NM_000999.3
*RPL4**	6124	Hs03044647_g1	NM_000968.2
*RPL8**	6132	Hs00361285_g1	NM_033301.1; NM_000973.3
*RPS29**	6235	Hs03004310_g1	NM_001030001.1; NM_001032.3
SDHA	6389	Hs00188166_m1	NM_004168.2
*SLC25A3**	6515	Hs00358082_m1	NM_213611.1; NM_002635.2; NM_005888.2
*TXNL1**	9352	Hs00355488_m1	NR_024546.1; NM_004786.2
*UBA52**	7311	Hs03004332_g1	NM_001033930.1; NM_003333.3
UBC	7316	Hs00824723_m1	NM_021009.4
YWHAZ	7534	Hs00237047_m1	NM_003406.3

* Subconjunto de genes de mantenimiento usados, por ejemplo, en el ejemplo 5.

En el caso de medir los niveles de ARN para los genes, un enfoque conveniente y sensible es el ensayo de PCR™ cuantitativa en tiempo real (qPCR), después de una reacción de transcripción inversa. Normalmente, se determina un umbral de ciclo (C_t) para cada gen de prueba y cada gen de normalización, es decir, el número de ciclos en los que la fluorescencia de una reacción de qPCR por encima del fondo es detectable.

La expresión global del uno o más genes de normalización puede representarse por un “valor de normalización” que puede generarse combinando la expresión de todos los genes de normalización, o bien ponderados igualmente (adición directa o promediado) o bien mediante coeficientes predefinidos diferentes. Por ejemplo, de la manera más sencilla, el valor de normalización C_{tH} puede ser el umbral de ciclo (C_t) de un único gen de normalización, o un promedio de los valores de C_t de 2 o más, preferiblemente 10 o más, o 15 o más genes de normalización, en cuyo caso, el coeficiente predefinido es 1/N, donde N es el número total de genes de normalización usados. Por tanto, $C_{tH} = (C_{tH1} + C_{tH2} + \cdots C_{tHn})/N$. Tal como resultará evidente para los expertos en la técnica, en función de los genes de normalización usados, y la ponderación deseada que va a darse a cada gen de normalización, pueden darse cualesquiera coeficientes (desde 0/N hasta N/N) a los genes de normalización en la ponderación de la expresión de tales genes de normalización. Es decir, $C_{tH} = xC_{tH1} + yC_{tH2} + \cdots zC_{tHn}$, donde $x + y + \cdots + z = 1$.

Tal como se comentó anteriormente, los métodos de la divulgación generalmente implican determinar el nivel de expresión de un panel de CCG. Con técnicas de alto rendimiento modernas, a menudo es posible determinar el nivel de expresión de decenas, cientos o miles de genes. De hecho, es posible determinar el nivel de expresión de todo el transcriptoma (es decir, cada secuencia transcrita en el genoma). Una vez que se ha realizado un ensayo global de este tipo, luego pueden analizarse de manera informática uno o más subconjuntos de transcritos (es decir, paneles o, tal como se usa a menudo en el presente documento, pluralidades de genes de prueba). Después de medir la expresión de cientos o miles de transcritos en una muestra, por ejemplo, puede analizarse (por ejemplo, de manera informática) la expresión de un panel o una pluralidad de genes de prueba que comprende principalmente CCG según la presente divulgación combinando los valores de nivel de expresión de los genes de prueba individuales para obtener un valor de prueba.

Tal como resultará evidente para un experto en la técnica, el valor de prueba proporcionado en la presente divulgación representa el nivel de expresión global de la pluralidad de genes de prueba que se compone sustancialmente de genes del ciclo celular. En una realización, para proporcionar un valor de prueba en los métodos de la divulgación, la expresión normalizada para un gen de prueba puede obtenerse normalizando el C_t medido para el gen de prueba frente al C_{tH}, es decir, $\Delta C_{t1} = (C_{t1} - C_{tH})$. Por tanto, el valor de prueba que representa la expresión global de la pluralidad de genes de prueba puede proporcionarse combinando la expresión normalizada de todos los genes de prueba, o bien mediante adición directa o promediado (es decir, ponderados igualmente) o bien mediante un coeficiente predefinido diferente. Por ejemplo, el enfoque más sencillo es promediar la expresión normalizada de todos los genes de prueba: valor de prueba = $(\Delta C_{t1} + \Delta C_{t2} + \cdots + \Delta C_{tn})/n$. Tal como resultará evidente para los expertos en la técnica, en función de los genes de prueba usados, también puede darse una ponderación diferente a genes de prueba diferentes en la presente divulgación. En cada caso en el que este documento divulga el uso de la expresión de una pluralidad de genes (por ejemplo, “determinar [en una muestra tumoral del paciente] la expresión de una pluralidad de genes de prueba” o “correlacionar la expresión aumentada de dicha pluralidad de genes de prueba con una probabilidad aumentada de recidiva”), esto incluye en algunas realizaciones usar un valor de prueba

que representa, correspondiente a o derivado o calculado de la expresión global de esta pluralidad de genes (por ejemplo, “determinar [en una muestra tumoral del paciente] un valor de prueba que representa la expresión de una pluralidad de genes de prueba” o “correlacionar un valor de prueba aumentado [o un valor de prueba por encima de algún valor de referencia] (que representa opcionalmente la expresión de dicha pluralidad de genes de prueba) con una probabilidad aumentada de respuesta”).

Se ha determinado que, una vez que se aprecia el fenómeno de CCP notificado en el presente documento, la elección de CCG individuales para un panel de prueba a menudo puede ser algo arbitraria. Dicho de otro modo, se ha hallado que muchos CCG son muy buenos sustitutos entre sí. Por tanto, puede usarse cualquier CCG (o panel de CCG) en las diversas realizaciones de la divulgación. En otras realizaciones de la divulgación, se usan CCG optimizados. Un modo de evaluar si los CCG particulares servirán bien en los métodos y las composiciones de la divulgación es mediante la evaluación de su correlación con la expresión media de los CCG (por ejemplo, todos los CCG conocidos, un conjunto específico de CCG, etc.). Se espera que esos CCG que se correlacionan particularmente bien con la media funcionen bien en los ensayos de la divulgación, por ejemplo, porque estos reducirán el ruido en el ensayo.

Se comparó la expresión de 126 CCG y 47 genes de mantenimiento con la media de CCG y de mantenimiento con el fin de determinar genes preferidos para su uso en algunas realizaciones de la divulgación. Se proporcionan clasificaciones de CCG seleccionados según su correlación con la expresión de CCG media así como su clasificación según el valor predictivo en las tablas 9-11 y 13-14.

Se proporciona, pero no forma parte de la invención, una pluralidad de genes de prueba. La pluralidad de genes de prueba comprende los primeros 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40 o más CCG enumerados en las tablas 9-11 y 13-14. La pluralidad de genes de prueba comprende al menos cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 ó 20 de los siguientes genes: *ASPM*, *BIRC5*, *BUB1B*, *CCNB2*, *CDC2*, *CDC20*, *CDCA8*, *CDKN3*, *CENPF*, *DLGAP5*, *FOXM1*, *KIAA0101*, *KIF11*, *KIF2C*, *KIF4A*, *MCM10*, *NUSAP1*, *PRC1*, *RACGAP1* y *TPX2*. La pluralidad de genes de prueba comprende al menos cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende cualesquiera dos, tres, cuatro, cinco, seis, siete, ocho, nueve o diez o la totalidad de los números de genes 1 y 2, 1 a 3, 1 a 4, 1 a 5, 1 a 6, 1 a 7, 1 a 8, 1 a 9 ó 1 a 10 de cualesquiera de las tablas 9-11 y 13-14. La pluralidad de genes de prueba comprende al menos cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende cualesquiera uno, dos, tres, cuatro, cinco, seis, siete, ocho o nueve o la totalidad de los números de genes 2 y 3, 2 a 4, 2 a 5, 2 a 6, 2 a 7, 2 a 8, 2 a 9 ó 2 a 10 de cualesquiera de las tablas 9-11 y 13-14. La pluralidad de genes de prueba comprende al menos cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende cualesquiera uno, dos, tres, cuatro, cinco, seis, siete u ocho o la totalidad de los números de genes 3 y 4, 3 a 5, 3 a 6, 3 a 7, 3 a 8, 3 a 9 ó 3 a 10 de cualesquiera de las tablas 9-11 y 13-14. La pluralidad de genes de prueba comprende al menos cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende cualesquiera uno, dos, tres, cuatro, cinco, seis o siete o la totalidad de los números de genes 4 y 5, 4 a 6, 4 a 7, 4 a 8, 4 a 9 ó 4 a 10 de cualesquiera de las tablas 9-11 y 13-14. La pluralidad de genes de prueba comprende al menos cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende cualesquiera uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, 10, 11, 12, 13, 14 ó 15 o la totalidad de los números de genes 1 y 2, 1 a 3, 1 a 4, 1 a 5, 1 a 6, 1 a 7, 1 a 8, 1 a 9, 1 a 10, 1 a 11, 1 a 12, 1 a 13, 1 a 14 ó 1 a 15 de cualesquiera de las tablas 9-11 y 13-14.

II. Pronóstico de cáncer

Se ha descubierto sorprendentemente que, en cánceres seleccionados tales como cáncer de próstata, cáncer de pulmón, cáncer de vejiga y cáncer de cerebro, pero no otros cánceres incluyendo determinado cáncer de colon, la expresión de genes del ciclo celular en células tumorales puede predecir con precisión el grado de agresión del cáncer y el riesgo de recidiva después del tratamiento (por ejemplo, extirpación quirúrgica de tejido canceroso, quimioterapia y radioterapia, etc.). Por tanto, el método anteriormente descrito para determinar expresión de genes del ciclo celular puede aplicarse en el pronóstico y tratamiento de tales cánceres. El tratamiento de cáncer no está abarcado por las reivindicaciones.

Generalmente, se proporciona un método para pronosticar cáncer seleccionado de cáncer de próstata, cáncer de pulmón, cáncer de vejiga o cáncer de cerebro, que comprende determinar, en una muestra tumoral de un paciente al que se le ha diagnosticado cáncer de próstata, cáncer de pulmón, cáncer de vejiga o cáncer de cerebro, la expresión de al menos 2, 4, 5, 6, 7 o al menos 8, 9, 10 ó 12 genes del ciclo celular, en el que una alta expresión (o expresión aumentada o sobreexpresión) de los al menos 4 genes del ciclo celular indica un mal pronóstico o una probabilidad aumentada de recidiva de cáncer en el paciente. La expresión puede determinarse según el método descrito anteriormente. En algunas realizaciones, el método comprende al menos una de las siguientes etapas: (a) correlacionar la alta expresión (o expresión aumentada o sobreexpresión) de los genes del ciclo celular con un mal pronóstico o una probabilidad aumentada de recidiva de cáncer en el paciente; (b) concluir que el paciente tiene

un mal pronóstico o una probabilidad aumentada de recidiva de cáncer basándose al menos en parte en la alta expresión (o expresión aumentada o sobreexpresión) de los genes del ciclo celular; o (c) comunicar que el paciente tiene un mal pronóstico o una probabilidad aumentada de recidiva de cáncer basándose al menos en parte en la alta expresión (o expresión aumentada o sobreexpresión) de los genes del ciclo celular.

En cada realización descrita en este documento que implica la correlación de un resultado de análisis o ensayo particular (por ejemplo, alta expresión de CCP, valor de prueba que incorpora una expresión de CCP mayor que algún valor de referencia, etc.) con cierta probabilidad (por ejemplo, aumentada, no aumentada, disminuida, etc.) de algún acontecimiento o desenlace clínico (por ejemplo, recidiva, progresión, muerte específica de cáncer, etc.), tal correlación puede comprender asignar un riesgo o una probabilidad de que se produzca el acontecimiento o desenlace clínico basándose al menos en parte en el resultado de análisis o ensayo particular. En algunas realizaciones, tal riesgo es una probabilidad en porcentaje de que se produzca el acontecimiento o desenlace. En algunas realizaciones, el paciente se asigna a un grupo de riesgo (por ejemplo, riesgo bajo, riesgo intermedio, riesgo alto, etc.). En algunas realizaciones, “riesgo bajo” es cualquier probabilidad en porcentaje por debajo del 5%, el 10%, el 15%, el 20%, el 25%, el 30%, el 35%, el 40%, el 45% o el 50%. En algunas realizaciones, “riesgo intermedio” es cualquier probabilidad en porcentaje por encima del 5%, el 10%, el 15%, el 20%, el 25%, el 30%, el 35%, el 40%, el 45% o el 50% y por debajo del 15%, el 20%, el 25%, el 30%, el 35%, el 40%, el 45%, el 50%, el 55%, el 60%, el 65%, el 70% o el 75%. En algunas realizaciones, “riesgo alto” es cualquier probabilidad en porcentaje por encima del 25%, el 30%, el 35%, el 40%, el 45%, el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, el 90%, el 95% o el 99%.

Tal como se usa en el presente documento, “comunicar” un fragmento de información particular significa hacer saber tal información a otra persona o transferir tal información a una cosa (por ejemplo, un ordenador). En algunos métodos de la divulgación, se comunica el pronóstico o riesgo de recidiva de un paciente. En algunas realizaciones, se comunica la información que se usa para llegar a un pronóstico o una predicción de riesgo de este tipo (por ejemplo, niveles de expresión de un panel de biomarcadores que comprende una pluralidad de CCG, factores clínicos o patológicos, etc.). Esta comunicación puede ser auditiva (por ejemplo, verbal), visual (por ejemplo, escrita), electrónica (por ejemplo, datos transferidos de un sistema informatizado a otro), etc. En algunas realizaciones, comunicar una clasificación de cáncer comprende generar un informe que comunica la clasificación de cáncer. En algunas realizaciones, el informe es un informe en papel, un informe auditivo o un registro electrónico. En algunas realizaciones, el informe se visualiza y/o almacena en un dispositivo informático (por ejemplo, dispositivo manual, ordenador de mesa, dispositivo inteligente, sitio web, etc.). En algunas realizaciones, la clasificación de cáncer se comunica a un médico (por ejemplo, se le proporciona al médico un informe que comunica la clasificación). En algunas realizaciones, la clasificación de cáncer se le comunica a un paciente (por ejemplo, se le proporciona al paciente un informe que comunica la clasificación). Comunicar una clasificación de cáncer también puede lograrse transfiriendo información (por ejemplo, datos) que representa la clasificación a un ordenador de servidor y permitiendo que un intermediario o usuario final acceda a tal información (por ejemplo, viendo la información tal como se visualiza a partir del servidor, descargando la información en forma de uno o más archivos transferidos a partir del servidor al dispositivo del intermediario o usuario final, etc.).

Siempre que una realización de la divulgación comprenda concluir algún hecho (por ejemplo, el pronóstico de un paciente o la probabilidad de recidiva de un paciente), esto puede incluir un programa informático que concluye tal hecho, normalmente después de realizar un algoritmo que aplica información sobre el estado de CCG, estado de *PTEN*, estado de *KLK3* y/o variables clínicas en una muestra de paciente (por ejemplo, tal como se muestra en la figura 7).

Se proporciona, pero no forma parte de la invención, un método de pronóstico que incluye (1) obtener una muestra tumoral de un paciente que se ha identificado que padece cáncer de próstata, cáncer de pulmón, cáncer de vejiga o cáncer de cerebro; (2) determinar la expresión de un panel de genes en la muestra tumoral que incluye al menos 2, 4, 6, 8 ó 10 genes del ciclo celular; y (3) proporcionar un valor de prueba mediante (a) ponderar la expresión determinada de cada uno de una pluralidad de genes de prueba seleccionados del panel de genes con un coeficiente predefinido, y (b) combinar la expresión ponderada para proporcionar dicho valor de prueba, en el que al menos el 20%, el 50%, al menos el 75% o al menos el 90% de dicha pluralidad de genes de prueba son genes del ciclo celular, y en el que la alta expresión (o expresión aumentada o sobreexpresión) de la pluralidad de genes de prueba indica un mal pronóstico o una probabilidad aumentada de recidiva de cáncer. En algunas realizaciones, el método comprende al menos una de las siguientes etapas: (a) correlacionar la alta expresión (o expresión aumentada o sobreexpresión) de la pluralidad de genes de prueba con un mal pronóstico o una probabilidad aumentada de recidiva de cáncer en el paciente; (b) concluir que el paciente tiene un mal pronóstico o una probabilidad aumentada de recidiva de cáncer basándose al menos en parte en la alta expresión (o expresión aumentada o sobreexpresión) de la pluralidad de genes de prueba; o (c) comunicar que el paciente tiene un mal pronóstico o una probabilidad aumentada de recidiva de cáncer basándose al menos en parte en la alta expresión (o expresión aumentada o sobreexpresión) de la pluralidad de genes de prueba.

En algunas realizaciones de la invención reivindicada, los niveles de expresión medidos en una muestra se usan para derivar o calcular un valor o una puntuación. Este valor puede derivarse únicamente a partir de estos niveles de expresión (por ejemplo, una puntuación de CCG) u opcionalmente derivarse a partir de una combinación del

valor/puntuación de expresión con otros componentes (por ejemplo, año de RP, márgenes quirúrgicos, extensión extracapsular, invasión de vesículas seminales, afectación ganglionar, puntuación primaria de Gleason, puntuación secundaria de Gleason o nivel de PSA preoperatorio, etc.) para dar un valor/puntuación más completo. Por tanto, en cada caso en el que una realización de la divulgación descrita en el presente documento implica determinar el estado de un biomarcador (por ejemplo, niveles de expresión de ARN de un CCG, *PTEN* o *KLK3*), las realizaciones relacionadas implican derivar o calcular un valor o una puntuación a partir del estado medido (por ejemplo, puntuación de expresión).

En algunas de tales realizaciones, pueden combinarse múltiples puntuaciones (por ejemplo, CCG, Gleason, PSA, *PTEN*, *KLK3*) en una puntuación más completa. Pueden compararse puntuaciones de prueba combinadas o de componente único (por ejemplo, CCG) para un paciente particular con puntuaciones combinadas o de componente único para poblaciones de referencia tal como se describe a continuación, correlacionándose las diferencias entre las puntuaciones de prueba y de referencia con, o siendo indicativas de, alguna característica clínica. Por tanto, en algunas realizaciones, la divulgación proporciona un método para determinar el pronóstico de un paciente con cáncer que comprende (1) obtener los niveles de expresión medidos de una pluralidad de genes que comprende una pluralidad de CCG en una muestra del paciente, (2) calcular un valor de prueba a partir de estos niveles de expresión medidos, (3) comparar dicho valor de prueba con un valor de referencia calculado a partir de niveles de expresión medidos de la pluralidad de genes en una población de referencia de pacientes, y (4)(a) correlacionar un valor de prueba mayor que el valor de referencia con un mal pronóstico o (4)(b) correlacionar un valor de prueba igual a o menor que el valor de referencia con un buen pronóstico.

En algunas de tales realizaciones, el valor de prueba se calcula promediando la expresión medida de la pluralidad de genes (tal como se comenta a continuación). En algunas realizaciones, el valor de prueba se calcula ponderando cada uno de la pluralidad de genes de un modo particular.

En algunas realizaciones, la pluralidad de CCG se ponderan de tal manera que contribuyen al menos cierta proporción del valor de prueba (por ejemplo, el 10%, el 20%, el 30%, el 40%, el 50%, el 60%, el 70%, el 80%, el 90%, el 95%, el 99%, el 100%). En algunas realizaciones, cada uno de la pluralidad de genes se pondera de tal manera que a todos no se les da una misma ponderación (por ejemplo, *KLK3* ponderado para contribuir más al valor de prueba que uno, algunos o la totalidad de CCG).

En algunas realizaciones, la expresión de CCP se pondera y combina con otros factores en una puntuación combinada (similar al valor de prueba comentado anteriormente). En algunas realizaciones, una puntuación combinada de este tipo se calcula sumando la puntuación de CCP y el/los otro(s) factor(es) linealmente según la siguiente fórmula:

(1) *Puntuación combinada = A*(puntuación de CCP) + B*(uno o más de los otros factores)*

Se apreciará que esta divulgación abarca otros medios de combinación (por ejemplo, multiplicación, logaritmos, exponentes, etc.). En algunas realizaciones, los otros factores son la expresión de otros genes, las características físicas del paciente (por ejemplo, altura, peso, etc.), las características clínicas del paciente (por ejemplo, variables clínicas tal como se comentan a continuación), etc. En algunas realizaciones, pueden combinarse una o más variables clínicas en una puntuación clínica, que luego puede combinarse con la puntuación de CCP para dar lugar a una puntuación combinada de la divulgación.

Por tanto, en algunas realizaciones, la divulgación proporciona un método para determinar el pronóstico de un paciente con cáncer que comprende: (1) obtener los niveles de expresión medidos de una pluralidad de genes que comprende una pluralidad de CCG en una muestra del paciente; (2) obtener una puntuación para el paciente que comprende uno o más de año de RP, márgenes quirúrgicos, extensión extracapsular, invasión de vesículas seminales, afectación ganglionar, puntuación primaria de Gleason, puntuación secundaria de Gleason o nivel de PSA preoperatorio; (3) derivar un valor de prueba combinado a partir de los niveles medidos obtenidos en (1) y la puntuación obtenida en (2); (4) comparar el valor de prueba combinado con un valor de referencia combinado derivado a partir de los niveles de expresión medidos de la pluralidad de genes y una puntuación que comprende uno o más de año de RP, márgenes quirúrgicos, extensión extracapsular, invasión de vesículas seminales, afectación ganglionar, puntuación primaria de Gleason, puntuación secundaria de Gleason o nivel de PSA preoperatorio en una población de referencia de pacientes; y (5)(a) correlacionar un valor de prueba combinado mayor que el valor de referencia combinado con un mal pronóstico o (5)(b) correlacionar un valor de prueba combinado igual a o menor que el valor de referencia combinado con un buen pronóstico.

En algunas realizaciones, la puntuación combinada incluye puntuación de CCP, PSA y puntuación de Gleason. La CCP puede ser una variable numérica continua. Las concentraciones de PSA (por ejemplo, ng/dl), sumando 1 para evitar valores nulos, puede transformarse mediante el algoritmo natural. Las puntuaciones de Gleason puede ser una variable numérica continua o puede categorizarse, por ejemplo, como < 7 (nivel de referencia), 7 y > 7. En algunas realizaciones, las puntuaciones de Gleason pueden introducirse como su valor numérico (en lugar de agruparse). En algunas realizaciones, una puntuación de Gleason de 7 puede delinearse adicionalmente mediante (3+4) frente a (4+3).

En algunas realizaciones, la puntuación combinada se calcula según la siguiente fórmula:

(2) *Puntuación combinada = A*(puntuación de CCP) + B*(ln(1+[PSA])) + {C (si Gleason = 7) o D (si Gleason > 7)}*

En algunas realizaciones, las variables clínicas (por ejemplo, PSA, Gleason, etc.) pueden combinarse en una puntuación clínica (por ejemplo, puntuación de nomograma), que luego puede combinarse con la puntuación de CCP para dar lugar a una puntuación combinada según la siguiente fórmula:

(3) *Puntuación combinada = A*(puntuación de CCP) + B*(puntuación clínica)*

En algunas realizaciones, la puntuación clínica es la puntuación de CAPRA o la puntuación de nomograma de Kattan-Stephenson. La puntuación de CAPRA puede calcularse tal como se comenta en el presente documento (véase especialmente el ejemplo 8 a continuación). En algunas realizaciones, la puntuación clínica no es una combinación de variables clínicas sino una puntuación que representa una variable (por ejemplo, puntuación de Gleason).

La puntuación combinada con CCP y otros componentes ponderados tal como se comenta en el presente documento abarca, haciendo los cambios necesarios, cualquier versión modificada o aumentada de escala de la misma. Por ejemplo, los elementos pueden multiplicarse por o dividirse entre un factor (por ejemplo, constante o nueva variable) y/o sumársele o restársele un factor (por ejemplo, constante o nueva variable). Como un ejemplo, una puntuación combinada según fórmula (3)

(3) *Puntuación combinada = A*(puntuación de CCP) + B*(puntuación clínica)*

abarca una versión de la misma que aumenta de escala en los factores C y D según la siguiente fórmula (3A)

(3A) *Puntuación combinada = C*(A*(puntuación de CCP) + B*(puntuación clínica)) + D*

En algunas realizaciones, se usa cualquiera de las fórmulas comentadas en el presente documento en los métodos, sistemas, etc., de la divulgación para determinar el pronóstico basándose en una muestra de prostatectomía radical del paciente. En algunas realizaciones, se usa cualquiera de las fórmulas comentadas en el presente documento en los métodos, sistemas, etc., de la divulgación para determinar el pronóstico basándose en la muestra de biopsia de próstata del paciente. En algunas realizaciones, la puntuación de CCP es la media no ponderada de valores de C_T para la expresión de los genes de CCP que están analizándose, opcionalmente normalizada mediante la media no ponderada de los genes HK de modo que mayores valores indican una mayor expresión (en algunas realizaciones, una unidad es equivalente a un cambio de dos veces en la expresión). En algunas realizaciones, la puntuación de CCP oscila desde -8 hasta 8 o desde -1,6 hasta 3,7.

En algunas realizaciones, A = 0,95, B = 0,61, C = 0,90 (cuando sea aplicable) y D = 1,00 (cuando sea aplicable); A 0,57 y B = 0,39; o A = 0,58 y B = 0,41. En algunas realizaciones, A, B, C y/o D están dentro del redondeo de estos valores (por ejemplo, A está entre 0,945 y 0,954, etc.). En algunos casos, una fórmula puede no tener todos los coeficientes especificados (y por tanto no incorporar la(s) variable(s) correspondiente(s)). Por ejemplo, la realización mencionada inmediatamente antes puede aplicarse a la fórmula (3), donde A en la fórmula (3) es 0,95 y B en la fórmula (3) es 0,61. C y D no serían aplicables como estos coeficientes y sus variables correspondientes no se encuentran en la fórmula (3) (aunque las variables clínicas pueden incorporarse en la puntuación clínica encontrada en la fórmula (3)). En algunas realizaciones, A está entre 0,9 y 1, 0,9 y 0,99, 0,9 y 0,95, 0,85 y 0,95, 0,86 y 0,94, 0,87 y 0,93, 0,88 y 0,92, 0,89 y 0,91, 0,85 y 0,9, 0,8 y 0,95, 0,8 y 0,9, 0,8 y 0,85, 0,75 y 0,99, 0,75 y 0,95, 0,75 y 0,9, 0,75 y 0,85 o entre 0,75 y 0,8. En algunas realizaciones, B está entre 0,40 y 1, 0,45 y 0,99, 0,45 y 0,95, 0,55 y 0,8, 0,55 y 0,7, 0,55 y 0,65, 0,59 y 0,63 o entre 0,6 y 0,62. En algunas realizaciones, C está, cuando sea aplicable, entre 0,9 y 1, 0,9 y 0,99, 0,9 y 0,95, 0,85 y 0,95, 0,86 y 0,94, 0,87 y 0,93, 0,88 y 0,92, 0,89 y 0,91, 0,85 y 0,9, 0,8 y 0,95, 0,8 y 0,9, 0,8 y 0,85, 0,75 y 0,99, 0,75 y 0,95, 0,75 y 0,9, 0,75 y 0,85 o entre 0,75 y 0,8. En algunas realizaciones, D está, cuando sea aplicable, entre 0,9 y 1, 0,9 y 0,99, 0,9 y 0,95, 0,85 y 0,95, 0,86 y 0,94, 0,87 y 0,93, 0,88 y 0,92, 0,89 y 0,91, 0,85 y 0,9, 0,8 y 0,95, 0,8 y 0,9, 0,8 y 0,85, 0,75 y 0,99, 0,75 y 0,95, 0,75 y 0,9, 0,75 y 0,85 o entre 0,75 y 0,8.

En algunas realizaciones, A está entre 0,1 y 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,2 y 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,3 y 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,4 y 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,5 y 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,6 y 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,7 y 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,8 y 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,9 y 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 1 y 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 1,5 y 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 2 y 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 2,5 y 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 3 y 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre

3,5 y 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 4 y 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 4,5 y 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 5 y 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 6 y 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 7 y 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 8 y 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 9 y 10, 11, 12, 13, 14, 15 ó 20; o entre 10 y 11, 12, 13, 14, 15 ó 20; o entre 11 y 12, 13, 14, 15 ó 20; o entre 12 y 13, 14, 15 ó 20; o entre 13 y 14, 15 ó 20; o entre 14 y 15 ó 20; o entre 15 y 20; B está entre 0,1 y 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,2 y 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,3 y 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,4 y 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,5 y 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,6 y 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,7 y 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,8 y 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,9 y 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 1 y 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 1,5 y 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 2 y 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 2,5 y 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 3 y 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 3,5 y 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 4 y 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 4,5 y 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 5 y 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 6 y 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 7 y 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 8 y 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 9 y 10, 11, 12, 13, 14, 15 ó 20; o entre 10 y 11, 12, 13, 14, 15 ó 20; o entre 11 y 12, 13, 14, 15 ó 20; o entre 12 y 13, 14, 15 ó 20; o entre 13 y 14, 15 ó 20; o entre 14 y 15 ó 20; o entre 15 y 20; C está, cuando sea aplicable, entre 0,1 y 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,2 y 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,3 y 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,4 y 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,5 y 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,6 y 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,7 y 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,8 y 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,9 y 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 1 y 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 1,5 y 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 2 y 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 2,5 y 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 3 y 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 3,5 y 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 4 y 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 4,5 y 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 5 y 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 6 y 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 7 y 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 8 y 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 9 y 10, 11, 12, 13, 14, 15 ó 20; o entre 10 y 11, 12, 13, 14, 15 ó 20; o entre 11 y 12, 13, 14, 15 ó 20; o entre 12 y 13, 14, 15 ó 20; o entre 13 y 14, 15 ó 20; o entre 14 y 15 ó 20; o entre 15 y 20; y D está, cuando sea aplicable, entre 0,1 y 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,2 y 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,3 y 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,4 y 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,5 y 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,6 y 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,7 y 0,8, 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,8 y 0,9, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 0,9 y 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 1 y 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 1,5 y 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 2 y 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 2,5 y 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 3 y 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 3,5 y 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 4 y 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 4,5 y 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 5 y 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 6 y 7, 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 7 y 8, 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 8 y 9, 10, 11, 12, 13, 14, 15 ó 20; o entre 9 y 10, 11, 12, 13, 14, 15 ó 20; o entre 10 y 11, 12, 13, 14, 15 ó 20; o entre 11 y 12, 13, 14, 15 ó 20; o entre 12 y 13, 14, 15 ó 20; o entre 13 y 14, 15 ó 20; o entre 14 y 15 ó 20; o entre 15 y 20. En algunas realizaciones, A, B y/o C están dentro del redondeo de cualquiera de estos valores (por ejemplo, A está entre 0,45 y 0,54, etc.).

En algunas realizaciones, el riesgo en porcentaje (absoluto o relativo) de un acontecimiento o desenlace clínico particular (por ejemplo, muerte específica de cáncer, recidiva después de cirugía, etc.) en el paciente se estima (por ejemplo, se calcula) según la divulgación (por ejemplo, según una o más de las fórmulas anteriores). Tal riesgo puede estimarse aplicando la razón de riesgos instantáneos para un parámetro particular (por ejemplo, puntuación de CCP, puntuación combinada) a ese parámetro para dar lugar al riesgo relativo de un desenlace clínico particular (por ejemplo, recidiva de cáncer o muerte específica de cáncer) en el paciente. En algunos casos, la razón de riesgos instantáneos representa el aumento de riesgo relativo por unidad del parámetro. En algunos de los ejemplos a continuación, por ejemplo, un aumento de una sola unidad en la puntuación de CCP (que representa el doble de expresión) representa un riesgo relativo aumentado en el múltiplo de la razón de riesgos instantáneos. Cuando la razón de riesgos instantáneos es igual a 2, por ejemplo, un aumento de una sola unidad en la puntuación de CCP corresponde al doble de riesgo relativo (es decir, un primer paciente con una puntuación de CCP que es una unidad mayor que un segundo paciente tiene el doble de riesgo de recidiva de cáncer o muerte específica de cáncer). Este

riesgo relativo puede usarse con el riesgo promedio en una población particular para determinar cómo se compara el riesgo de un paciente específico con tal población. Un modo de hacer esto es establecer el valor promedio del parámetro en la población como cero y luego comparar los valores del paciente específico para el parámetro, lo que significa que un paciente con, por ejemplo, una puntuación de CCP de 1 tendría el doble de riesgo promedio de recidiva de cáncer o muerte específica de cáncer.

Por tanto, algunas realizaciones de la divulgación proporcionan un método para calcular el riesgo de recidiva de cáncer o muerte específica de cáncer en un paciente que comprende (1) obtener los niveles de expresión medidos de una pluralidad de genes que comprende al menos 3 genes elegidos de cualesquiera de las tablas 1, 2, 7-11, 13-14 y/o Y o de los paneles A I en una muestra del paciente, (2) calcular un valor de prueba a partir de estos niveles de expresión medidos tal como se comenta en el presente documento (por ejemplo, genes de CCP que contribuyen al menos cierta ponderación, al menos cierto número de genes de CCP, etc.), y (3) calcular dicho riesgo de recidiva de cáncer o muerte específica de cáncer en el paciente multiplicando el número de unidades aumentadas del valor de prueba con respecto a cierto valor de referencia (por ejemplo, valores promedio en una población particular) por la razón de riesgos instantáneos para el valor. En algunas realizaciones, el valor de prueba comprende la puntuación de CCP o una puntuación combinada tal como se describe en el presente documento y la razón de riesgos instantáneos es cualquiera de las razones de riesgos instantáneos notificadas en el presente documento.

Un riesgo más absoluto (en lugar de relativo) de recidiva también puede estimarse recopilando datos a partir de una cohorte de pacientes de estudio y correlacionando valores y puntuaciones de prueba para tales pacientes con sus acontecimientos o desenlaces clínicos eventuales. Tales datos pueden usarse en forma gráfica (figuras 23 y 24), tabular o tal como se representa en una fórmula (por ejemplo, fórmula (4) a continuación) para, basándose en la puntuación de prueba de un paciente, determinar el riesgo de acontecimiento o desenlace clínico particular en tal paciente que se somete a la prueba. Por ejemplo, puede medirse la puntuación de CCP del paciente como 1,7, calcularse la puntuación de CAPRA del paciente como 5 y luego calcularse la puntuación combinada del paciente (por ejemplo, biopsia) usando la fórmula (2), con A = 0,57 y B = 0,39, como 2,92. Luego puede usarse esta puntuación combinada para calcular el riesgo de muerte específica de cáncer en tal paciente en el plazo de 10 años, por ejemplo, usando una fórmula preespecificada (por ejemplo, fórmula (4)), usando una curva (por ejemplo, figura 23 ó 24) o una tabla de riesgos. En algunas realizaciones, el riesgo es cualquiera dentro del intervalo de confianza del 95%, por ejemplo, los intervalos mostrados en la figura 23 ó 24.

Usando estos datos empíricos (por ejemplo, los representados en las figuras 23 y 24), pueden derivarse las fórmulas de la siguiente forma general para estimar el riesgo:

(4) *Riesgo estimado de A = $B*e^{C*D}$*

A menudo, estas fórmulas se derivarán representando gráficamente el riesgo frente a la puntuación de CCP o la puntuación combinada para un conjunto de muestras de pacientes de estudio y ajustando una curva a la línea resultante (por ejemplo, una de las líneas mostradas en la figura 24). En algunas realizaciones, A es muerte específica de cáncer o recidiva bioquímica. En algunas realizaciones, D es puntuación de CCP o una puntuación combinada (cada una tal como se describe en los diversos aspectos y realizaciones en el presente documento). En algunas realizaciones, B = 0,0155 y C = 0,0054 (tal como se derivan a partir de, por ejemplo, una línea tal como se muestra en la figura 24). En un caso de este tipo, la fórmula (4) se modificaría de la siguiente manera:

(5) *Riesgo estimado de muerte específica de cáncer = $0,0155*e^{0,0054*(puntuación\ combinada)}$*

donde la puntuación combinada se calcula según cualquiera de las fórmulas comentadas en el presente documento. En algunas de tales realizaciones específicas, la puntuación combinada se calcula según la fórmula (2) con A = 0,57 y B = 0,39 y la puntuación clínica es la puntuación de CAPRA.

Por tanto, en algunas realizaciones, la divulgación proporciona un método para calcular el riesgo de muerte específica de cáncer de próstata en un paciente que comprende: (1) obtener los niveles de expresión medidos de una pluralidad de genes que comprende al menos 4 genes de cualesquiera de las tablas 1, 2, 7-11, 13-14 y/o Y o de los paneles A I; (2) opcionalmente obtener una puntuación para el paciente calculada a partir de al menos una variable clínica (por ejemplo, uno o más de año de RP, márgenes quirúrgicos, extensión extracapsular, invasión de vesículas seminales, afectación ganglionar, puntuación primaria de Gleason, puntuación secundaria de Gleason o nivel de PSA preoperatorio); (3) opcionalmente derivar un valor de prueba combinado a partir de los niveles medidos obtenidos en (1) y la puntuación obtenida en (2); y (4) calcular el riesgo de muerte específica de cáncer de próstata en el paciente según la siguiente fórmula:

Riesgo de muerte específica de cáncer = $0,0155*e^{0,0054([niveles\ de\ expresión\ medidos\ en\ (1)]\ o\ [puntuación\ combinada\ obtenida\ en\ (2)])}$

En algunas realizaciones, la divulgación proporciona un método para calcular el riesgo de muerte específica de cáncer de próstata en un paciente que comprende: (1) obtener los niveles de expresión medidos de una pluralidad de genes que comprende al menos 4 genes de cualesquiera de las tablas 1, 2, 7-11, 13-14 y/o Y o de los paneles A I; (2) opcionalmente obtener una puntuación de CAPRA para la muestra de paciente tal como se comenta en el

presente documento; (3) opcionalmente derivar un valor de prueba combinado a partir de los niveles medidos obtenidos en (1) y la puntuación obtenida en (2) según la fórmula: Puntuación combinada = (0,57*(puntuación en (2)) + (0,39*puntuación en (1)); y (4) calcular el riesgo de muerte específica de cáncer de próstata en el paciente según la tabla en la figura 26.

En algunas realizaciones, el valor de prueba derivado o calculado a partir de un gen particular (por ejemplo, *KLK3*) o a partir de la expresión global de la pluralidad de genes de prueba (por ejemplo, CCG) se compara con uno o más valores de referencia (o valores indicadores), y el valor de prueba se correlaciona opcionalmente con el pronóstico, el riesgo de progresión de cáncer, el riesgo de recidiva de cáncer o el riesgo de muerte específica de cáncer si difiere del valor indicador.

Por ejemplo, el valor indicador puede derivarse o calcularse a partir de los niveles de expresión de genes hallados en una muestra normal obtenida del paciente de interés, en cuyo caso un valor de prueba (derivado o calculado a partir de un nivel de expresión en la muestra tumoral) significativamente mayor que este valor indicador indicaría, por ejemplo, un mal pronóstico o una probabilidad aumentada de recidiva de cáncer o muerte específica de cáncer o la necesidad de tratamiento agresivo. En algunas realizaciones, el valor de prueba se considera “mayor que” el valor de referencia (por ejemplo, el valor indicador de umbral) y, por tanto, se correlaciona con una probabilidad aumentada de respuesta al tratamiento que comprende quimioterapia, si el valor de prueba supera el valor de referencia en al menos cierta cantidad (por ejemplo, al menos 0,5, 0,75, 0,85, 0,90, 0,95, 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más veces o desviaciones estándar).

Alternativamente, el valor indicador puede derivarse o calcularse a partir del nivel de expresión promedio para un conjunto de individuos de una diversa población con cáncer o un subconjunto de la población. Por ejemplo, puede determinarse el nivel de expresión promedio de un gen o panel de genes en una toma de muestras aleatoria de pacientes con cáncer (por ejemplo, cáncer de próstata, vejiga, cerebro, mama o pulmón). Este nivel de expresión promedio puede denominarse “valor indicador de umbral”, esperándose que los pacientes que tienen una expresión de CCG mayor que este valor tengan un pronóstico peor que aquellos que tienen una expresión menor que este valor.

Alternativamente, el valor indicador puede representar el nivel de expresión promedio de un marcador o una pluralidad de marcadores génicos particulares en una pluralidad de pacientes de entrenamiento (por ejemplo, pacientes con cáncer de próstata) con desenlaces similares cuyos datos clínicos y de seguimiento están disponibles y son suficientes para definir y categorizar a los pacientes por desenlace de enfermedad, por ejemplo, recidiva o pronóstico. Véanse, por ejemplo, los ejemplos a continuación. Por ejemplo, un “valor indicador de buen pronóstico” puede generarse a partir de una pluralidad de pacientes con cáncer de entrenamiento caracterizados como que tienen un “buen desenlace”, por ejemplo, aquellos que no han tenido recidiva de cáncer cinco años (o diez años o más) después del tratamiento inicial, o aquellos que no han tenido progresión en su cáncer cinco años (o diez años o más) después del diagnóstico inicial. Un “valor indicador de mal pronóstico” puede generarse a partir de una pluralidad de pacientes con cáncer de entrenamiento definidos como que tienen un “mal desenlace”, por ejemplo, aquellos que han tenido recidiva de cáncer en el plazo de cinco años (o diez años, etc.) después del tratamiento inicial, o aquellos que han tenido progresión en su cáncer en el plazo de cinco años (o diez años, etc.) después del diagnóstico inicial. Por tanto, un valor indicador de buen pronóstico de un gen particular puede representar el nivel promedio de expresión del gen particular en pacientes que tienen un “buen desenlace”, mientras que un valor indicador de mal pronóstico de un gen particular representa el nivel promedio de expresión del gen particular en pacientes que tienen un “mal desenlace”.

Por tanto, un aspecto de la divulgación proporciona un método para clasificar el cáncer que comprende determinar el estado de un panel de genes que comprende al menos dos CCG, en una muestra tisular o celular, particularmente una muestra tumoral, de un paciente, en el que un estado anómalo indica una clasificación de cáncer negativa. Tal como se usa en el presente documento, “determinar el estado” de un gen se refiere a determinar la presencia, la ausencia o el grado/nivel de alguna característica física, química o genética del gen o de sus producto(s) de expresión. Tales características incluyen, pero no se limitan a, niveles de expresión, niveles de actividad, mutaciones, número de copias, estado de metilación, etc.

En el contexto de los CCG que se usan para determinar el riesgo de recidiva o progresión de cáncer o la necesidad de tratamiento agresivo, las características particularmente útiles incluyen los niveles de expresión (por ejemplo, niveles de ARNm o proteína) y los niveles de actividad. Las características pueden someterse a ensayo directamente (por ejemplo, sometiendo a ensayo el nivel de expresión de un CCG) o determinarse indirectamente (por ejemplo, sometiendo a ensayo el nivel de un gen o genes cuyo nivel de expresión se correlaciona con el nivel de expresión del CCG). Por tanto, algunas realizaciones de la divulgación proporcionan un método para clasificar el cáncer que comprende determinar el nivel de expresión, particularmente el nivel de ARNm, de un panel de genes que comprende al menos dos CCG, en una muestra tumoral, en el que una alta expresión (o expresión aumentada o sobreexpresión) indica una clasificación de cáncer negativa, o un riesgo aumentado de recidiva o progresión de cáncer, o la necesidad de tratamiento agresivo. En algunas realizaciones, el método comprende al menos una de las siguientes etapas: (a) correlacionar la alta expresión (o expresión aumentada o sobreexpresión) del panel de genes con una clasificación de cáncer negativa, un riesgo aumentado de recidiva o progresión de cáncer, o la necesidad de

tratamiento agresivo; (b) concluir que el paciente tiene una clasificación de cáncer negativa, un riesgo aumentado de recidiva o progresión de cáncer, o la necesidad de tratamiento agresivo basándose al menos en parte en la alta expresión (o expresión aumentada o sobreexpresión) del panel de genes; o (c) comunicar que el paciente tiene una clasificación de cáncer negativa, un riesgo aumentado de recidiva o progresión de cáncer, o la necesidad de tratamiento agresivo basándose al menos en parte en la alta expresión (o expresión aumentada o sobreexpresión) del panel de genes.

"Estado anómalo" significa el estado de un marcador en una muestra particular que difiere del estado generalmente hallado en muestras regulares (por ejemplo, muestras sanas o muestras con enfermedad regulares). Los ejemplos incluyen mutado, elevado, disminuido, presente, ausente, etc. Un "estado elevado" significa que una o más de las características anteriores (por ejemplo, expresión o nivel de ARNm) es mayor que los niveles normales. Generalmente, esto significa un aumento en la característica (por ejemplo, expresión o nivel de ARNm) en comparación con un valor indicador. A la inversa, un "estado bajo" significa que una o más de las características anteriores (por ejemplo, expresión de genes o nivel de ARNm) es menor que los niveles normales. Generalmente, esto significa una disminución en la característica (por ejemplo, expresión) en comparación con un valor indicador. En este contexto, un "estado negativo" generalmente significa que la característica está ausente o es indetectable. Por ejemplo, el estado de *PTEN* es negativo si el ácido nucleico de *PTEN* y/o la proteína PTEN está ausente o es indetectable en una muestra. Sin embargo, estado de *PTEN* negativo también incluye una mutación o una reducción del número de copias en *PTEN*.

En algunas realizaciones de la divulgación, los métodos comprenden determinar la expresión de uno o más CCG, y si esta expresión está "aumentada", el paciente tiene un mal pronóstico. En el contexto de la divulgación, expresión "aumentada" de un CCG significa que el nivel de expresión del paciente o bien está elevado por encima de un valor indicador normal o de umbral (por ejemplo, en al menos cierta cantidad de umbral) o bien es más próximo al "valor indicador de mal pronóstico" que al "valor indicador de buen pronóstico".

Por tanto, cuando el nivel determinado de expresión de un marcador génico relevante es más próximo al valor indicador de buen pronóstico del gen que al valor indicador de mal pronóstico del gen, entonces puede concluirse que es más probable que el paciente tenga un buen pronóstico, es decir, una probabilidad baja (o no aumentada) de recidiva de cáncer. Por otro lado, si el nivel determinado de expresión de un marcador génico relevante es más próximo al valor indicador de mal pronóstico del gen que al valor indicador de buen pronóstico del gen, entonces puede concluirse que es más probable que el paciente tenga un mal pronóstico, es decir, una probabilidad aumentada de recidiva de cáncer.

Alternativamente, los valores indicadores pueden determinarse de la siguiente manera: con el fin de asignar a los pacientes a grupos de riesgo, se establecerá un valor de umbral para la media de ciclo celular. El valor de umbral óptimo se selecciona basándose en la curva característica operativa del receptor (ROC), que representa gráficamente la sensibilidad frente a (1-especificidad). Para cada incremento de la media de ciclo celular, la sensibilidad y la especificidad de la prueba se calculan usando ese valor como umbral. El umbral real será el valor que optimiza estos parámetros según requisitos de los expertos (por ejemplo, qué grado de sensibilidad o especificidad se desea, etc.). El ejemplo 5 demuestra la determinación de un valor de umbral determinado y validado experimentalmente.

Los paneles de CCG (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más CCG) pueden predecir con precisión el pronóstico, tal como se muestra en el ejemplo 3. Los expertos en la técnica están familiarizados con diversos modos de determinar la expresión de un panel de genes (es decir, una pluralidad de genes). Puede determinarse la expresión de un panel de genes determinando el nivel de expresión promedio (normalizado o absoluto) de la totalidad de genes de panel en una muestra obtenida de un paciente particular (o bien en toda la muestra o bien en un subconjunto de células de la muestra o bien en una única célula). Expresión aumentada en este contexto significará que la expresión promedio es mayor que el nivel de expresión promedio de estos genes en pacientes normales (o mayor que cierto valor indicador que se ha determinado que representa el nivel de expresión promedio en una población de referencia tal como pacientes con el mismo cáncer). Alternativamente, puede determinarse la expresión de un panel de genes determinando el nivel de expresión promedio (normalizado o absoluto) de al menos un determinado número (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30 o más) o de al menos una determinada proporción (por ejemplo, el 10%, el 20%, el 30%, el 40%, el 50%, el 60%, el 70%, el 80%, el 90%, el 95%, el 99%, el 100%) de los genes en el panel. Alternativamente, puede determinarse la expresión de un panel de genes determinando el número de copias absoluto del ARNm (o proteína) de todos los genes en el panel y o bien sumando o promediando los mismos a lo largo de los genes.

Tal como se usa en el presente documento, "clasificar un cáncer" y "clasificación de cáncer" se refieren a determinar una o más características clínicamente relevantes de un cáncer y/o determinar un pronóstico particular de un paciente que tiene dicho cáncer. Por tanto "clasificar un cáncer" incluye, pero no se limita a: (i) evaluar el potencial metastásico, el potencial para metastatizarse a órganos específicos, el riesgo de recidiva y/o la evolución del tumor; (ii) evaluar el estadio tumoral; (iii) determinar el pronóstico del paciente en ausencia de tratamiento del cáncer; (iv) determinar el pronóstico de respuesta del paciente (por ejemplo, reducción del volumen tumoral o supervivencia libre de progresión) al tratamiento (por ejemplo, quimioterapia, radioterapia, cirugía para extirpar el tumor, etc.); (v)

diagnóstico de la respuesta real del paciente al tratamiento actual y/o anterior; (vi) determinar un transcurso de tratamiento preferido para el paciente; (vii) pronóstico de la recaída del paciente después del tratamiento (o bien tratamiento en general o bien algún tratamiento particular); (viii) pronóstico de la esperanza de vida del paciente (por ejemplo, pronóstico de la supervivencia global), etc.

Por tanto, una "clasificación negativa" significa una característica clínica desfavorable del cáncer (por ejemplo, un mal pronóstico). Los ejemplos incluyen (i) un potencial metastásico, potencial para metastatizarse a órganos específicos y/o riesgo de recidiva aumentados; (ii) un estadio tumoral avanzado; (iii) un mal pronóstico del paciente en ausencia de tratamiento del cáncer; (iv) un mal pronóstico de la respuesta del paciente (por ejemplo, reducción del volumen tumoral o supervivencia libre de progresión) a un tratamiento particular (por ejemplo, quimioterapia, radioterapia, cirugía para extirpar el tumor, etc.); (v) un mal pronóstico de la recaída del paciente después del tratamiento (o bien tratamiento en general o bien algún tratamiento particular); (vi) un mal pronóstico de la esperanza de vida del paciente (por ejemplo, pronóstico de la supervivencia global), etc. En algunas realizaciones, un parámetro clínico asociado a recidiva (o una alta puntuación de nomograma) y una expresión aumentada de un CCG indican una clasificación negativa en el cáncer (por ejemplo, probabilidad aumentada de recidiva o progresión).

Un paciente con una muestra que muestra una alta puntuación o un alto valor de CCP (o expresión de CCP aumentada) tiene una probabilidad aumentada de recidiva después del tratamiento (por ejemplo, las células cancerosas no destruidas o eliminadas mediante el tratamiento volverán a crecer rápidamente). Un paciente de este tipo también tiene una probabilidad aumentada de progresión de cáncer para una progresión más rápida (por ejemplo, las células que proliferan rápidamente provocarán que cualquier tumor crezca rápidamente, aumente su virulencia y/o se metastatice). Un paciente de este tipo también puede requerir un tratamiento relativamente más agresivo. Por tanto, en algunas realizaciones, la divulgación proporciona un método para clasificar el cáncer que comprende determinar el estado de un panel de genes que comprende al menos dos CCG, en el que un estado anómalo indica una probabilidad aumentada de recidiva o progresión. En algunas realizaciones, el método comprende al menos una de las siguientes etapas: (a) correlacionar el estado anómalo del panel de genes con una probabilidad aumentada de recidiva o progresión; (b) concluir que el paciente tiene una probabilidad aumentada de recidiva o progresión basándose al menos en parte en el estado anómalo del panel de genes; o (c) comunicar que el paciente tiene una probabilidad aumentada de recidiva o progresión basándose al menos en parte en el estado anómalo del panel de genes. Tal como se comentó anteriormente, en algunas realizaciones, el estado que va a determinarse es los niveles de expresión de genes. Por tanto, en algunas realizaciones, la divulgación proporciona un método para determinar el pronóstico del cáncer de un paciente que comprende determinar el nivel de expresión de un panel de genes que comprende al menos dos CCG, en el que la alta expresión (o expresión aumentada o sobreexpresión) indica una probabilidad aumentada de recidiva o progresión del cáncer. En algunas realizaciones, el método comprende al menos una de las siguientes etapas: (a) correlacionar la alta expresión (o expresión aumentada o sobreexpresión) del panel de genes con una probabilidad aumentada de recidiva o progresión; (b) concluir que el paciente tiene una probabilidad aumentada de recidiva o progresión basándose al menos en parte en la alta expresión (o expresión aumentada o sobreexpresión) del panel de genes; o (c) comunicar que el paciente tiene una probabilidad aumentada de recidiva o progresión basándose al menos en parte en la alta expresión (o expresión aumentada o sobreexpresión) del panel de genes.

"Recidiva" y "progresión" son términos bien conocidos en la técnica y se usan en el presente documento según sus significados conocidos. Debido a que los métodos de la divulgación pueden predecir o determinar la probabilidad de cada una de "recidiva", "progresión", "muerte específica de cáncer" y "respuesta a un tratamiento particular" en un paciente, dichos términos se usan de manera intercambiable, a menos que se especifique lo contrario, en el sentido de que la referencia a una se aplica igualmente a las demás. Como un ejemplo, el significado de "progresión" puede ser dependiente del tipo de cáncer, significando progresión en el cáncer de pulmón algo diferente de la progresión en el cáncer de próstata. Sin embargo, dentro de cada tipo y subtipo de cáncer, los expertos en la técnica entienden claramente "progresión". Debido a que predecir la recidiva y predecir la progresión son labores de pronóstico, "predecir el pronóstico" a menudo se usará en el presente documento para referirse a una cualquiera o a ambas. En estos casos, un "mal pronóstico" se referirá generalmente a una probabilidad aumentada de recidiva, progresión o ambas.

"Respuesta" (por ejemplo, respuesta a un régimen de tratamiento particular) es un término bien conocido en la técnica y se usa en el presente documento según su significado conocido. Como un ejemplo, el significado de "respuesta" puede ser dependiente del tipo de cáncer, significando respuesta en el cáncer de pulmón algo diferente de respuesta en el cáncer de próstata. Sin embargo, dentro de cada tipo y subtipo de cáncer, los expertos en la técnica entienden claramente "respuesta". Por ejemplo, algunos criterios objetivos de respuesta incluyen los criterios de evaluación de respuesta en tumores sólidos (RECIST), un conjunto de reglas publicadas (por ejemplo, cambios en el tamaño tumoral, etc.) que definen cuándo los pacientes con cáncer mejoran ("responden"), permanecen igual ("se estabilizan") o empeoran ("progresión") durante los tratamientos. Véase, por ejemplo, Eisenhauer *et al*., EUR. J. CANCER (2009) 45:228-247. "Respuesta" también puede incluir parámetros de supervivencia (por ejemplo, "supervivencia libre de enfermedad" (DFS), "supervivencia global" (OS), etc.). En algunos casos, los criterios RECIST pueden incluir: (a) respuesta completa (CR): desaparición de todas las metástasis; (b) respuesta parcial (PR): una disminución de al menos el 30% en la suma del diámetro más grande (LD) de las lesiones metastásicas, tomando como referencia el LD de la suma de nivel inicial; (c) enfermedad estable (SD): ni suficiente reducción del

volumen para calificar como PR ni suficiente aumento para calificar como PD tomando como referencia el LD de la suma más pequeño desde que se inició el tratamiento; (d) progresión (PD): un aumento de al menos el 20% en la suma del LD de las lesiones metastásicas diana tomando como referencia el LD de la suma más pequeño desde que se inició el tratamiento o apareció una o más lesiones nuevas.

Tal como se usa en el presente documento, un paciente tiene una “probabilidad aumentada” de alguna característica o algún desenlace clínico (por ejemplo, recidiva o progresión) si la probabilidad del paciente de tener la característica o el desenlace supera alguna probabilidad o algún valor de referencia. La probabilidad de referencia puede ser la probabilidad de la característica o del desenlace a lo largo de la población general de pacientes relevantes. Por ejemplo, si la probabilidad de recidiva en la población general con cáncer de próstata es del X% y se ha determinado mediante los métodos de la presente divulgación que un paciente particular tiene una probabilidad de recidiva del Y%, y si Y > X, entonces el paciente tiene una “probabilidad aumentada” de recidiva. Alternativamente, tal como se comentó anteriormente, puede determinarse un valor de umbral o de referencia y puede compararse la probabilidad de recidiva en un paciente particular con ese umbral o esa referencia.

En algunas realizaciones, el método correlaciona la puntuación específica del paciente (por ejemplo, puntuación de CCP, puntuación combinada de CCP con variables clínicas) con una probabilidad específica (por ejemplo, el 10%, el 15%, el 20%, el 25%, el 30%, el 35%, el 40%, el 45%, el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, el 90%, el 95%, el 99%, el 100%) de recidiva, progresión o muerte específica de cáncer (cada una opcionalmente dentro de un marco temporal específico, por ejemplo, 5 años, 10 años) o respuesta a un tratamiento particular. En algunas realizaciones, la divulgación proporciona un método para determinar el pronóstico de un paciente con cáncer de próstata que comprende: (1) determinar, en una muestra de paciente, los niveles de expresión de una pluralidad de genes de prueba, en el que la pluralidad de genes de prueba comprende al menos 5 de los genes en uno cualquiera de los paneles A a G; (2) derivar un valor de prueba a partir de los niveles de expresión determinados en (1), en el que los al menos 5 genes en uno cualquiera de los paneles A a G contribuyen al menos el 25% al valor de prueba; (3) comparar el valor de prueba con un valor de referencia; y (4) asignar una probabilidad de recidiva, progresión, muerte específica de cáncer o respuesta a un tratamiento particular basándose al menos en parte en la comparación en (3).

En algunas realizaciones, la muestra de paciente es de una biopsia de próstata, el valor de prueba es el C_T medio para los genes en el panel F normalizados frente a los genes en la tabla 3 y la probabilidad de muerte específica de cáncer de próstata en el plazo de 10 años del diagnóstico se calcula de la siguiente manera:

Tabla 4

Valor de prueba	Probabilidad de muerte específica de cáncer
-1	5,9%
0	11,6%
1	22%
2	39,5%
3	63,8%
4	87,2%

En algunas realizaciones, la muestra de paciente es de una prostatectomía, el valor de prueba es el C_T medio para los genes en el panel F normalizados frente a los genes en la tabla A y la probabilidad de recidiva de cáncer de próstata en el plazo de 10 años de la cirugía se calcula de la siguiente manera:

Tabla 5

Valor de prueba	Probabilidad de recidiva
-1	12,6%
0	24,9%
1	45,5%
2	72,5%
3	93,6%

En algunas realizaciones, la muestra de paciente es de una prostatectomía, el valor de prueba es una puntuación combinada calculada tal como se muestra en los párrafos [0067] y [0068] anteriores y la probabilidad de recidiva de cáncer de próstata en el plazo de 10 años de la cirugía se calcula de la siguiente manera:

Tabla 6

Valor de prueba	Probabilidad de recidiva
0	11,5%
1	25%

2	49,3%
3	79,8%
4	97,7%

Tal como se muestra en el ejemplo 3, los CCG individuales pueden predecir el pronóstico bastante bien. Por tanto, la divulgación proporciona un método para predecir el pronóstico que comprende determinar la expresión de al menos un CCG enumerado en la tabla 1 o en los paneles A a G.

El ejemplo 3 también muestra que los paneles de CCG (por ejemplo, 2, 3, 4, 5 ó 6 CCG) pueden predecir con precisión el pronóstico. Por tanto, en algunos aspectos, la divulgación proporciona un método para clasificar un cáncer que comprende determinar el estado de un panel de genes (por ejemplo, una pluralidad de genes de prueba) que comprende una pluralidad de CCG. Por ejemplo, expresión aumentada en un panel de genes (o pluralidad de genes de prueba) puede referirse al nivel de expresión promedio de la totalidad de genes de panel o de prueba en un paciente particular que es mayor que el nivel de expresión promedio de estos genes en pacientes normales (o mayor que cierto valor indicador que se ha determinado que representa el nivel de expresión promedio normal). Alternativamente, expresión aumentada en un panel de genes puede referirse a expresión aumentada en al menos un determinado número (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30 o más) o al menos una determinada proporción (por ejemplo, el 10%, el 20%, el 30%, el 40%, el 50%, el 60%, el 70%, el 80%, el 90%, el 95%, el 99%, el 100%) de los genes en el panel en comparación con el nivel de expresión normal promedio.

En algunas realizaciones, el panel comprende al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 70, 80, 90, 100, 200 o más CCG. En algunas realizaciones, el panel comprende al menos 10, 15, 20 o más CCG. En algunas realizaciones, el panel comprende entre 5 y 100 CCG, entre 7 y 40 CCG, entre 5 y 25 CCG, entre 10 y 20 CCG o entre 10 y 15 CCG. En algunas realizaciones, los CCG comprenden al menos una determinada proporción del panel. Por tanto, en algunas realizaciones, el panel comprende al menos el 25%, el 30%, el 40%, el 50%, el 60%, el 70%, el 75%, el 80%, el 85%, el 90%, el 95%, el 96%, el 97%, el 98% o el 99% de los CCG. En algunas realizaciones preferidas, el panel comprende al menos 10, 15, 20, 25, 30, 35, 40, 45, 50, 70, 80, 90, 100, 200 o más CCG, y tales CCG constituyen al menos el 50%, el 60%, el 70%, preferiblemente al menos el 75%, el 80%, el 85%, más preferiblemente al menos el 90%, el 95%, el 96%, el 97%, el 98% o el 99% o más del número total de genes en el panel. En algunas realizaciones, los CCG se eligen del grupo que consiste en los genes en la tabla 1 y en los paneles A a G. En algunas realizaciones, el panel comprende al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30 o más de los genes en cualquiera de la tabla 1 y los paneles A a G. En algunas realizaciones, la divulgación proporciona un método para predecir el pronóstico que comprende determinar el estado de los CCG en los paneles A a G, en el que el estado anómalo indica un mal pronóstico. En algunas realizaciones, el método comprende al menos una de las siguientes etapas: (a) correlacionar el estado anómalo (por ejemplo, expresión alta o aumentada) de los CCG en los paneles A a G con un mal pronóstico; (b) concluir que el paciente tiene un mal pronóstico basándose al menos en parte en el estado anómalo (por ejemplo, expresión alta o aumentada) de los CCG en los paneles A a G; o (c) comunicar que el paciente tiene un mal pronóstico basándose al menos en parte en el estado anómalo (por ejemplo, expresión alta o aumentada) de los CCG en los paneles A a G.

En algunas de estas realizaciones, una expresión elevada indica una probabilidad aumentada de recidiva o progresión. Por tanto, en una realización preferida, la divulgación proporciona un método para predecir el riesgo de recidiva o progresión de cáncer en un paciente que comprende determinar el estado de un panel de genes, en el que el panel comprende entre aproximadamente 10 y aproximadamente 15 CCG, los CCG constituyen al menos el 90% del panel, y un estado elevado para los CCG indica una probabilidad aumentada de recidiva o progresión. En algunas realizaciones, el método comprende al menos una de las siguientes etapas: (a) correlacionar el estado elevado (por ejemplo, expresión alta o aumentada) del panel de genes con un mal pronóstico; (b) concluir que el paciente tiene un mal pronóstico basándose al menos en parte en el estado elevado (por ejemplo, expresión alta o aumentada) del panel de genes; o (c) comunicar que el paciente tiene un mal pronóstico basándose al menos en parte en el estado elevado (por ejemplo, expresión alta o aumentada) del panel de genes.

Se han evaluado varios paneles de CCG (tabla 2 anteriormente y tablas 7 y 8 a continuación) para determinar su capacidad para predecir el pronóstico en varios cánceres diferentes. Los resultados de estos estudios se describen en los ejemplos 1 a 6 a continuación.

Tabla 7: “Panel C” evaluado en los ejemplos 1 a 4

Símbolo de gen	ID de gen de Entrez	Símbolo de gen	ID de gen de Entrez	Símbolo de gen	ID de gen de Entrez
AURKA	6790	*DTL**	51514	*PTTG1**	9232
*BUB1**	699	*FOXM1**	2305	*RRM2**	6241
*CCNB1**	891	*HMMR**	3161	*TIMELESS**	8914
*CCNB2**	9133	*KIF23**	9493	*TPX2**	22974
*CDC2**	983	*KPNA2*	3838	*TRIP13**	9319
*CDC20**	991	*MAD2L1**	4085	*TTK**	7272
*CDC45L**	8318	*MELK*	9833	*UBE2C*	11065
*CDCA8**	55143	*MYBL2**	4605	*UBE2S**	27338
CENPA	1058	*NUSAP1**	51203	*ZWINT**	11130
*CKS2**	1164	*PBK**	55872		
*DLG7**	9787	*PRC1**	9055		

* Estos genes se usaron como un panel de subconjuntos de 26 genes (“panel D”) en la rama de validación del experimento descrito en el ejemplo 2.

Tabla 8: “Panel E”

Nombre	ID de gen	Nombre	ID de gen	Nombre	ID de gen
*ASF1B**	55723	*CENPM**	79019	*ORC6L**	23594
*ASPM**	259266	*CEP55**	55165	*PBK**	55872
*BIRC5**	332	*DLGAP5**	9787	*PLK1**	5347
*BUB1B**	701	*DTL**	51514	*PRC1**	9055
*C18orf24**	220134	*FOXM1**	2305	*PTTG1**	9232
*CDC2**	983	*KIAA0101**	9768	*RAD51**	5888
*CDC20**	991	*KIF11**	3832	*RAD54L**	8438
*CDCA3**	83461	*KIF20A**	10112	*RRM2**	6241
*CDCA8**	55143	*KIF4A*	24137	*TK1**	7083
*CDKN3**	1033	*MCM10**	55388	*TOP2A**	7153
*CENPF**	1063	*NUSAP1**	51203		

* Estos genes se usaron como un panel de subconjuntos de 31 genes (“panel F”) en el experimento descrito en el ejemplo 5.

Se ha determinado que la elección de los CCG individuales para un panel a menudo puede ser relativamente arbitraria. Dicho de otro modo, se ha hallado que la mayoría de los CCG son muy buenos sustitutos entre sí. Un modo para evaluar si los CCG particulares servirán bien en los métodos y las composiciones de la divulgación es evaluando su correlación con la expresión media de los CCG (por ejemplo, todos los CCG conocidos, un conjunto específico de CCG, etc.). Se espera que esos CCG que se correlacionan particularmente bien con la media funcionen bien en los ensayos de la divulgación, por ejemplo, porque estos reducirán el ruido en el ensayo. Se proporciona una clasificación CCG seleccionados según su correlación con la expresión media de CCG en las tablas 9-11.

En las firmas de CCG, los CCG particulares sometidos a ensayo a menudo no es tan importante como el número total de CCG. El número de CCG sometidos a ensayo puede variar en función de muchos factores, por ejemplo,

restricciones técnicas, consideraciones de costes, la clasificación que está realizándose, el cáncer que está sometiéndose a prueba, el nivel deseado de potencia predictiva, etc. Aumentar el número de CCG sometidos a ensayo en un panel según la divulgación es, como cuestión general, ventajoso porque, por ejemplo, un mayor grupo de ARNm que van a someterse a ensayo significa menos "ruido" provocado por valores atípicos y menor oportunidad de que un error de ensayo altere la potencia predictiva global de la prueba. Sin embargo, el coste y otras consideraciones limitarán generalmente este número y es deseable hallar el número óptimo de CCG para una firma.

Se ha descubierto que la potencia predictiva de una firma de CCG a menudo deja de aumentar significativamente más allá un determinado número de CCG (véase la figura 1; ejemplo 1). Más específicamente, el número óptimo de CCG en una firma (n_O) puede hallarse siempre que sea cierto lo siguiente

$$(P_{n+1} - P_n) < C_O,$$

donde P es la potencia predictiva (es decir, P_n es la potencia predictiva de una firma con n genes y P_{n+1} es la potencia predictiva de una firma con n genes más uno) y C_O es cierta constante de optimización. La potencia predictiva puede definirse de muchos modos conocidos por los expertos en la técnica, incluyendo, pero sin limitarse a, el valor de p de la firma. C_O puede elegirla el experto basándose en sus restricciones específicas. Por ejemplo, si el coste no es un factor crítico y se desean niveles extremadamente altos de sensibilidad y especificidad, puede establecerse una C_O muy baja de tal manera que sólo se menosprecian aumentos triviales en la potencia predictiva. Por otro lado, si el coste es decisivo y son aceptables niveles moderados de sensibilidad y especificidad, puede establecerse una C_O más alta de tal manera que sólo aumentos significativos en la potencia predictiva justifican el aumento del número de genes en la firma.

Alternativamente, puede representarse gráficamente un gráfico de potencia predictiva en función del número de genes (como en la figura 1) y tomarse la segunda derivada de esta representación gráfica. El punto en el que la segunda derivada disminuye hasta cierto valor predeterminado (C_O') puede ser el número óptimo de genes en la firma.

Los ejemplos 1 y 3 y las figuras 1 y 17 ilustran la determinación empírica de los números óptimos de CCG en paneles de CCG de la divulgación. Se sometieron a prueba subconjuntos seleccionados aleatoriamente de los 31 CCG enumerados en la tabla 7 como firmas de CCG distintas y se determinó la potencia predictiva (es decir, el valor de p) para cada una. Tal como muestra la figura 1, los valores de p dejaron de mejorar significativamente entre aproximadamente 10 y aproximadamente 15 CCG, lo que indica que un número óptimo de CCG en un panel de pronóstico es de desde aproximadamente 10 hasta aproximadamente 15. Por tanto, algunas realizaciones de la divulgación proporcionan un método para predecir el pronóstico en un paciente que padece cáncer de próstata que comprende determinar el estado de un panel de genes, en el que el panel comprende entre aproximadamente 10 y aproximadamente 15 CCG y un estado elevado para los CCG indica un mal pronóstico. En algunas realizaciones, el método comprende al menos una de las siguientes etapas: (a) correlacionar el estado elevado (por ejemplo, expresión alta o aumentada) del panel de genes con un mal pronóstico; (b) concluir que el paciente tiene un mal pronóstico basándose al menos en parte en el estado elevado (por ejemplo, expresión alta o aumentada) del panel de genes; o (c) comunicar que el paciente tiene un mal pronóstico basándose al menos en parte en el estado elevado (por ejemplo, expresión alta o aumentada) del panel de genes. En algunas realizaciones, el panel comprende entre aproximadamente 10 y aproximadamente 15 CCG y los CCG constituyen al menos el 90% del panel. En otras realizaciones, el panel comprende CCG más uno o más marcadores adicionales que aumentan significativamente la potencia predictiva del panel (es decir, hacen que la potencia predictiva sea significativamente mejor que si el panel consistiera sólo en los CCG). Puede usarse cualquier otra combinación de CCG (incluyendo cualquiera de los enumerados en la tabla 1 o en los paneles A a G) para poner en práctica la divulgación.

Se ha descubierto que los CCG son particularmente predictivos en determinados cánceres. Por ejemplo, se ha determinado que los paneles de CCG son precisos para predecir la recidiva en el cáncer de próstata (Ejemplos 1 a 5). Además, los CCG pueden determinar el pronóstico en cánceres de vejiga, cerebro, mama y pulmón, tal como se resume en el ejemplo 6 a continuación.

Por tanto, la divulgación proporciona un método que comprende determinar el estado de un panel de genes que comprende al menos dos CCG, en el que un estado anómalo indica un mal pronóstico. En algunas realizaciones, el panel comprende al menos 2 genes elegidos del grupo de genes en al menos uno de los paneles A a G. En algunas realizaciones, el panel comprende al menos 10 genes elegidos del grupo de genes en al menos uno de los paneles A a G. En algunas realizaciones, el panel comprende al menos 15 genes elegidos del grupo de genes en al menos uno de los paneles A a G. En algunas realizaciones, el panel comprende la totalidad de los genes en al menos uno de los paneles A a G. La divulgación también proporciona un método para determinar el pronóstico de cáncer de vejiga, que comprende determinar el estado de un panel de genes que comprende al menos dos CCG (por ejemplo, al menos dos de los genes en cualquiera de los paneles B, C y F), en el que un estado anómalo indica un mal pronóstico. La divulgación también proporciona un método para determinar el pronóstico de cáncer de cerebro, que comprende determinar el estado de un panel de genes que comprende al menos dos CCG (por ejemplo, al menos dos de los genes en cualquiera de los paneles B, C y F), en el que un estado anómalo indica un mal pronóstico. La

divulgación proporciona además un método para determinar el pronóstico de cáncer de mama, que comprende determinar el estado de un panel de genes que comprende al menos dos CCG (por ejemplo, al menos dos de los genes en cualquiera de los paneles B, C y F), en el que un estado anómalo indica un mal pronóstico. La divulgación también proporciona un método para determinar el pronóstico de cáncer de pulmón, que comprende determinar el estado de un panel de genes que comprende al menos dos CCG (por ejemplo, al menos dos de los genes en cualquiera de los paneles B, C y F), en el que un estado anómalo indica un mal pronóstico. En algunas realizaciones, el método comprende al menos una de las siguientes etapas: (a) correlacionar el estado anómalo (por ejemplo, expresión alta o aumentada) del panel de genes con un mal pronóstico; (b) concluir que el paciente tiene un mal pronóstico basándose al menos en parte en el estado anómalo (por ejemplo, expresión alta o aumentada) del panel de genes; o (c) comunicar que el paciente tiene un mal pronóstico basándose al menos en parte en la alta expresión (o expresión aumentada o sobreexpresión) del panel de genes.

En algunas realizaciones, el panel comprende al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG. En algunas realizaciones, el panel comprende entre 5 y 100 CCG, entre 7 y 40 CCG, entre 5 y 25 CCG, entre 10 y 20 CCG o entre 10 y 15 CCG. En algunas realizaciones, los CCG comprenden al menos una determinada proporción del panel. Por tanto, en algunas realizaciones, el panel comprende al menos el 25%, el 30%, el 40%, el 50%, el 60%, el 70%, el 75%, el 80%, el 85%, el 90%, el 95%, el 96%, el 97%, el 98% o el 99% de los CCG. En algunas realizaciones, los CCG se eligen del grupo que consiste en los genes enumerados en las tablas 1, 2, 7-11, 13-14 y/o Y y los paneles A a I. En algunas realizaciones, el panel comprende al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más genes elegidos del grupo de genes en cualesquiera de las tablas 1, 2, 7-11, 13-14 y/o Y o de los paneles A a I. En algunas realizaciones, el panel comprende la totalidad de los genes en cualesquiera de las tablas 1, 2, 7-11, 13-14 y/o Y o de los paneles A a I.

Tal como se mencionó anteriormente, se han analizado muchos de los CCG de la divulgación para determinar su correlación con la media de CCG y también para determinar su valor predictivo relativo dentro de un panel (véanse las tablas 9-11 y 13-14). Las siguientes tablas clasifican los CCG según estos criterios.

Las tablas 9-11 a continuación proporcionan clasificaciones de CCG seleccionados según su correlación con la expresión media de CCG. La tabla 9 proporciona una clasificación de genes de control seleccionados según su correlación con la expresión media de control.

Tabla 9

N.º de gen	Símbolo de gen	Correl. con media	N.º de gen	Símbolo de gen	Correl. con media	N.º de gen	Símbolo de gen	Correl. con media
1	*TPX2*	0.931	44	*PBK*	0.805	87	*KIF18A*	0.6987
2	*CCNB2*	0.9287	45	*ESPL1*	0.805	88	*DONSON*	0.688
3	*KIF4A*	0.9163	46	*MKI67*	0.7993	89	*MCM4*	0.686
4	*KIF2C*	0.9147	47	*SPAG5*	0.7993	90	*RAD54B*	0.679
5	*BIRC5*	0.9077	48	*MCM10*	0.7963	91	*RNASEH2A*	0.6733
6	*BIRC5*	0.9077	49	*MCM6*	0.7957	92	*TUBA1C*	0.6697
7	*RACGAP1*	0.9073	50	*OIP5*	0.7943	93	*C18orf24*	0.6697

8	*CDC2*	0.906	51	*CDC45L*	0.7937	94	*SMC2*	0.6697
9	*PRC1*	0.9053	52	*KIF23*	0.7927	95	*CENPI*	0.6697
10	*DLGAP5* (*DLG7*)	0.9033	53	*EZH2*	0.789	96	*GMPS*	0.6683
11	*CEP55*	0.903	54	*SPC25*	0.7887	97	*DDX39*	0.6673
12	*CCNB1*	0.9	55	*STIL*	0.7843	98	*POLE2*	0.6583
13	*TOP2A*	0.8967	56	*CENPN*	0.783	99	*APOBEC3B*	0.6513
14	*CDC20*	0.8953	57	*GTSE1*	0.7793	100	*RFC2*	0.648
15	*KIF20A*	0.8927	58	*RAD51*	0.779	101	*PSMA7*	0.6473
16	*BUB1B*	0.8927	59	*CDCA3*	0.7783	102	*KIF20B* (*MPHOSPH1*)	0.6457
17	*CDKN3*	0.8887	60	*TACC3*	0.778	103	*CDT1*	0.645
18	*NUSAP1*	0.8873	61	*PLK4*	0.7753	104	*H2AFX*	0.6387
19	*CCNA2*	0.8853	62	*ASF1B*	0.7733	105	*ORC6L*	0.634
20	*KIF11*	0.8723	63	*DTL*	0.769	106	*C1orf135*	0.6333
21	*CDCA8*	0.8713	64	*CHEK1*	0.7673	107	*PSRC1*	0.633
22	*NCAPG*	0.8707	65	*NCAPG2*	0.7667	108	*VRK1*	0.6323
23	*ASPM*	0.8703	66	*PLK1*	0.7657	109	*CKAP2*	0.6307
24	*FOXM1*	0.87	67	*TIMELESS*	0.762	110	*CCDC99*	0.6303
25	*NEK2*	0.869	68	*E2F8*	0.7587	111	*CCNE1*	0.6283
26	*ZWINT*	0.8683	69	*EXO1*	0.758	112	*LMNB2*	0.625
27	*PTTG1*	0.8647	70	*ECT2*	0.744	113	*GPSM2*	0.625
28	*RRM2*	0.8557	71	*STMN1*	0.737	114	*PAICS*	0.6243
29	*TTK*	0.8483	72	*STMN1*	0.737	115	*MCAM*	0.6227
30	*TRIP13*	0.841	73	*RFC4*	0.737	116	*DSN1*	0.622
31	*GINS1*	0.841	74	*CDC6*	0.7363	117	*NCAPD2*	0.6213
32	*CENPF*	0.8397	75	*CENPM*	0.7267	118	*RAD54L*	0.6213
33	*HMMR*	0.8367	76	*MYBL2*	0.725	119	*PDSS1*	0.6203
34	*NCAPH*	0.8353	77	*SHCBP1*	0.723	120	*HN1*	0.62
35	*NDC80*	0.8313	78	*ATAD2*	0.723	121	*C21orf45*	0.6193
36	*KIF15*	0.8307	79	*KIFC1*	0.7183	122	*CTSL2*	0.619
37	*CENPE*	0.8287	80	*DBF4*	0.718	123	*CTPS*	0.6183
38	*TYMS*	0.8283	81	*CKS1B*	0.712	124	*MCM7*	0.618
39	*KIAA0101*	0.8203	82	*PCNA*	0.7103	125	*ZWILCH*	0.618
40	*FANCI*	0.813	83	*FBXO5*	0.7053	126	*RFC5*	0.6177
41	*RAD51AP1*	0.8107	84	*C12orf48*	0.7027			
42	*CKS2*	0.81	85	*TK1*	0.7017			
43	*MCM2*	0.8063	86	*BLM*	0.701			

Tabla 10

N.º de gen	Símbolo de gen	Correl. con media de CCG	N.º de gen	Símbolo de gen	Correl. con media de CCG	N.º de gen	Símbolo de gen	Correl. con media de CCG
1	*DLGAP5*	0.931	12	*C18orf24*	0.885	22	*TOP2A*	0.852
2	*ASPM*	0.931	13	*PLK1*	0.879	23	*KIF20A*	0.851
3	*KIF11*	0.926	14	*CDKN3*	0.874	24	*KIAA0101*	0.839
4	*BIRC5*	0.916	15	*RRM2*	0.871	25	*CDCA3*	0.835
5	*CDCA8*	0.902	16	*RAD51*	0.864	26	*ASF1B*	0.797
6	*CDC20*	0.9	17	*CEP55*	0.862	27	*CENPM*	0.786
7	*MCM10*	0.899	18	*ORC6L*	0.86	28	*TK1*	0.783
8	*PRC1*	0.895	19	*RAD54L*	0.86	29	*PBK*	0.775
9	*BUB1B*	0.892	20	*CDC2*	0.858	30	*PTTG1*	0.751
10	*FOXM1*	0.889	21	*CENPF*	0.855	31	*DTL*	0.737
11	*NUSAP1*	0.888						

Tabla 11: 56 CCG clasificados por correlación con la media en el ejemplo 5 (“panel G”)

N.º de gen	Símbolo de gen	Correl. con media de CCG	N.º de gen	Símbolo de gen	Correl. con media de CCG	N.º de gen	Símbolo de gen	Correl. con media de CCG
1	*FOXM1*	0.908	20	*C18orf24*	0.817	39	*FANCI*	0.702
2	*CDC20*	0.907	21	*RAD54L*	0.816	40	*KIF15*	0.701
3	*CDKN3*	0.9	22	*PTTG1*	0.814	41	*PLK4*	0.688
4	*CDC2*	0.899	23	*KIF4A*	0.814	42	*APOBEC3B*	0.67
5	*KIF11*	0.898	24	*CDCA3*	0.811	43	*NCAPG*	0.667
6	*KIAA0101*	0.89	25	*MCM10*	0.802	44	*TRIP13*	0.653
7	*NUSAP1*	0.887	26	*PRC1*	0.79	45	*KIF23*	0.652
8	*CENPF*	0.882	27	*DTL*	0.788	46	*NCAPH*	0.649
9	*ASPM*	0.879	28	*CEP55*	0.787	47	*TYMS*	0.648
10	*BUB1B*	0.879	29	*RAD51*	0.783	48	*GINS1*	0.639
11	*RRM2*	0.876	30	*CENPM*	0.781	49	*STMN1*	0.63
12	*DLGAP5*	0.875	31	*CDCA8*	0.774	50	*ZWINT*	0.621
13	*BIRC5*	0.864	32	*OIP5*	0.773	51	*BLM*	0.62
14	*KIF20A*	0.86	33	*SHCBP1*	0.762	52	*TTK*	0.62
15	*PLK1*	0.86	34	*ORC6L*	0.736	53	*CDC6*	0.619
16	*TOP2A*	0.851	35	*CCNB1*	0.727	54	*KIF2C*	0.596
17	*TK1*	0.837	36	*CHEK1*	0.723	55	*RAD51AP1*	0.567
18	*PBK*	0.831	37	*TACC3*	0.722	56	*NCAPG2*	0.535
19	*ASF1B*	0.827	38	*MCM4*	0.703			

Tabla 12: 15 genes de mantenimiento (HK) clasificados por correlación con la media en el ejemplo 5

Símbolo de gen	Correlación con media de HK
RPL38	0.989
UBA52	0.986
PSMC1	0.985
RPL4	0.984
RPL37	0.983
RPS29	0.983
SLC25A3	0.982
CLTC	0.981
TXNL1	0.98
PSMA1	0.98
RPL8	0.98
MMADHC	0.979
RPL13A; LOC728658	0.979
PPP2CA	0.978
MRFAP1	0.978

La tabla 13 a continuación proporciona una clasificación de los CCG en el panel F según su valor predictivo relativo en el ejemplo 5.

Tabla 13

N.º de gen	Símbolo de gen	Valor de p	N.º de gen	Símbolo de gen	Valor de p	N.º de gen	Símbolo de gen	Valor de p
1	*MCM10*	8.60E-10	12	*BUB1B*	1.10E-05	23	*C18orf24*	0.0011
2	*ASPM*	2.30E-09	13	*RAD54L*	1.40E-05	24	*BIRC5*	0.00118
3	*DLGAP5*	1.20E-08	14	*CEP55*	2.60E-05	25	*RRM2*	0.00255
4	*CENPF*	1.40E-08	15	*CDCA8*	3.10E-05	26	*CENPM*	0.0027
5	*CDC20*	2.10E-08	16	*TK1*	3.30E-05	27	*RAD51*	0.0028
6	*FOXM1*	3.40E-07	17	*DTL*	3.60E-05	28	*KIAA0101*	0.00348
7	*TOP2A*	4.30E-07	18	*PRC1*	3.90E-05	29	*CDCA3*	0.00863
8	*NUSAP1*	4.70E-07	19	*PTTG1*	4.10E-05	30	*PBK*	0.00923
9	*CDKN3*	5.50E-07	20	*CDC2*	0.00013	31	*ASF1B*	0.00936
10	*KIF11*	6.30E-06	21	*ORC6L*	0.00017			
11	*KIF20A*	6.50E-06	22	*PLK1*	0.0005			

La tabla 14 a continuación proporciona una clasificación de los CCG en el panel C según su valor predictivo relativo en el ejemplo 3.

Tabla 14

N.º de gen	Símbolo de gen	Valor de p*	N.º de gen	Símbolo de gen	Valor de p*	N.º de gen	Símbolo de gen	Valor de p*
1	*NUSAP1*	2.8E-07	12	*BUB1*	8.3E-05	23	*KPNA2*	2.0E-02
2	*DLG7*	5.9E-07	13	*PBK*	1.2E-04	24	*UBE2C*	2.2E-02
3	*CDC2*	6.0E-07	14	*TTK*	3.2E-04	25	*MELK*	2.5E-02

4	*FOXM1*	1.1E-06	15	*CDC45L*	7.7E-04	26	*CENPA*	2.9E-02
5	*MYBL2*	1.1E-06	16	*PRC1*	1.2E-03	27	*CKS2*	5.7E-02
6	*CDCA8*	3.3E-06	17	*DTL*	1.4E-03	28	*MAD2L1*	1.7E-01
7	*CDC20*	3.8E-06	18	*CCNB1*	1.5E-03	29	*UBE2S*	2.0E-01
8	*RRM2*	7.2E-06	19	*TPX2*	1.9E-03	30	*AURKA*	4.8E-01
9	*PTTG1*	1.8E-05	20	*ZWINT*	9.3E-03	31	*TIMELESS*	4.8E-01
10	*CCNB2*	5.2E-05	21	*KIF23*	1.1E-02			
11	*HMMR*	5.2E-05	22	*TRIP13*	1.7E-02			

* Valor de p para la prueba de razón de probabilidades del modelo de RP de Cox completo (puntuación de nomograma posterior a RP + expresión de ciclo celular + nomograma:ciclo celular) frente a reducido (puntuación de nomograma posterior a RP sola) del tiempo hasta la recidiva.

Por tanto, en algunas realizaciones, la pluralidad de genes de prueba comprende al menos cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende los primeros 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40 o más CCG enumerados en las tablas 9-11 y 13-14. En algunas realizaciones, la pluralidad de genes de prueba comprende al menos cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 ó 20 de los siguientes genes: *ASPM*, *BIRC5*, *BUB1B*, *CCNB2*, *CDC2*, *CDC20*, *CDCA8*, *CDKN3*, *CENPF*, *DLGAP5*, *FOXM1*, *KIAA0101*, *KIF11*, *KIF2C*, *KIF4A*, *MCM10*, *NUSAP1*, *PRC1*, *RACGAP1* y *TPX2*. En algunas realizaciones, la pluralidad de genes de prueba comprende al menos cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende cualesquiera uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve o diez o la totalidad de los números de genes 1 y 2, 1 a 3, 1 a 4, 1 a 5, 1 a 6, 1 a 7, 1 a 8, 1 a 9 ó 1 a 10 de cualesquiera de las tablas 9-11 y 13-14. En algunas realizaciones, la pluralidad de genes de prueba comprende al menos cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende cualesquiera uno, dos, tres, cuatro, cinco, seis, siete, ocho o nueve o la totalidad de los números de genes 2 y 3, 2 a 4, 2 a 5, 2 a 6, 2 a 7, 2 a 8, 2 a 9 ó 2 a 10 de cualesquiera de las tablas 9-11 y 13-14. En algunas realizaciones, la pluralidad de genes de prueba comprende al menos cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende cualesquiera uno, dos, tres, cuatro, cinco, seis, siete u ocho o la totalidad de los números de genes 3 y 4, 3 a 5, 3 a 6, 3 a 7, 3 a 8, 3 a 9 ó 3 a 10 de cualesquiera de las tablas 9-11 y 13-14. En algunas realizaciones, la pluralidad de genes de prueba comprende al menos cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende cualesquiera uno, dos, tres, cuatro, cinco, seis, o siete o la totalidad de los números de genes 4 y 5, 4 a 6, 4 a 7, 4 a 8, 4 a 9 ó 4 a 10 de cualesquiera de las tablas 9-11 y 13-14. En algunas realizaciones, la pluralidad de genes de prueba comprende al menos de cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende cualesquiera uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, 10, 11, 12, 13, 14 ó 15 o la totalidad de los números de genes 1 y 2, 1 a 3, 1 a 4, 1 a 5, 1 a 6, 1 a 7, 1 a 8, 1 a 9, 1 a 10, 1 a 11, 1 a 12, 1 a 13, 1 a 14 ó 1 a 15 de cualesquiera de las tablas 9-11 y 13-14.

Además, se ha descubierto que el estado de CCG se añade sinérgicamente a parámetros clínicos en el pronóstico de cáncer. En el caso de cáncer de próstata, por ejemplo, se ha descubierto que un alto nivel de expresión de genes de uno cualquiera de los genes en los paneles C a F está asociado con un riesgo aumentado de recidiva o progresión de cáncer de próstata en pacientes cuya puntuación de nomograma clínico indica un riesgo relativamente bajo de recidiva o progresión. Debido a que evaluar los niveles de expresión de CCG puede detectar, por tanto, un riesgo aumentado no detectado usando los parámetros clínicos solos, la divulgación proporciona generalmente métodos que combinan evaluar al menos un parámetro clínico con evaluar el estado de al menos un CCG.

Tal como muestra el ejemplo 3, incluso los CCG individuales se añaden a parámetros clínicos para predecir la recidiva de cáncer. Por tanto, un aspecto de la divulgación proporciona un método de diagnóstico *in vitro* que comprende determinar al menos un parámetro clínico para un paciente con cáncer y determinar el estado de al menos un CCG en una muestra obtenida del paciente. Sin embargo, evaluar el estado de múltiples CCG mejora la potencia predictiva incluso más (también mostrado en el ejemplo 1). Por tanto, en algunas realizaciones, se determina el estado de una pluralidad de CCG (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50 o más). En algunas realizaciones, el estado anómalo indica una probabilidad aumentada de recidiva o progresión. En algunas realizaciones, el paciente padece cáncer de próstata. En algunas realizaciones, el paciente padece cáncer de pulmón. A menudo, el parámetro clínico es al menos algo independientemente predictivo de recidiva o progresión y la adición del estado de CCG mejora la potencia predictiva. Tal como se usa en el presente documento, “parámetro clínico” y “medida clínica” se refieren a características de enfermedad o de paciente que normalmente se aplican para evaluar la evolución de la enfermedad y/o predecir el desenlace. Los ejemplos en cáncer generalmente incluyen estadio tumoral, grado tumoral, estado ganglionar, histología, estado de rendimiento, tipo de cirugía, márgenes quirúrgicos, tipo de tratamiento y edad de aparición. En el cáncer de próstata, los médicos a menudo usan

los niveles de PSA en sangre antes de la cirugía, el estadio (definido por el tamaño del tumor y la evidencia de metástasis) y la puntuación de Gleason (similar al concepto de grado). Después de la intervención quirúrgica, los parámetros clínicos importantes en el cáncer de próstata incluyen el margen y el estado ganglionar. En el cáncer de mama, los médicos a menudo usan el tamaño de lesión indicadora en cm, la invasión, el número de ganglios afectados y el grado.

A menudo, determinados parámetros clínicos se correlacionan con un carácter de enfermedad particular. Por ejemplo, en el cáncer generalmente así como en cánceres específicos, determinados parámetros clínicos se correlacionan con, por ejemplo, la probabilidad de recidiva o metástasis, el pronóstico de la supervivencia durante un determinado periodo de tiempo, la probabilidad de respuesta al tratamiento generalmente o a un tratamiento específico, etc. En el cáncer de próstata, algunos parámetros clínicos son de tal manera que su estado (presencia, ausencia, nivel, etc.) está asociado con la probabilidad aumentada de recidiva. Los ejemplos de tales parámetros asociados a la recidiva (algunos de los cuales, pero no todos, son específicos del cáncer de próstata) incluyen altos niveles de PSA (por ejemplo, mayores de 4 ng/ml), alta puntuación de Gleason, tamaño tumoral grande, evidencia de metástasis, estadio tumoral avanzado, grado nuclear, afectación ganglionar, edad temprana de aparición. Otros tipos de cáncer pueden tener parámetros diferentes correlacionados con la probabilidad de recidiva o progresión, y el estado de CCG, como medida de actividad proliferativa, se añade a estos parámetros para predecir el pronóstico en estos cánceres. Tal como se usa en el presente documento, “parámetro clínico asociado a recidiva” tiene su significado convencional para cada cáncer específico, con lo que los expertos en la técnica están bastante familiarizados. De hecho, los expertos en la técnica están familiarizados con diversos parámetros clínicos asociados a la recidiva más allá de los enumerados en el presente documento.

A menudo, un médico evaluará más de un parámetro clínico en un paciente y realizará una evaluación más completa para los caracteres de enfermedad de interés. El ejemplo 5 muestra cómo puede añadirse el estado de CCG a una agrupación particular de parámetros clínicos usados para determinar el riesgo de recidiva en el cáncer de próstata. Los parámetros clínicos en el ejemplo 5 incluyen variables binarias para la enfermedad confinada en órgano y la puntuación de Gleason menor de o igual a 6, y una variable continua para el PSA logarítmico (tabla 1). Este modelo incluye todos los parámetros clínicos incorporados en el nomograma posterior a RP (es decir, nomograma de Kattan-Stephenson) excepto para el año de RP y los dos componentes de la puntuación de Gleason. Por tanto, en algunas realizaciones, se evalúan al menos dos parámetros clínicos (por ejemplo, dos de los parámetros anteriormente enumerados) junto con el nivel de expresión de al menos un CCG.

Un modo en el que los médicos utilizan parámetros clínicos individuales, pero más a menudo múltiples, es con la ayuda de nomogramas. En el ámbito clínico, los nomogramas son representaciones (a menudo visuales) de una correlación entre uno o más parámetros y uno o más caracteres de paciente o de enfermedad. Un ejemplo de un nomograma clínico prevalente usado para determinar la probabilidad de recidiva en un paciente con cáncer de próstata se describe en Kattan *et al*., J. CLIN. ONCOL. (1999) 17:1499-1507, y se actualiza en Stephenson *et al*., J. CLIN. ONCOL. (2005) 23:7005-7012 (“nomograma de Kattan-Stephenson”). Este nomograma evalúa un paciente asignando un valor de punto a cada uno de varios parámetros clínicos (año de RP, márgenes quirúrgicos, extensión extracapsular, invasión de vesículas seminales, afectación ganglionar, puntuación primaria de Gleason, puntuación secundaria de Gleason y nivel de PSA preoperatorio), sumando los puntos para un paciente en una puntuación de nomograma, y luego prediciendo la probabilidad del paciente de estar libre de recidiva a intervalos de tiempo variables (hasta 10 años) basándose en esta puntuación de nomograma. Un ejemplo de un nomograma clínico prevalente usado para determinar el pronóstico de la supervivencia en un paciente con cáncer de mama es el índice de pronóstico de Nottingham (NPI). Véase, por ejemplo, Galea *et al*., BREAST CANCER RES. & TREAT. (1992) 22:207-19.

Se ha descubierto que determinar el estado de un CCG en una muestra obtenida de un paciente con cáncer de próstata, junto con la puntuación de nomograma de Kattan-Stephenson del paciente, es un mejor predictor de la supervivencia libre de recidiva a los 10 años que la puntuación de nomograma sola. Véanse, por ejemplo, los ejemplos 2 y 5 a continuación. Específicamente, añadir el estado de CCG al nomograma de Kattan-Stephenson detecta pacientes con riesgo de recidiva significativamente aumentado que el nomograma por sí solo no detecta. La tabla 7 anterior proporciona un panel a modo de ejemplo de 31 CCG (panel C) y un panel de subconjuntos de 26 CCG (panel D, mostrado con *) determinados en el ejemplo 2 para mostrar la sinergia predictiva con el nomograma de Kattan-Stephenson en el pronóstico de cáncer de próstata. También se ha descubierto que determinar el estado de un CCG en una muestra obtenida de un paciente con cáncer de mama, junto con la puntuación de NPI del paciente, es un mejor predictor de pronóstico que la puntuación de NPI sola. Véase, por ejemplo, el ejemplo 6 a continuación. Específicamente, añadir el estado de CCG al nomograma de NPI detecta pacientes con riesgo de recidiva significativamente aumentado que el nomograma por sí solo no detecta. Se determinó en el ejemplo 2 que los paneles B, C y D muestran una sinergia predictiva con el nomograma de NPI en el pronóstico de cáncer de mama.

Por tanto, otro aspecto de la divulgación proporciona un método *in vitro* que comprende determinar una puntuación de nomograma clínico (por ejemplo, puntuación de nomograma de Kattan-Stephenson o de NPI) para un paciente con cáncer y determinar el estado de al menos un CCG en una muestra obtenida del paciente. El ejemplo 3 ilustra la determinación empírica de la potencia predictiva de CCG individuales y de varios paneles de CCG de tamaño

variable a lo largo del nomograma de Kattan-Stephenson. Se sometieron a prueba subconjuntos seleccionados aleatoriamente de los 31 CCG enumerados en la tabla 7 como firmas de CCG distintas y se determinó la potencia predictiva (es decir, el valor de p) para cada una. Tal como muestra la figura 1, las firmas de CCG de 2, 3, 4, 5, 6, 10, 15, 20, 25 y 26 genes añaden, cada una, potencia predictiva al nomograma. Por tanto, la divulgación proporciona un método para determinar si un paciente con cáncer de próstata tiene una probabilidad aumentada de recidiva que comprende determinar el estado de un panel de genes que comprende al menos 2, 3, 4, 5, 6, 10, 15, 20, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 40, 45, 50, 60, 70, 80, 90 ó 100 o más CCG, en el que un estado elevado (por ejemplo, expresión aumentada) para los CCG indica una probabilidad aumentada de recidiva. En algunas realizaciones, el método comprende además determinar una puntuación de nomograma clínico del paciente. La divulgación proporciona además un método para determinar si un paciente con cáncer de mama tiene una probabilidad aumentada de recidiva que comprende determinar el estado de un panel de genes que comprende al menos 2, 3, 4, 5, 6, 10, 15, 20, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 40, 45, 50, 60, 70, 80, 90 ó 100 o más CCG, en el que un estado elevado (por ejemplo, expresión aumentada) para los CCG indica una probabilidad aumentada de recidiva. En algunas realizaciones, el método comprende además determinar una puntuación de nomograma clínico del paciente. En algunas realizaciones, el método comprende al menos una de las siguientes etapas: (a) correlacionar el estado elevado (por ejemplo, expresión alta o aumentada) del panel de genes con una probabilidad aumentada de recidiva; (b) concluir que el paciente tiene una probabilidad aumentada de recidiva basándose al menos en parte en el estado elevado (por ejemplo, expresión alta o aumentada) del panel de genes; o (c) comunicar que el paciente tiene una probabilidad aumentada de recidiva basándose al menos en parte en el estado elevado (por ejemplo, expresión alta o aumentada) del panel de genes.

A menudo, los nomogramas clínicos para el cáncer se diseñan de tal manera que un valor particular (por ejemplo, alta puntuación) se correlaciona con un riesgo aumentado de recidiva. El estado de CCG elevado (por ejemplo, expresión o actividad aumentada) también se correlaciona con un riesgo aumentado. Por tanto, en algunas realizaciones, la divulgación proporciona un método para determinar si un paciente con cáncer tiene una probabilidad aumentada de recidiva o progresión que comprende determinar una puntuación de nomograma clínico para el paciente y determinar el estado de al menos un CCG en una muestra obtenida del paciente, en el que una alta puntuación de nomograma y/o un estado de CCG elevado indican que el paciente tiene una probabilidad aumentada de recidiva o progresión. En algunas realizaciones, el cáncer es cáncer de próstata. En algunas realizaciones, el cáncer es cáncer de pulmón. En algunas realizaciones, el método comprende al menos una de las siguientes etapas: (a) correlacionar una alta puntuación de nomograma y/o un estado de CCG elevado (por ejemplo, expresión alta o aumentada) con una probabilidad aumentada de recidiva o progresión; (b) concluir que el paciente tiene una probabilidad aumentada de recidiva o progresión basándose al menos en parte en una alta puntuación de nomograma y/o un estado de CCG elevado (por ejemplo, expresión alta o aumentada); o (c) comunicar que el paciente tiene una probabilidad aumentada de recidiva o progresión basándose al menos en parte en una alta puntuación de nomograma y/o un estado de CCG elevado (por ejemplo, expresión alta o aumentada).

En algunas realizaciones, esta evaluación se realiza antes de la prostatectomía radical (por ejemplo, usando una muestra de biopsia de próstata), mientras que en algunas realizaciones se realiza después (por ejemplo, usando la muestra de próstata resecada). En algunas realizaciones, una muestra de una o más células se obtiene de un paciente con cáncer de próstata antes o después del tratamiento para su análisis según la presente divulgación. El tratamiento de cáncer de próstata aplicado actualmente en la técnica incluye, por ejemplo, prostatectomía, radioterapia, terapia hormonal (por ejemplo, usando antagonistas de GnRH, agonistas de GnRH, antiandrógenos), quimioterapia y ultrasonidos enfocados de alta intensidad. En algunas realizaciones, una o más células tumorales de próstata de tejido de cáncer de próstata se obtienen de un paciente con cáncer de próstata durante la biopsia o prostatectomía y se usan para su análisis en el método de la presente divulgación.

La presente divulgación también se basa en el descubrimiento de que el estado de *PTEN* predice cáncer de próstata agresivo. El estado de *PTEN* se añade tanto a los parámetros clínicos (por ejemplo, nomograma de Kattan-Stephenson) como a los CCG (por ejemplo, los genes en la tabla 1 o en los paneles A a G). Tal como se describe con más detalle en el ejemplo 4 a continuación, se determinó el estado de *PTEN* en 191 muestras de pacientes con cáncer de próstata con datos de antecedentes clínicos y datos de firma de CCG complementarios. Se halló que un estado de *PTEN* negativo era un predictor significativo del riesgo de recidiva (valor de p de 0,031). *PTEN* siguió siendo un predictor significativo de recidiva después de ajustar para parámetros clínicos posteriores a la cirugía y la firma de CCG mostrada en la tabla 7 (valor de p de 0,026). Además, y de manera importante, la combinación de *PTEN* y la firma de CCG parece ser un mejor predictor de recidiva que los parámetros clínicos posteriores a la cirugía (valor de p de 0,0002).

Debido a que *PTEN* es un predictor independiente de recidiva de cáncer de próstata, un aspecto de la divulgación proporciona un método para predecir la probabilidad de recidiva de cáncer de próstata en un paciente que comprende determinar el estado de *PTEN* en una muestra del paciente, en el que un estado de *PTEN* bajo o negativo indica que el paciente tiene una probabilidad aumentada de recidiva. En algunas realizaciones, el método comprende al menos una de las siguientes etapas: (a) correlacionar el estado de *PTEN* bajo o negativo (por ejemplo, expresión baja o negativa) con una probabilidad aumentada de recidiva; (b) concluir que el paciente tiene una probabilidad aumentada de recidiva basándose al menos en parte en el estado de *PTEN* bajo o negativo (por ejemplo, expresión baja o negativa); o (c) comunicar que el paciente tiene una probabilidad aumentada de recidiva

basándose al menos en parte en el estado de *PTEN* bajo o negativo (por ejemplo, expresión baja o negativa). El estado de *PTEN* puede determinarse mediante cualquier técnica conocida en la técnica, incluyendo, pero sin limitarse a, aquellas comentadas en el presente documento.

Debido a que *PTEN* se añade al estado de CCG para predecir la recidiva de cáncer de próstata, otro aspecto de la divulgación proporciona un método *in vitro* que comprende determinar el estado de *PTEN* y determinar el estado de una pluralidad de CCG en una muestra obtenida de un paciente. Pueden usarse diferentes combinaciones de técnicas para determinar el estado de los diversos marcadores. Por ejemplo, en una realización, el estado de *PTEN* se determina mediante inmunohistoquímica (IHC) mientras que el estado de la pluralidad de CCG se determina mediante reacción en cadena de la polimerasa cuantitativa (qPCR™), por ejemplo, TaqMan™. Algunas realizaciones de la divulgación proporcionan un método para determinar la probabilidad de recidiva en un paciente con cáncer de próstata que comprende determinar el estado de *PTEN* en una muestra obtenida del paciente, determinar el estado de una pluralidad de CCG en una muestra obtenida del paciente, en el que el estado de *PTEN* bajo o negativo y/o el estado de CCG elevado indican que el paciente tiene una probabilidad aumentada de recidiva. En algunas realizaciones, el método comprende al menos una de las siguientes etapas: (a) correlacionar el estado de *PTEN* bajo o negativo (por ejemplo, expresión baja o negativa) y/o el estado de CCG elevado (por ejemplo, expresión alta o aumentada) con una probabilidad aumentada de recidiva; (b) concluir que el paciente tiene una probabilidad aumentada de recidiva basándose al menos en parte en el estado de *PTEN* bajo o negativo (por ejemplo, expresión baja o negativa) y/o el estado de CCG elevado (por ejemplo, expresión alta o aumentada); o (c) comunicar que el paciente tiene una probabilidad aumentada de recidiva basándose al menos en parte en el estado de *PTEN* bajo o negativo (por ejemplo, expresión baja o negativa) y/o el estado de CCG elevado (por ejemplo, expresión alta o aumentada).

Debido a que el estado de *PTEN* añade valor predictivo a parámetros clínicos para predecir la recidiva de cáncer de próstata, aún otro aspecto de la divulgación proporciona un método *in vitro* que comprende determinar el estado de *PTEN* y determinar al menos un parámetro clínico para un paciente con cáncer. A menudo, el parámetro clínico es al menos algo independientemente predictivo de recidiva y la adición del estado de *PTEN* mejora la potencia predictiva. En algunas realizaciones, la divulgación proporciona un método para determinar si un paciente con cáncer tiene una probabilidad aumentada de recidiva que comprende determinar el estado de *PTEN* en una muestra obtenida del paciente y determinar una puntuación de nomograma clínico para el paciente, en el que un estado de *PTEN* bajo o negativo y/o una puntuación de nomograma desfavorable (por ejemplo, alta) indican que el paciente tiene una probabilidad aumentada de recidiva. En algunas realizaciones, el método comprende al menos una de las siguientes etapas: (a) correlacionar el estado de *PTEN* bajo o negativo (por ejemplo, expresión baja o negativa) y/o la puntuación de nomograma desfavorable (por ejemplo, alta) con una probabilidad aumentada de recidiva; (b) concluir que el paciente tiene una probabilidad aumentada de recidiva basándose al menos en parte en el estado de *PTEN* bajo o negativo (por ejemplo, expresión baja o negativa) y/o la puntuación de nomograma desfavorable (por ejemplo, alta); o (c) comunicar que el paciente tiene una probabilidad aumentada de recidiva basándose al menos en parte en el estado de *PTEN* bajo o negativo (por ejemplo, expresión baja o negativa) y/o la puntuación de nomograma desfavorable (por ejemplo, alta).

Debido a que los tres marcadores anteriores son aditivos, algunas realizaciones de la divulgación proporcionan un método para determinar si un paciente con cáncer tiene una probabilidad aumentada de recidiva que comprende determinar el estado de *PTEN* en una muestra obtenida del paciente, determinar una puntuación de nomograma clínico para el paciente y determinar el estado de al menos un CCG en una muestra obtenida del paciente, en el que un estado de *PTEN* bajo o negativo, una puntuación de nomograma desfavorable (por ejemplo, alta) y/o un estado de CCG elevado indican que el paciente tiene una probabilidad aumentada de recidiva. En algunas realizaciones, el método comprende al menos una de las siguientes etapas: (a) correlacionar el estado de *PTEN* bajo o negativo (por ejemplo, expresión baja o negativa), la puntuación de nomograma desfavorable (por ejemplo, alta) y/o el estado de CCG elevado (por ejemplo, expresión alta o aumentada) con una probabilidad aumentada de recidiva; (b) concluir que el paciente tiene una probabilidad aumentada de recidiva basándose al menos en parte en el estado de *PTEN* bajo o negativo (por ejemplo, expresión baja o negativa), la puntuación de nomograma desfavorable (por ejemplo, alta) y/o el estado de CCG elevado (por ejemplo, expresión alta o aumentada); o (c) comunicar que el paciente tiene una probabilidad aumentada de recidiva basándose al menos en parte en el estado de *PTEN* bajo o negativo (por ejemplo, expresión baja o negativa), la puntuación de nomograma desfavorable (por ejemplo, alta) y/o el estado de CCG elevado (por ejemplo, expresión alta o aumentada).

La presente divulgación también se basa en el descubrimiento de que el estado de ARN de peptidasa relacionada con calicreína 3 (*KLK3*) predice cáncer de próstata agresivo. *KLK3* (ID de gen de Entrez n.º 354) es el gen que codifica para la proteína PSA. El estado de *KLK3* se añade tanto a parámetros clínicos (por ejemplo, nomograma de Kattan-Stephenson) como a los CCG (por ejemplo, los genes en la tabla 1 o en los paneles A a G). Tal como se describe con más detalle en los ejemplos 7 y 9 a continuación, se midió la expresión de ARN de *KLK3* en muestras de pacientes con cáncer de próstata con datos de antecedentes clínicos y datos de firma de CCG complementarios. Cabe destacar que la expresión de ARN de *KLK3* en FFPE no se correlacionó bien con los niveles de proteína PSA en suero. Se halló que la expresión de *KLK3* disminuida era un predictor significativo del riesgo de recidiva (valor de $p < 0,0005$). *KLK3* fue un predictor significativo de recidiva independiente de los parámetros clínicos posteriores a la cirugía (por ejemplo, puntuación de Gleason, PSA) y la firma de CCG mostrada en el panel F (valor de p de 2×10^{-6}).

Se analizaron otros genes específicamente para determinar su capacidad para añadir potencia de pronóstico más allá de la puntuación de CCP y las variables clínicas. Estos genes se encuentran en las tablas R, S e Y a continuación y forman los paneles H e I de la divulgación. Cada uno de estos genes puede usarse independientemente para diagnosticar el pronóstico de recidiva de cáncer o muerte específica de cáncer en un paciente según los métodos, sistemas, kits, etc., de la divulgación comentados en el presente documento. O uno o más de estos genes pueden añadirse a un panel de la divulgación que comprende genes de CCP para formar un panel más grande con potencia predictiva mejorada.

Debido a que *KLK3* y los genes del panel H o I son predictores independientes de recidiva de cáncer y muerte específica de cáncer, un aspecto de la divulgación proporciona un método para predecir el pronóstico de un paciente (por ejemplo, probabilidad de recidiva de cáncer de próstata o muerte específica de cáncer) que comprende determinar el estado de *KLK3* y/o el estado de uno o más genes en el panel H o I en una muestra del paciente, en el que un estado anómalo (por ejemplo, expresión disminuida, expresión aumentada) indica que el paciente tiene un mal pronóstico (por ejemplo, alta probabilidad de recidiva o muerte específica de cáncer). En algunas realizaciones, el método comprende al menos una de las siguientes etapas: (a) correlacionar el estado anómalo (por ejemplo, expresión de ARNm disminuida) con un mal pronóstico (por ejemplo, alta probabilidad de recidiva o muerte específica de cáncer); (b) concluir que el paciente tiene un mal pronóstico (por ejemplo, alta probabilidad de recidiva o muerte específica de cáncer) basándose al menos en parte en el estado anómalo (por ejemplo, expresión de ARNm disminuida); o (c) comunicar que el paciente tiene un mal pronóstico (por ejemplo, alta probabilidad de recidiva o muerte específica de cáncer) basándose al menos en parte en el estado anómalo (por ejemplo, expresión de ARNm disminuida). El estado de *KLK3* o el estado de uno o más genes en el panel H o I puede determinarse aplicando y adaptando técnicas conocidas en la técnica, incluyendo, pero sin limitarse a, aquellas comentadas en el presente documento. En algunas realizaciones, la expresión de ARN se mide, por ejemplo, midiendo directamente los niveles de ARN o midiendo los niveles de ADNc derivado de ARN.

Debido a que *KLK3* y cada uno de los genes en el panel H o I se añade al estado de CCG para predecir la recidiva de cáncer y la muerte específica de cáncer, otro aspecto de la divulgación proporciona un método *in vitro* que comprende determinar el estado de *KLK3* y/o el estado de uno o más genes en el panel H o I y determinar el estado de una pluralidad de CCG en una muestra obtenida de un paciente. Algunas realizaciones de la divulgación proporcionan un método para determinar el pronóstico de un paciente con cáncer de próstata que comprende determinar la expresión de *KLK3* y/o la expresión de uno o más genes en el panel H o I en una muestra obtenida del paciente, determinar la expresión de una pluralidad de CCG en una muestra obtenida del paciente, en el que el estado de *KLK3* anómalo (por ejemplo, expresión de ARNm disminuida) y/o el estado anómalo de uno o más genes en el panel H o I (por ejemplo, expresión de ARNm aumentada) y/o el estado de CCG elevado indican que el paciente tiene un mal pronóstico. En algunas realizaciones, el método comprende al menos una de las siguientes etapas: (a) correlacionar el estado de *KLK3* anómalo (por ejemplo, expresión de ARNm disminuida) y/o el estado anómalo de uno o más genes en el panel H o I y/o el estado de CCG elevado (por ejemplo, expresión alta o aumentada) con un mal pronóstico (por ejemplo, alta probabilidad de recidiva o muerte específica de cáncer); (b) concluir que el paciente tiene un mal pronóstico (por ejemplo, alta probabilidad de recidiva o muerte específica de cáncer) basándose al menos en parte en el estado de *KLK3* anómalo (por ejemplo, expresión de ARNm disminuida) y/o el estado anómalo de uno o más genes en el panel H o I y/o el estado de CCG elevado (por ejemplo, expresión alta o aumentada); o (c) comunicar que el paciente tiene un mal pronóstico (por ejemplo, alta probabilidad de recidiva o muerte específica de cáncer) basándose al menos en parte en el estado de *KLK3* anómalo (por ejemplo, expresión de ARNm disminuida) y/o el estado anómalo de uno o más genes en el panel H o I y/o el estado de CCG elevado (por ejemplo, expresión alta o aumentada).

Debido a que el estado de *KLK3* y cada uno de los genes en el panel H o I añade valor predictivo a parámetros clínicos para predecir la recidiva de cáncer de próstata, aún otro aspecto de la divulgación proporciona un método *in vitro* que comprende determinar el estado de *KLK3* y/o el estado de uno o más genes en el panel H o I y determinar al menos un parámetro clínico para un paciente con cáncer. A menudo, el parámetro clínico es al menos algo independientemente predictivo de recidiva y la adición del estado de *KLK3* y/o el estado de uno o más genes en el panel H o I mejora la potencia predictiva. En algunas realizaciones, la divulgación proporciona un método para predecir el pronóstico de un paciente (por ejemplo, probabilidad de recidiva de cáncer de próstata o muerte específica de cáncer) que comprende determinar la expresión de *KLK3* en una muestra obtenida del paciente y/o el estado de uno o más genes en el panel H o I y determinar una puntuación clínica para el paciente, en el que un estado de *KLK3* anómalo (por ejemplo, expresión de ARNm disminuida) y/o un estado anómalo de uno o más genes en el panel H o I y/o una puntuación clínica desfavorable (por ejemplo, alta) indican que el paciente tiene un mal pronóstico (por ejemplo, probabilidad aumentada de recidiva de cáncer de próstata o muerte específica de cáncer). En algunas realizaciones, el método comprende al menos una de las siguientes etapas: (a) correlacionar el estado de *KLK3* anómalo (por ejemplo, expresión de ARNm disminuida) y/o el estado anómalo de uno o más genes en el panel H o I y/o la puntuación clínica desfavorable (por ejemplo, alta) con un mal pronóstico (por ejemplo, alta probabilidad de recidiva o muerte específica de cáncer); (b) concluir que el paciente tiene un mal pronóstico (por ejemplo, alta probabilidad de recidiva o muerte específica de cáncer) basándose al menos en parte en el estado de *KLK3* anómalo (por ejemplo, expresión de ARNm disminuida) y/o el estado anómalo de uno o más genes en el panel H o I y/o la puntuación clínica desfavorable (por ejemplo, alta); o (c) comunicar que el paciente tiene un mal

pronóstico (por ejemplo, alta probabilidad de recidiva o muerte específica de cáncer) basándose al menos en parte en el estado de *KLK3* anómalo (por ejemplo, expresión de ARNm disminuida) y/o el estado anómalo de uno o más genes en el panel H o I y/o la puntuación clínica desfavorable (por ejemplo, alta).

Debido a que los cuatro marcadores anteriores son aditivos, algunas realizaciones de la divulgación proporcionan un método para predecir el pronóstico de un paciente (por ejemplo, probabilidad de recidiva de cáncer de próstata o muerte específica de cáncer) que comprende determinar el estado de *PTEN* en una muestra obtenida del paciente, determinar la expresión de *KLK3* en una muestra obtenida del paciente, determinar una puntuación de nomograma clínico para el paciente y determinar el estado de una pluralidad de CCG (por ejemplo, panel F) en una muestra obtenida del paciente, en el que cualquiera de (1) un estado de *PTEN* bajo o negativo, (2) un estado de *KLK3* anómalo (por ejemplo, expresión de ARNm disminuida), (3) una puntuación de nomograma desfavorable (por ejemplo, alta) y/o (4) un estado de CCG elevado indican que el paciente tiene un mal pronóstico (por ejemplo, probabilidad aumentada de recidiva de cáncer de próstata o muerte específica de cáncer). En algunas realizaciones, el método comprende al menos una de las siguientes etapas: (a) correlacionar el estado de *PTEN* bajo o negativo, el estado de *KLK3* anómalo (por ejemplo, expresión de ARNm disminuida), la puntuación de nomograma desfavorable (por ejemplo, alta) y/o el estado de CCG elevado con un mal pronóstico (por ejemplo, alta probabilidad de recidiva o muerte específica de cáncer); (b) concluir que el paciente tiene un mal pronóstico (por ejemplo, alta probabilidad de recidiva o muerte específica de cáncer) basándose al menos en parte en el estado de *PTEN* bajo o negativo, el estado de *KLK3* anómalo (por ejemplo, expresión de ARNm disminuida), la puntuación de nomograma desfavorable (por ejemplo, alta) y/o el estado de CCG elevado; o (c) comunicar que el paciente tiene un mal pronóstico (por ejemplo, alta probabilidad de recidiva o muerte específica de cáncer) basándose al menos en parte en el estado de *PTEN* bajo o negativo, el estado de *KLK3* anómalo (por ejemplo, expresión de ARNm disminuida), la puntuación de nomograma desfavorable (por ejemplo, alta) y/o el estado de CCG elevado. Determinar el estado de uno o más genes en el panel H o I también puede añadirse a cualquiera de estos análisis, indicando un estado anómalo (por ejemplo, alta expresión) un mal pronóstico.

Los genes en las tablas R, S e Y se clasifican según su valor de p (por ejemplo, después de ajustar para la puntuación de CCP). Por tanto, los diversos aspectos de la divulgación que implican estos genes (por ejemplo, los diversos párrafos anteriores) pueden incorporar estos genes según esta clasificación. En algunas realizaciones, la pluralidad de genes de prueba comprende al menos cierto número de genes de cualesquiera de las tablas R, S o Y (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 o más) y esta pluralidad de genes comprende los primeros 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35 o más genes enumerados en la tabla Y. En algunas realizaciones, la pluralidad de genes de prueba comprende al menos cierto número de genes de cualesquiera de las tablas R, S o Y (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 o más) y esta pluralidad de genes comprende cualesquiera uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve o diez o la totalidad de los números de genes 1 y 2, 1 a 3, 1 a 4, 1 a 5, 1 a 6, 1 a 7, 1 a 8, 1 a 9 ó 1 a 10 de cualesquiera de las tablas R, S o Y. En algunas realizaciones, la pluralidad de genes de prueba comprende al menos cierto número de genes de cualesquiera de las tablas R, S o Y (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 o más) y esta pluralidad de genes comprende cualesquiera uno, dos, tres, cuatro, cinco, seis, siete, ocho o nueve o la totalidad de los números de genes 2 y 3, 2 a 4, 2 a 5, 2 a 6, 2 a 7, 2 a 8, 2 a 9 ó 2 a 10 de cualesquiera de las tablas R, S o Y. En algunas realizaciones, la pluralidad de genes de prueba comprende al menos cierto número de genes de cualesquiera de las tablas R, S o Y (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 o más) y esta pluralidad de genes comprende cualesquiera uno, dos, tres, cuatro, cinco, seis, siete u ocho o la totalidad de los números de genes 3 y 4, 3 a 5, 3 a 6, 3 a 7, 3 a 8, 3 a 9 ó 3 a 10 de cualesquiera de las tablas R, S o Y. En algunas realizaciones, la pluralidad de genes de prueba comprende al menos cierto número de genes de cualesquiera de las tablas R, S o Y (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 o más) y esta pluralidad de genes comprende cualesquiera uno, dos, tres, cuatro, cinco, seis o siete o la totalidad de los números de genes 4 y 5, 4 a 6, 4 a 7, 4 a 8, 4 a 9 ó 4 a 10 de cualesquiera de las tablas R, S o Y. En algunas realizaciones, la pluralidad de genes de prueba comprende al menos cierto número de genes de cualesquiera de las tablas R, S o Y (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 o más) y esta pluralidad de genes comprende cualesquiera uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, 10, 11, 12, 13, 14 ó 15 o la totalidad de los números de genes 1 y 2, 1 a 3, 1 a 4, 1 a 5, 1 a 6, 1 a 7, 1 a 8, 1 a 9, 1 a 10, 1 a 11, 1 a 12, 1 a 13, 1 a 14 ó 1 a 15 de cualesquiera de las tablas R, S o Y.

A menudo, los resultados de cualquier análisis según la divulgación se comunicarán a los médicos, asesores genéticos y/o pacientes (u otras partes interesadas tales como investigadores) en una forma transmisible que puede comunicarse o transmitirse a cualquiera de las partes anteriores. Una forma de este tipo puede variar y puede ser tangible o intangible. Los resultados pueden representarse en declaraciones descriptivas, diagramas, fotografías, esquemas, imágenes o cualquier otra forma visual. Por ejemplo, para explicar los resultados pueden usarse gráficos que muestran el nivel de expresión o de actividad o la información de variación de secuencia para diversos genes. Los diagramas que muestran tal información para gen(es) diana adicional(es) también son útiles para indicar algunos resultados de prueba. Las declaraciones y formas visuales pueden grabarse en un medio tangible tal como papeles, medios legibles por ordenador tales como discos flexibles, discos compactos, etc., o en un medio intangible, por ejemplo, un medio electrónico en forma de correo electrónico o sitio web en Internet o Intranet. Además, los resultados también pueden grabarse en forma de sonido y transmitirse a través de cualquier medio adecuado, por ejemplo, líneas de cable analógicas o digitales, cables de fibra óptica, etc., a través del teléfono,

facsímil, teléfono móvil inalámbrico, teléfono por Internet, y similares.

Por tanto, la información y los datos sobre un resultado de prueba pueden producirse en cualquier parte del mundo y transmitirse a una ubicación diferente. Como un ejemplo ilustrativo, cuando se lleva a cabo un ensayo de nivel de expresión, nivel de actividad o secuenciación (o genotipado) fuera de los Estados Unidos, la información y los datos sobre un resultado de prueba pueden generarse, emitirse en una forma transmisible tal como se describió anteriormente y luego importarse a los Estados Unidos. Por consiguiente, la presente divulgación también abarca un método para producir una forma transmisible de información sobre al menos uno de (a) nivel de expresión o (b) nivel de actividad para al menos una muestra de paciente. El método comprende las etapas de (1) determinar al menos uno de (a) o (b) anteriores según los métodos de la presente divulgación; y (2) incluir el resultado de la etapa de determinación en una forma transmisible. La forma transmisible es el producto de un método de este tipo.

A menudo se implementarán técnicas para analizar tales datos de expresión, actividad y/o secuencia (de hecho, cualquier dato obtenido según la divulgación) usando hardware, software o una combinación de los mismos en uno o más sistemas informáticos u otros sistemas de procesamiento capaces de efectuar tal análisis.

Por tanto, la presente divulgación proporciona además un sistema para determinar la expresión de genes en una muestra tumoral, que comprende: (1) un analizador de muestras para determinar los niveles de expresión de un panel de genes en una muestra tumoral que incluye al menos 2, 4, 6, 8 ó 10 genes del ciclo celular, en el que el analizador de muestras contiene la muestra tumoral que es de un paciente que se ha identificado que padece cáncer de próstata, cáncer de pulmón, cáncer de vejiga o cáncer de cerebro, o moléculas de ADNc del ARNm expresado a partir del panel de genes; (2) un primer programa informático para (a) recibir datos de expresión de genes sobre al menos 4 genes de prueba seleccionados del panel de genes, (b) ponderar la expresión determinada de cada uno de los genes de prueba y (c) combinar la expresión ponderada para proporcionar un valor de prueba, en el que al menos el 20%, el 50%, al menos el 75% o al menos el 90% de los genes de prueba son genes del ciclo celular; y opcionalmente (3) un segundo programa informático para comparar el valor de prueba con uno o más valores de referencia asociados, cada uno, con un grado predeterminado de riesgo de recidiva o progresión de cáncer del cáncer de próstata, cáncer de pulmón, cáncer de vejiga o cáncer de cerebro. En algunas realizaciones, el sistema comprende además un módulo de visualización que visualiza la comparación entre el valor de prueba con el uno o más valores de referencia, o que visualiza un resultado de la etapa de comparación.

En una realización preferida, se mide la cantidad de ARN transcrito a partir del panel de genes que incluye genes de prueba en la muestra tumoral. Además, también se mide la cantidad de ARN de uno o más genes de mantenimiento en la muestra tumoral y se usa para normalizar o calibrar la expresión de los genes de prueba, tal como se describió anteriormente.

En algunas realizaciones, la pluralidad de genes de prueba incluye al menos 2, 3 ó 4 genes del ciclo celular, que constituyen al menos el 50%, el 75% o el 80% de la pluralidad de genes de prueba, y preferiblemente el 100% de la pluralidad de genes de prueba. En algunas realizaciones, la pluralidad de genes de prueba incluye al menos 5, 6 ó 7, o al menos 8, genes del ciclo celular, que constituyen al menos el 20%, el 25%, el 30%, el 40%, el 50%, el 60%, el 70%, el 75%, el 80% o el 90% de la pluralidad de genes de prueba, y preferiblemente el 100% de la pluralidad de genes de prueba.

En algunas otras realizaciones, la pluralidad de genes de prueba incluye al menos 8, 10, 12, 15, 20, 25 ó 30 genes del ciclo celular, que constituyen al menos el 20%, el 25%, el 30%, el 40%, el 50%, el 60%, el 70%, el 75%, el 80% o el 90% de la pluralidad de genes de prueba, y preferiblemente el 100% de la pluralidad de genes de prueba.

El analizador de muestras puede ser cualquier instrumento útil para determinar la expresión de genes, incluyendo, por ejemplo, una máquina de secuenciación, una máquina de PCR en tiempo real y un instrumento de microalineamientos.

La función de análisis basada en ordenador puede implementarse en cualquier lenguaje y/o navegador adecuado. Por ejemplo, puede implementarse con lenguaje C y preferiblemente usando lenguajes de programación de alto nivel orientados a objetos tales como Visual Basic, Small Talk, C++, y similares. La aplicación puede escribirse para adecuarse a entornos tales como el entorno de Microsoft Windows™ incluyendo Windows™ 98, Windows™ 2000, Windows™ NT, y similares. Además, la aplicación también puede escribirse para el entorno de MacIntosh™, SUN™, UNIX o LINUX. Además, las etapas funcionales también pueden implementarse usando un lenguaje de programación universal o independiente de plataforma. Los ejemplos de tales lenguajes de programación de múltiples plataformas incluyen, pero no se limitan a, lenguaje de marcas de hipertexto (HTML), JAVA™, JavaScript™, lenguaje de programación Flash, interfaz de entrada común/lenguaje de consulta estructurado (CGI/SQL), lenguaje de extracción práctica e informe (PERL), AppleScript™ y otros lenguajes de guiones de sistema, lenguaje de programación/lenguaje de consulta estructurado (PL/SQL), y similares. Pueden usarse navegadores habilitados para Java™ o JavaScript™ tales como HotJava™, Microsoft™ Explorer™ o Netscape™. Cuando se usan páginas web de contenido activo, pueden incluir miniprogramas (*applets*) de Java™ o controles de ActiveX™ u otras tecnologías de contenido activo.

La función de análisis también puede incluirse en productos de programa informático y usarse en los sistemas descritos anteriormente u otros sistemas basados en ordenador o en Internet. Por consiguiente, otro aspecto de la presente divulgación se refiere a un producto de programa informático que comprende un medio utilizable con ordenador que tiene códigos o instrucciones de programa legibles por ordenador incluidos en el mismo para permitir que un procesador lleve a cabo el análisis del estado de genes. Estas instrucciones de programa informático pueden cargarse en un ordenador u otro aparato programable para producir una máquina, de tal manera que las instrucciones que se ejecutan en el ordenador u otro aparato programable crean medios para implementar las funciones o etapas descritas anteriormente. Estas instrucciones de programa informático también pueden almacenarse en un medio o una memoria legible por ordenador que puede dirigir un ordenador u otro aparato programable a funcionar de una manera particular, de tal manera que las instrucciones almacenadas en el medio o la memoria legible por ordenador producen un artículo de fabricación que incluye instrucciones que implementan el análisis. Las instrucciones de programa informático también pueden cargarse en un ordenador u otro aparato programable para hacer que se realicen una serie de etapas de operación en el ordenador u otro aparato programable para producir un procedimiento implementado por ordenador de tal manera que las instrucciones que se ejecutan en el ordenador u otro aparato programable proporcionan etapas para implementar las funciones o etapas descritas anteriormente.

Por tanto, un aspecto de la presente divulgación proporciona un sistema para determinar si un paciente tiene una probabilidad aumentada de recidiva. En términos generales, el sistema comprende (1) programa informático para recibir, almacenar y/o recuperar datos del estado de genes de un paciente (por ejemplo, nivel de expresión, nivel de actividad, variantes) y opcionalmente datos de parámetros clínicos (por ejemplo, puntuación de Gleason, puntuación de nomograma); (2) programa informático para consultar estos datos de paciente; (3) programa informático para concluir si existe una probabilidad aumentada de recidiva basándose en estos datos de paciente; y opcionalmente (4) programa informático para emitir/visualizar esta conclusión. En algunas realizaciones, este programa informático para emitir la conclusión puede comprender un programa informático para informar a un profesional sanitario sobre la conclusión.

Un ejemplo de un sistema informático de este tipo es el sistema [600] informático ilustrado en la figura 6. El sistema [600] informático puede incluir al menos un módulo [630] de entrada para introducir datos de paciente en el sistema [600] informático. El sistema [600] informático puede incluir al menos un módulo [624] de salida para indicar si un paciente tiene una probabilidad aumentada o disminuida de respuesta y/o indicar tratamientos sugeridos determinados por el sistema [600] informático. El sistema [600] informático puede incluir al menos un módulo [606] de memoria en comunicación con el al menos un módulo [630] de entrada y el al menos un módulo [624] de salida.

El al menos un módulo [606] de memoria puede incluir, por ejemplo, un dispositivo [608] de almacenamiento extraíble que puede estar en diversas formas, incluyendo, pero sin limitarse a, un dispositivo de banda magnética, un dispositivo de disco flexible, un dispositivo de VCD, un dispositivo de DVD, un dispositivo de disco óptico, etc. El dispositivo [608] de almacenamiento extraíble puede ser compatible con una unidad [610] de almacenamiento extraíble de tal manera que puede leer a partir de y/o escribir en la unidad [610] de almacenamiento extraíble. La unidad [610] de almacenamiento extraíble puede incluir un medio de almacenamiento utilizable con ordenador que tiene almacenados en el mismo códigos o instrucciones de programa legibles por ordenador y/o datos legibles por ordenador. Por ejemplo, la unidad [610] de almacenamiento extraíble puede almacenar datos de paciente. En la técnica se conocen bien ejemplos de unidades [610] de almacenamiento extraíbles, incluyendo, pero sin limitarse a, discos flexibles, bandas magnéticas, discos ópticos, y similares. El al menos un módulo [606] de memoria también puede incluir un dispositivo [612] de disco duro que puede usarse para almacenar códigos o instrucciones de programa legibles por ordenador y/o datos legibles por ordenador.

Además, tal como se muestra en la figura 1, el al menos un módulo [606] de memoria puede incluir además una interfaz [614] y una unidad [616] de almacenamiento extraíble que es compatible con la interfaz [614] de tal manera que pueden transferirse software, códigos o instrucciones legibles por ordenador desde la unidad [616] de almacenamiento extraíble al sistema [600] informático. Los ejemplos de pares de interfaz [614] y unidad [616] de almacenamiento extraíble incluyen, por ejemplo, chips de memoria extraíbles (por ejemplo, EPROM o PROM) y zócalos asociados con los mismos, cartuchos de programa e interfaz de cartucho, y similares. El sistema [600] informático también puede incluir un módulo [618] de memoria secundario tal como memoria de acceso aleatorio (RAM).

El sistema [600] informático puede incluir al menos un módulo [602] de procesador. Debe entenderse que el al menos un módulo [602] de procesador puede consistir en cualquier número de dispositivos. El al menos un módulo [602] de procesador puede incluir un dispositivo de procesamiento de datos, tal como un microprocesador o microcontrolador o una unidad de procesamiento central. El al menos un módulo [602] de procesador puede incluir otro dispositivo lógico tal como un procesador de DMA (acceso directo a memoria), un dispositivo de procesador de comunicación integrada, un dispositivo de VLSI (integración a escala muy grande) personalizado o un dispositivo de ASIC (circuito integrado específico de aplicación). Además, el al menos un módulo [602] de procesador puede incluir cualquier otro tipo de circuitos analógicos o digitales que están diseñados para realizar las funciones de procesamiento descritas en el presente documento.

Tal como se muestra en la figura 6, en el sistema [600] informático, el al menos un módulo [606] de memoria, el al menos un módulo [602] de procesador y el módulo [618] de memoria secundario están todos ellos unidos operativamente entre sí a través de una infraestructura [620] de comunicación, que puede ser un bus de comunicaciones, un tablero de sistema, una barra cruzada, etc.). A través de la infraestructura [620] de comunicación, pueden transferirse e intercambiarse códigos o instrucciones del programa informático o datos legibles por ordenador. La interfaz [626] de entrada puede conectar de manera operable el al menos un módulo [626] de entrada a la infraestructura [620] de comunicación. Del mismo modo, la interfaz [622] de salida puede conectar de manera operable el al menos un módulo [624] de salida a la infraestructura [620] de comunicación.

El al menos un módulo [630] de entrada puede incluir, por ejemplo, un teclado, un ratón, una pantalla táctil, un escáner y otros dispositivos de entrada conocidos en la técnica. El al menos un módulo [624] de salida puede incluir, por ejemplo, una pantalla de visualización, tal como un monitor de ordenador, un monitor de TV o la pantalla táctil del al menos un módulo [630] de entrada; una impresora; y altavoces. El sistema [600] informático también puede incluir módems, puertos de comunicación, tarjetas de red tales como tarjetas de Ethernet, y dispositivos recién desarrollados para acceder a Intranet o Internet.

El al menos un módulo [606] de memoria puede estar configurado para almacenar datos de paciente introducidos a través del al menos un módulo [630] de entrada y procesados a través del al menos un módulo [602] de procesador. Los datos de paciente relevantes para la presente divulgación pueden incluir el nivel de expresión, el nivel de actividad, el número de copias y/o la información de secuencia para *PTEN* y/o un CCG. Los datos de paciente relevantes para la presente divulgación también pueden incluir parámetros clínicos relevantes para la enfermedad del paciente. También pueden introducirse al sistema cualesquiera otros datos de paciente que un médico podría encontrar útiles al tomar decisiones/hacer recomendaciones de tratamiento, incluyendo, pero sin limitarse a, la edad, el sexo y la raza/etnia y datos del estilo de vida tales como información sobre la dieta. Otros posibles tipos de datos de paciente incluyen síntomas experimentados en la actualidad o anteriormente, antecedentes del paciente de enfermedades, medicamentos y procedimientos médicos.

El al menos un módulo [606] de memoria puede incluir un método implementado por ordenador almacenado en el mismo. El al menos un módulo [602] de procesador puede usarse para ejecutar software o códigos de instrucción legibles por ordenador del método implementado por ordenador. El método implementado por ordenador puede estar configurado, basándose en los datos de paciente, para indicar si el paciente tiene una probabilidad aumentada de recidiva, progresión o respuesta a cualquier tratamiento particular, para generar una lista de posibles tratamientos, etc.

En determinadas realizaciones, el método implementado por ordenador puede estar configurado para identificar si un paciente tiene o no una probabilidad aumentada de recidiva o progresión. Por ejemplo, el método implementado por ordenador puede estar configurado para informar a un médico que un paciente particular tiene una probabilidad aumentada de recidiva. De manera alternativa o adicional, el método implementado por ordenador puede estar configurado para sugerir realmente un transcurso de tratamiento particular basándose en respuestas a/resultados para diversas consultas.

La figura 7 ilustra una realización de un método [700] implementado por ordenador de la divulgación que puede implementarse con el sistema [600] informático de la divulgación. El método [700] comienza con una de tres consultas ([710], [711], [712]), o bien de manera secuencial o bien de manera sustancialmente simultánea. Si la respuesta a/resultado para cualquiera de estas consultas es “Sí” [720], el método concluye [730] que el paciente tiene una probabilidad aumentada de recidiva. Si la respuesta a/resultado para cualquiera de estas consultas es “No” [721], el método concluye [731] que el paciente no tiene una probabilidad aumentada de recidiva. El método [700] puede proceder entonces con más consultas, hacer una recomendación ([740], [741]) de tratamiento particular o simplemente terminar.

Cuando las consultas se realizan de manera secuencial, pueden realizarse en el orden sugerido por la figura 7 o en cualquier otro orden. Si se realizan consultas posteriores también puede depender de los resultados/respuestas para consultas previas. En algunas realizaciones del método ilustrado en la figura 7, por ejemplo, el método pregunta en primer lugar sobre parámetros clínicos [712], y si el paciente tiene uno o más parámetros clínicos que identifican que el paciente tiene un riesgo aumentado de recidiva, entonces el método concluye tal [730] u opcionalmente confirma consultando el estado de CCG, mientras que si el paciente no tiene tales parámetros clínicos, entonces el método procede a preguntar sobre el estado de CCG [711]. Opcionalmente, si el estado de CCG no está elevado, entonces el método puede continuar preguntando sobre el estado de *PTEN* [710]. Tal como se mencionó anteriormente, el orden previo de consultas puede modificarse. En algunas realizaciones, una respuesta de “Sí” a una consulta (por ejemplo, [712]) provoca que una o más de las consultas restantes confirmen que el paciente tiene un riesgo aumentado de recidiva.

En algunas realizaciones, el método [700] implementado por ordenador de la divulgación es abierto. Dicho de otro modo, la primera etapa [710, 711 y/o 712] aparente en la figura 7 en realidad puede formar parte de un procedimiento más largo y, dentro de este procedimiento más largo, no es necesario que sea la primera etapa/consulta. También pueden añadirse etapas adicionales a los métodos principales comentados anteriormente.

Estas etapas adicionales incluyen, pero no se limitan a, informar a un profesional sanitario (o al propio paciente) sobre la conclusión alcanzada; combinar la conclusión alcanzada mediante el método [700] ilustrado con otros hechos u otras conclusiones para llegar a cierta conclusión adicional o refinada sobre el diagnóstico, pronóstico, tratamiento, etc., del paciente; hacer una recomendación para el tratamiento (por ejemplo, "el paciente debe/no debe someterse a prostatectomía radical"); consultas adicionales sobre biomarcadores adicionales, parámetros clínicos u otra información útil del paciente (por ejemplo, edad de diagnóstico, salud general del paciente, etc.).

Con respecto al método [700] implementado por ordenador anterior, las respuestas a las consultas pueden determinarse mediante el método que establece una búsqueda de datos de paciente para la respuesta. Por ejemplo, para responder a las consultas [710, 711, 712] respectivas, puede realizarse una búsqueda en los datos de paciente para el estado de *PTEN* (por ejemplo, cribado de mutaciones o IHC de *PTEN*), el estado de CCG (por ejemplo, datos de nivel de expresión de CCG) o parámetros clínicos (por ejemplo, puntuación de Gleason, puntuación de nomograma, etc.). Si aún no se ha realizado una comparación de este tipo, el método puede comparar estos datos con cierta referencia con el fin de determinar si el paciente tiene un estado anómalo (por ejemplo, elevado, bajo, negativo) status. De manera adicional o alternativa, el método puede presentar una o más de las consultas [710, 711, 712] a un usuario (por ejemplo, un médico) del sistema [100] informático. Por ejemplo, las consultas [710, 711, 712] pueden presentarse a través de un módulo [624] de salida. El usuario puede responder entonces "Sí" o "No" a través de un módulo [630] de entrada. El método puede proceder entonces basándose en la respuesta recibida. Del mismo modo, las conclusiones [730, 731] pueden presentarse a un usuario del método implementado por ordenador a través de un módulo [624] de salida.

Por tanto, en algunas realizaciones, la divulgación proporciona un método que comprende: acceder a información sobre el estado de CCG, parámetros clínicos y/o el estado de *PTEN* de un paciente almacenados en un medio legible por ordenador; consultar esta información para determinar al menos uno de si una muestra obtenida del paciente muestra expresión aumentada de al menos un CCG, si el paciente tiene un parámetro clínico asociado a recidiva y/o si el paciente tiene un estado de *PTEN* bajo/negativo; emitir [o visualizar] el estado de expresión de CCG de la muestra, el estado de parámetro clínico asociado a recidiva del paciente y/o el estado de *PTEN* de la muestra. Tal como se usa en el presente documento en el contexto de realizaciones implementadas por ordenador de la divulgación, "visualizar" significa comunicar cualquier información de manera sensorial. Los ejemplos incluyen, pero no se limitan a, visualizaciones visuales, por ejemplo, en una pantalla de ordenador o en una hoja de papel impresa en el comando del ordenador, y visualizaciones auditivas, por ejemplo, expresión auditiva generada o registrada por ordenador del genotipo de un paciente.

Tal como se comentó con detalle anteriormente, los parámetros clínicos asociados a recidiva o el estado de *PTEN* en combinación con el estado de CCG elevado indican una probabilidad significativamente aumentada de recidiva. Por tanto, algunas realizaciones proporcionan un método implementado por ordenador para determinar si un paciente tiene una probabilidad aumentada de recidiva que comprende acceder a información sobre el estado de *PTEN* de un paciente (por ejemplo, a partir de una muestra tumoral obtenida del paciente) o parámetros clínicos y el estado de CCG (por ejemplo, a partir de una muestra tumoral obtenida del paciente) almacenados en un medio legible por ordenador; consultar esta información para determinar al menos uno de si el paciente tiene un estado de *PTEN* bajo/negativo o si el paciente tiene un parámetro clínico asociado a recidiva; consultar esta información para determinar si una muestra obtenida del paciente muestra expresión aumentada de al menos un CCG; emitir (o visualizar) una indicación de que el paciente tiene una probabilidad aumentada de recidiva si el paciente tiene un estado de *PTEN* bajo/negativo o un parámetro clínico asociado a recidiva y la muestra muestra expresión aumentada de al menos un CCG. Algunas realizaciones comprenden además visualizar *PTEN*, parámetros clínicos (o sus valores) y/o los CCG y su estado (incluyendo, por ejemplo, niveles de expresión), opcionalmente junto con una indicación de si el estado de *PTEN* o CCG y/o el parámetro clínico indican una probabilidad aumentada de riesgo.

La práctica de la presente divulgación también puede emplear métodos de biología, software y sistemas convencionales. Los productos de software informático de la divulgación incluyen normalmente medios legibles por ordenador que tienen instrucciones ejecutables por ordenador para realizar las etapas lógicas del método de la divulgación. Los medios legibles por ordenador adecuados incluyen disco flexible, CD-ROM/DVD/DVD-ROM, dispositivo de disco duro, memoria flash, ROM/RAM, bandas magnéticas, etc. Se describen métodos de biología computacionales básicos en, por ejemplo, Setubal *et al*., INTRODUCTION TO COMPUTATIONAL BIOLOGY METHODS (PWS Publishing Company, Boston, 1997); Salzberg *et al*. (Ed.), COMPUTATIONAL METHODS IN MOLECULAR BIOLOGY (Elsevier, Ámsterdam, 1998); Rashidi y Buehler, BIOINFORMATICS BASICS: APPLICATION IN BIOLOGICAL SCIENCE AND MEDICINE (CRC Press, Londres, 2000); y Ouelette y Bzevanis, BIOINFORMATICS: A PRACTICAL GUIDE FOR ANALYSIS OF GENE AND PROTEINS (Wiley & Sons, Inc., 2ª ed., 2001); véase también la patente estadounidense n.º 6.420.108.

La presente divulgación también puede hacer uso de diversos productos de programa informático y software para una variedad de propósitos, tales como diseño de sondas, gestión de datos, análisis y operación instrumental. Véanse las patentes estadounidenses n.ºs 5.593.839; 5.795.716; 5.733.729; 5.974.164; 6.066.454; 6.090.555; 6.185.561; 6.188.783; 6.223.127; 6.229.911 y 6.308.170. Además, la presente divulgación puede tener realizaciones que incluyen métodos para proporcionar información genética sobre redes tales como Internet tal como se muestra

en las patentes estadounidenses con n.os de serie 10/197.621 (publicación estadounidense n.º 20030097222); 10/063.559 (publicación estadounidense n.º 20020183936), 10/065.856 (publicación estadounidense n.º 20030100995); 10/065.868 (publicación estadounidense n.º 20030120432); 10/423.403 (publicación estadounidense n.º 20040049354).

Las técnicas para analizar tales datos de expresión, actividad y/o secuencia (de hecho, cualquier dato obtenido según la divulgación) a menos se implementarán usando hardware, software o una combinación de los mismos en uno o más sistemas informáticos u otros sistemas de procesamiento capaces de efectuar tal análisis.

Por tanto, un aspecto de la presente divulgación proporciona sistemas relacionados con los métodos anteriores de la divulgación. En una realización, la divulgación proporciona un sistema para determinar la expresión de genes en una muestra tumoral, que comprende:

(1) un analizador de muestras para determinar los niveles de expresión en una muestra de un panel de genes que incluye al menos 4 CCG, en el que el analizador de muestras contiene la muestra, ARN de la muestra y expresado a partir del panel de genes, o ADN sintetizado a partir de dicho ARN;

(2) un primer programa informático para

(a) recibir datos de expresión de genes sobre al menos 4 genes de prueba seleccionados del panel de genes,

(b) ponderar la expresión determinada de cada uno de los genes de prueba con un coeficiente predefinido, y

(c) combinar la expresión ponderada para proporcionar un valor de prueba, en el que la ponderación combinada dada a dichos al menos 4 ó 5 ó 6 CCG es al menos el 40% (o el 50%, el 60%, el 70%, el 80%, el 90%, el 95% o el 100%) de la ponderación total dada a la expresión de la totalidad de dicha pluralidad de genes de prueba; y opcionalmente

(3) un segundo programa informático para comparar el valor de prueba con uno o más valores de referencia, cada uno asociado con un grado predeterminado de riesgo de cáncer.

En algunas realizaciones, al menos el 20%, el 50%, el 75% o el 90% de dicha pluralidad de genes de prueba son CCG. En algunas realizaciones, el analizador de muestras contiene reactivos para determinar los niveles de expresión en la muestra de dicho panel de genes que incluye al menos 4 CCG. En algunas realizaciones, el analizador de muestras contiene reactivos específicos de CCG tal como se describe a continuación.

En otra realización, la divulgación proporciona un sistema para determinar la expresión de genes en una muestra tumoral, que comprende: (1) un analizador de muestras para determinar los niveles de expresión de un panel de genes en una muestra tumoral que incluye al menos 4 CCG, en el que el analizador de muestras contiene la muestra tumoral que es de un paciente que se ha identificado que padece cáncer de próstata, cáncer de mama, cáncer de cerebro, cáncer de vejiga o cáncer de pulmón, ARN de la muestra y expresado a partir del panel de genes, o ADN sintetizado a partir de dicho ARN; (2) un primer programa informático para (a) recibir datos de expresión de genes sobre al menos 4 genes de prueba seleccionados del panel de genes, (b) ponderar la expresión determinada de cada uno de los genes de prueba con un coeficiente predefinido y (c) combinar la expresión ponderada para proporcionar un valor de prueba, en el que la ponderación combinada dada a dichos al menos 4 ó 5 ó 6 CCG es al menos el 40% (o el 50%, el 60%, el 70%, el 80%, el 90%, el 95% o el 100%) de la ponderación total dada a la expresión de la totalidad de dicha pluralidad de genes de prueba; y opcionalmente (3) un segundo programa informático para comparar el valor de prueba con uno o más valores de referencia, cada uno asociado con un grado predeterminado de riesgo de recidiva o progresión de cáncer del cáncer de próstata, cáncer de mama, cáncer de cerebro, cáncer de vejiga o cáncer de pulmón. En algunas realizaciones, al menos el 20%, el 50%, el 75% o el 90% de dicha pluralidad de genes de prueba son CCG. En algunas realizaciones, el sistema comprende un programa informático para determinar el pronóstico del paciente y/o determinar (incluyendo cuantificar) el grado de riesgo de recidiva o progresión de cáncer en el paciente basándose al menos en parte en la comparación del valor de prueba con dichos uno o más valores de referencia.

En algunas realizaciones, el sistema comprende además un módulo de visualización que visualiza la comparación entre el valor de prueba y el uno o más valores de referencia, o que visualiza un resultado de la etapa de comparación, o que visualiza el pronóstico y/o el grado de riesgo de recidiva o progresión de cáncer del paciente.

En una realización preferida, se mide la cantidad de ARN transcrito a partir del panel de genes que incluye genes de prueba (y/o ADN transcrito de manera inversa a partir del mismo) en la muestra. Además, también se mide la cantidad de ARN de uno o más genes de mantenimiento en la muestra (y/o ADN transcrito de manera inversa a partir del mismo) y se usa para normalizar o calibrar la expresión de los genes de prueba, tal como se describió anteriormente.

En algunas realizaciones, la pluralidad de genes de prueba incluye al menos 2, 3 ó 4 CCG, que constituyen al

menos el 50%, el 75% o el 80% de la pluralidad de genes de prueba, y preferiblemente el 100% de la pluralidad de genes de prueba. En algunas realizaciones, la pluralidad de genes de prueba incluye al menos 5, 6 ó 7 o al menos 8 CCG, que constituyen al menos el 20%, el 25%, el 30%, el 40%, el 50%, el 60%, el 70%, el 75%, el 80% o el 90% de la pluralidad de genes de prueba, y preferiblemente el 100% de la pluralidad de genes de prueba. Por tanto, en algunas realizaciones, la pluralidad de genes de prueba comprende al menos cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende los primeros 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40 o más CCG enumerados en las tablas 9-11 y 13-14. En algunas realizaciones, la pluralidad de genes de prueba comprende al menos cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 ó 20 de los siguientes genes: *ASPM*, *BIRC5*, *BUB1B*, *CCNB2*, *CDC2*, *CDC20*, *CDCA8*, *CDKN3*, *CENPF*, *DLGAP5*, *FOXM1*, *KIAA0101*, *KIF11*, *KIF2C*, *KIF4A*, *MCM10*, *NUSAP1*, *PRC1*, *RACGAP1* y *TPX2*. En algunas realizaciones, la pluralidad de genes de prueba comprende al menos cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende cualesquiera uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve o diez o la totalidad de los números de genes 1 y 2, 1 a 3, 1 a 4, 1 a 5, 1 a 6, 1 a 7, 1 a 8, 1 a 9 ó 1 a 10 de cualesquiera de las tablas 9-11 y 13-14. En algunas realizaciones, la pluralidad de genes de prueba comprende al menos cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende cualesquiera uno, dos, tres, cuatro, cinco, seis, siete, ocho o nueve o la totalidad de los números de genes 2 y 3, 2 a 4, 2 a 5, 2 a 6, 2 a 7, 2 a 8, 2 a 9 ó 2 a 10 de cualesquiera de las tablas 9-11 y 13-14. En algunas realizaciones, la pluralidad de genes de prueba comprende al menos cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende cualesquiera uno, dos, tres, cuatro, cinco, seis, siete u ocho o la totalidad de los números de genes 3 y 4, 3 a 5, 3 a 6, 3 a 7, 3 a 8, 3 a 9 ó 3 a 10 de cualesquiera de las tablas 9-11 y 13-14. En algunas realizaciones, la pluralidad de genes de prueba comprende al menos cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende cualesquiera uno, dos, tres, cuatro, cinco, seis o siete o la totalidad de los números de genes 4 y 5, 4 a 6, 4 a 7, 4 a 8, 4 a 9 ó 4 a 10 de cualesquiera de las tablas 9-11 y 13-14. En algunas realizaciones, la pluralidad de genes de prueba comprende al menos cierto número de CCG (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más CCG) y esta pluralidad de CCG comprende cualesquiera uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, 10, 11, 12, 13, 14 ó 15 o la totalidad de los números de genes 1 y 2, 1 a 3, 1 a 4, 1 a 5, 1 a 6, 1 a 7, 1 a 8, 1 a 9, 1 a 10, 1 a 11, 1 a 12, 1 a 13, 1 a 14 ó 1 a 15 de cualesquiera de las tablas 9-11 y 13-14.

En algunas otras realizaciones, la pluralidad de genes de prueba incluye al menos 8, 10, 12, 15, 20, 25 ó 30 CCG, que constituyen al menos el 20%, el 25%, el 30%, el 40%, el 50%, el 60%, el 70%, el 75%, el 80% o el 90% de la pluralidad de genes de prueba, y preferiblemente el 100% de la pluralidad de genes de prueba.

El analizador de muestras puede ser cualquier instrumento útil para determinar la expresión de genes, incluyendo, por ejemplo, una máquina de secuenciación (por ejemplo, HiSeq™ de Illumina, PGM de Ion Torrent, secuenciador SOLiD™ de ABI, *RS* de PacBio, Heliscope™ de Helicos, etc.), una máquina de PCR en tiempo real (por ejemplo, 7900 de ABI, BioMark™ de Fluidigm, etc.), un instrumento de microalineamiento, etc.

En un aspecto, que no está abarcado por las reivindicaciones, la presente divulgación proporciona métodos para tratar un paciente con cáncer que comprenden obtener información de estado de CCG (por ejemplo, los CCG en la tabla 1 o en los paneles A a G), y recomendar, recetar o administrar un tratamiento para el paciente con cáncer basándose en el estado de CCG. En algunas realizaciones, el método incluye además obtener información de parámetros clínicos y/u obtener información de estado de *PTEN* a partir de una muestra del paciente y tratar el paciente con un tratamiento particular basándose en información de estado de CCG, parámetro clínico y/o estado de *PTEN*. Por ejemplo, la divulgación proporciona un método para tratar un paciente con cáncer que comprende:

(1) determinar el estado de al menos un CCG;

(2) determinar el estado de al menos un parámetro clínico;

(3) determinar el estado de *PTEN* en una muestra obtenida del paciente; y

(4) recomendar, recetar o administrar o bien

(a) un tratamiento activo (incluyendo agresivo) si el paciente tiene al menos uno de expresión aumentada del CCG, parámetro clínico asociado a recidiva o estado de *PTEN* bajo/negativo, o bien

(b) un tratamiento pasivo (o menos agresivo) si el paciente no tiene ninguno de expresión aumentada del CCG, parámetro clínico asociado a recidiva o estado de *PTEN* bajo/negativo.

En algunas realizaciones, las etapas de determinación comprenden recibir un informe que comunica el estado relevante (por ejemplo, estado de CCG). En algunas realizaciones, este informe comunica tal estado de manera cualitativa (por ejemplo, expresión “alta” o “aumentada”). En algunas realizaciones, este informe comunica tal estado

indirectamente comunicando una puntuación (por ejemplo, puntuación de pronóstico, puntuación de recidiva, puntuación combinada tal como se comentó anteriormente, etc.) que incorpora tal estado.

Si un tratamiento es agresivo o no dependerá generalmente del tipo de cáncer, la edad del paciente, etc. Por ejemplo, en el cáncer de mama, la quimioterapia complementaria es un tratamiento agresivo habitual proporcionado para complementar los tratamientos habituales menos agresivos de cirugía y terapia hormonal. Los expertos en la técnica están familiarizados con otros diversos tratamientos agresivos y menos agresivos para cada tipo de cáncer. Los expertos en la técnica entienden bien "tratamiento activo" en el cáncer de próstata y, tal como se usa en el presente documento, tiene el significado convencional en la técnica. En términos generales, tratamiento activo en el cáncer de próstata es cualquiera otra cosa distinta de "conducta expectante". El tratamiento activo aplicado actualmente en la técnica del tratamiento de cáncer de próstata incluye, por ejemplo, prostatectomía, radioterapia, terapia hormonal (por ejemplo, antagonistas de GnRH, agonistas de GnRH, antiandrógenos), quimioterapia, ultrasonidos enfocados de alta intensidad ("HIFU"), etc. Cada opción de tratamiento conlleva ciertos riesgos así como efectos secundarios de gravedad variable, por ejemplo, impotencia, incontinencia urinaria, etc. Por tanto, es habitual que los médicos, dependiendo de la edad y la salud general del varón diagnosticado con cáncer de próstata, recomienden un régimen de "conducta expectante".

"Conducta expectante", también denominada "vigilancia activa", también tiene su significado convencional en la técnica. Esto significa generalmente observación y monitorización regular sin tratamiento invasivo. Conducta expectante se usa algunas veces, por ejemplo, cuando se halla un cáncer de próstata de estadio temprano y de crecimiento lento en un varón más anciano. Conducta expectante también puede sugerirse cuando los riesgos de cirugía, radioterapia o terapia hormonal superan los posibles beneficios. Pueden iniciarse otros tratamientos si se desarrollan síntomas o si existen signos de que el crecimiento del cáncer está acelerándose (por ejemplo, aumento rápido de PSA, aumento en la puntuación de Gleason en biopsia repetida, etc.).

Aunque los varones que eligen conducta expectante evitan los riesgos de cirugía y radiación, la conducta expectante conlleva sus propios riesgos, por ejemplo, riesgo aumentado de metástasis. Para varones más jóvenes, un ensayo de vigilancia activa puede no significar evitar el tratamiento completamente, sino que puede permitir de manera razonable un retraso de algunos años o más, tiempo durante el que puede evitarse el impacto del tratamiento activo sobre la calidad de vida. Los datos publicados hasta la fecha sugieren que los varones cuidadosamente seleccionados no se perderán un margen para la cura con este enfoque. Los problemas de salud adicionales que se desarrollan con el paso de la edad durante el periodo de observación también pueden hacer que sea más difícil someterse a cirugía y radioterapia. Por tanto, es clínicamente importante determinar cuidadosamente qué pacientes con cáncer de próstata son buenos candidatos para conducta expectante y qué pacientes deben recibir tratamiento activo.

Por tanto, la divulgación proporciona un método para tratar un paciente con cáncer de próstata, que no está abarcado por las reivindicaciones, o proporcionar pautas para el tratamiento de un paciente. En este método, se determina el estado de al menos un CCG (por ejemplo, aquellos en la tabla 1 o en los paneles A a G), al menos un parámetro clínico asociado a recidiva y/o el estado de *PTEN* y (a) se recomienda, inicia o continúa el tratamiento activo si una muestra del paciente tiene un estado elevado para al menos un CCG, el paciente tiene al menos un parámetro clínico asociado a recidiva y/o un estado de *PTEN* bajo/negativo, o (b) se recomienda/inicia/continúa la conducta expectante si el paciente no tiene ninguno de un estado elevado para al menos un CCG, un parámetro clínico asociado a recidiva y estado de *PTEN* bajo/negativo. En determinadas realizaciones, el estado de CCG, el/los parámetro(s) clínico(s) y el estado de *PTEN* pueden indicar no sólo qué tratamiento activo se recomienda, sino qué tratamiento activo particular es preferible para el paciente (incluyendo tratamientos relativamente agresivos tales como, por ejemplo, RP y/o terapia complementaria).

En general, la terapia complementaria (por ejemplo, quimioterapia, radioterapia, HIFU, terapia hormonal, etc., después de la prostatectomía o radioterapia) no es el tratamiento habitual en el cáncer de próstata. Sin embargo, según la presente divulgación, los médicos pueden ser capaces de determinar qué pacientes con cáncer de próstata tienen una enfermedad particularmente agresiva y, por tanto, deben recibir terapia complementaria. Por tanto, en una realización, que no está abarcada por las reivindicaciones, la divulgación proporciona un método para tratar un paciente (por ejemplo, un paciente con cáncer de próstata) que comprende determinar el estado de al menos un CCG (por ejemplo, aquellos en la tabla 1 o en los paneles A a G), el estado de al menos un parámetro clínico asociado a recidiva y/o el estado de *PTEN* e iniciar la terapia complementaria después de la prostatectomía o radioterapia si una muestra del paciente tiene un estado elevado para al menos un CCG, el paciente tiene al menos un parámetro clínico asociado a recidiva y/o el paciente tiene un estado de *PTEN* bajo/negativo.

En un aspecto, la divulgación proporciona composiciones para su uso en los métodos anteriores. Tales composiciones incluyen, pero no se limitan a, sondas de ácido nucleico que se hibridan con *PTEN* o un CCG (o con cualquier ácido nucleico codificado por los mismos o complementario a los mismos); cebadores y pares de cebadores de ácido nucleico adecuados para amplificar la totalidad o una porción de *PTEN* o un CCG o cualquier ácido nucleico codificado por los mismos; anticuerpos que se unen inmunológicamente a un polipéptido codificado por *PTEN* o un CCG; conjuntos de sondas que comprenden una pluralidad de dichas sondas de ácido nucleico, cebadores de ácido nucleico, anticuerpos y/o polipéptidos; micromatrices que comprenden cualesquiera de estos;

kits que comprenden cualesquiera de estos; etc. En algunos aspectos, la divulgación proporciona métodos, sistemas, software y/o módulos informáticos para su uso en los métodos anteriores.

En algunas realizaciones, la divulgación proporciona una sonda que comprende un oligonucleótido aislado capaz de hibridarse selectivamente con *PTEN* o al menos uno de los genes en la tabla 1 o en los paneles A a G. Los términos "sonda" y "oligonucleótido" (también "oligo"), cuando se usan en el contexto de ácidos nucleicos, se refieren de manera intercambiable a una secuencia o un fragmento de ácido nucleico relativamente corto. La divulgación también proporciona cebadores útiles en los métodos de la divulgación. En las condiciones adecuadas y con los reactivos acompañantes adecuados, los "cebadores" son sondas capaces de amplificar selectivamente un ácido nucleico diana (por ejemplo, un gen diana). En el contexto de ácidos nucleicos, "sonda" se usa en el presente documento para abarcar "cebador" puesto que los cebadores generalmente también pueden servir como sondas.

La sonda puede tener generalmente cualquier tamaño/longitud adecuados. En algunas realizaciones, la sonda tiene una longitud de aproximadamente 8 a 200, de 15 a 150, de 15 a 100, de 15 a 75, de 15 a 60 o de 20 a 55 bases de longitud. Pueden marcarse con marcadores detectables con cualquier marcador de detección adecuado, incluyendo, pero sin limitarse a, isótopos radiactivos, fluoróforos, biotina, enzimas (por ejemplo, fosfatasa alcalina), sustratos enzimáticos, ligandos y anticuerpos, etc. Véanse Jablonski *et al*., NUCLEIC ACIDS RES. (1986) 14:6115-6128; Nguyen *et al*., BIOTECHNIQUES (1992) 13:116-123; Rigby *et al*., J. MOL. BIOL. (1977) 113:237-251. De hecho, las sondas pueden modificarse de cualquier manera convencional para diversas aplicaciones de biología molecular. Las técnicas para producir y usar tales sondas oligonucleotídicas son convencionales en la técnica.

Las sondas según la divulgación pueden usarse en las técnicas de hibridación/amplificación/detección comentadas anteriormente. Por tanto, algunas realizaciones de la divulgación comprenden conjuntos de sondas adecuados para su uso en una micromatriz para detectar, amplificar y/o cuantificar *PTEN* y/o una pluralidad de CCG. En algunas realizaciones, los conjuntos de sondas tienen una determinada proporción de sus sondas dirigidas a los CCG, por ejemplo, un conjunto de sondas que consiste en el 10%, el 20%, el 30%, el 40%, el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, el 90%, el 95%, el 96%, el 97%, el 98%, el 99% o el 100% de sondas específicas de los CCG. En algunas realizaciones, el conjunto de sondas comprende sondas dirigidas a al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 700 u 800 o más, o la totalidad, de los genes en la tabla 1 o en los paneles A a G. Tales conjuntos de sondas pueden incorporarse en matrices de alta densidad que comprenden 5.000, 10.000, 20.000, 50.000, 100.000, 200.000, 300.000, 400.000, 500.000, 600.000, 700.000, 800.000, 900.000 ó 1.000.000 o más de sondas diferentes. En otras realizaciones, los conjuntos de sondas comprenden cebadores (por ejemplo, pares de cebadores) para amplificar ácidos nucleicos que comprenden al menos una porción de *PTEN* o de uno o más de los CCG en la tabla 1 o en los paneles A a G.

En otro aspecto de la presente divulgación, se proporciona un kit para poner en práctica el método de pronóstico de la presente divulgación. El kit puede incluir un portador para los diversos componentes del kit. The portador puede ser un recipiente o soporte, en forma de, por ejemplo, bolsa, caja, tubo, rejilla, y opcionalmente está compartimentalizado. El portador puede definir un confinamiento encerrado con propósitos de seguridad durante el envío y almacenamiento. El kit incluye diversos componentes útiles para determinar el estado de uno o más CCG y uno o más marcadores de genes de mantenimiento, usando las técnicas de detección anteriormente comentadas. Por ejemplo, el kit puede incluir oligonucleótidos que se hibridan específicamente bajo alta rigurosidad con ARNm o ADNc de los genes en la tabla 1 o en los paneles A a G. Tales oligonucleótidos pueden usarse como cebadores de PCR en reacciones de RT-PCR o como sondas de hibridación. En algunas realizaciones, el kit comprende reactivos (por ejemplo, sondas, cebadores y/o anticuerpos) para determinar el nivel de expresión de un panel de genes, en el que dicho panel comprende al menos el 25%, el 30%, el 40%, el 50%, el 60%, el 75%, el 80%, el 90%, el 95%, el 99% o el 100% de los CCG (por ejemplo, CCG en la tabla 1 o en cualquiera de los paneles A a G). En algunas realizaciones, el kit consiste en reactivos (por ejemplo, sondas, cebadores y/o anticuerpos) para determinar el nivel de expresión de no más de 2500 genes, en el que al menos 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 150, 200, 250 o más de estos genes son CCG (por ejemplo, los CCG en la tabla 1 o en cualquiera de los paneles A a G).

Los oligonucleótidos en el kit de detección pueden marcarse con cualquier marcador de detección adecuado, incluyendo, pero sin limitarse a, isótopos radiactivos, fluoróforos, biotina, enzimas (por ejemplo, fosfatasa alcalina), sustratos enzimáticos, ligandos y anticuerpos, etc. Véanse Jablonski *et al*., Nucleic Acids Res., 14:6115-6128 (1986); Nguyen *et al*., Biotechniques, 13:116-123 (1992); Rigby *et al*., J. Mol. Biol., 113:237-251 (1977). Alternativamente, los oligonucleótidos incluidos en el kit no están marcados, y en su lugar, se proporcionan uno o más marcadores en el kit de modo que los usuarios pueden marcar los oligonucleótidos en el momento de uso.

En otra realización de la divulgación, el kit de detección contiene uno o más anticuerpos selectivamente inmunorreactivos con una o más proteínas codificadas por *PTEN* o uno o más CCG u opcionalmente cualquier marcador adicional. Los ejemplos incluyen anticuerpos que se unen inmunológicamente a *PTEN* o una proteína codificada por un gen en la tabla 1 o en los paneles A a G. Los métodos para producir y usar tales anticuerpos se han descrito anteriormente con detalle.

También pueden incluirse otros diversos componentes útiles en las técnicas de detección en el kit de detección de

esta divulgación. Los ejemplos de tales componentes incluyen, pero no se limitan a, polimerasa Taq, desoxirribonucleótidos, didesoxirribonucleótidos, otros cebadores adecuados para la amplificación de una secuencia de ADN diana, ARNasa A, y similares. Además, el kit de detección incluye preferiblemente instrucciones para usar el kit para poner en práctica el método de pronóstico de la presente divulgación usando muestras humanas.

Realizaciones específicas

A continuación se muestran dos realizaciones específicas de la divulgación para su uso en muestras de biopsia y prostatectomía. Los expertos en la técnica comprenderán que cada elemento de estos procedimientos puede modificarse conservando las características esenciales y logrando los mismos objetivos.

BIOPSIA

Indicaciones y uso

Puede usarse tejido fijado en formol e incrustado en parafina (FFPE) procedente de bloques o portaobjetos de biopsias de adenocarcinoma prostático. Los bloques pueden incluir al menos 2 mm de tumor en portaobjetos de H&E de diagnóstico para el procesamiento de muestras y la extracción de ARN. En los casos en los que no estén disponibles bloques, puede aceptarse un portaobjetos de H&E de 3-5 μm seguido de diez portaobjetos consecutivos de 10 μm sin teñir y un portaobjetos de H&E final. Los códigos de barras de las muestras, que se escanean y rastrean, pueden aplicarse a cada bloque (o portaobjetos). Los portaobjetos de H&E de cada caso pueden evaluarse, por ejemplo, por un patólogo, para determinar la localización y la cantidad de tumor por portaobjetos. Usando los portaobjetos teñidos con H&E como guía, puede extraerse tejido tumoral de diez portaobjetos sin teñir y puede extraerse ARN total del tejido. La expresión de los genes en cualquiera de los paneles A-F, normalizada con respecto a la de los genes de mantenimiento, puede medirse entonces por triplicado para generar un valor de prueba (por ejemplo, la puntuación de CCP).

Como medida opcional de control de calidad, en cada serie de muestras puede analizarse un control sin ARN y un control con ARN humano normal con una puntuación de CCP determinada previamente. Los controles pueden analizarse para verificar los resultados esperados.

Características de rendimiento/Limitaciones

La puntuación de CCP puede usarse sola o en combinación con información clínica para llegar a un pronóstico clínico. La puntuación de CCP puede combinarse con la puntuación de nomograma de CAPRA del paciente (véase Cooperberg *et al.*, J. Natl. Cancer Inst. (2009) 101(12):878-887 para más detalles sobre el nomograma de CAPRA) según la siguiente ecuación: puntuación combinada = (0,58*CCP +0,41*CAPRA).

Intervalo clínicamente notificable

Puede aplicarse un intervalo de puntuación de CCP clínicamente notificable de -1,3 a 4,7. Puede notificarse una escala de puntuaciones de CCP para la categoría de riesgo de la Asociación Urológica Americana (AUA) del paciente individual. La escala puede consistir en cinco intervalos de 1 unidad, centrándose el intervalo central en la mediana de la puntuación de CCP para esa categoría de riesgo específica según la AUA en la población estadounidense. Puede haber un cambio de aproximadamente 2 veces en la mortalidad por cáncer de próstata entre intervalos, que sería la razón de riesgos instantáneos correspondiente a un cambio de 1 unidad en la puntuación de CCP.

Límite de detección/Linealidad

Las puntuaciones de CCP entre -3,0 y +7,0 pueden representar el intervalo de puntuaciones detectables por el ensayo. La linealidad puede establecerse dentro de este intervalo tal como sigue: los genes relevantes pueden amplificarse previamente, diluirse hasta 7 concentraciones diferentes y añadirse en adiciones conocidas a una muestra de ADNc de concentración conocida. Cada muestra con adiciones conocidas puede luego someterse a ensayo por triplicado, y los 3 puntos de datos resultantes pueden promediarse para cada concentración para generar una puntuación de CCP.

Interferencia

En algunos casos, el tratamiento de terapia hormonal complementaria y radioterapia puede afectar a las puntuaciones de CCP. Por tanto, en algunas realizaciones el método se aplica sólo a pacientes que no han recibido estos tratamientos antes de la biopsia.

Limitaciones

En algunas realizaciones sólo se analizan muestras de tumores de próstata humanos FFPE.

Criterios de interpretación

Puntuaciones de CCP dentro del intervalo técnico del ensayo y dentro del intervalo de puntuaciones para las que la predicción clínica está validada (por ejemplo, entre -1,3 y 4,7)

El riesgo estimado de mortalidad específica de cáncer de próstata puede proporcionarse para cada puntuación de CCP dentro de este intervalo y, en algunos casos, puede mostrar cómo la puntuación de CCP diferencia entre pacientes con la misma puntuación de CAPRA. Además, puede proporcionarse el percentil de distribución estadounidense para las puntuaciones de CCP, por ejemplo, para pacientes en la misma categoría de riesgo de CAPRA (bajo, intermedio o alto). Aunque el porcentaje de riesgo puede darse en todo el intervalo, a continuación se dan ejemplos de puntuaciones de riesgo:

Puntuación de CCP	Probabilidad de muerte específica de cáncer
-1	5,9%
0	11,6%
1	22%
2	39,5%
3	63,8%
4	87,2%

Puntuaciones de CCP dentro del intervalo técnico del ensayo pero fuera del intervalo de puntuaciones para las que la predicción clínica está validada (por ejemplo, -1,3 pero mayor de -3,0)

Si se ha establecido la linealidad de las puntuaciones de CCP dentro de un intervalo de este tipo, puede notificarse entonces la puntuación de CCP calculada, pero en algunos casos no puede proporcionarse el riesgo relativo de mortalidad específica de cáncer de próstata estimado (en algunos casos puede notificarse el percentil de distribución estadounidense para las puntuaciones de CCP, por ejemplo, para pacientes en la misma categoría de riesgo de CAPRA (bajo, intermedio o alto)).

Puntuaciones de CCP fuera del intervalo técnico del ensayo (por ejemplo, mayor de 4,7 o de menos de 7,0)

Estas puntuaciones pueden estar fuera de los límites de detección verificados de este ensayo y pueden representar un artefacto o error técnico. Por tanto, en algunos casos estas puntuaciones pueden no ser notificables.

POST-PROSTATECTOMÍA

Indicaciones y uso

Puede usarse tejido fijado en formol e incrustado en parafina (FFPE) procedente de bloques de prostatectomía de adenocarcinoma prostático. Los bloques pueden incluir al menos 5 mm de tumor en portaobjetos de H&E de diagnóstico para el procesamiento de muestras y la extracción de ARN. En los casos en los que no se disponga de bloques, puede aceptarse un portaobjetos de H&E de 3-5 μm seguido de cinco portaobjetos consecutivos de 10 μm sin teñir y un portaobjetos de H&E final. Los códigos de barras de las muestras, que se escanean y rastrean, pueden aplicarse a cada bloque (o portaobjetos). Los portaobjetos de H&E de cada caso pueden evaluarse, por ejemplo, por un patólogo, para determinar la localización y la cantidad de tumor por portaobjetos. Usando los portaobjetos teñidos con H&E como guía, puede extraerse tejido tumoral de cinco portaobjetos sin teñir y puede extraerse ARN total del tejido.

La expresión de los genes en cualquiera de los paneles A-F, normalizada con respecto a la de los genes de mantenimiento, puede medirse entonces por triplicado para generar un valor de prueba (por ejemplo, puntuación de CCP). Esta puntuación de CCP puede usarse para estimar la probabilidad de recidiva (por ejemplo, recidiva bioquímica) dentro de un periodo de tiempo determinado (por ejemplo, dentro de los 10 años posteriores a la cirugía). La puntuación de CCP de un paciente también puede compararse con las puntuaciones de CCP de otros pacientes dentro de una distribución estadounidense de puntuaciones previamente observada. Para una estimación más precisa del riesgo de recidiva bioquímica a los 10 años, puede usarse la información clínica proporcionada por el profesional sanitario para calcular una puntuación de nomograma. Luego, la puntuación de CCP puede combinarse con la puntuación de nomograma para generar una puntuación combinada. Esta puntuación combinada puede usarse para estimar el riesgo de recidiva bioquímica a los 10 años, y puede compararse con las puntuaciones combinadas de otros pacientes dentro de una distribución de puntuaciones estadounidense. En algunas

realizaciones, la puntuación combinada sólo se comunica al profesional sanitario si se ha facilitado toda la información clínica requerida y, si no se facilitan todos los parámetros clínicos requeridos en el formulario de solicitud de la prueba, sólo se comunica la puntuación Prolaris. En algunas de tales realizaciones, la puntuación combinada puede obtenerse introduciendo la información clínica requerida después de notificar la puntuación de CCP introduciendo los parámetros clínicos requeridos para el nomograma junto con la puntuación de CCP del paciente.

Como medida opcional de control de calidad, en cada serie de muestras puede analizarse un control sin ARN y un control con ARN humano normal con una puntuación de CCP determinada previamente. Los controles pueden analizarse para verificar los resultados esperados.

Características de rendimiento/Limitaciones

Intervalos clínicamente notificables

Puede aplicarse un intervalo de puntuación de CCP clínicamente notificable de -1,6 a 3,7. Puede considerarse que las personas con una puntuación de CCP de 1,2 o superior tienen una probabilidad prevista de recidiva bioquímica a los 10 años superior al 50%. Las puntuaciones de CCP fuera del intervalo de -1,6 a 3,7 pueden notificarse, pero pueden matizarse con la información de que se encuentran fuera del intervalo del modelo de predicción.

De manera similar, puede aplicarse a las puntuaciones combinadas un intervalo de puntuación combinada clínicamente notificable de -0,9 a 4,5. Puede considerarse que las personas con una puntuación combinada de 2,0 o superior tienen una probabilidad prevista de recidiva bioquímica a los 10 años superior al 50%. Las puntuaciones combinadas fuera de este intervalo pueden notificarse, pero pueden matizarse con la información de que se encuentran fuera del intervalo del modelo de predicción.

Límite de detección/Linealidad

Las puntuaciones de CCP entre -8 y 8 son técnicamente detectables por el ensayo. La linealidad puede establecerse dentro de este intervalo tal como sigue: Los genes relevantes pueden amplificarse previamente, diluirse a diferentes concentraciones y añadirse en adiciones conocidas a una muestra de ADNc de concentración conocida. Cada muestra con adiciones conocidas puede someterse a ensayo por triplicado, y los 3 puntos de datos resultantes pueden promediarse para cada concentración para generar una puntuación de recidiva. La linealidad puede establecerse para las puntuaciones de CCP que van desde 0 hasta 8 usando este método. La linealidad para las puntuaciones de CCP que van desde -8 hasta 0 puede establecerse de manera similar.

Interferencia

En algunos casos, el tratamiento de terapia hormonal prequirúrgica y radioterapia pueden afectar a las puntuaciones de CCP. Por tanto, en algunas realizaciones el método se aplica sólo a pacientes que no han recibido estos tratamientos antes de la cirugía.

Limitaciones

En algunas realizaciones sólo se analizan muestras de tumores de próstata humanos FFPE. En algunas realizaciones sólo se analizan muestras de pacientes con niveles de PSA ≤100 ng/ml. En algunas realizaciones sólo se analizan muestras que produzcan al menos 125 ng de ARN.

Criterios de interpretación

Puntuaciones dentro del intervalo técnico del ensayo y dentro del intervalo de puntuaciones para las que la predicción clínica está validada (por ejemplo, de 1,6 a 3,7 para las puntuaciones de CCP: de -0,9 a 4,5 para las puntuaciones combinadas)

Tanto las puntuaciones de CCP como las puntuaciones combinadas dentro de este intervalo pueden notificarse junto con la probabilidad prevista de recidiva. Aunque el porcentaje de riesgo puede darse de manera continua en todo el intervalo, a continuación se dan ejemplos de puntuaciones de riesgo:

Puntuación de CCP	Probabilidad de recidiva
-1	12,6%
0	24,9%
1	45,5%

2	72,5%
3	93,6%

Puntuación combinada	Probabilidad de recidiva
0	11,5%
1	25%
2	49,3%
3	79,8%
4	97,7%

Puntuaciones dentro del intervalo técnico del ensayo pero fuera del intervalo de puntuaciones para las que la predicción clínica está validada (por ejemplo, de -8 a 1,7 y de 3,8 a 8 para CCP: de -8 a -1,0 y de 4,6 a 8 para puntuaciones combinadas)

Si se ha establecido la linealidad de las puntuaciones CCP y de las puntuaciones combinadas dentro de un intervalo de este tipo, puede notificarse entonces la puntuación de CCP o la puntuación combinada calculadas, pero en algunos casos puede que no se facilite el riesgo de recidiva estimado.

Puntuaciones fuera del intervalo técnico del ensayo (por ejemplo, menos de -8 o más de 8 para las puntuaciones de CCP o combinadas)

Estas puntuaciones pueden estar fuera de los límites de detección verificados de este ensayo y pueden representar un artefacto o error técnico. Por tanto, en algunos casos estas puntuaciones pueden no ser notificables.

EJEMPLO 1

Se sometió a prueba la siguiente firma de genes del ciclo celular (CCG) para predecir el tiempo hasta la recidiva química después de la prostatectomía radical.

Tabla A

Firma de recidiva de próstata de 31 CCG		
AURKA	*DTL*	*PTTG1*
BUB1	*FOXM1*	*RRM2*
CCNB1	*HMMR*	*TIMELESS*
CCNB2	*KIF23*	*TPX2*
CDC2	*KPNA2*	*TRIP13*
CDC20	*MAD2L1*	*TTK*
CDC45L	*MELK*	*UBE2C*
CDCA8	*MYBL2*	*UBE2S*
CENPA	*NUSAP1*	*ZWINT*
CKS2	*PBK*	
DLG7	*PRC1*	

La expresión media de ARNm para los 31 CCG anteriores se sometió a prueba en 440 muestras FFPE de tumores de próstata usando un modelo de riesgos proporcionales de Cox en Splus 7.1 (Insightful, Inc., Seattle WA). El valor de p de la prueba de razón de probabilidades fue de $3,98 \times 10^{-5}$.

La media de la expresión de CCG es robusta al error de medición y a la variación individual entre genes. Con el fin de determinar el número óptimo de genes del ciclo celular para la firma, se sometió a prueba el poder predictivo de la media para conjuntos seleccionados aleatoriamente de desde 1 hasta 30 de los CCG enumerados anteriormente. Esta simulación demostró que existe un número umbral de CCG en un panel que proporciona un poder predictivo significativamente mejorado.

EJEMPLO 2

En un análisis univariante, se encontró que un conjunto de 31 CCG (tabla 7) era un predictor significativo de la

recidiva bioquímica (valor de p = 1,8 x 10^{-9}) después de RP en pacientes con cáncer de próstata. Esta firma se evaluó adicionalmente para determinar si se sumaba a un nomograma clínico establecido para la recidiva del cáncer de próstata (el nomograma de Kattan-Stephenson). En resumen, el nomograma fue un predictor altamente significativo de recidiva (valor de p de 1,6 x 10^{-10}) y, después de ajustar para el nomograma, la firma de CCG fue un predictor significativo de recidiva bioquímica (valor de p de 4,8 x 10^{-5}, tabla C).

Pacientes y métodos

Se realizó un seguimiento de ochocientos cuatro pacientes consecutivos de RP durante una mediana de 9,5 años. Las características de los pacientes y los desenlaces del tratamiento de toda la cohorte se han notificado anteriormente (Swanson *et al.*, UROL ONCOL. (2007) 25: 110-114). Se disponía de bloques de tejido y/o portaobjetos de la evaluación patológica final con suficiente tejido para el análisis de 430 pacientes. La cohorte se dividió aleatoriamente en 212 pacientes utilizados para un conjunto de entrenamiento y 199 muestras de pacientes como conjunto de validación.

Expresión génica (métodos estadísticos): la asociación entre la recidiva bioquímica y la expresión de CCG se evaluó usando modelos de RP de Cox para el tiempo hasta la recidiva. Todos los valores de p notificados en este estudio se derivaron de una prueba de razón de probabilidades que comparaba el modelo nulo con el modelo que contenía la variable de prueba. Se seleccionó aleatoriamente un conjunto de 31 CCG (tabla 7, anterior). Los ensayos se usaron para generar datos de expresión de 212 pacientes en el conjunto de entrenamiento. Todos los datos de expresión se generaron por triplicado. Los datos de expresión se combinaron en una firma calculando el nivel de expresión medio para 26 CCG. La asociación entre la recidiva bioquímica y la expresión de CCG se evaluó usando modelos de RP de Cox para el tiempo hasta la recidiva.

Preparación de las muestras y diseño del estudio: se aisló ARN de cortes tumorales FFPE derivadas de 411 pacientes con cáncer de próstata tratados con RP. Se usaron cortes tumorales representativas de 10 μm de grosor para aislar el ARN. Cuando fue necesario, se usó una macro o microdisección de la muestra guiada por un patólogo para enriquecer el tejido tumoral antes del aislamiento del ARN. Ninguna de las muestras de la cohorte de validación se sometió a microdisección. Antes de cualquier análisis, la cohorte se dividió en 212 pacientes para la caracterización inicial de la firma (“conjunto de entrenamiento”) y 199 pacientes para la validación. Las características clínicas de la cohorte de entrenamiento y validación se enumeran en la tabla B.

Tabla B

	Entrenamiento	Validación	valor de p	Estadística
Edad en años a RP, media (des. est.)	67,3 (5,9)	66,8 (5,8)	0,355	Prueba de la t
Origen étnico (% de no blancos)	2,80%	7,60%	0,042	Prueba exacta de Fisher
Método de disección (% lcm)	24%	0%	NA	NA
Recidiva (%)	71/212 (33,5%)	72/199 (36,2%)	0,605	Prueba exacta de Fisher
Días hasta la recidiva, mediana	910	822	0,463	Prueba de la t
Días hasta el seguimiento, mediana	3373	3387	0,173	Prueba de la t
PSA preoperatorio (mediana)	7,3	6,8	0,163	Prueba de la t de log
Vesícula seminal	23/212 (10,8%)	28/199 (14,1%)	0,37	Prueba exacta de Fisher
Vejiga	12/212 (5,7%)	17/199 (8,5%)	0,335	Prueba exacta de Fisher
Ganglio linfático	8/212 (3,8%)	10/199 (5,0%)	0,632	Prueba exacta de Fisher
Capsular	100/212 (47,2%)	101/199 (50,8%)	0,49	Prueba exacta de Fisher
A través de la cápsula	59/212 (27,8%)	66/199 (33,2%)	0,283	Prueba exacta de Fisher
Márgenes positivos	43/212 (20,3%)	57/199 (28,6%)	0,051	Prueba exacta de Fisher
Puntuación de Gleason posterior a	80/212 (37,7%)	66/199 (33,2%)	0,354	Prueba exacta

RP > 6				de Fisher
Nomograma posterior a RP, media (des. est.)	137 (19,5)	138 (23,0)	0,424	Prueba de la t

Resultados

La firma de expresión de CCG (tabla 7, anterior) fue predictiva de la recidiva de la enfermedad en un análisis univariante (valor de p = 1,8 x 10^{-9}, tabla C). La distribución de la puntuación de firma estaba sesgada hacia valores más altos (menor expresión). El valor mediano de la puntuación de firma se usó para dividir la cohorte de entrenamiento en dos grupos que contenían muestras con expresión de CCG alta o baja. En la figura 2 se muestra la supervivencia frente al tiempo de ambos grupos.

También se evaluó el poder predictivo de la firma de CCG después de tener en cuenta las variables clínicas que se incluyen normalmente en un nomograma posquirúrgico (el nomograma de Kattan-Stephenson). El nomograma fue un predictor muy significativo de la recidiva (valor de p de 1,6 x 10^{-10}). Después de ajustar para el nomograma, la firma de CCG fue un predictor significativo de la recidiva bioquímica (figura 3) en la cohorte de descubrimiento (valor de p de 0,03) y en la cohorte de validación clínica (valor de p de 4,8 x 10^{-5}).

Tabla C

	N	Co-variables	Valor de p medio de CCG*	Razón de riesgos instantáneos de recidiva
ENTRENAMIENTO (31 CCG)	212	ninguno	0,00404	1,24
	204	nomograma posterior a la cirugía	0,03320	1,16
VALIDACIÓN (subconjunto de 26 CCG)	199	ninguno	1,8 x 10^{-9}	2,68
	197	nomograma posterior a la cirugía	4,8 x 10^{-5}	1,94

*Media de la expresión de genes del ciclo celular con imputación de valores faltantes, prueba de razón de probabilidades para el modelo de riesgos proporcionales de Cox.

Para ayudar a comprender la interacción entre el nomograma y la firma de expresión de CCG, se generó un gráfico de dispersión que comparó estos predictores (figura 4) (las estrellas gris claro representan los pacientes cuyo cáncer recidivó, mientras que las estrellas negras representan a los pacientes cuyo cáncer no recidivó). El análisis del gráfico de dispersión mediante medios de KM dividió las muestras en tres agrupaciones basándose únicamente en la puntuación de nomograma. Posteriormente, se descubrió que las agrupaciones se basaban en parámetros clínicos bien comprendidos. Los pacientes en la agrupación de puntuación más baja (116/117) presentaban una enfermedad confinada en órgano. Los pacientes en la agrupación de puntuación media (48/60) tenían al menos un parámetro posquirúrgico que se sabe que está asociado a un mal desenlace (es decir, enfermedad a través de la cápsula, ganglios linfáticos positivos para la enfermedad y/o vesículas seminales positivas para la enfermedad) y un PSA prequirúrgico bajo (<10 ng/ml). Los pacientes en la agrupación de puntuación más alta tenían al menos un parámetro posquirúrgico desfavorable y un PSA prequirúrgico alto. A continuación, los pacientes en las agrupaciones de puntuaciones baja y media se dividieron entre la media de la puntuación de CCG. Los desenlaces para los pacientes en la agrupación de puntuación más alta se predicen adecuadamente mediante el nomograma y, por tanto, no se dividieron adicionalmente. Como resultado, el gráfico de dispersión define cinco grupos de pacientes con tasas de recidiva de la enfermedad del 2%, del 40% (para dos grupos), del 65% y del 80% (tabla D). La tasa de recidiva de los cinco grupos frente al tiempo se muestra en la figura 5.

Tabla D

	Nomograma posterior a RP		
Puntuación de CCG	Bajo	Medio	Alta
Bajo	1/62 (1,6%)	13/31 (41,9%)	16/20 (80%)
Alta	21/55 (38,2%)	19/29 (65,5%)	

El gráfico de dispersión mostrado en la figura 4 sugiere que existe una interacción no lineal entre la firma de CCG y el nomograma posquirúrgico. Es decir, la firma de CCG es un mejor predictor en pacientes con puntuaciones de nomograma bajas. Por tanto, el estudio buscó pruebas estadísticas de una interacción entre estas variables en un

modelo multivariante para predecir la recidiva de la enfermedad (tabla E). Hubo pruebas significativas de una interacción favorable tanto en los estudios de entrenamiento como en los de validación. La inclusión del término de interacción en el modelo mejoró drásticamente la significación pronóstica de la firma de CCG después de ajustar para el nomograma (valores de p de 0,0015 en la cohorte de entrenamiento y 1,2 x 10^{-8} en la de validación).

Tabla E

Resumen estadístico						
Cohorte	N	Variables independientes	Co-variables	Valor de p de interacción	Valor de p variable	Razón de riesgos instantáneos de recidiva
Entrenamiento	204	nomograma	ninguna	NA	1,6 x 10^{-10}	
	212	Firma de CCG	ninguna	NA	0,004	1,24
	204	Firma de CCG	nomograma	0,021	0,0015	
Validación	197	nomograma	ninguna	NA	7,7 x 10^{-13}	
	199	Firma de CCG	ninguna	NA	1,8 x 10^{-9}	2,68
	197	Firma de CCG	nomograma	0,0001	1,28 x 10^{-8}	

EJEMPLO 3

El siguiente estudio tenía como objetivo determinar el número óptimo de CCG que debían incluirse en la firma. Tal como se mencionó anteriormente, los niveles de expresión de CCG están correlacionados entre sí, por lo que era posible que la medición de un pequeño número de genes fuera suficiente para predecir el desenlace de la enfermedad. De hecho, los CCG individuales del conjunto de 31 genes de la tabla 7 (panel C) se añaden significativamente al nomograma de Kattan-Stephenson, tal como se muestra en la tabla F a continuación (después de ajustar para el nomograma y un término de interacción entre el nomograma y la expresión de CCG).

Tabla F

N.º de gen	Símbolo de gen	Valor de p*	N.º de gen	Símbolo de gen	Valor de p*	N.º de gen	Símbolo de gen	Valor de p*
1	*NUSAP1*	2.8E-07	12	*BUB1*	8.3E-05	23	*KPNA2*	2.0E-02
2	*DLG7*	5.9E-07	13	*PBK*	1.2E-04	24	*UBE2C*	2.2E-02
3	*CDC2*	6.0E-07	14	*TTK*	3.2E-04	25	*MELK*	2.5E-02
4	*FOXM1*	1.1E-06	15	*CDC45L*	7.7E-04	26	*CENPA*	2.9E-02
5	*MYBL2*	1.1E-06	16	*PRC1*	1.2E-03	27	*CKS2*	5.7E-02
6	*CDCA8*	3.3E-06	17	*DTL*	1.4E-03	28	*MAD2L1*	1.7E-01
7	*CDC20*	3.8E-06	18	*CCNB1*	1.5E-03	29	*UBE2S*	2.0E-01
8	*RRM2*	7.2E-06	19	*TPX2*	1.9E-03	30	*AURKA*	4.8E-01
9	*PTTG1*	1.8E-05	20	*ZWINT*	9.3E-03	31	*TIMELESS*	4.8E-01
10	*CCNB2*	5.2E-05	21	*KIF23*	1.1E-02			
11	*HMMR*	5.2E-05	22	*TRIP13*	1.7E-02			

* Valor de p para la prueba de razón de probabilidades del modelo de RP de Cox completo (puntuación de nomograma posterior a RP + expresión de ciclo celular + nomograma: ciclo celular) frente a reducido (puntuación de nomograma posterior a RP sola) del tiempo hasta la recidiva.

Para evaluar el rendimiento de subconjuntos más pequeños del conjunto de CCG más grande (es decir, paneles de CCG más pequeños), el estudio también comparó lo bien que la firma predice el desenlace en función del número de CCG incluidos en la firma (figura 1). El tiempo hasta la recidiva química después de la cirugía de próstata se regresionó sobre la media de CCG ajustada por la puntuación de nomograma posterior a RP. Los datos consisten en ensayos de TLDA expresados como deltaCT para 199 muestras de tumores de próstata FFPE y 26 CCG y se analizaron mediante un modelo de RP de Cox multivariante. Los valores de p corresponden a la prueba de razón de probabilidades del modelo completo (nomograma + media del ciclo celular, incluida la interacción) frente al modelo reducido (sólo nomograma). Tal como se muestra en la tabla G a continuación y en la figura 1, las firmas de CCG pequeñas (por ejemplo, 2, 3, 4, 5, 6 CCG, etc.) se añaden significativamente al nomograma de Kattan-Stephenson:

Tabla G

n.º de CCG	Media de log10 (valor de p)*
1	-3,579
2	-4,279
3	-5,049
4	-5,473
5	-5,877
6	-6,228

* Para 1000 subconjuntos extraídos aleatoriamente, de tamaño 1 a 6, de genes del ciclo celular.

EJEMPLO 4

El objetivo de este experimento fue evaluar la asociación entre las mutaciones de *PTEN* y la recidiva bioquímica en pacientes con cáncer de próstata después de la prostatectomía radical. Se encontró que las mutaciones somáticas de *PTEN* estaban significativamente asociadas con la recidiva y, lo que es más importante, añadían información pronóstica más allá del nomograma clínico establecido para la recidiva del cáncer de próstata (el nomograma de Kattan-Stephenson) y la puntuación de firma de CCG (descrita en los ejemplos 1 y 2, anteriormente).

Pacientes y métodos

Se realizó un seguimiento de ochocientos cuatro pacientes consecutivos de RP durante una mediana de 9,5 años. Las características de los pacientes y los desenlaces del tratamiento de toda la cohorte se han descrito anteriormente (Swanson *et al.*, UROL. ONCOL. (2007) 25: 110-114). Se disponía de bloques de tejido y/o portaobjetos de la evaluación patológica final con suficiente tejido para el análisis de 430 pacientes. De ellos, se seleccionaron 191 para el cribado de mutaciones de *PTEN* en función de la cantidad de tumor disponible.

Se aisló ADN genómico de las muestras tumorales FFPE para el cribado de mutaciones de PTEN usando el kit de tejido FFPE para ADN QIAamp (Qiagen, Valencia, CA) según el protocolo del kit. Los portaobjetos FFPE se tiñeron en primer lugar con hematoxilina y eosina y se examinaron por un patólogo para identificar la región tumoral. Después de la desparafinización, se recortó el tejido tumoral de los portaobjetos con una cuchilla de afeitar. En algunas muestras, la disección se facilitó mediante microscopía de captura láser (LCM), debido a la dispersión de las células tumorales.

Las mutaciones se detectaron diseñando cebadores de secuenciación para interrogar la secuencia genómica de *PTEN*. Los cebadores contenían colas M13 directas e inversas para facilitar la secuenciación. Después de la amplificación, la secuencia de ADN se determinó en un equipo Mega BASE 4500 (GE healthcare) usando la química de colorante-cebador tal como se describe en Frank *et al.*, J. CLIN. ONCOL. (2002) 20:1480-1490. Debido a las dificultades técnicas asociadas con la secuenciación del ADN derivado de material FFPE, cada mutación se detectó mediante al menos dos reacciones independientes de amplificación y secuenciación.

Métodos estadísticos: a menos que se especifique lo contrario, la asociación entre la recidiva bioquímica y las mutaciones de *PTEN* se evaluó usando modelos de RP de Cox para el tiempo hasta la recidiva. Los valores de p resultantes se derivaron de una prueba de razón de probabilidades que comparaba el modelo nulo con el modelo que contenía la variable de prueba. En este ejemplo (ejemplo 4), la firma de CCG se derivó de 26 CCG (panel D de la tabla 2, anteriormente). Todos los datos de expresión se generaron por triplicado. Los datos de expresión se combinaron en una firma calculando el nivel de expresión medio para 26 CCG. Los datos clínicos fueron las variables incluidas en el nomograma de Kattan-Stephenson.

Resultados

Se encontraron mutaciones de *PTEN* en 13 individuos (13/191). En este subconjunto de 191 pacientes, *PTEN* fue un predictor significativo de recidiva bioquímica (valor de p = 0,031). La tasa de recidiva en portadores de mutaciones fue del 69% (9/13) en comparación con el 36% (64/178) en pacientes no mutantes. La diferencia en la tasa de recidiva también es significativa usando una prueba exacta de Fisher (valor de p = 0,034). En el subconjunto de pacientes con datos de parámetros clínicos, puntuación de firma de CCG y mutaciones de *PTEN*, el estado de *PTEN* fue un predictor significativo de recidiva bioquímica después de ajustar tanto para los parámetros clínicos como para la firma de CCG (valor de p de 0,024). Finalmente, la combinación de la mutación de *PTEN* con la firma de CCG fue un mejor predictor del desenlace después de ajustar para parámetros clínicos que el uso de la firma de CCG después de ajustar para los parámetros clínicos (valor de p = 0,0002 para la combinación en comparación con 0,0028 para CCG solamente). Estos resultados muestran que las mutaciones de *PTEN* proporcionan información sobre la probabilidad de recidiva que no está correlacionada con los parámetros clínicos ni con la firma de CCG, y que el uso de los tres parámetros para evaluar el riesgo de recidiva proporciona una estimación más precisa de la probabilidad de recidiva de lo que era posible anteriormente.

EJEMPLO 5

Este ejemplo describe estudios adicionales para validar y refinar algunas realizaciones de las firmas de CCG de la divulgación.

Pacientes y métodos

Se realizó un seguimiento de ochocientos y cuatro pacientes consecutivos sometidos a prostatectomía radical durante una mediana de 9,5 años. La mediana de edad fue de 67 años. El estadio clínico fue Tl 34%, T2 66% y T3 <1%. La mediana de PSA preoperatorio fue de 6,6 ng/ml, con un 72% < 10 ng/ml y un 28% > 10 ng/ml. Se entintaron las muestras y se registraron los parámetros clínicos en cuanto a margen del cuello vesical o de la uretra positivo, invasión de la cápsula, extensión a través de la cápsula, márgenes positivos y afectación de las vesículas seminales. La recidiva bioquímica se definió como un PSA >0,3 ng/ml. Para este estudio había un acceso a los datos clínicos de 690 pacientes. Se disponía de bloques de tejido y/o portaobjetos de la evaluación patológica final con suficiente tejido para el análisis de 442 pacientes. La cohorte se dividió en 195 pacientes para una cohorte de entrenamiento y 247 pacientes para la validación.

Selección de genes

Se realizaron ensayos de 126 CCG y 47 genes HK (de mantenimiento) contra 96 muestras FFPE de tumores de próstata anónimas obtenidas comercialmente, sin desenlaces ni otros datos clínicos. La hipótesis de trabajo era que los ensayos medirían con distintos grados de precisión el mismo fenómeno subyacente (proliferación del ciclo celular dentro del tumor para los CCG y concentración de la muestra para los genes HK). Los ensayos se clasificaron según el coeficiente de correlación de Pearson entre el gen individual y la media de todos los genes candidatos, que es la mejor estimación disponible de la actividad biológica. Los resultados de la correlación de cada uno de los 126 CCG con la media se presentan en la tabla 23. Sin incluir los CCG con una expresión promedio baja, o los ensayos que produjeron fallos en las muestras, aproximadamente la mitad de los CCG tuvieron correlaciones inferiores a 0,58, y una cuarta parte de los genes HK tuvieron correlaciones inferiores a 0,95. Estos ensayos se eliminaron, dejando un subconjunto de 56 CCG (panel G) y 36 genes HK candidatos (tablas 11 y 12). Se volvieron a calcular los coeficientes de correlación en este subconjunto y se realizó la selección final a partir de la lista clasificada.

Expresión génica

Se extrajo ARN total de cortes tumorales FFPE representativas de 5 μm de grosor. Las muestras se desparafinizaron usando un baño de xileno y posteriormente se hidrataron en series graduales de baños de etanol. Después de eso, la región tumoral se diseccionó del portaobjetos con una cuchilla de afeitar siguiendo las instrucciones del patólogo. Alternativamente, la región tumoral se diseccionó directamente en un tubo Eppendorf y se eliminó la parafina usando xileno y se lavó con etanol. Después, las muestras se trataron durante la noche con digestión con proteinasa K a 55ºC. El ARN total se extrajo usando RNeasy FFPE o miRNeasy (Qiagen) tal como describe el fabricante (con la única excepción de la digestión prolongada con proteinasa K descrita anteriormente). El ARN total aislado se trató con ADNasa I (Sigma) antes de la síntesis de ADNc. Posteriormente, se empleó el kit de archivo de ADNc de alta capacidad (Applied Biosystems) para convertir el ARN total en ADNc monocatenario, tal como describe el fabricante. Se necesitó un mínimo de 200 ng de ARN para la reacción de RT.

Antes de medir los niveles de expresión, el ADNc se amplificó previamente con una reacción conjunta que contenía ensayos TaqMan. Las condiciones de la reacción de amplificación previa fueron: 14 ciclos de 95ºC durante 15 s y 60ºC durante 4 minutos. El primer ciclo se modificó para incluir una incubación de 10 minutos a 95ºC. La reacción de amplificación se diluyó 1:20 usando el tampón TE 1X antes de su carga en los matrices de baja densidad de TaqMan™ (TLDA, Applied Biosystems) para medir la expresión génica.

Puntuación de CCG

La puntuación de CCG se calcula a partir de la expresión de ARN de 31 CCG (panel F) normalizada por 15 genes de mantenimiento (HK). Los números relativos de CCG (31) y genes HK (15) se optimizaron para minimizar la varianza de la puntuación de CCG. La puntuación de CCG es la media no ponderada de los valores CT para la expresión de los CCG, normalizada por la media no ponderada de los genes HK, de modo que los valores más altos indican una mayor expresión. Una unidad equivale a un cambio de dos veces en la expresión. Los valores faltantes se imputaron usando la expresión media para cada gen determinada en el conjunto de entrenamiento usando únicamente muestras de buena calidad. Las puntuaciones de CCG se centraron en el valor medio, también determinado en el conjunto de entrenamiento.

Se realizó un experimento de dilución con cuatro de las muestras comerciales de próstata para estimar el error de medición de la puntuación de CCG (ee = 0,10) y el efecto de los valores faltantes. Se comprobó que la puntuación de CCG permanecía estable a medida que disminuía la concentración hasta el punto de 5 fallos del total de 31 CCG. Basándose en este resultado, a las muestras con más de 4 valores faltantes no se les asignó una puntuación de CCG.

El umbral de puntuación de CCG para determinar el bajo riesgo se basó en la puntuación de CCG más baja de recidivas en el conjunto de entrenamiento. Luego, el umbral se ajustó a la baja en 1 desviación estándar para optimizar el valor predictivo negativo de la prueba.

Modelo de riesgo clínico

Se usó un modelo de riesgos proporcionales de Cox para resumir los datos de parámetros clínicos disponibles y estimar el riesgo clínico previo de recidiva bioquímica para cada paciente. El conjunto de datos consistió en 195 casos del conjunto de entrenamiento y otros 248 casos con información de parámetros clínicos pero muestra insuficiente para medir la expresión de ARN. Se realizaron pruebas univariantes de los parámetros clínicos que se sabe que están asociados con el desenlace (véase la Tabla H a continuación). Los parámetros no significativos se excluyeron del modelo. Se creó una variable compuesta para la enfermedad confinada en órgano, con invasión definida como márgenes quirúrgicos, extensión extracapsular o afectación de cualquiera de las vesículas seminales, márgenes del cuello vesical/uretra o ganglios linfáticos. La variable compuesta de enfermedad confinada en órgano resultó más significativa en el modelo que cualquiera de sus cinco componentes, algunos de los cuales estaban correlacionados unos con otros o no eran prevalentes. El ajuste del modelo se realizó utilizando los criterios AIC para las covariables posoperatorias.

Tabla H: Análisis univariante de los parámetros clínicos y asociación con la recidiva bioquímica.

Variable clínica	valor de p*	n.º de apariciones	Total	Frecuencia
VEJIGA	0,0002	36	443	0,081
CAPSULAR	$1,1 \times 10^{-9}$	194	443	0,438
ETNIA (BLANCA)	0,6741	416	439	0,948
GANGL.LINF.	0,0009	33	443	0,074
MARG.POS	$6,1 \times 10^{-11}$	83	443	0,187
PATHGLEA	$6,7 \times 10^{-16}$	NA	443	NA
PATHGRAD	$2,4 \times 10^{-11}$	NA	443	NA
PATHSTAG	$3,1 \times 10^{-15}$	NA	443	NA
PRE.PSA.LOG10	$6,2 \times 10^{-12}$	NA	443	NA
VES.SEM.	$3,0 \times 10^{-8}$	56	443	0,126

CIRUGÍA.AÑO	0,0803	NA	443	NA
A TRAVÉS.CAP	$1,3 \times 10^{-10}$	114	443	0,257
*Valor de p de RP de Cox para la prueba de razón de probabilidades				

El modelo final (es decir, el nomograma) tiene variables binarias para la enfermedad confinada en órgano y la puntuación de Gleason menor que o igual a 6, y una variable continua para el PSA logarítmico (tabla I). Este modelo incluye todos los parámetros clínicos incorporados en el nomograma posterior a la RP (es decir, el nomograma de Kattan-Stephenson), excepto el año de la RP y los dos componentes de la puntuación de Gleason. La distribución del riesgo clínico previo muestra tres ganglios distintos (figura 8). Se usó la agrupación de *k*-medias con 3 centros para establecer el umbral de la agrupación de bajo riesgo, que comprende aproximadamente el 50% de la muestra.

Tabla I: modelo clínico

Parámetro clínico	Coeficiente	RRHH	valor de p*
Enfermedad confinada en órgano	-0,827	0,44	$3,4 \times 10^{-6}$
Puntuación de Gleason ≤ 6	-0,8734	0,42	$4,2 \times 10^{-7}$
log PSA	0,6678	1,95	$2,0 \times 10^{-4}$
Valor de p de RP de Cox para la prueba de razón de probabilidades			

Análisis estadístico

Los parámetros clínicos se compararon entre los conjuntos de entrenamiento y de validación usando la prueba de la t de Student para los parámetros continuos y la prueba exacta de Fisher para los parámetros categóricos. El riesgo clínico previo de recidiva bioquímica de los pacientes después de la cirugía se estimó mediante una puntuación de nomograma posterior a RP que resumía 7 covariables. Se usó la agrupación de *k*-medias de la puntuación de nomograma para clasificar a los pacientes como de bajo o alto riesgo clínico previo. Los datos de expresión se expresaron como CT (el ciclo de PCR en el que la intensidad de fluorescencia supera un umbral predeterminado) de cada CCG normalizado por la media de los 15 genes de mantenimiento (tabla 12 anterior).

Las muestras de mala calidad se excluyeron del análisis para eliminar las muestras de mala calidad o las lecturas dudosas sin comprometer la integridad de la firma al excluir inadvertidamente muestras con baja expresión de CCG. Por consiguiente, los umbrales de limpieza o filtrado de los datos se fijaron de manera conservadora. Los niveles medios de expresión de los genes HK de cada muestra, que eran superiores a los de los CCG, se usaron para identificar las muestras de mala calidad. Las métricas técnicas de la eficiencia de amplificación y las desviaciones estándar excesivamente altas de las réplicas se usaron para identificar mediciones de CT poco fiables. No se permitieron fallos de genes HK, ni más de 1 fallo de 3 réplicas para los CCG.

La asociación entre la recidiva bioquímica y la expresión de CCG después de ajustar para el riesgo clínico predicho por los parámetros clínicos se evaluó usando un modelo de riesgos proporcionales de Cox para el tiempo hasta la recidiva. El supuesto de riesgos proporcionales de no dependencia temporal se sometió a prueba para el modelo completo de la firma de CCG más la puntuación binaria de parámetros clínicos con un término de interacción, y para la firma de CCG sólo en los subconjuntos de riesgo clínico. No fue significativo ni en el entrenamiento ni en la validación, lo que indica que no hay pruebas de dependencia temporal. Todos los valores de p indicados proceden de una prueba de razón de probabilidades que compara el modelo reducido o nulo con el modelo que contiene la variable de prueba. Los gráficos de Kaplan-Meier se usan para mostrar las probabilidades de supervivencia estimadas para subconjuntos de pacientes; sin embargo, los valores de p proceden de la prueba de razón de probabilidades de Cox para los valores continuos de la variable. Todos los análisis estadísticos se realizaron en S+ versión 8.1.1 para Linux (TIBCO Spotfire) o R 2.9.0 (http://www.r-project.org).

Resultados

Se aisló ARN de cortes tumorales FFPE derivadas de 442 pacientes con cáncer de próstata tratados con RP. La cohorte se dividió en 195 pacientes para la caracterización inicial de la firma (“conjunto de entrenamiento”) y 247 pacientes para la validación. Los parámetros clínicos de la cohorte de entrenamiento y de validación se enumeran en la Tabla J. No hubo diferencias significativas después de ajustar para comparaciones múltiples.

Tabla J: parámetros clínicos de las cohortes de pacientes de entrenamiento y de validación

Parámetro clínico	Entrenamiento	Validación	Valor de p	Análisis estadístico
Edad en años en RP, media (des. est.)	67,5 (6,2)	66,8 (5,6)	0,204	Prueba de la t
Origen étnico (% de no blancos)	3,10% (2 negros, 3 hispanos, 1 otro)	7,30% (10 negros, 7 hispanos, 1 otro)	0,058	Prueba exacta de Fisher
Recidiva	73/195 (37,4%)	90/247 (36,4%)	0,843	Prueba exacta de Fisher
Días hasta la recidiva, mediana	839	736	0,308	Prueba de la t
Días hasta el seguimiento, mediana	3300	3332	0,556	Prueba de la t
PSA antes de la cirugía RP, mediana	7,4	6,4	0,022	Prueba de la t de log
Vesículas seminales	23/195 (11,8%)	33/247 (13,4%)	0,668	Prueba exacta de Fisher
Margen del cuello vesical/uretra	12/195 (6,2%)	16/247 (6,5%)	1	Prueba exacta de Fisher
Ganglios linfáticos	8/195 (4,1%)	12/247 (4,9%)	0,819	Exacta de Fisher
Penetración capsular	104/195 (53,3%)	115/247 (46,6%)	0,18	Prueba exacta de Fisher
A través de la cápsula	66/195 (33,8%)	73/247 (29,6%)	0,354	Prueba exacta de Fisher
Márgenes positivos	51/195 (26,2%)	61/247 (24,7%)	0,742	Prueba exacta de Fisher
Puntuación de Gleason posterior a RP < 7	114/195 (58,5%)	166/247 (67,2%)	0,06	Prueba exacta de Fisher
Enfermedad confinada en órgano	108/195 (55,4%)	156/247 (63,2%)	0,118	Prueba exacta de Fisher
PFP a los 10 años (IC del 95%)	61% (52%, 69%)	67% (60%, 73%)	0,905	Prueba de rangos logarítmicos

Para analizar la firma de CCG para este estudio, se sometieron a prueba 126 CCG en ARN derivado de 96 tumores de próstata (tabla 11). Las muestras tumorales eran anónimas y no estaban asociadas con datos clínicos. De este conjunto de genes, se seleccionaron 31 genes (panel F) para su inclusión en la firma (tabla K). Los genes se seleccionaron en función de su rendimiento técnico y de lo bien que se correlacionada cada gen con el nivel de expresión medio de todo el conjunto de CCG en las 96 muestras anónimas.

Tabla K: firma de CCG del conjunto de entrenamiento (panel F)

Símbolo	ID de gen	Símbolo	ID de gen	Símbolo	ID de gen
ASF1B	55723	*CENPM*	79019	*ORC6L*	23594
ASPM	259266	*CEP55*	55165	*PBK*	55872
BIRC5	332	*DLGAP5*	9787	*PLK1*	5347
BUB1B	701	*DTL*	51514	*PRC1*	9055
C18orf24	220134	*FOXM1*	2305	*PTTG1*	9232
CDC2	983	*KIAA0101*	9768	*RAD51*	5888
CDC20	991	*KIF11*	3832	*RAD54L*	8438
CDCA3	83461	*KIF20A*	10112	*RRM2*	6241
CDCA8	55143	*MCM10*	55388	*TK1*	7083
CDKN3	1033	*NUSAP1*	51203	*TOP2A*	7153
CENPF	1063				

Para evaluar la utilidad pronóstica de la firma de CCG, se generaron datos de expresión en 195 pacientes en el conjunto de entrenamiento. Dado que los niveles de expresión génica individuales estaban correlacionados, se combinaron en una puntuación de firma calculando la expresión media para el conjunto completo de 31 genes (panel F), normalizada por 15 genes de mantenimiento (tabla 12). La distribución de la puntuación de CCG se centró en cero, y cada unidad de puntuación corresponde a un cambio de 2 veces en el nivel de expresión. Las muestras de mala calidad se identificaron observando una baja expresión de los genes de mantenimiento o un número inaceptable de fallos de CCG, y se excluyeron del análisis. Después de aplicar las reglas de exclusión, quedaban 140 muestras disponibles para el análisis. La asociación entre la recidiva bioquímica y la expresión de CCG se evaluó mediante modelos de RP de Cox para el tiempo hasta la recidiva. Un valor elevado de expresión de CCG fue predictivo de recidiva de la enfermedad en un análisis univariante (valor de p= 0,01, tabla 17).

A continuación, se evaluó la utilidad pronóstica de la firma de CCG después de tener en cuenta los parámetros clínicos que se sabe que están asociados con la recidiva después de la RP. Para tener en cuenta las medidas clínicas en el análisis, se creó un modelo/nomograma que incluía el PSA preoperatorio, la puntuación de Gleason y la evidencia de enfermedad fuera de la próstata (es decir, cualquier extensión extracapsular o patología posquirúrgica positiva en ganglios linfáticos, márgenes, cuello vesical, margen de la uretra o vesículas seminales). El modelo se optimizó en 443 pacientes (tablas 13 y 14), incluyendo todos los pacientes de los que se tenían datos clínicos pero que no estaban en el conjunto de validación, y fue un predictor altamente significativo de recidiva en la cohorte de entrenamiento (valor de $p = 2,5 \times 10^{-11}$). La distribución de las puntuaciones del modelo clínico contenía varios modos (figura 8), que separaban los grupos de pacientes de alto y bajo riesgo. Por tanto, la puntuación se usó posteriormente como variable binaria (alto o bajo riesgo). La agrupación de bajo riesgo se correlacionó con un conjunto coherente de parámetros clínicos. Específicamente, la gran mayoría (215/218) tenía enfermedad confinada en órgano y puntuación de Gleason < 7. Además, el 80% tenía un PSA prequirúrgico bajo (< 10 ng/ml). Los pacientes de la agrupación de alto riesgo (N=225) eran más heterogéneos, pero tendían a presentar características clínicas que se sabe que están asociadas con un mal desenlace (por ejemplo, Gleason > 6 y/o enfermedad a través de la cápsula).

El análisis multivariante del conjunto de entrenamiento que incorporaba el modelo clínico binario mostró evidencias de una interacción no lineal entre la firma de expresión y los parámetros clínicos (tabla L). Para ayudar a comprender la naturaleza de esta interacción, se generó un gráfico de dispersión comparando estos predictores (figura 8). Tal como resulta evidente a partir de la figura, la puntuación de CCG resultó útil para evaluar el riesgo de recidiva en pacientes definidos como de bajo riesgo por los parámetros clínicos. De hecho, incluso después de ajustar para el modelo clínico dentro de los pacientes de bajo riesgo, la firma de CCG fue un fuerte predictor de recidiva bioquímica (valor de p = 0,0071).

Tabla L: resumen estadístico

		Entrenamiento de 31 genes N=195			Validación de 31 genes N=247		
	Subconjunto basado en el modelo clín.	Valor de p del efecto principal	Valor de p de interacción	n	Valor de p del efecto principal	Valor de p de interacción	n
Puntuación de CCG		0,01		140	$5,8 \times 10^{-8}$		218

Riesgo clínico binario (bajo frente a alto)		$5,1 \times 10^{6}$		133	$1,1 \times 10^{-10}$		215
Puntuación de CCG ajustada para riesgo clín. binario + interacción		0,018	0,032	133	$8,3 \times 10^{-7}$	0,026	215
Puntuación de CCG sólo	bajo riesgo	0,0038		54	$7,5 \times 10^{-5}$		112
Puntuación de riesgo clín.	bajo riesgo	0,22		54	0,044		112
Puntuación de CCG ajustada para riesgo clín. (riesgo clín. frente a riesgo clín. + CCG)	bajo riesgo	0,0071		54	0,00019		112
Puntuación de CCG	alto riesgo	0,48		79	$5,8 \times 10^{-4}$		103
Puntuación de riesgo clínico	alto riesgo	$2,8 \times 10^{-6}$		79	0,0076		103
Puntuación de CCG ajustada para riesgo clín. (riesgo clín. frente a riesgo clín. + CCG)	alto riesgo	0,51		79	0,0026		103

Se usaron los datos de entrenamiento en el gráfico de dispersión para establecer una puntuación de umbral optimizada de -0,16 para la firma de CCG (la puntuación de CCG media es cero). La figura 12 muestra este umbral aplicado a los 443 pacientes estudiados en este ejemplo. El cuarenta por ciento de los pacientes de bajo riesgo se sitúan por debajo de este umbral, y se seleccionó de forma que no hubiera recidivas a los 10 años después de la RP (es decir, un valor predictivo negativo [VPN] del 100%). Como resultado del establecimiento de valores de umbral tanto para el modelo clínico como para la puntuación de CCG, el gráfico de dispersión se dividió en cuatro secciones con tasas de recidiva del 0% (CCG bajo) y del 26% (CCG alto) para los pacientes de bajo riesgo; y del 60% (CCG bajo) y del 50% para los pacientes de alto riesgo.

A continuación, se generaron datos de expresión de CCG en 247 pacientes en la cohorte de validación. Treinta y dos muestras se eliminaron de los análisis adicionales según las reglas de exclusión desarrolladas en la cohorte de entrenamiento. El panel F fue un predictor significativo de recidiva bioquímica en un análisis univariante (valor de p = $5,8 \times 10^{-8}$, tabla 1). Después de ajustar para el modelo clínico binario, la firma de CCG fue altamente predictiva de recidiva en la cohorte de validación (valor de p de $8,3 \times 10^{-7}$), y como en el conjunto de entrenamiento, hubo evidencia significativa de una interacción no lineal entre variables. La firma de CCG fue informativa en todo el espectro de riesgo definido clínicamente (tabla 17). En cuanto a validación de los resultados de entrenamiento, el valor de p para la asociación entre recidiva y firma de CCG en pacientes de bajo riesgo fue de $1,9 \times 10^{-4}$.

Se aplicó el umbral de CCG derivado del presente análisis de la cohorte de entrenamiento al conjunto de datos de validación (figura 9). Los pacientes de bajo riesgo con puntuaciones de CCG por debajo del umbral tenían una tasa de recidiva prevista a los 10 años del 5% (equivalente a un VPN validado de 0,95). En general, la combinación de la puntuación de CCG y los parámetros clínicos dividió la cohorte en cuatro grupos con tasas de recidiva previstas a los 10 años del 5%, del 22%, del 36% y del 70% (tabla M). En las figuras 10 y 11 se muestra la tasa de recidiva prevista frente a la puntuación de CCG para los pacientes de la cohorte de validación.

Tabla M: resumen de las tasas de recidiva en la cohorte de validación definida por riesgo clínico y puntuación de CCG

Riesgo clínico	Puntuación de CCG	Tasa de recidiva a los 10 años	
		Estimación de Kaplan-Meier	n
bajo	bajo	0,05	39
bajo	alta	0,22	73
alta	bajo	0,36	27
alta	alta	0,7	76

Se sometió a prueba el umbral validado frente a diversas definiciones de pacientes de bajo riesgo (tabla N). La puntuación de firma fue un indicador pronóstico significativo en una variedad de definiciones clínicas de bajo riesgo y, dependiendo de la definición, generó una tasa de recidiva predicha a los 10 años de 0,05 a 0,10.

Tabla N: VPN de la firma de CCG en otras definiciones de pacientes de bajo riesgo

Definición clínica de bajo riesgo	CCP bajo*		valor de p***
	Predicción de recidiva a loa 10 años**	n	
Enfermedad confinada en órgano y puntuación de Gleason < 7 y PSA < 10	0,05	39	$9,4 \times 10^{-4}$
Enfermedad confinada en órgano y puntuación de Gleason < 7	0,08	40	$5,8 \times 10^{-3}$
Enfermedad confinada en órgano y puntuación de Gleason < 8 y PSA < 10	0,07	42	$8,7 \times 10^{-4}$
Enfermedad confinada en órgano y puntuación de Gleason < 8	0,1	43	$4,1 \times 10^{-3}$
Enfermedad confinada en órgano sola	0,1	44	$2,4 \times 10^{-3}$
*- definido por umbral validado			
**- Estimaciones de Kaplan-Meier			
***- para la diferencia entre las estimaciones de KM para riesgo bajo y alto ajustadas por la varianza de Greenwood.			

Comentario

Los presentes inventores han desarrollado y validado una firma molecular pronóstica para el cáncer de próstata. La firma se basa en la medición de los niveles de expresión de ARNm de los genes del ciclo celular (CCG). Por definición, la expresión de los CCG se regula en función de la fase del ciclo celular. Es decir, se activan en fases específicas del ciclo celular, de modo que las células en crecimiento activo tienen mayores niveles de expresión de CCG que las células quiescentes. Presumiblemente, este hecho subyace a la capacidad de la firma para predecir la progresión del cáncer. Sin querer limitarse a la teoría, se cree que al medir los niveles de expresión de CCG se están midiendo indirectamente la tasa de crecimiento y la agresividad inherente del tumor, lo que en última instancia repercute en la probabilidad de recidiva del cáncer de próstata después de la prostatectomía.

Existe una distinción importante entre este estudio y muchos otros que han intentado generar firmas moleculares pronósticas. A menudo, los estudios similares comienzan con un gran número de biomarcadores candidatos (que a veces superan los 1.000 genes) que luego se evalúan en busca de su asociación con un fenotipo clínico de interés. En ocasiones, este enfoque puede adolecer de pruebas múltiples inherentes, lo que puede hacer que la importancia de la firma derivada sea incierta. En este caso se ha sometido a prueba una única hipótesis: el CCG sería pronóstico en el cáncer de próstata (de hecho, se seleccionaron los genes en función de su correlación con la expresión de la CCG, no en función de su asociación con la recidiva). Y puesto que la expresión de CCG está correlacionada, se combinaron los datos de expresión en una firma predictiva determinando el valor medio de expresión de todos los

genes de la firma. La simplicidad de este enfoque, desde el punto de vista biológico y computacional, respalda la opinión de que es probable que la afirmación principal de este estudio sea muy sólida y se reproduzca en estudios posteriores.

La firma de CCG (panel F) es predictiva de manera independiente y se añade significativamente al poder predictivo de los parámetros clínicos normalmente empleados para predecir la recidiva de la enfermedad después de la cirugía. Esto es cierto tanto en la cohorte de entrenamiento como en la de validación.

La firma es inmediatamente útil para definir el riesgo de los pacientes que presentan parámetros clínicos de bajo riesgo. En este caso, se definió esencialmente el bajo riesgo como Gleason <7, PSA<10 y enfermedad confinada en órgano. La puntuación de firma de CCG subdivide eficazmente el grupo de bajo riesgo en pacientes con tasas de recidiva muy bajas (5%) y un mayor riesgo de recidiva (22%) (figura 9 y tabla M). Este es el efecto más drástico de la firma molecular, redefinir con precisión el riesgo de los pacientes previamente definidos como de bajo riesgo basándose en parámetros clínicos. Cabe destacar que dentro de esta subpoblación de pacientes (es decir, pacientes definidos como de bajo riesgo basándose en parámetros clínicos) los parámetros clínicos no son especialmente pronósticos (véase la tabla L). Por tanto, como prueba diagnóstica, la firma podría ser útil para un gran número de pacientes. En este estudio, casi el 60% de la cohorte se caracterizó como de bajo riesgo y se esperaba que el 40% de ellos tuviera puntuaciones de CCG bajas. Por tanto, la firma de CCG puede predecir enfermedad indolente en una cuarta parte de los pacientes que han sido identificados previamente como de alto riesgo (y, por tanto, identificados como candidatos a prostatectomía radical). Finalmente, los datos de validación en particular sugieren que la firma de CCG puede ser útil para definir el riesgo en todos los pacientes. Específicamente, ayudó a dividir a los pacientes definidos como de alto riesgo según los parámetros clínicos en aquellos con tasas de recidiva del 30% y del 70% (tabla M).

La combinación de los parámetros clínicos y la firma de CCG permite a los médicos predecir con mayor precisión el riesgo de fallo quirúrgico y, por tanto, identificar el transcurso de intervención terapéutica adecuado. Tal como se demostró, la firma mejora drásticamente la predicción de recidiva en pacientes que presentan parámetros clínicos generales de enfermedad no agresiva (tabla N). Dentro de este subgrupo clínico, los pacientes con puntuaciones de CCG bajas se beneficiarían de la seguridad absoluta de que no está indicado ningún tratamiento adicional. Por el contrario, el grupo con CCG alto puede requerir una intervención inmediata. Los pacientes con parámetros clínicos posquirúrgicos desfavorables se benefician de la radioterapia complementaria. Por tanto, la firma de CCG debería predecir la eficacia de la radiación complementaria para pacientes con características clínicas de bajo riesgo y puntuaciones de CCG altas. En la cohorte de validación, los pacientes con puntuaciones de CCG altas y enfermedad más allá de la próstata tienen una tasa de recidiva del 70%, lo que debería identificar claramente a los pacientes que son buenos candidatos para la radiación complementaria. Por tanto, la combinación de parámetros clínicos y la firma de CCG conduce claramente a una definición más precisa del riesgo del paciente, lo que debería permitir una evaluación más inteligente de la necesidad de tratamiento adicional.

EJEMPLO 6

Algunos de los paneles de CCG descritos en el presente documento se evaluaron adicionalmente para determinar su capacidad de pronóstico en otros tipos de cáncer. Se encontró que los paneles C, D y F fueron pronósticos en diversos grados en el cáncer de vejiga, cerebro, mama y pulmón.

Métodos

Los datos de expresión génica y de pacientes se obtuvieron de los siguientes conjuntos de datos disponibles públicamente: GSE7390 (Desmedt *et al.*, CLIN. CANCER RES. (2007) 13:3207-14; PMID 17545524); GSE11121 (Schmidt *et al.*, CANCER RES. (2008) 68:5405-13; PMID 18593943); GSE8894 (Son *et al.*; sin publicar); Shedden (Shedden *et al.*, NATURE MED. (2008) 14: 822; PMID 18641660); GSE4412 (Freije *et al.*, CANCER RES. (2004) 64:6503-10; PMID 15374961); GSE4271 (Phillips *et al.*, CANCER CELL (2006) 9:157-73; PMID 16530701); GSE5287 (Als *et al.*, CLIN. CANCER RES. (2007) 13:4407-14; PMID 17671123). Cada uno de estos conjuntos de datos tiene asociada una descripción detallada de los procedimientos experimentales usados en la recopilación de datos de expresión y de pacientes. Los micromatrices de expresión usados para generar cada conjunto de datos se resumen a continuación en la tabla O.

Tabla O

Conjunto de datos	Matriz
GSE7390	Affymetrix U133 A
GSE11121	Affymetrix U133 A
GSE8894	Affymetrix U133 plus 2.0

Shedden	Affymetrix U133 A
GSE4412	Affymetrix U133 A y B
GSE4271	Affymetrix U133 A y B
GSE5287	Affymetrix U133 A

Los datos de expresión de cada uno de los genes en los paneles C, D y F se recopilaron a partir de estos conjuntos de datos y se determinó el nivel de expresión medio de cada panel para cada paciente, cuyo desenlace clínico se conocía (por ejemplo, recidiva, progresión, supervivencia libre de progresión, supervivencia global, etc.). La puntuación de CCG es una expresión promedio de los genes de un panel. Si un gen está representado por más de un conjunto de sondas en la matriz, la expresión génica es una expresión promedio de todos los conjuntos de sondas que representan al gen. La asociación entre la puntuación de CCG y la supervivencia o la recidiva de la enfermedad se sometió a prueba usando un modelo de riesgos proporcionales de Cox univariante y multivariante. El análisis multivariante se realizó cuando se disponía de parámetros clínicos relevantes (grado en el cáncer cerebral, estadio en el cáncer de pulmón, NPI en el cáncer de mama).

Resultados

Tal como se muestra en la tabla P a continuación, cada panel, en el análisis univariante, fue un factor pronóstico en cada uno de los cánceres analizados.

Tabla P

Tipo de cáncer	Conjunto de datos	Valor de p		
		Panel C	Panel F	Panel B
Cáncer de mama positivo para ER	GSE7390	$2,4 \times 10^{-3}$	$2,3 \times 10^{-3}$	$4,3 \times 10^{-3}$
Cáncer de mama positivo para ER	GSE11121	$1,2 \times 10^{-5}$	$8,7 \times 10^{-6}$	$1,5 \times 10^{-5}$
Adenocarcinoma de pulmón	GSE8894	$2,0 \times 10^{-3}$	$2,5 \times 10^{-3}$	$5,6 \times 10^{-3}$
Adenocarcinoma de pulmón	Shedden	$1,3 \times 10^{-7}$	$2,6 \times 10^{-7}$	$2,2 \times 10^{-7}$
Cáncer cerebral	GSE4412	$3,2 \times 10^{-5}$	$2,2 \times 10^{-5}$	$9,0 \times 10^{-5}$
Cáncer cerebral	GSE4271	$1,3 \times 10^{-3}$	$1,0 \times 10^{-3}$	$2,8 \times 10^{-4}$
Cáncer de vejiga	GSE5287	$6,4 \times 10^{-2}$	$5,0 \times 10^{-2}$	$8,6 \times 10^{-2}$

Tal como se muestra en la tabla Q a continuación, cada panel fue también pronóstico en el análisis multivariante cuando se combinó con al menos un parámetro clínico (o nomograma).

Tabla Q

Tipo de cáncer	Conjunto de datos	Valor de p			Variable/nomograma clínico adicional
		Panel C	Panel F	Panel B	
Cáncer de cerebro	GSE4271	0,022	0,017	0,0065	grado
Adenocarcinoma de pulmón	Shedden	1×10^{-6}	$2,1 \times 10^{-6}$	$1,4 \times 10^{-6}$	estadio
Cáncer de mama positivo para ER	GSE7390	0,0077	0,0064	0,011	índice de pronóstico de Nottingham (NPI)
Cáncer de mama positivo para ER	GSE11121	0,0041	0,0027	0,0045	NPI

EJEMPLO 7

Para el presente experimento, los casos se definieron como varones que murieron de cáncer de próstata en un plazo de 5 años. Los controles se definieron como varones que vivieron al menos 10 años. A continuación, los casos y los controles se ordenaron por clasificación mediante la puntuación combinada (tal como se explica en los párrafos [0066]-[0068] anteriores). En la figura 18 se muestra la distribución de casos y controles por puntuación combinada.

Se seleccionaron 25 casos con las puntuaciones combinadas más bajas y 31 controles con las puntuaciones combinadas más altas para el análisis de expresión del transcriptoma usando Illumina™ Hi-Seq 2000™. El aislamiento del ARN y la construcción de bibliotecas se realizaron según el protocolo del fabricante.

Análisis estadístico

La expresión de ARN se mide para todos los productos de transcripción (TP). Los recuentos brutos se normalizaron mediante el percentil 75º de todos los TP de cada muestra y serie, y luego se convirtieron al logaritmo de base 2. Los múltiples TP para el mismo locus génico se combinan en un gen unificado (UG).

Los datos de expresión de ARN para cada muestra se recopilaron a partir de los TP del conjunto de locus con TP único y de los UG del conjunto de locus con múltiples TP. Para poder incluir los transcritos con recuentos nulos, se usó el logaritmo de base 2 de los recuentos normalizados + 1 para el análisis.

El análisis primario se diseñó para hallar asociaciones entre la expresión del ARN y el estado de casos y controles. Se realizaron pruebas de Kolmogorov-Smirnov en cada locus, y los resultados se clasificaron según el valor de p. Además, cada uno de los genes candidatos se sometió a prueba en un modelo de regresión logística que incluía la puntuación de CCP, el PSA sérico y Gleason.

Resultados

Basándose en el análisis del transcriptoma completo, se identificó que seis genes candidatos (tabla R) estaban asociados con la muerte por cáncer de próstata después del ajuste para la puntuación de CCP y los parámetros clínicos. Todos estos genes tenían un valor de p inferior a 0,001 en el modelo multivariante. En la figura 19 se muestra la distribución de los valores de p observados en comparación con los esperados (al considerar que no existe asociación). Algunos valores de p estaban asociados de manera más significativa con la muerte por cáncer de próstata de lo esperado por azar.

Tabla R: Panel H

N.º de gen	Símbolo de gen	ID de gen	Valor de p
1	*KLK3*	354	2×10^{-6}
2	*STX4*	6810	$1{,}7 \times 10^{-5}$
3	*TAF5L*	27097	$3{,}3 \times 10^{-5}$
4	*GTPB5*	26164	$5{,}2 \times 10^{-5}$
5	*SIRT3*	23410	$9{,}7 \times 10^{-5}$
6	*EIF3D*	8664	$1{,}97 \times 10^{-4}$

Con parámetros ligeramente diferentes, se identificaron los siguientes genes como los mejores predictores (tabla S):

Tabla S: Panel H

N.º de gen	Símbolo de gen	ID de gen	Valor de p
1	*KLK3*	354	$1{,}7 \times 10^{-5}$
2	*GTPBP5*	26164	$1{,}4 \times 10^{-4}$
3	*LOC100126784*	100126784	$2{,}2 \times 10^{-4}$
4	*ABCG1*	9619	$3{,}93 \times 10^{-4}$

5	*CYP1B1-AS1*	1545	$3,96 \times 10^{-4}$
6	*CECR6*	27439	$4,8 \times 10^{-4}$

En las figuras 20 y 22 se muestran los perfiles de expresión de ARN subyacentes a los valores de p significativos. El gen mejor clasificado fue *KLK3*. Los niveles bajos de ARN de *KLK3* estaban asociados con un mal pronóstico. Los niveles de expresión de ARN de *KLK3* no estaban correlacionados con el PSA sérico (coeficiente de correlación de Pearson con el logaritmo del PSA sérico = 0,13, valor de p=0,33). La expresión de ARN de *KLK3* predice el estado de casos y controles independientemente de Gleason (figura 21).

EJEMPLO 8

El panel F se combinó con determinadas características clínicas y/o estratificadores de riesgo clínico y se demostró que la(s) combinación/combinaciones predecía(n) el riesgo de muerte específica de cáncer de próstata en muestras de biopsia.

Pacientes y muestras

Las muestras se analizaron siguiendo el procedimiento descrito en los ejemplos anteriores (en algunos casos se usaron datos del análisis de muestras real descrito en los ejemplos anteriores). Se combinaron datos de muestras de pacientes con cáncer de próstata de seis cohortes diferentes, designadas como P1, P2, P3, P4, P5A, P5B y P7 (P5A y P5B eran subconjuntos distinguibles (prostatectomía radical frente a radiación) de una cohorte P5 más grande). Las características de estas cohortes se proporcionan en la tabla T.

Tabla T

N.º de cohorte	Pacientes clínicos	Pacientes totales	Tipo de muestra	Tratamiento	Medida del desenlace
P1	200	337	TURP	conservador (vigilancia activa)	muerte por cáncer de próstata
P2	180	349	biopsia por punción	conservador (vigilancia activa)	muerte por cáncer de próstata
P3	353	353	tumor quirúrgico	prostatectomía radical	recidiva bioquímica
P4	388	413	tumor quirúrgico	prostatectomía radical	recidiva bioquímica
P5A	131	179	biopsia por punción	prostatectomía radical	recidiva bioquímica
P5B	118	142	biopsia por punción	radiación	recidiva bioquímica
P7	272	281	biopsia por punción	desconocido	desconocida

Sólo se incluyó a los pacientes que disponían de toda la información clínica necesaria para calcular la puntuación de CAPRA. También se excluyeron de cualquier análisis los pacientes a los que podía asignarse una categoría de riesgo según la AUA pero que no disponían de una puntuación de CAPRA. En todas las cohortes, los datos de tiempo transcurrido hasta el acontecimiento se censuraron a los 10 años. El inicio fue la fecha de diagnóstico para los pacientes con muestras de TURP y biopsia por punción y la fecha de cirugía para los pacientes con muestras de tumor quirúrgico. Los tiempos se registraron en días para todos los estudios excepto P1, P5A y P5B, que fueron en meses, y se convirtieron a días mediante un factor de 365,25/12. Para los pacientes cuyo estadio clínico no incluía el subestadio, se realizaron las siguientes conversiones: T1 a T1A, T2 a T2A y T3 a T3A.

Se definieron tres subconjuntos diferentes para entrenar la puntuación combinada (entrenamiento), validar la puntuación combinada y estimar el riesgo de muerte por cáncer de próstata (validación) y caracterizar la distribución de la puntuación de CCP en la población clínica estadounidense (clínica estadounidense). Estos se presentan en la tabla U.

Tabla U

Entrenamiento	Validación	Clínica

		estadounidense
P1 P3 P4 P5B	P2	 P3 P4 P5A P5B P7
N=1059	N=180	N=1219

Estratificación del riesgo según la AUA

El nomograma/directrices según la AUA estratifica el riesgo de fallo del PSA y la mortalidad específica de cáncer de próstata tras prostatectomía radical, radioterapia de haz externo o braquiterapia intersticial de próstata. Véase American Urological Association, Guideline for the Management of Clinically Localized Prostate Cancer: 2007 Update (disponible en el sitio web de la AUA). La categoría de riesgo de cada paciente se determinó según las directrices interpretadas a continuación (Directrices de la AUA 2007, página 10). Los individuos con estadio clínico T1A o T1B, o estadio clínico T3, se asignaron a categorías de riesgo a pesar de que las directrices según la AUA no son técnicamente aplicables en estos estadios.

Bajo: PSA ≤ 10 ng/ml Y puntuación de Gleason ≤ 6 Y estadio clínico ≤ T2A

Intermedio: (PSA > 10 y ≤ 20 ng/ml O puntuación de Gleason = 7 O estadio clínico = T2B) Y no reúne los requisitos para alto riesgo

Alto: PSA > 20 ng/ml O puntuación de Gleason > 7 O estadio clínico ≥ T2C

Nomograma de CAPRA

CAPRA es un predictor preoperatorio de recidiva de enfermedad después de la prostatectomía radical. Véase Cooperberg *et al.*, J. UROL. (2005) 173: 1938-1942. Las puntuaciones (de 0 a 10) se calcularon según un sistema de puntos, que se interpreta a continuación (véase Cooperberg *et al.* en la tabla 1). A los pacientes con estadio clínico T3B o superior se les asignó una puntuación, aunque técnicamente no se aplicó el sistema de puntuación. Los pacientes con PSA ≤ 2 ng/ml se incluyeron en el intervalo más bajo, que es (2, 6] en Cooperberg *et al.* Se usaron el estadio clínico y los grados de Gleason en lugar de datos patológicos en las cohortes posteriores a prostatectomía, aunque la puntuación de CCP procedía de una muestra del tumor quirúrgico en lugar de una biopsia.

PSA (ng/ml)
- 0 [0, 6]
- 1 (6, 10]
- 2 (10, 20]
- 3 (20, 30]
- 4 (30, 100]
- Pacientes exluidos con PSA > 100

segun los criterios de exclusión del estudio.

Estadio clínico
- 0 T1/T2
- 1 ≥ T3A

Porcentaje de núcleos positivos
- 0 < 34%
- 1 ≥ 34%

Puntuación de Gleason
- 0 Primario 1-3; Secundario 1-3
- 1 Primario 1-3; Secundario 4-5
- 3 Primario 4-5; Secundario 1-5

Puntuación de Gleason para pacientes con grados de componente
- 0 Gleason < 7
- 1 Gleason 7
- 3 Gleason > 7

Edad de diagnóstico
- 0 < 50 años
- 1 ≥ 50 años

Desarrollo de la puntuación combinada

La puntuación combinada se ajustó en el conjunto de entrenamiento mediante un modelo de riesgos proporcionales de Cox estratificado por cohorte. La estratificación por cohortes se ajustó para tener en cuenta las diferencias en los perfiles de supervivencia que podrían producirse por diversos regímenes de tratamiento y criterios de valoración en cada cohorte. CAPRA se trató como una variable de valor entero (0-10) y la puntuación de CCP como una variable numérica continua. Para garantizar que CAPRA era un predictor aproximadamente lineal, se sometió a prueba el término cuadrático. Fue significativo (X^2 = 8; valor de p=0,0041), pero menor en comparación con el término lineal (X^2 = 53, valor de p < 10^{-12}).

Las interacciones con la cohorte se sometieron a prueba en un modelo preliminar para confirmar que el pronóstico de la puntuación de CCP y CAPRA no dependía de la cohorte. La interacción tuvo un valor de p de 0,059 con CAPRA; y un valor de p de 0,050 con la puntuación de CCP, y no se incluyó en el modelo final (tabla V).

Tabla V

Variable	Coeficiente	HR (IC del 95%)	X^2 (1*df*)	valor de p
CAPRA	0,394	1,48 (1,38, 1,59)	116	< 10^{-26}
Puntuación de CCP	0,567	1,76 (1,51, 2,05)	48	< 10^{-11}

Basándose en este modelo, la puntuación combinada se definió como

*Puntuación combinada = 0,39*CAPRA + 0,57*Puntuación de CCP*

Validación de la puntuación combinada

La puntuación combinada se validó en P2, una cohorte de biopsia por punción de pacientes tratados de manera conservadora (vigilancia activa/conducto expectante) con muerte por cáncer de próstata como desenlace (tabla W). Hubo 33 (18%) muertes entre los 180 pacientes. En un modelo multivariante en el que se añadió CAPRA a la puntuación combinada, el valor de p para la puntuación combinada fue de 0,0028 y el valor de p para CAPRA fue de 0,58, lo que confirma que la puntuación combinada tuvo en cuenta adecuadamente tanto CAPRA como la puntuación de CCP en la cohorte de validación.

Tabla W

Variable	Coeficiente	HR (IC del 95%)	X^2 (1*df*)	valor de p
Puntuación combinada	0,82	2,27 (1,63, 3,16)	28	$< 10^{-6}$
CAPRA	0,35	1,42 (1,20, 1,68)	19	$< 10^{-4}$
Puntuación de CCP	0,75	2,12 (1,49, 3,03)	18	$< 10^{-4}$

Predicción del riesgo de mortalidad

El riesgo predicho de muerte por cáncer de próstata en los 10 años siguientes al diagnóstico se estimó en la cohorte de validación P2. Los tiempos se censuraron a los 10 años (120 meses) y el riesgo predicho se estimó en el momento del último acontecimiento (118,1109 meses). El intervalo de las puntuaciones de CCP en el conjunto de validación (n=180) fue de -0,8 a 4,1. Los riesgos de mortalidad de la muestra se muestran en los párrafos [00186] y [00197] anteriores.

Reestratificación del riesgo según la AUA basándose en la puntuación de CCP o puntuación combinada

Se determinó el percentil correspondiente a cada incremento de 0,1 de la puntuación de CCP para las muestras clínicas estadounidenses de cada cohorte dentro de cada categoría de riesgo según la AUA. Se excluyeron los individuos con estadios clínicos de T3, quedando 1219 (97%) de los 1262 disponibles. El percentil para cada paciente fue la fracción de pacientes dentro de esa categoría de riesgo según la AUA que tenían una puntuación de CCP más baja. Cuando varios pacientes compartían la misma puntuación de CCP, redondeada a la décima parte, el percentil sería el mismo; es decir, la fracción de pacientes con puntuaciones de CCP inferiores a la puntuación redondeada.

La mediana de la puntuación de CCP de las muestras clínicas estadounidenses de cada cohorte dentro de cada categoría de riesgo según la AUA se usó para evaluar la agresividad del cáncer, tal como se describe en la sección siguiente. Se comparó la puntuación de CCP de cada individuo con la puntuación de CCP promedio de los pacientes de la misma categoría de riesgo según la AUA para ofrecer una evaluación relativa de la agresividad del cáncer y para modificar o confirmar la predicción del riesgo dada usando las directrices según la AUA.

La escala de las puntuaciones de CCP para cada categoría de riesgo según la AUA consistió en cinco intervalos de 1 unidad, centrándose el intervalo central en la mediana de la puntuación de CCP para esa categoría en la presente cohorte de muestras. Hubo aproximadamente un cambio de 2 veces en el riesgo entre intervalos, que era la razón de riesgos instantáneos correspondiente a un cambio de 1 unidad en la puntuación de CCP. A cada sección se le proporcionó una etiqueta cualitativa (Tabla X).

Tabla X

		Rango de CCP y clasificación de CCP en relación con la categoría según la AUA				
		Considerablemente menos agresivo	Menos agresivo	Consistente	Más agresivo	Considerablemente más agresivo
Categoría de riesgo según la AUA	Bajo	[-2,7, -1,7]	(-1,7, -0,7]	(-0,7, 0,3)	[0,3, 1,3)	[1,3, 2,3]
	Intermedio	[-2,6, -1,6]	(-1,6, -0,6]	(-0,6, 0,4)	[0,4, 1,4)	[1,4, 2,4]
	Alto	[-2,5, -1,5]	(-1,5, -0,5]	(-0,5, 0,5)	[0,5, 1,5)	[1,5, 2,5]

EJEMPLO 9

Se evaluó la utilidad pronóstica de los genes de CCP (en este caso, el panel F) y de varios genes candidatos, incluyendo *KLK3*, todos ellos mediante técnicas de ensayo tales como las comentadas anteriormente.

Pacientes

Los pacientes cuyas muestras se usaron en este experimento fueron diagnosticados incidentalmente de cáncer de próstata tras someterse a una TURP y fueron tratados de forma conservadora. La cohorte se ha descrito previamente (en los ejemplos anteriores y en Cuzick *et al.*, Long-term outcome among men with conservatively treated localised prostate cancer, BR. J. CANCER (2006) 95:1186-1194). Una parte de esta cohorte (pero sin incluir

a ninguno de los varones de este experimento) se usó previamente para evaluar la utilidad clínica de la puntuación de CCP, tal como se ha comentado en los ejemplos anteriores. Los pacientes únicos de este experimento, y que no formaban parte de ninguna evaluación previa de la puntuación de CCP o *KLK3*) se denominan en el presente documento TURP1B. Los pacientes fueron excluidos del presente análisis si en sus historiales clínicos faltaba información sobre los niveles de PSA, la puntuación de Gleason o la extensión de la enfermedad. Se obtuvieron datos clínicos y moleculares de 303 individuos con 66 muertes específicas de cáncer de próstata para los análisis con la puntuación de CCP y de 291 individuos con 61 muertes específicas de cáncer de próstata para los presentes análisis con *KLK3*. Los datos de los pacientes se censuraron a los 10 años.

Estadística

La asociación entre los niveles de expresión y la mortalidad específica de cáncer de próstata se sometió a prueba mediante modelos de riesgos proporcionales de Cox univariantes y multivariantes. El análisis multivariante se realizó usando los parámetros clínicos relevantes indicados a continuación. Las razones de riesgos instantáneos se indican por unidad de aumento en la puntuación de expresión génica (equivalente a una duplicación de la expresión génica).

Resultados

La puntuación de CCP predijo significativamente la mortalidad específica de cáncer de próstata en las muestras de TURP1B. A continuación se presentan las estadísticas univariantes y multivariantes resumidas:

Univariante: Valor de p de CCP < 10^{-15}; HR = 3,3 (2,5, 4,3)

Multivariante:

Valor de p de CCP < 10^{-5}; HR = 2,1 (1,5, 2,8)

Valor de p de Gleason = 0,00035

Log del valor de p de PSA = 0,013

También se validó la utilidad de pronóstico de añadir *KLK3* a la puntuación de CCP. En estos análisis se usó el negativo de la expresión de *KLK3*, de modo que los valores más altos corresponderían a un mayor riesgo, al igual que en el caso de CCP. Esto se debe a que, para *KLK3*, una menor expresión predice un mayor riesgo de recidiva o mortalidad específica de cáncer de próstata. A continuación se presentan las estadísticas univariantes y multivariantes resumidas:

Univariante: Valor de p de *KLK3* < 10^{-6}; HR = 1,8 (IC del 95% 1,5, 2,2)

Multivariante:

Valor de p de CCP < 10^{-5}; HR = 2,2 (IC del 95% 1,6, 3,0)

Valor de p de *KLK3* de 0,00019; HR = 1,6 (IC del 93% 1,3, 2,1)

Valor de p de Gleason = 0,0055

Log de valor de p de PSA = 0,055

Además de *KLK3*, se evaluó la capacidad de otros genes candidatos para añadir información pronóstica independiente a la puntuación de CCP. Los resultados se resumen en la figura 25, que muestra los valores de p univariantes de la asociación con la mortalidad específica de cáncer de próstata (eje x) y los valores de p después de ajustar para la puntuación de CCP (eje y). Los genes adicionales se enumeran en la tabla Y a continuación (un subconjunto de los cuales forma el panel I de la divulgación), clasificados según el valor de p después de ajustar para la puntuación de CCP.

Tabla Y

		Independiente				Ajustado para puntuación de CCP				
N.º de gen	Símbolo de gen	Valor de p	hr	hr.lcl	hr.ucl	Valor de p	hr	hr.lcl	hr.ucl	ID de ensayo de ABI de ejemplo
1	*SLC45A3*^	3.01E-06	1.72	1.39	2.13	6.47E-07	1.82	1.45	2.28	Hs00263832_m1
2	*ACPP*^	1.14E-09	1.71	1.46	1.99	3.09E-06	1.59	1.32	1.91	Hs00173475_m1
3	*TRPM8*^	5.54E-08	1.47	1.29	1.67	4.02E-06	1.39	1.22	1.59	Hs00375481_m1
4	*MSMB*^	1.65E-07	1.41	1.25	1.61	1.28E-05	1.33	1.17	1.51	Hs00159303_m1
5	*KLK3*^	8.48E-07	1.77	1.44	2.17	2.21E-05	1.67	1.34	2.08	Hs03063374_m1
6	*GTPBP5*	9.79E-01	1.00	0.75	1.34	5.42E-05	2.04	1.44	2.88	Hs00534998_m1
7	*IRF1*	1.04E-02	1.50	1.10	2.04	1.43E-04	1.64	1.28	2.10	Hs00971965_m1
8	*FOXA1*	5.03E-01	1.13	0.80	1.58	2.50E-04	2.12	1.44	3.11	Hs00270129_m1
9	*SLC30A4*	2.24E-03	1.56	1.19	2.04	2.91E-04	1.78	1.31	2.41	Hs00203308_m1
10	*AZGP1*	1.18E-07	1.45	1.28	1.63	2.97E-04	1.32	1.15	1.52	Hs00426651_m1
11	*MARC1*	4.20E-01	1.15	0.82	1.60	3.10E-04	1.82	1.32	2.50	Hs00224227_m1
12	*PTPRC*	2.25E-02	1.35	1.04	1.74	3.51E-04	1.57	1.23	2.00	Hs00894732_m1
13	*PCA3*	2.53E-03	1.12	1.05	1.20	3.68E-04	1.14	1.07	1.22	Hs01371939_g1
14	*PMEPA1*	5.43E-04	1.84	1.30	2.61	4.38E-04	1.77	1.29	2.43	Hs00375306_m1
15	*TMPRSS2*	3.41E-04	1.39	1.20	1.61	6.13E-04	1.37	1.18	1.59	Hs01120965_m1
16	*CDH1*	5.57E-01	1.10	0.80	1.51	1.13E-03	1.67	1.25	2.24	Hs01023894_m1
17	*NKX3*	4.24E-02	1.46	1.03	2.08	1.75E-03	1.89	1.29	2.77	Hs00171834_m1
18	*KLK2*	1.57E-04	1.68	1.31	2.15	3.07E-03	1.58	1.19	2.10	Hs00428383_m1
19	*SORD*	8.49E-05	1.79	1.36	2.37	4.73E-03	1.51	1.14	2.01	Hs00973148_m1
20	*IRF4*	2.07E-01	1.14	0.93	1.41	8.35E-03	1.33	1.07	1.65	Hs00180031_m1
21	*TARP; TRGC2*	1.18E-03	1.50	1.19	1.88	1.66E-02	1.32	1.06	1.65	Hs00827007_m1
22	*STX4*	2.61E-01	1.40	0.78	2.54	1.93E-02	2.09	1.14	3.84	Hs00190266_m1
23	*KLK4*	8.50E-06	1.70	1.36	2.13	5.23E-02	1.30	1.00	1.69	Hs00191772_m1
24	*KLK3* (ensayo alt.)	3.56E-03	1.43	1.13	1.81	7.27E-02	1.28	0.98	1.66	Hs02576345_m1
25	*HLA-DRA*	2.13E-02	1.42	1.06	1.89	1.11E-01	1.24	0.96	1.61	Hs00219575_m1
26	*SIRT3*	2.32E-01	1.21	0.89	1.65	1.55E-01	1.27	0.92	1.77	Hs00202030_m1
27	*EIF3D*	1.44E-02	1.55	1.11	2.17	1.94E-01	1.24	0.90	1.71	Hs00388727_m1
28	*IGJ*	4.28E-01	1.05	0.92	1.20	2.69E-01	1.08	0.94	1.24	Hs00950678_g1
29	*HLA-DPA1*	1.47E-01	1.40	0.89	2.22	3.18E-01	1.24	0.81	1.91	Hs01072899_m1
30	*IGLL5; CKAP2*	1.58E-01	0.85	0.67	1.06	3.49E-01	1.15	0.86	1.53	Hs00382306_m1
31	*HOXB13*	4.16E-01	0.86	0.60	1.24	5.42E-01	0.88	0.58	1.33	Hs00197189_m1
32	*STEAP2*	5.33E-02	1.66	0.99	2.78	7.87E-01	1.08	0.60	1.96	Hs00537786_m1
33	*AR*	5.77E-01	0.91	0.64	1.29	8.47E-01	1.04	0.73	1.47	Hs00171172_m1
34	*CREB3L4*	4.80E-01	1.10	0.85	1.43	9.24E-01	1.01	0.77	1.34	Hs00370116_m1
35	*MKI67*	2.97E-12	0.40	0.32	0.52	9.83E-01	0.99	0.60	1.65	Hs01032443_m1
36	*HLA-E*	8.24E-02	1.34	0.96	1.87	9.85E-01	1.00	0.74	1.36	Hs03045171_m1

^ Estos genes forman el panel I de la divulgación

* “hr” = razón de riesgos instantáneos; “hr.lcl” = límite inferior de confianza del 95% de la razón de riesgos instantáneos; “hr.ucl” = límite superior de confianza del 95% de la razón de riesgos instantáneos

La tabla 1 a continuación proporciona una lista larga, pero no exhaustiva, de CCG.

Tabla 1

	Gen (nombre y/o símbolo) o número (EST, clon de ADNc o registro)

1	STK15: serina/treonina cinasa 15 Hs.48915 R11407
2	PLK: cinasa similar a polo (*Drosophia*) Hs.77597 AA629262
3	UBCH10: proteína portadora de ubiquitina E2-C Hs.93002 AA430504
4	MAPK13: proteína cinasa activada por mitógenos 13 Hs.178695 AA157499 p38delta ARNm=proteína cinasa activada por estrés 4
5	CDC2: ciclo de división celular 2, de G1 a S y de G2 a M Hs.184572 AA598974
6	TOP2A: topoisomerasa (ADN) II alfa (170 kD) Hs.156346 AA504348
7	CENPE: proteína centromérica E (312 kD) Hs.75573 AA402431 CENP-E=supuesto motor de cinetocoro que se acumula justo antes
8	TOP2A: topoisomerasa (ADN) II alfa (170 kD) Hs.156346 AA026682
9	KPNA2: carioferina alfa 2 (cohorte 1 de RAG, importina alfa 1) Hs.159557 AA676460
10	FLJ10468: proteína hipotética FLJ10468 Hs.48855 N63744
11	CCNF: ciclina F Hs.1973 AA676797
12	DKFZp762E1312: proteína hipotética DKFZp762E1312 Hs.104859 T66935
13	CKS2: proteína asociada a CDC2 CKS2 Hs.83758 AA292964
14	C20ORF1: marco de lectura abierto 1 del cromosoma 20 Hs.9329 H73329
15	BUB1: gemación no inhibida por bencimidazoles 1 (homólogo de levaduras) Hs.98658 AA430092 BUB1=supuesta proteína ser/thr cinasa de punto de control mitótico
16	TOP2A: **topoisomerasa (ADN) II alfa (170 kD) Hs.156346 AI734240
17	CKS2: proteína asociada a CDC2 CKS1 Hs.83758 0065 ckshs2=homólogo de Cks1=proteína asociada a Cdc2/p34Cdc28
18	ARL6IP: proteína de interacción 6 similar a factor de ribosilación de ADP Hs.75249 H20558
19	L2DTL: proteína L2DTL Hs.126774 R06900
20	STK15: **serina/treonina cinasa 15 Hs.48915 H63492 cinasa relacionada con IPL1/aurora
21	E2-EPF: proteína portadora de ubiquitina Hs.174070 AA464019
22	UBCH10: proteína portadora de ubiquitina E2-C Hs.93002 R80790
23	KNSL5: similar a cinesina 5 (proteína similar a cinesina mitótica 1) Hs.270845 AA452513 proteína similar a cinesina mitótica 1
24	CENPF: proteína centromérica F (350/40010, mitosina) Hs.77204 AA701455
25	CCNA2: ciclina A2 Hs.85137 AA608568 ciclina A
26	CDC2: ciclo de división celular 2, de G1 a S y de G2 a M Hs.184572 AA278152 CDC2=homólogo de proteína de control de división celular 2=proteína cinasa P34
27	HMMR: **receptor de motilidad mediado por hialuronano (RHAMM) Hs.72550 AA171715
28	KIAA0008: producto génico de KIAA0008 Hs.77695 AA262211
29	HSPC145: proteína HSPC145 Hs.18349 R22949
30	FLJ20510: proteína hipotética FLJ20510 Hs.6844 N53214
31	ARNm de NUF2R de *Homo sapiens*, cds completa Hs.234545 AA421171:

32	HSPC216: proteína hipotética Hs.13525 T87341
33	P37NB: proteína de repeticiones ricas en leucina de 37 kDa (LRR) Hs.155545 AA423870
34	CDC20:
35	CCNE1: ciclina E1 Hs.9700 T54121
36	ESTs: Hs.221754 R84407
37	FLJ11252: proteína hipotética FLJ11252 Hs.23495 N30185
38	LOC51203: clon HQ0310 PRO0310p1 Hs.279905 AA620485
39	FLJ10491: proteína hipotética FLJ10491 Hs.274283 AA425404
40	KNSL1: similar a cinesina 1 Hs.8878 AA504625
41	CENPA: proteína centromérica A (17 kD) Hs.1594 AI369629
42	*Homo sapiens*, clon IMAGE:2823731, ARNm, cds parcial Hs.70704 R96941:
43	CDC6: homólogo de CDC6 (ciclo de división celular 6, *S. cerevisiae*) Hs.69563 H59203
44	ARNm de homólogo de ADN helicasa de *Homo sapiens* (PIF1), cds parcial Hs.112160 AA464521:
45	ESTs: Hs.48480 AA135809
46	TSN: translina Hs.75066 AA460927
47	KPNA2: carioferina alfa 2 (cohorte 1 de RAG, importina alfa 1) Hs.159557 AA489087
48	RRM2: polipéptido de ribonucleótido reductasa M2 Hs.75319 AA187351
49	ESTs: Hs.14119 AA204830
50	CCNB1: ciclina B1 Hs.23960 R25788
51	GTSE1: expresada en fase G-2 y S 1 Hs.122552 AI369284
52	C20ORF1: marco de lectura abierto 1 del cromosoma 20 Hs.9329 AA936183
53	TACC3: proteína ácida que contiene hélice superenrollada transformante 3 Hs.104019 AA279990 ARNm de JkR1 regulado por disminución tras activación de células T
54	E2F1: factor de transcripción de E2F 1 Hs.96055 H61303
55	BUB1B: gemación no inhibida por bencimidazoles 1 (homólogo de levaduras), beta Hs.36708 AA488324
56	ESTs: débilmente similar a precursor de cadena de CGHU7L del colágeno alfa 1(III) [*H. sapiens*] Hs.19322 AA088457
57	KIAA0074: proteína KIAA0074 Hs.1192 N54344
58	MPHOSPH1: fosfoproteína de fase M 1 Hs.240 AA282935
59	ANLN: anilina (homólogo de restos de *Drosophila*), proteína de unión a actina Hs.62180 R12261
60	BIRC5: proteína baculoviral que contiene repeticiones IAP 5 (survivina) Hs.1578 AA460685 survivina=inhibidor de apoptosis=proteasa de células efectoras EPR-1
61	PTTG1: transformante de tumor de hipófisis 1 Hs.252587 AA430032
62	KIAA0159: proteína asociada a SMC relacionada con la condensación de cromosomas 1 Hs.5719 AA668256

63	ESTs: débilmente similar a proteína OS-4 [*H. sapiens*] Hs.18714 W93120
64	HMMR: receptor de motilidad mediado por hialuronano (RHAMM) Hs.72550 R10284
65	DKFZp762E1312: proteína hipotética DKFZp762E1312 Hs.104859 AA936181
66	CKAP2: proteína asociada al citoesqueleto 2 Hs.24641 T52152
67	RAMP: proteína asociada a la matriz nuclear regulada por RA
68	SMAP: proteína coactivadora del receptor de la hormona tiroidea Hs.5464 AA481555
69	FLJ22624: proteína hipotética FLJ22624 Hs.166425 AA488791
70	CKS1: proteína asociada a CDC2 CKS1 Hs.77550 N48162
71	NEK2: cinasa relacionada con NIMA (nunca en gen a de mitosis) 2 Hs.153704 W93379
72	MK167: antígeno identificado por anticuerpo monoclonal Ki-67
73	TTK: proteína cinasa TTK Hs.169840 AI337292
74	VEGFC: factor de crecimiento endotelial vascular C Hs.79141 H07899 proteína relacionada con factor de crecimiento endotelial vascular VRP
75	CDKN3: inhibidor de cinasa dependiente de ciclina 3 (fosfatasa de doble especificidad asociada a CDK2) Hs.84113 AA284072 CIP2=Cdi1=fosfatasa KAP1=gen del ciclo celular G1/S
76	ARNm de NUF2R de *Homo sapiens*, cds completa Hs.234545 R92435:
77	ADNc FLJ10325 fis de *Homo sapiens*, clon NT2RM2000569 Hs.245342 AA235662:
78	HSPC145: proteína HSPC145 Hs.18349 AA628867
79	HSU54999: proteína LGN Hs.278338 W92010
80	FLJ20333: proteína hipotética FLJ20333 Hs.79828 R27552
81	KNSL2: similar a cinesina 2 Hs.20830 N69491
82	ESTs: Hs.133294 AI053446
83	**ESTs: Hs.41294 H95819
84	SMTN: esmotelina Hs.149098 AA449234
85	FLJ23311: proteína hipotética FLJ23311 Hs.94292 N73916
86	USF1: factor de transcripción aguas arriba 1 Hs.247842 AA719022
87	LOC51203: clon HQ0310 PRO0310p1 Hs.279905 AA779949
88	ADH4: alcohol deshidrogenasa 4 (clase II), polipéptido pi Hs.1219 AA007395
89	ESTs: Hs.186579 AA960844
90	CCNB2: ciclina B2 Hs.194698 AA774665
91	*Homo sapiens*, similar a agrupación ricas en genes, gen C8, clon MGC:2577, ARNm, cds completa Hs.30114 AA634371:
92	ESTs: Hs.99480 AA485454
93	ARNm de IRE1b de *Homo sapiens* para proteína cinasa/ribonucleasa IRE1 beta, cds completa Hs.114905 AA088442:

94	PCNA: antígeno nuclear de células en proliferación Hs.78996 AA450264 PCNA=antígeno nuclear de células en proliferación
95	AA075920:
96	GTSE1: expresado en fase G-2 y S 1 Hs.122552 AA449474
97	CKS1: proteína asociada a CDC2 CKS1 Hs.77550 AA278629
98	CDC25B: proteína 25B de ciclo de división celular Hs.153752 AA448659 cdc25B=fosfatasa inductora de fase M 2
99	ESTs: débilmente similar a producto de proteína sin nombre [*H. sapiens*] Hs.99807 AA489023 UG desconocido Hs.99807 ESTs sc_id384
100	PCNA: antígeno nuclear de células en proliferación Hs.78996 H05891
101	LTBP3: **proteína de unión a factor de crecimiento transformante latente beta 3 Hs.289019 R60197
102	ARNm de *Homo sapiens*; ADNc DKFZp434D0818 (del clon DKFZp434D0818) Hs.5855 N95578:
103	ESTs: Hs.126714 AA919126
104	CIT: citrón (serina/treonina cinasa de interacción con rho, 21) Hs.15767 H10788
105	LBR: receptor de lámina B Hs.152931 AA099136
106	E2F1: factor de transcripción de E2F 1 Hs.96055 AA424949
107	AA699928:
108	CDKN2C: inhibidor de cinasa dependiente de ciclinas 2C (p18, inhibe CDK4) Hs.4854 N72115 p18-INK6=inhibidor de cinasa dependiente de ciclinas 6
109	STK12: serina/treonina cinasa 12 Hs.180655 H81023 ARK2=cinasa relacionada con aurora 2
110	ESTs: Hs.111471 AA682533
111	ESTs: Hs.44269 AA465090
112	MCM4: proteína deficiente en el mantenimiento de minicromosomas (*S. cerevisiae*) 4 Hs.154443 AA485983
113	PMSCL1: **autoantígeno de polimiositis/esclerodermia 1 (75 kD) Hs.91728 AA458994 ciclina A
114	MKI67: antígeno identificado por anticuerpo monoclonal Ki-67 Hs.80976 AA425973 Ki67 (tipo largo)
115	ESTs: Hs.133294 AI144063
116	CDC25B: ciclo de división celular 25B Hs.153752 HI 4343 cdc25B=fosfatasa inductora de fase M 2
117	FOXM1: caja de cabeza de horquilla M1 Hs.239 AA129552 MPP2=supuesta fosfoproteína de fase M 2
118	FLJ11029: proteína hipotética FLJ11029 Hs.274448 AI124082
119	H2AFX: familia de histonas H2A, miembro X Hs.147097 H95392
120	FLJ20333: proteína hipotética FLJ20333 Hs.79828 AA147792
121	SLC17A2: familia de portadores de soluto 17 (fosfato de sodio), miembro 2 Hs.19710 H60423

122	ARNm de IRE1b de *Homo sapiens* para proteína cinasa/ribonucleasa IRE1 beta, cds completa Hs.114905 AA102368:
123	ESTs: Hs.163921 AA573689
124	MCM5: proteína deficiente en el mantenimiento de minicromosomas (*S. cerevisiae*) 5 (ciclo de división celular 46) Hs.77171 AA283961
125	CDKN1B: inhibidor de cinasa dependiente de ciclinas 1B (p27, Kip1) Hs.238990 AA630082
126	AA779865:
127	PTTG1: transformante de tumor de hipófisis 1 Hs.252587 AI362866
128	RAD21: homólogo de RAD21 (*S. pombe*) Hs.81848 AA683102
129	ADNc FLJ10325 fis de *Homo sapiens*, clon NT2RM2000569 Hs.245342 AA430511:
130	NEK2: cinasa relacionada con NIMA (nunca en gen a de mitosis) 2 Hs.153704 AA682321
131	FLJ20101: proteína de interacción con LIS1 NUDE1, homólogo de rata Hs.263925 N79612
132	FZR1: proteína Fzr1 Hs.268384 AA621026
133	ESTs: Hs.120605 AI220472
134	KIAA0855: golgina-67 Hs.I82982 AA098902
135	SRD5A1: esteroide-5-alfa-reductasa, polipéptido alfa 1 (3-oxo-5 alfa-esteroide-delta 4-deshidrogenasa alfa 1) Hs.552 H16833
136	RAD51: homólogo de RAD51 (*S. cerevisiae*) (homólogo de RecA de *E. coli*) Hs.23044 N70010
137	KNSL2: similar a cinesina 2 Hs.20830 R11542
138	KIAA0097: producto génico de KIAA0097 Hs.76989 AA598942
139	TUBB: tubulina, polipéptido beta Hs.179661 AA427899
140	HEC: altamente expresada en cáncer, rico en repeticiones de heptadas de leucina Hs.58169 W72679
141	TROAP: proteína asociada a trofinina (tastina) Hs.171955 H94949
142	ESTs: Hs.49047 N64737
143	ESTs: Hs.15091 AA678348
144	ESTs: Hs.133431 AI061169
145	KIAA0042: producto génico de KIAA0042 Hs.3104 AA477501
146	FZR1: proteína Fzr1 Hs.268384 AA862886
147	FEN1: endonucleasa específica de estructura Flap 1 Hs.4756 AA620553
148	CKS1: proteína asociada a CDC2 CKS1 Hs.77550 AA459292 ckshs1=homólogo de Cks1=proteína asociada a Cdc2/p34Cdc28
149	ESTs: Hs.193379 N57936
150	CASP8AP2: proteína asociada a CASP8 2 Hs.122843 H50582
151	BIRC2: proteína baculoviral que contiene repeticiones de IAP 2 Hs.289107 R19628 c-IAP1=MIHB=homólogo B de IAP
152	CKAP2: proteína asociada al citoesqueleto 2 Hs.24641 AA504130
153	HLA-DRA: complejo mayor de histocompatibilidad, clase II, DR alfa Hs.76807 R47979

154	HBP: proteína de unión en horquilla, histona Hs.75257 AA629558
155	FLJ10483: proteína hipotética FLJ10483 Hs.6877 H12254
156	CASP3: caspasa 3, cisteína proteasa relacionada con la apoptosis Hs.74552 R14760 CASPASA-3=isoforma alfa de CPP32=yama=cisteína proteasa
157	**ESTs: débilmente similar a proteína que es inmunorreactiva con anticuerpos policlonales anti-PTH [*H. sapiens*] Hs.301486 AA088258
158	HMG2: proteína (cromosómica distinta de histona) con grupos de alta movilidad 2 Hs.80684 AA019203
159	PRO2000: proteína PRO2000 Hs.46677 H58234
160	FLJ20333: proteína hipotética FLJ20333 Hs.79828 T48760
161	T56726:
162	TIMP1: inhibidor tisular de metaloproteinasa 1 (actividad potenciadora de eritroides, inhibidor de colagenasa) Hs.5831 H80214
163	ESTs: Hs.102004 R94281
164	FLJ10858: proteína hipotética FLJ10858 Hs.134403 AA677552
165	ADNc FLJ11883 fis de *Homo sapiens*, clon HEMBA1007178 Hs.157148 N62451:
166	RFC4: factor de replicación C (activador 1) 4 (37 kD) Hs.35120 N93924 factor de replicación C
167	PRO2000: proteína PRO2000 Hs.46677 N47113
168	ECT2: oncogén de secuencia transformante de células epiteliales 2 Hs.132808 AI031571
169	ESTs: Hs.165909 AA629538
170	PCF11: homólogo de PCF11p Hs.123654 AA053411
171	BIRC3: proteína baculoviral que contiene repeticiones de IAP 3 Hs.127799 H48533 c-IAP2=MIHC=homólogo C de IAP=prot. de complejo de señalización de TNFR2-TRAF
172	EST: débilmente similar a dJ45P21.2 [*H. sapiens*] Hs.326451 AA931528
173	KIAA0952: proteína KIAA0952 Hs.7935 AA454989
174	KIF5B: familia de cinesina, miembro 5B Hs.149436 AA608707
175	DKFZP566C134: proteína DKFZP566C134 Hs.20237 N39306
176	ANLN: anilina (homólogo de restos de *Drosophila*), proteína de unión a actina Hs.62180 R17092
177	ORC1L: complejo de reconocimiento de origen, similar a subunidad 1 (homólogo de levaduras) Hs.17908 H51719
178	ESTs: Hs.14139 T77757
179	IFIT1: proteína inducida por interferón con repeticiones tetratricopeptídicas 1 Hs.20315 AA074989
180	MGC5338: proteína hipotética MGC5338 Hs.99598 AA463627
181	COPEB: proteína de unión a elemento promotor central Hs.285313 AA013481
182	UK114: proteína inhibidora de la traducción P14.5 Hs.18426 N72715
183	ESTs: Hs.265592 H67282
184	HMG4: proteína (cromosómica distinta de histona) con grupos de alta movilidad 4 Hs.19114 AA670197

185	MDS025: proteína hipotética MDS025 Hs.154938 AI225067
186	DKFZP564A122: proteína DKFZP564A122 Hs.187991 N53236
187	TSC22: proteína estimulada por factor de crecimiento transformante beta TSC-22 Hs.114360 AA664389
188	AAAS: aladina Hs.125262 AA916726
189	PLAG1: **gen de adenoma pleiomórfico 1 Hs.14968 AA418251
190	FLJ23293: **proteína hipotética FLJ23293 similar a proteína de interacción con ARL-6 2 Hs.31236 R91583
191	H11: proteína cinasa H11; proteína similar a proteína de estrés pequeña HSP22 Hs.111676 AA010110
192	POLD3: polimerasa (dirigida por ADN), delta 3 Hs.82502 AA504204
193	SERPINB3: inhibidor de serina (o cisteína) proteinasa, clado B (ovoalbúmina), miembro 3 Hs.227948 AA292860
194	DNAJB1: homólogo de DnaJ (Hsp40), subfamilia B, miembro 1 Hs.82646 AA435948
195	ESTs: Hs.99480 AA458886
196	BUB3: homólogo de BUB3 (gemación no inhibida por bencimidazoles 3, levadura) Hs.40323 AA405690
197	TUBB2: tubulina, beta, 2 Hs.251653 AI000256
198	ARNm de proteína SNC73 de *Homo sapiens* (SNC73), cds completa Hs.293441 H28469:
199	BUB3: homólogo de BUB3 (gemación no inhibida por bencimidazoles 3, levadura) Hs.40323 H38804
200	FLJ20699: proteína hipotética FLJ20699 Hs.15125 AA459420
201	KIAA0013: producto génico de KIAA0013 Hs.172652 N63575
202	ESTs: Hs.20575 N20305
203	CDC25C: ciclo de división celular 25C Hs.656 W95000 cdc25C=fosfatasa inductora de fase M 3
204	FLJ11186: proteína hipotética FLJ11186 Hs.89278 AA394225
205	TOPK: cinasa de unión a PDZ; proteína cinasa originada por células T Hs.104741 AA448898
206	KIAA0165: polos adicionales del husillo, *S. cerevisiae*, homólogo de Hs.153479 AA948058
207	LOC51659: proteína HSPC037 Hs.108196 AA961752
208	ESTs: Hs.10338 AA436456
209	SUCLG2: succinato-CoA ligasa, formador de GDP, subunidad beta Hs.247309 AA465233
210	ZNF265: proteína de dedos de zinc 265 Hs.194718 AA452256
211	SKP2: proteína asociada a cinasa de fase S 2 (p45) Hs.23348 R22188
212	NS1-BP: proteína de unión a NS1 Hs.197298 AA486796
213	C21ORF50: marco de lectura abierto 50 del cromosoma 21 Hs.4055 AA416628
214	BIRC2: proteína baculoviral que contiene repeticiones de IAP 2 Hs.289107 AA702174

215	BIRC3: proteína baculoviral que contiene repeticiones de IAP 3 Hs.127799 AA002125 c-IAP2=MIHC=homólogo C de IAP=prot. de complejo de señalización de TNFR2-TRAF
216	INDO: indoleamina-pirrol-2,3-dioxigenasa Hs.840 AA478279
217	DEEPEST: proteína relacionada con la hélice superenrollada del husillo mitótico Hs.16244 T97349
218	ESTs: Hs.105826 AA534321
219	C20ORF1: marco de lectura abierto 1 del cromosoma 20 Hs.9329 AI654707
220	ADNc de *Homo sapiens*: FLJ21869 fis, clon HEP02442 Hs.28465 R63929:
221	RGS3: regulador de señalización de proteína G 3 Hs.82294 AI369623
222	ARNm de DC29 de *Homo sapiens*, cds completa Hs.85573 AA186460
223	MCM6: proteína deficiente en el mantenimiento de minicromosomas (mis5, *S. pombe*) 6 Hs.155462 AA663995
224	NPAT: proteína nuclear, locus de ataxia-telangiectasia Hs.89385 AA284172 NPAT=E14=gen en el locus de ATM
225	KNSL6: similar a cinesina 6 (cinesina asociada con centrómeros mitóticos) Hs.69360 AA400450
226	HN1: de expresión hematológica y neurológica 1 Hs.109706 AA459865
227	TUBA3: tubulina, alfa, específica de cerebro Hs.272897 AA865469
228	ESTs: Hs.221197 N55457
229	KIAA0175: producto génico de KIAA0175 Hs.184339 AA903137
230	CLASPIN: homólogo de claspina de *Xenopus* Hs.175613 AA857804
231	CTNNA1: **catenina (proteína asociada a cadherina), alfa 1 (102 kD) Hs.178452 AA026631
232	ESTs: Hs.221962 AA229644
233	SMC4L1: similar a SMC4 (mantenimiento estructural de cromosomas 4, levadura) 1 Hs.50758 AA452095
234	ICBP90: factor de transcripción Hs.108106 AA026356
235	EXO1: exonucleasa 1 Hs.47504 AA703000
236	ARNm de factor asociado a TRAF4 de *Homo sapiens* 1, cds parcial Hs.181466 T84975:
237	ESTs: Hs.186814 AA700879
238	FLJ11269: proteína hipotética FLJ11269 Hs.25245 R-37817
239	SFPQ: factor de corte y empalme, rico en prolina/glutamina (asociado a proteína de unión al tracto de polipirimidina) Hs.180610 AA425258
240	ZF: factor de transcripción de unión a HCF Zhangfei Hs.29417 AA164474
241	TUBA2: tubulina, alfa 2 Hs.98102 AA626698
242	ARNm de *Homo sapiens*; ADNc DKFZp434M0435 (del clon DKFZp434M0435) Hs.25700 N94435:
243	FLJ20530: **proteína hipotética FLJ20530 Hs.279521 AA425442

244	BTEB1: proteína de unión a elemento de transcripción básico 1 Hs.150557 N80235
245	LOC51053: geminina Hs.234896 H51100
246	D21S2056E: segmento de ADN en la secuencia expresada 2056 del cromosoma 21 (único) Hs.110757 AI362799
247	HDAC3: histona desacetilasa 3 Hs.279789 H88540
248	USP1: proteasa específica de ubiquitina 1 Hs.35086 AA099033
249	C21ORF50: marco de lectura abierto 50 del cromosoma 21 Hs.4055 AA135912
250	FLJ13046: **proteína hipotética FLJ13046 similar a exportina 4 Hs.117102 T95333
251	ESTs: Hs.181059 AA912032
252	FLJ22009: proteína hipotética FLJ22009 Hs.123253 AA401234
253	ESTs: Hs.62711 AA056377
254	RAD51C: homólogo C de RAD51 (*S. cerevisiae*) Hs.11393 R37145 RAD51C=proteína relacionada con Rad51 de recombinación/reparación
255	ESTs: Hs.268919 H53508
256	ADNc FLJ11381 fis de *Homo sapiens*, clon HEMBA1000501 Hs.127797 AA885096:
257	SAP30: polipéptido asociado a sin3, 30 kD Hs.20985 AA126982
258	H4FG: familia de histonas H4, miembro G Hs.46423 AA868008
259	TUBA1: tubulina, alfa 1 (específica de testículo) Hs.75318 AA180742 tubulin-alfa-4
260	DHFR: dihidrofolato reductasa Hs.83765 R00884 DHFR=dihidrofolato reductasa
261	DHFR: dihidrofolato reductasa Hs.83765 N52980
262	MGC5528: proteína hipotética MGC5528 Hs.315167 AA934904
263	NNMT: nicotinamida-N-metiltransferasa Hs.76669 T72089
264	TUBB: tubulina, polipéptido beta Hs.179661 AI672565
265	HSPA1L: similar a proteína de 70 kD de choque térmico 1 Hs.80288 H17513 HSP70-HOM=proteína de 70 kD de choque térmico 1
266	TUBA1: **tubulina, alfa 1 (específica de testículo) Hs.75318 R36063
267	PRO1073: **proteína PRO1073 Hs.6975 AA176999 CIP4=proteína de interacción con Cdc42 4
268	POLD3: polimerasa (dirigida por ADN), delta 3 Hs.82502 AI017254
269	ESTs: moderadamente similar a proteína hipotética T50635, DKFZp762L0311.1, [*H. sapiens*] Hs.47378 N38809
270	DKFZP564A122: proteína DKFZP564A122 Hs.187991 N57723
271	LRRFIP1: **proteína de interacción con repeticiones ricas en leucina (en FLII) 1 Hs.326159 T84633
272	ESTs: Hs.55468 AA165312
273	ESTs: Hs.31444 H16772
274	AFAP: proteína asociada a filamentos de actina Hs.80306 R69355

275	CXCR4: quimiocina (motivo C-X-C), receptor 4 (fusina) Hs.89414 T62491 receptor de quimiocina CXC 4=fusina=receptor Y de neuropéptido=L3
276	MSH2: **homólogo 2 de mutS (*E. coli*) (cáncer de colon, tipo 1 distinto de poliposis) Hs.78934 AA679697
277	ESTs: Hs.48474 N62074
278	AA677337:
279	ESTs: moderadamente similar a TBB2, CADENA BETA-2 DE TUBULINA HUMANA [*H. sapiens*] Hs.23189 AA629908
280	HP1-BP74: HP1-BP74 Hs.142442 H79795
281	FLJ20101: proteína de interacción con LIS1 NUDE1, homólogo de rata Hs.263925 AA459394
282	ARNm de *Homo sapiens*; ADNc DKFZp434D1428 (del clon DKFZp434D1428); cds completa Hs.321775 AA431268:
283	ESTs: Hs.265592 AA992658
284	ESTs:
285	DDX11: polipéptido 11 de caja DEAD/H (Asp-Glu-Ala-Asp/His) (helicasa similar a CHL1 de *S. cerevisiae*) Hs.27424 AA402879
286	CDC27: ciclo de división celular 27 Hs.72405 T81764
287	ARGBP2: proteína de interacción con Arg/Ab1 ArgBP2 Hs.278626 N89738
288	DKFZP564A122: proteína DKFZP564A122 Hs.187991 AA025807
289	OPN3: opsina 3 (encefalopsina) Hs.279926 AA150060
290	DKFZP566C134: proteína DKFZP566C134 Hs.20237 AA456319
291	KIAA0855: golgina-67 Hs.182982 H15101
292	PIN: dineína, citoplásmico, polipéptido ligero Hs.5120 AA644679
293	ESTs: débilmente similar a LIP1, PRECURSOR DE PROTEÍNA 1 RELACIONADA CON LIPASA PANCREÁTICA HUMANA [*H. sapiens*] Hs.68864 AA088857
294	HDAC3: histona desacetilasa 3 Hs.279789 AA973283
295	DONSON: vecino aguas abajo de SON Hs.17834 AA417895
296	LOC51053: geminina Hs.234896 AA447662
297	FLJ10545: proteína hipotética FLJ10545 Hs.88663 AA460110
298	MAD2L1: similar a MAD2 (deficiente en detención mitótica, levadura, homólogo) 1 Hs.79078 AA481076 homólogo de proteína de control de retroalimentación mitótica Madp2
299	TASR2: proteína de serina-arginina asociada a TLS 2 Hs.3530 H11042
300	MCM6: proteína deficiente en el mantenimiento de minicromosomas (mis5, *S. pombe*) 6 Hs.155462 N57722
301	CIT: citrón (serina/treonina cinasa 21 de interacción con rho) Hs.15767 W69425
302	**ESTs: Hs.205066 AA284803

303	ICAM1: molécula de adhesión intercelular 1 (CD54), receptor de rinovirus humano Hs.168383 R77293 CD54=ICAM-1
304	KIAA0855: golgina-67 Hs.182982 AA456818
305	ESTs: débilmente similar a supuesta p150 [*H. sapiens*] Hs.300070 R10422
306	DEEPEST: proteína relacionada con la hélice superenrollada del husillo mitótico Hs.16244 AI652290
307	MCM2: proteína deficiente en el mantenimiento de minicromosomas (*S. cerevisiae*) 2 (mitotina) Hs.57101 AA454572
308	ADNc de *Homo sapiens*: FLJ22272 fis, clon HRC03192 Hs.50740 AA495943:
309	WISP1: **proteína WNT1 de la ruta de señalización inducible 1 Hs.194680 T54850
310	KIAA0855: golgina-67 Hs.182982 AA280248
311	TEM8: marcador endotelial tumoral 8 Hs.8966 H58644
312	BITE: proteína de unión a p10 Hs.42315 H96392
313	RAN: RAN, miembro de la familia de oncogenes de RAS Hs.10842 AA456636
314	EZH2: homólogo de potenciador de zeste (*Drosophila*) 2 Hs.77256 AA428252
315	MCM4: proteína deficiente en el mantenimiento de minicromosomas (*S. cerevisiae*) 4 Hs.154443 W74071
316	DKFZp434J0310: proteína hipotética Hs.278408 AA279657 UG desconocido Hs.23595 ESTs sc_id6950
317	PPP1R10: proteína fosfatasa 1, subunidad reguladora 10 Hs.106019 AA071526
318	H11: proteína cinasa H11; proteína similar a proteína de estrés pequeña HSP22 Hs.111676 H57493
319	ESTs: débilmente similar a proteína KIAA1074 [*H. sapiens*] Hs.200483 AA463220
320	ESTs: débilmente similar a ALU8, ENTRADA DE ADVERTENCIA DE CONTAMINACIÓN DE SECUENCIA SX DE SUBFAMILIA DE ALU HUMANA [*H. sapiens*] Hs.226414 N72576
321	AA775033:
322	LOC51004: proteína CGI-10 Hs.12239 AA677920
323	ESTs: Hs.150028 AI292036
324	MCM6: proteína deficiente en el mantenimiento de minicromosomas (mis5, *S. pombe*) 6 Hs.155462 AA976533
325	ESTs: moderadamente similar a proteína hipotética T50635, DKFZp762L0311.1, [*H. sapiens*] Hs.47378 AA406348
326	UCP4: proteína de desacoplamiento 4 Hs.40510 H60279
327	MSH5: homólogo 5 de mutS (*E. coli*) Hs.112193 AA621155
328	ROCK 1: proteína cinasa asociada a Rho que contiene hélice superenrollada 1 Hs.17820 AA872143
329	KIAA0855: golgina-67 Hs.182982 AA694481
330	AA705332:
331	CDC27: ciclo de división celular 27 Hs.172405 N47994
332	DONSON: vecino aguas abajo de SON Hs.17834 AI732249

333	SH3GL2: similar a GRB2 de dominio SH3 2 Hs.75149 R12817
334	PRC1: proteína reguladora de citocinesis 1 Hs.5101 AA449336
335	ESTs: débilmente similar a producto de proteína sin nombre [*H. sapiens*] Hs.99807 AA417744 UG desconocido Hs.119424 ESTs sc_id2235
336	Humano: secuencia de ARNm del clon 23719 Hs.80305 AA425722
337	ARNm de *Homo sapiens*; ADNc DKFZp564O2364 (del clon DKFZp564O2364) Hs.28893 W90240:
338	ESTs: débilmente similar a LIP1, PRECURSOR DE PROTEÍNA 1 RELACIONADA CON LIPASA PANCREÁTICA HUMANA [*H. sapiens*] Hs.68864 AA132858
339	TUBA3: tubulina, alfa, específica de cerebro Hs.272897 AA864642
340	A1283530:
341	ESTs: Hs.302878 R92512
342	PPP1R10: proteína fosfatasa 1, subunidad reguladora 10 Hs.106019 T75485
343	SFRS5: factor de corte y empalme, rico en arginina/serina 5 Hs.166975 R73672
344	SFRS3: factor de corte y empalme, rico en arginina/serina 3 Hs.167460 AA598400
345	PRIM1: primasa, polipéptido 1 (49 kD) Hs.82741 AA025937 ADN primasa (subunidad P48)
346	FLJ20333: proteína hipotética FLJ20333 Hs.79828 H66982
347	HSPA8: proteína de 70 kD de choque térmico 8 Hs.180414 AA620511
348	C4A: componente del complemento 4A Hs.170250 AA664406
349	DKC1: disqueratosis congénita 1, disquerina Hs.4747 AA052960
350	HP1-BP74: HP1-BP74 Hs.142442 T84669
351	ETV4: gen 4 de variante de ets (proteína de unión a potenciador E1A, E1AF) Hs.77711 AA010400 E1A-F=proteína de unión a potenciador E1A=variante de translocación de ETS
352	ADNc de *Homo sapiens*: FLJ23037 fis, clon LNG02036, altamente similar a HSU68019, ARNm de homólogo de proteína mad de *Homo sapiens* (hMAD-3) Hs.288261 W42414 Smad3=hMAD-3=homólogo de madres contra decapentapléjico (M:
353	KIAA0952: proteína KIAA0952 Hs.7935 AA679150
354	STK9: serina/treonina cinasa 9 Hs.50905 N80713
355	NXF1: **factor de exportación de ARN nuclear 1 Hs.323502 R01238
356	FLJ12892: proteína hipotética FLJ12892 Hs.17731 AA449357
357	UNG: uracil-ADN glicosilasa Hs.78853 H15111
358	STK17B: **serina/treonina cinasa 17b (de inducción de apoptosis) Hs.120996 AA419485
359	YWHAH: proteína de activación de tirosina 3-monooxigenasa/triptófano-5-monooxigenasa, polipéptido eta Hs.75544 N69107
360	FLJ13154: proteína hipotética FLJ13154 Hs.25303 AA923560

361	LOC51116: proteína CGI-91 Hs.20776 AA459419
362	SSXT: sarcoma sinovial, translocada en el cromosoma X Hs.153221 N59206
363	KIAA0978: proteína KIAA0978 Hs.3686 AA485878
364	EST: Hs.147907 AI223432
365	FLJ23468: proteína hipotética FLJ23468 Hs.38178 AA431741
366	FLJ10339: **proteína hipotética FLJ10339 Hs.203963 N95450
367	BMP2: proteína morfogenética ósea 2 Hs.73853 AA011061
368	PIR51: proteína de interacción con RAD51 Hs.24596 AI214426
369	FLJ20364: proteína hipotética FLJ20364 Hs.32471 AA676296
370	EIF4A2: **factor de inicio de traducción eucariota 4A, isoforma 2 Hs.173912 H54751
371	ESTs: débilmente similar a MCAT, PROTEÍNA PORTADORA DE CARNITINA/ACILCARNITINA MITOCONDRIAL HUMANA [*H. sapiens*] Hs.27769 AA469975
372	FLJ11323: proteína hipotética FLJ11323 Hs.25625 AA775600
373	DKFZP564D0764: proteína DKFZP564D0764 Hs.26799 AA460732
374	CTL2: gen de CTL2 Hs.105509 AA454710
375	ESTs: Hs.293419 AA775845
376	IFIT1: proteína inducida por interferón con repeticiones tetratricopeptídicas 1 Hs.20315 AA489640 proteína de 56 KDa inducida por interferón
377	RBBP8: proteína de unión a retinoblastoma 8 Hs.29287 H23021
378	**Secuencia de ARNm del clon 25061 de *Homo sapiens* Hs.183475 R38944
379	Humano: secuencia de ADN del clon RP3-383J4 en el cromosoma 1q24.1-24.3 que contiene parte de un gen que codifica para una proteína que contiene motivos kelch, parte de un nuevo gen que codifica para una proteína similar a aspartil-TRNA sy Hs.117305 N29457
380	FLJ12888: proteína hipotética FLJ12888 Hs.284137 N68390
381	ESTs: débilmente similar a IF38, SUBUNIDAD 8 DEL FACTOR DE INICIO DE TRADUCCIÓN EUCARIOTA HUMANO 3 [*H. sapiens*] Hs.222088 AI139629
382	ESTs: Hs.241101 AA133590
383	H4FI: familia de histonas H4, miembro 1 Hs.143080 AI218900
384	SP38: proteína de unión a la zona pelúcida Hs.99875 AA400474
385	GABPB1: factor de transcripción de proteína de unión a GA, subunidad beta 1 (53 kD) Hs.78915 1191651
386	LCHN: proteína LCHN Hs.12461 AA029330
387	DKFZP564D0462: proteína hipotética DKFZp564D0462 Hs.44197 N32904
388	LENG8: nuevo gen codificado por agrupación del receptor de leucocitos (LRC) 8 Hs.306121 AA464698

389	HIF1A: factor inducible por hipoxia 1, subunidad alfa (factor de transcripción de hélice-bucle-hélice básico) Hs.197540 AA598526
390	ESTs: Hs.93714 R09201
391	FLJ23468: proteína hipotética FLJ23468 Hs.38178 AA454949
392	DKFZP566C134: proteína DKFZP566C134 Hs.20237 AA448164
393	PPP3CA: proteína fosfatasa 3 (anteriormente 2B), subunidad catalítica, isoforma alfa (calcineurina A alfa) Hs.272458 W60310
394	HMGE: proteína cochaperona similar a GrpE Hs.151903 H55907
395	CDK7: cinasa dependiente de ciclinas 7 (homólogo de cinasa de activación de cdk MO15 de *Xenopus*) Hs.184298 R22624 CAK=cdk7=NRTALRE=sdk=cinasa de activación de CDK
396	ABCC5: **casete de unión a ATP, subfamilia C (CFTR/MRP), miembro 5 Hs.108660 AA186613
397	AA477707:
398	**ESTs: Hs.15607 R92899
399	LOC57209: proteína de dedos de zinc de tipo Kruppel Hs.25275 N50827
400	FLJ20101: proteína de interacción con LIS1 NUDE1, homólogo de rata Hs.263925 R87716
401	KNSL4: similar a cinesina 4 Hs.119324 AA430503
402	E2F5: factor 5 de transcripción de E2F, de unión a p130 Hs.2331 AA455521 E2F-5=factor de transcripción de unión a pRB
403	TMPO: timopoyetina Hs.11355 T63980
404	POLQ: polimerasa (dirigida por ADN), zeta Hs.241517 AI057325
405	TGIF: factor de interacción con TG (homeosecuencia de la familia de TALE) Hs.90077 H51705
406	TRIP13: interactor del receptor de la hormona tiroidea 13 Hs.6566 AA630784
407	GAS6: proteína específica de la detención del crecimiento 6 Hs.78501 AA461110
408	HN1: de expresión hematológica y neurológica 1 Hs.109706 AA035429
409	BARD1: dominio 1 de RING asociado a BRCA1 Hs.54089 AA558464
410	DHFR: dihidrofolato reductasa Hs.83765 AA424790
411	AA490946:
412	ESTs: Hs.130435 AA167114
413	HSPA8: proteína de 70 kD de choque térmico 8 Hs.180414 AA629567
414	RRM2: polipéptido M2 de ribonucleótido reductasa Hs.75319 AA826373
415	FLJ20036: proteína hipotética FLJ20036 Hs.32922 H59114
416	COPEB: proteína de unión a elemento promotor central Hs.285313 AA055584 CPBP=CBA1=proteína de unión a ADN
417	FLJ10604: proteína hipotética FLJ10604 Hs.26516 N72697
418	ESTs: débilmente similar a ADNc EST yk415c12.5 procede de este gen [*C. elegans*] Hs.108824 H97880

419	UBE2D3: **enzima de conjugación con ubiquitina E2D 3 (homóloga a UBC4/5 de levadura) Hs.118797 AA017199
420	FLJ10890: **proteína hipotética FLJ10890 Hs.17283 AA004210
421	ESTs: Hs.214410 AA579336
422	OLR1: receptor de lipoproteína de baja densidad oxidada (similar a lectina) 1 Hs.77729 AA682386
423	FLJ13231: proteína hipotética FLJ13231 Hs.156148 W92787
424	EST: Hs.323101 W40398
425	ESTs: débilmente similar a R06F6.5b [*C. elegans*] Hs.180591 N59330
426	ADNc de *Homo sapiens*: FLJ23285 fis, clon HEP09071 Hs.90424 N26163:
427	ADNc de inserto de longitud completa de ARNm de *Homo sapiens*, clon EUROIMAGE 42408 Hs.284123 AA211446:
428	NFKB1: factor nuclear de potenciador de genes de polipéptido ligero kappa en células B 1 (p105) Hs.83428 AA451716 NFkB1=NF-kappaB p105=p50
429	LOC58486: proteína similar a transposasa Buster1 derivada de transposones Hs.25726 AA630256
430	ADNc FLJ10976 fis de *Homo sapiens*, clon PLACE1001399 Hs.296323 AA424756:
431	KIAA0182: proteína KIAA0182 Hs.75909 AI023801
432	RANGAP1: proteína de activación de GTPasa Ran 1 Hs.183800 AA991855
433	PKMYT1: cinasa inhibidora de cdc2 específica de tirosina y treonina asociada a la membrana Hs.77783 AA478066 cinasa Myt1
434	HSPA8: proteína de 70 kD de choque térmico 8 Hs.180414 H64096
435	LUC7A: proteína sobreexpresada asociada a resistencia a cisplatino Hs.3688 AA411969
436	RRM1: polipéptido M1 de ribonucleótido reductasa Hs.2934 AA633549
437	SET07: proteína que contiene dominios PR/SET 7 Hs.111988 AA421470
438	**ESTs: débilmente similar a ALU1, ENTRADA DE ADVERTENCIA DE CONTAMINACIÓN DE SECUENCIA J DE LA SUBFAMILIA ALU HUMANA [*H. sapiens*] Hs.193452 W96179
439	Secuencia de ARNm del clon 25058 de *Homo sapiens* Hs.179397 R38894:
440	ESTs: débilmente similar a proteína KIAA0973 [*H. sapiens*] Hs.14014 AA780791
441	EST: Hs.105298 AA489813
442	CTCF: factor de unión a CCCTC (proteína de dedos de zinc) Hs.57419 H89996
443	HRB: proteína de unión a Rev de VIH-1 Hs.171545 AA485958
444	**ESTs: Hs.294083 AA447679
445	KIAA0878: proteína KIAA0878 Hs.188006 AA599094
446	ESTs: débilmente similar a ALUB !!!! ENTRADA DE ADVERTENCIA DE CLASE B DE ALU HUMANA!!! [*H. sapiens*] Hs.180552 AA481283
447	OGT: N-acetilglucosamina (GlcNAc) transferasa unida a O (UDP-N-acetilglucosamina:polipéptido-N-acetilglucosaminiltransferasa) Hs.100293 AA425229

448	ARNm de *Homo sapiens* para proteína KIAA1700, cds parcial Hs.20281 N40952:
449	Humano: secuencia de ADN del clon RP1-187J11 en el cromosoma 6q11.1-22.33. Contiene el gen para una nueva proteína similar a proteínas predichas de *S. pombe* y *S. cerevisiae*, el gen para una nueva proteína similar Hs.72325 AA159962
450	KIAA1265: proteína KIAA1265 Hs.24936 AA479302
451	HIF0: familia de histonas H1, miembro 0 Hs.226117 H57830
452	ARGBP2: proteína de interacción con Arg/Abl ArgBP2 Hs.278626 H02525
453	ODF2: fibra densa exterior de colas de espermatozoides 2 Hs.129055 AA149882
454	CD97: antígeno CD97 Hs.3107 AI651871
455	BMI1: **homólogo de oncogén viral de leucemia murina (bmi-1) Hs.431 AA193573
456	POLG: polimerasa (dirigida por ADN), gamma Hs.80961 AA188629
457	XPR1: receptor de retrovirus xenotrópico y politrópico Hs.227656 AA453474
458	ESTs: Hs.128096 AA971179
459	DNAJB1: homólogo de DnaJ (Hsp40), subfamilia B, miembro 1 Hs.82646 AA481022
460	ARL4: similar a factor de ribosilación de ADP 4 Hs.201672 AI142552
461	SFRS5: factor de corte y empalme, rico en arginina/serina 5 Hs.166975 AA598965
462	ESTs: Hs.25933 R11605
463	RIG-I: ARN helicasa Hs.145612 AA126958
464	FLJ10339: proteína hipotética FLJ10339 Hs.203963 AA628231
465	DR1: regulador por disminución de la transcripción 1, de unión a TBP (cofactor negativo 2) Hs.16697 AA043503
466	*Homo sapiens*, similar a proteína hipotética FLJ20093, clon MGC:1076, ARNm, cds completa Hs.298998 AA703249:
467	HSPC163: proteína HSPC163 Hs.108854 H98963
468	DKFZP564A122: proteína DKFZP564A122 Hs.187991 R27345
469	FLJ10128: autoantígeno uveal con dominios de hélice superenrollada y repeticiones de anquirina Hs.49753 T47624
470	DSCR1: gen 1 de región crítica de síndrome de Down Hs.184222 AA629707
471	FLJ10342: proteína hipotética FLJ10342 Hs.101514 AA490935
472	ARNm de *Homo sapiens*; ADNc DKFZp586N1323 (del clon DKFZp586N1323) Hs.24064 R26176:
473	ESTs: Hs.4983 H59921
474	ESTs: débilmente similar a ALUB, ENTRADA DE ADVERTENCIA DE CLASE B DE ALU HUMANA [*H. sapiens*] Hs.117949 H91167
475	CDC45L: similar a CDC45 (ciclo de división celular 45, *S. cerevisiae*, homólogo) Hs.114311 AA700904
476	STAT5B: transductor de señales y activador de la transcripción 5B Hs.244613 AA280647 STAT5A/5B

477	ADNc FLJ14028 fis de *Homo sapiens*, clon HEMBA1003838 Hs.281434 AA454682:
478	KIAA1524: proteína KIAA1524 Hs.151343 AI248987
479	CTSD: catepsina D (aspartil proteasa lisosómica) Hs.79572 AA485373
480	*Homo sapiens*, similar a proteína hipotética FLJ20093, clon MGC:1076, ARNm, cds completa Hs.298998 AA682274:
481	GTPBP2: proteína de unión a GTP 2 Hs.13011 T67069
482	LOC51003: proteína CGI-125 Hs.27289 AA485945
483	VCL: vinculina Hs.75350 AA486727
484	KIF5B: miembro 5B de la familia de cinesinas Hs.149436 AA046613
485	CDC25A: ciclo de división celular 25A Hs.1634 AA071514
486	LOC51141: proteína inducida por insulina 2 Hs.7089 AA045308
487	**ESTs: moderadamente similar a CALD, CALDESMONA HUMANA [*H. sapiens*] Hs.117774 H48508
488	TBX3-iso: proteína TBX3-iso Hs.267182 T48941
489	KIAA0176: proteína KIAA0176 Hs.4935 R44371
490	PRKAR1A: proteína cinasa, dependiente de cAMP, reguladora, tipo I, alfa (extinguidor específico de tejido 1) Hs.183037 N25969 PKA-R1 alfa=proteína cinasa dependiente de cAMP de tipo I-alfa-cata
491	ESTs: Hs.268991 H77818
492	ESTs: débilmente similar a isopentenil-difosfato delta-isomerasa A53028 [*H. sapiens*] Hs.9270 R17362
493	ESTs: débilmente similar a proteína hipotética B34087 [*H. sapiens*] Hs.120946 H50656
494	TRN2: carioferina beta 2b, transportina Hs.278378 R08897
495	LMNA: lamina A/C Hs.77886 AA489582
496	NFE2L2: similar a factor nuclear 2 (derivado de eritroides 2) Hs.155396 AA629687
497	DKFZp762L0311: proteína hipotética DKFZp762L0311 Hs.16520 AA486418
498	ESTs: débilmente similar a S71752, proteína gigante p619 [*H. sapiens*] Hs.14870 T96829
499	ARNm de *Homo sapiens*; ADNc DKFZp434A1315 (del clon DKFZp434A1315); cds completa Hs.298312 AA991355:
500	E2IG4: proteína hipotética, inducida por estradiol Hs.8361 R13844
501	RANGAP1: proteína de activación de GTPasa Ran 1 Hs.183800 AA485734
502	HIF0: familia de histonas H1, miembro 0 Hs.226117 W69399
503	KIAA0239: proteína KIAA0239 Hs.9729 AA454740
504	ESTs: débilmente similar a ALU7, ENTRADA DE ADVERTENCIA DE CONTAMINACIÓN DE SECUENCIA SQ DE LA SUBFAMILIA DE ALU HUMANA [*H. sapiens*] Hs.68647 R96804
505	PRO0650: proteína PRO0650 Hs.177258 N54333

506	DNAJB9: homólogo de DnaJ (Hsp40), subfamilia B, miembro 9 Hs.6790 AA045792
507	ADNc de *Homo sapiens*: FLJ21971 fis, clon HEP05790 Hs.71331 AA774678
508	LOC56996: **proteína de interacción con cotransportador de cationes-cloruro Hs.119178 AA037466
509	AP3D1: complejo 3 de proteína relacionada con adaptador, subunidad delta 1 Hs.75056 AA630776
510	SGK: cinasa regulada por suero/glucocorticoides Hs.159640 AA486082 sgk=supuesta serina/treonina proteína cinasa transcripcional
511	HSPC148: proteína hipotética Hs.42743 R23666
512	MRPL19: proteína ribosómica mitocondrial L19 Hs.75574 AA521243 KIAA0104
513	AA455102:
514	ESTs: Hs.150325 AI278813
515	.**ESTs: Hs.40527 AA029844
516	HSPC145: proteína HSPC145 Hs.18349 AI271431
517	KIAA0170: producto génico de KIAA0170 Hs.277585 H68789
518	FLJ11127: proteína hipotética Hs.91165 T98200
519	KIAA0182: proteína KIAA0182 Hs.75909 H05099
520	FLJ23151: proteína hipotética FLJ23151 Hs.137260 AA284259
521	AMD1: S-adenosilmetionina descarboxilasa 1 Hs.262476 AA425692
522	FLJ10342: **proteína hipotética FLJ10342Hs.101514 AA934516
523	SPS: SELENOFOSFATO SINTETASA; proteína donadora de selenio humana Hs.124027 AA486372
524	KIAA1586: proteína KIAA1586 Hs.180663 AA779733
525	ICBP90: factor de transcripción Hs.108106 AA908902
526	ADNc de *Homo sapiens*: FLJ21971 fis, clon HEP05790 Hs.71331 AI002036:
527	ABCC2: casete de unión a ATP, subfamilia C (CFTR/MRP), miembro 2 Hs.193852 R91502
528	ARHGDIB: inhibidor de disociación de Rho GDP (GDI) beta Hs.83656 AA487426 LyGDI=inhibidor de disociación de Rho GDP 2=RHO GDI 2
529	RAD53: proteína cinasa Chk2 Hs.146329 AI653182
530	R96880:
531	TNFAIP3: factor de necrosis tumoral, proteína alfa-inducida 3 Hs.211600 AA433807
532	ESTs: Hs.26979 H23469
533	AOC2: amina oxidasa, que contiene cobre 2 (específica de retina) Hs.143102 N50959
534	ARNm de *Homo sapiens*; ADNc DKFZp586N1323 (del clon DKFZp586N1323) Hs.24064 R30941:
535	AA452872:
536	ESTs: Hs.124169 R58970

537	ACYP1: acilfosfatasa 1, de tipo eritrocitos (habitual) Hs.18573 W78754
538	SIL: locus interruptor de TAL1 (SCL) Hs.323032 AA704809
539	AA016234:
540	ARNm de *Homo sapiens*; ADNc DKFZp566P1124 (del clon DKFZp566P1124) Hs.321022 N50895:
541	KIAA1067: proteína KIAA1067 Hs.325530 AA099138
542	SMC4L1: similar a SMC4 (mantenimiento estructural de cromosomas 4, levadura) 1 Hs.50758 AA283006
543	ESTs: Hs.29074 R70174
544	SNK: cinasa inducible por suero Hs.3838 AA460152
545	FANCG: anemia de Fanconi, grupo de complementación G Hs.8047 AA427484
546	ADNc de *Homo sapiens*: FLJ21531 fis, clon COL06036 Hs.102941 N95440:
547	ARNm de *Homo sapiens*; ADNc DKFZp547B086 (del clon DKFZp547B086) Hs.36606 N48700:
548	C1ORF2: marco de lectura abierto 2 del cromosoma 1 Hs.19554 H11464 cote1=ORF en el locus de glucocerebrosidasa
549	HTF9C: locus de diminutos fragmentos de Hpall 9C Hs.63609 H17888
550	ATF4: factor de transcripción activante 4 (elemento potenciador de respuesta a tax B67) Hs.181243 AA600217
551	ESTs: Hs.101014 AA194941
552	CDC25A: ciclo de división celular 25A Hs.1634 AA913262
553	TOPK: cinasa de unión a PDZ; proteína cinasa originada por células T Hs.104741 AI002631
554	ASIP: proteína de señalización de agutí (ratón) Hs.37006 AI220203
555	DKFZP564F013: **proteína hipotética DKFZp564F013 Hs.128653 R14908
556	ZNF265: proteína de dedos de zinc 265 Hs.194718 N66014
557	SLC30A1: familia de portadores de soluto 30 (transportador de zinc), miembro 1 Hs.55610 AA195463
558	ESTs: Hs.28462 R63922
559	ESTs: Hs.114055 R27431
560	IL6: interleucina 6 (interferón, beta 2) Hs.93913 N98591 IL-6
561	H3F3B: histona H3, familia 3B (H3.3B) Hs.180877 AA608514
562	ESTs: Hs.81263 W81524
563	ADNc de *Homo sapiens*: FLJ23538 fis, clon LNG08010, altamente similar a ARNm del clon épsilon/beta de la región de MEN1 humana BETA2 Hs.240443 AA400234:
564	AMD1: S-adenosilmetionina descarboxilasa 1 Hs.262476 R82299
565	MAP3K2: proteína cinasa activada por mitógenos 2 Hs.28827 AA447971
566	NET1: gen transformante de células neuroepiteliales 1 Hs.25155 R24543

567	CHAF1A: factor de ensamblaje de cromatina 1, subunidad A (p150) Hs.79018 AA704459
568	MGC5585: proteína hipotética MGC5585 Hs.5152 H50655
569	KIAA1598: proteína KIAA1598 Hs.23740 HI 7868
570	PNN: pinina, proteína asociada a desmosomas Hs.44499 W86139
571	ESTs: Hs.238797 N70848
572	ESTs: débilmente similar a ALUB, !!!! ENTRADA DE ADVERTENCIA DE CLASE B HUMANA!!! [*H. sapiens*] Hs.180552 AA600192
573	PDGFA: polipéptido alfa de factor de crecimiento derivado de plaquetas Hs.37040 AA701502
574	ARNm del clon FLC0675 PRO2870 de *Homo sapiens*, cds completa Hs.306117 AA443127:
575	ESTs: Hs.143375 AA001841
576	TUBB: tubulina, polipéptido beta Hs.179661 H37989
577	MSH2: homólogo 2 de mutS (*E. coli*) (cáncer de colon, tipo 1 distinto de poliposis) Hs.78934 AA219060 MSH2=homólogo de mutS de reparación de apareamiento erróneo de ADN
578	TOPBP1: proteína de unión a topoisomerasa (ADN) II Hs.91417 R97785
579	KIAA0869: proteína KIAA0869 Hs.21543 R43798
580	H4FH: familia de histonas H4, miembro H Hs.93758 AA702781
581	FLJ23293: proteína hipotética FLJ23293 similar a proteína de interacción con ARL-6 2 Hs.31236 AA629027
582	**ADNc de *Homo sapiens*: FLJ23538 fis, clon LNG08010, altamente similar a ARNm del clon épsilon/beta de la región de MEN1 humana BETA2 Hs.240443 AA053165:
583	KIAA0978: proteína KIAA0978 Hs.3686 N64780
584	KIAA1547: proteína KIAA1547 Hs.31305 AA057737
585	DKFZP761C169: proteína hipotética DKFZp761C169 Hs.71252 AA608709
586	WS-3: nueva proteína que contiene RGD Hs.39913 AA449975
587	FRZB: proteína relacionada con Frizzled Hs.153684 H87275
588	BRCA1: cáncer de mama 1, aparición temprana Hs.194143 H90415 BRCA1=mutado en cáncer de mama y de ovario
589	ESTs: Hs.4983 H22936
590	HSPC150: proteína HSPC150 similar a enzima de conjugación con ubiquitina Hs.5199 AA460431
591	ARNm de *Homo sapiens* para proteína KIAA1712, cds parcial Hs.29798 H54592:
592	FLJ11186: proteína hipotética FLJ11186 Hs.89278 AA504111 UG desconocido Hs.89278 ESTs
593	ESTs: débilmente similar a producto de proteína sin nombre [*H. sapiens*] Hs.118338 R25481
594	APEXL2: proteína similar a endonucleasa apurínica/apirimidínica (nucleasa APEX) 2 Hs.154149 AI674393
595	CDR2: proteína relacionada con la degeneración cerebelosa (62 kD) Hs.75124 AA074613

596	ESTs: Hs.69662 AA459724
597	PSCD2L: homólogo de pleckstrina, similar a dominios de Sec7 y de hélice superenrollada 2 Hs.8517 AA464957
598	CRK: homólogo de oncogén de CT10 del virus de sarcoma aviar v-crk Hs.306088 H75530
599	CCNE2: ciclina E2 Hs.30464 AA520999 UG desconocido Hs.30464 ciclina E2
600	LOC51240: proteína hipotética Hs.7870 AA988037
601	FLJ11259: proteína hipotética FLJ11259 Hs.184465 AA485877
602	PTP4A1: proteína tirosina fosfatasa de tipo IVA, miembro 1 Hs.227777 AA482193
603	ADNc de *Homo sapiens*: FLJ22355 fis, clon HRC06344 Hs.288283 AA026375:
604	Humano: secuencia de ARNm del clon 23719 Hs.80305 H43437
605	ARNm del clon FLC0675 PRO2870 de *Homo sapiens*, cds completa Hs.306117 AA485453:
606	MSE55: proteína constituyente del suero Hs.148101 H73234
607	CFLAR: regulador de apoptosis similar a CASP8 y FADD Hs.195175 AA453766
608	ADNc de *Homo sapiens*: FLJ22844 fis, clon KAIA5181 Hs.296322 AA975103:
609	Humano: secuencia de ADN del clon RP11-371L19 en el cromosoma 20 que contiene dos nuevos genes, el gen para una nueva proteína similar a proteína ribosómica 40S S10 (RPS10), ESTs, STSs, GSSs y cinco islas de CpG Hs.19002 R00846
610	ESTs: Hs.60054 R26390
611	ESTs: débilmente similar a ALU7, ENTRADA DE ADVERTENCIA DE CONTAMINACIÓN DE SECUENCIA SQ DE LA SUBFAMILIA DE ALU HUMANA [*H. sapiens*] Hs.325158 AA032084
612	FLJ10980: proteína hipotética FLJ10980 Hs.29716 N45467
613	IFIT1: **proteína inducida por interferón con repeticiones tetratricopeptídicas 1 Hs.20315 AA157787
614	ESTs: Hs.21734 AA429809
615	DKFZP434C245: proteína DKFZP434C245 Hs.59461 AA705518
616	RNPS1: proteína de unión a ARN S1, dominio rico en serina Hs.75104 AA496837
617	FLJ13639: proteína hipotética FLJ13639 Hs.101821 AA131681
618	PCF11: homólogo de PCF11p Hs.123654 W73749
619	EIF4G3: factor de inicio de traducción eucariota 4 gamma, 3 Hs.25732 N92469
620	ADNc de *Homo sapiens*: FLJ21971 fis, clon HEP05790 Hs.71331 AA130595:
621	STAT1: transductor de señales y activador de la transcripción 1, 91 kD Hs.21486 AA079495
622	BIRC3: proteína baculoviral que contiene repeticiones de IAP 3 Hs.127799 R07870
623	HP1-BP74: HP1-BP74 Hs.142442 N20589
624	HSPC228: proteína hipotética Hs.267288 AI734268
625	KIAA0675: producto génico de KIAA0675 Hs.165662 AA454867
626	AMD1: S-adenosilmetionina descarboxilasa 1 Hs.262476 AA504772
627	EST: Hs.149338 AI249089

628	PWP1: fosfoproteína nuclear similar a PWP1 de *S. cerevisiae* Hs.172589 AA485992
629	AI336973:
630	DUSP4: fosfatasa de doble especificidad 4 Hs.2359 AA444049
631	FLJ12788: proteína hipotética FLJ12788 Hs.20242 AA497041
632	HSPC150: proteína HSPC150 similar a enzima de conjugación con ubiquitina Hs.5199 AA985450
633	FLJ11729: proteína hipotética FLJ11729 Hs.286212 W15533
634	KLF4: factor similar a Kruppel 4 (intestino) Hs.7934 H45668
635	FLJ11058: proteína hipotética FLJ11058 Hs.180817 N63911
636	FLJ23468: proteína hipotética FLJ23468 Hs.38178 AA460299
637	ESTs: Hs.115315 AI278336
638	EBI3: gen 3 inducido por virus de Epstein-Barr Hs.185705 AA425028 EBI3=receptor de citocinas
639	ESTs: Hs.293797 N63988
640	MGAT2: manosil (alfa-1,6-)-glicoproteína beta-1,2-N-acetilglucosaminiltransferasa Hs.172195 AA485653
641	H2BFQ: familia de histonas H2B, miembro Q Hs.2178 AA456298
642	NMB: neuromedina B Hs.83321 AI650675
643	SSR3: receptor de secuencia señal, gamma (proteína asociada a translocones gamma) Hs.28707 AA453486
644	HSPC196: proteína hipotética Hs.239938 R78498
645	EST: Hs.44522 N33610
646	BRF1: factor de respuesta a butirato 1 (factor de respuesta a EGF 1) Hs.85155 AA723035
647	MAN1A2: manosidasa, alfa, clase 1A, miembro 2 Hs.239114 H97940
648	KIAA1201: proteína KIAA1201 Hs.251278 AA427719
649	NUCKS: similar a caseína cinasa ubicua nuclear de rata 2 Hs.118064 AA158345
650	MAGEF1: proteína MAGEF1 Hs.306123 AA425302
651	Humano: cromosoma 16 de BAC, clon CIT987SK-A-362G6 Hs.6349 N75498
652	R40377:
653	AP3M2: complejo 3 de proteína relacionada con adaptador, subunidad mu 2 Hs.77770 R14443
654	ESTs: débilmente similar a proteína relacionada con transcriptasa inversa 1207289A [*H. sapiens*] Hs.272135 AA705010
655	ARNm de *Homo sapiens* para proteína FLJ00116, cds parcial Hs.72363 AA159893:
656	EIF4E: factor de inicio de la traducción eucariota 4E Hs.79306 AA193254
657	ARNm de *Homo sapiens* para proteína hipotética (gen de TR2/D15) Hs.180545 N47285:
658	ESTs: Hs.99542 AA461474
659	CTNND1: catenina (proteína asociada a cadherina), delta 1 Hs.166011 AA024656

660	ESTs: Hs.188554 R75884
661	ZNF217: proteína de dedos de zinc 217 Hs.155040 R81830
662	FLJ12892: proteína hipotética FLJ12892 Hs.17731 AI243595
663	ETV5: gen 5 de variante de ets (molécula relacionada con ets) Hs.43697 AA460265
664	EST: Hs.251574 T54821
665	RPS25: proteína ribosómica S25 Hs.113029 T98662
666	CNN2: calponina 2 Hs.169718 AA284568
667	ESTs: débilmente similar a placofilina 2b [*H. sapiens*] Hs.12705 AA485365
668	PAPPA: proteína plasmática asociada al embarazo A Hs.75874 AA609463
669	TFF3: factor de trébol 3 (intestinal) Hs.82961 N74131
670	A1204264:
671	DJ328E19.C1.1: proteína hipotética Hs.218329 AA486041
672	ME3: enzima málica 3, dependiente de NADP(+), mitocondrial Hs.2838 AA779401
673	ESTs: débilmente similar a IEFS, PROTEÍNA SENSIBLE A TRANSFORMACIÓN HUMANA IEF SSP 3521 [*H. sapiens*] Hs.43213 AA490554
674	FLJ13181: proteína hipotética FLJ13181 Hs.301526 AA057266
675	KIAA1547: proteína KIAA1547 Hs.31305 AA136692
676	ZNF281: proteína de dedos de zinc 281 Hs.59757 N47468
677	ADNc de *Homo sapiens*: FLJ23260 fis, clon COL05804, altamente similar a secuencia de ARNm del clon 23652 humano HSU90911 Hs.13996 AA463961
678	ESTs: Hs.25933 AA411392
679	NCBP1: subunidad 1 de proteína de unión a caperuza nuclear, 80 kD Hs.89563 AA278749 proteína de unión a caperuza nuclear
680	H2BFL: familia de histonas H2B, miembro L Hs.239884 H70774
681	DKFZP564A122: proteína DKFZP564A122 Hs.187991 H66150
682	NASP: proteína espermática autoantigénica nuclear (de unión a histonas) Hs.243886 AA644128
683	**ESTs: débilmente similar a proteína KIAA0822 [*H. sapiens*] Hs.98368 AA422008
684	MAP2K6: proteína cinasa activada por mitógenos 6 Hs.1 18825 H07920
685	ESTs: Hs.158357 AA865842
686	GADD45A: detención del crecimiento e inducible por daño al ADN, alfa Hs.80409 AA147214 GADD45 alfa=proteína de detención del crecimiento e inducible por daño al ADN
687	DHFR: dihidrofolato reductasa Hs.83765 AA488803
688	AA151930:
689	ARNm de *Homo sapiens*; ADNc DKFZp434P116 (del clon DKFZp434P116); cds completa Hs.103378 AA431133:

690	ARNm de *Homo sapiens*; ADNc DKFZp564D156 (del clon DKFZp564D156) Hs.9927 T55704:
691	ESTs: Hs.32204 R93719
692	PRPSAP1: proteína asociada a fosforribosil pirofosfato sintetasa 1 Hs.77498 R20005
693	ZNF42: proteína de dedos de zinc 42 (sensible a ácido retinoico específico de mieloide) Hs.169832 AA987906
694	.**ESTs: Hs.43712 N25936
695	RUNX1: factor de transcripción relacionado con Runt 1 (leucemia mieloide aguda 1; oncogén de aml1) Hs.129914 AA146826
696	ARNm de *Homo sapiens*; ADNc DKFZp547C244 (del clon DKFZp547C244) Hs.9460 T64452:
697	TYMS: timidilato sintetasa Hs.82962 AA663310
698	MGC5528: proteína hipotética MGC5528 Hs.315167 AA843451
699	ESTs: Hs.268685 R22952
700	SFPQ: factor de corte y empalme, rico en prolina/glutamina (asociado a proteína de unión al tracto de polipirimidina) Hs.180610 AA418910
701	ESTs: Hs.155105 AI221390
702	FLJ10624: proteína hipotética FLJ10624 Hs.306000 AA489592
703	TRIP8: interactor del receptor de la hormona tiroidea 8 Hs.6685 AA425205
704	DNAJB6: homólogo de DnaJ (Hsp40), subfamilia B, miembro 6 Hs.181195 AA496105
705	ESTs: Hs.18331 T98244
706	RBM14: proteína de motivos de unión a ARN 14 Hs.11170 AA421233
707	SCYA2: citocina inducible pequeña A2 (proteína quimiotáctica de monocitos 1, homóloga a Sig-je de ratón) Hs.303649 AA425102 MCP-1=MCAF=citocina inducible pequeña A2=JE=quimiocina
708	MGC4161: proteína hipotética MGC4161 Hs.177688 AI224867
709	TUBB2: tubulina, beta, 2 Hs.251653 AA888148
710	FLJ20280: proteína hipotética FLJ20280 Hs.270134 N74086
711	TERA: proteína TERA Hs.180780 AA465096
712	CPS1: **carbamoil-fosfato sintetasa 1, mitocondrial Hs.50966 N68399
713	KIAA0802: proteína KIAA0802 Hs.27657 W55875
714	FYN: oncogén de FYN relacionado con SRC, FGR, YES Hs.169370 N22980
715	ARNm de PRO2751 de *Homo sapiens*, cds completa Hs.283978 H12784:
716	CLTH: gen de leucemia mieloide linfoide de ensamblaje de clatrina Hs.7885 AA441930
717	CHMP1.5: proteína CHMP1.5 Hs.42733 W85875

718	SMARCB1: regulador de la cromatina dependiente de actina, relacionado con SWI/SNF, asociado a matriz, subfamilia b, miembro 1 Hs.159971 AA446018
719	AA487823: SRF= factor de transcripción de unión al elemento de respuesta sérico de c-fos
720	**ESTs: Hs.130741 AA608725
721	ADNc FLJ10976 fis de *Homo sapiens*, clon PLACE1001399 Hs. 296323 R36085:
722	FLJ20036: proteína hipotética FLJ20036 Hs.32922 N91145
723	C11ORF5: marco de lectura abierto 5 del cromosoma 11 Hs.121025 AA776702
724	AF3P21: proteína SH3 Hs.102929 N94372
725	LOC54104: proteína hipotética Hs.12871 H05934
726	DF: componente D del complemento (adipsina) Hs.155597 AA233549
727	CEP4: proteína efectora de Cdc42 4; grupo de unión de GTPasas Rho 4 Hs.3903 AA449061
728	KIF5B: miembro 5B de familia de cinesinas Hs.149436 AA644218
729	MGC5627: proteína hipotética MGC5627 Hs.237971 H02336
730	G3BP: proteína de unión a dominio SH3 de proteína de activación de Ras-GTPasa Hs.220689 AA449834
731	ESTs: Hs.293987 AA229758
732	ESTs: Hs.36828 AA194796
733	ARNm de *Homo sapiens* para proteína FLJ00101, cds parcial Hs.221600 W92262:
734	ADNc de *Homo sapiens*: FLJ21288 fis, clon COL01927 Hs.6019 R07184:
735	ESTs: débilmente similar a proteína relacionada con transcriptasa inversa 1207289A [*H. sapiens*] Hs.250594 H86813
736	ADNc FLJ11941 fis de *Homo sapiens*, clon HEMBB1000649 Hs.124106 AI301573:
737	ESTs: Hs.24908 H77726
738	TOB2: transductor de ERBB2, 2 Hs.4994 AA486088
739	ESTs: Hs.143900 AI193212
740	ARNm del clon FLC0675 PRO2870 de *Homo sapiens*, cds completa Hs.306117 H16589:
741	ESTs: débilmente similar a proteína KIAA0638 [*H. sapiens*] Hs.296288 T83657
742	FLJ20039: proteína hipotética FLJ20039 Hs.267448 AA448268
743	RPA2: proteína de replicación A2 (32 kD) Hs.79411 R13557
744	GAS1: específica de detención del crecimiento 1 Hs.65029 AA025819
745	Humano: secuencia de ADN del clon 967N21 en el cromosoma 201p12.3-13. Contiene el gen de CHGB para cromogranina B (secretogranina 1, SCG1), un pseudogén similar a parte de KIAA0172, el gen para una nueva proteína Hs.88959 R56678
746	ESTs: Hs.21175 AI341642
747	LBC: oncogén de crisis blástica linfoide Hs.301946 AA135716
748	ESTs: Hs.194595 R06761
749	MGC4707: proteína hipotética MGC4707 Hs.291003 R14653

750	ZNF183: proteína de dedos de zinc 183 (dedos de RING, tipo C3HC4) Hs.64794 AA132766
751	RAD18: proteína de reparación posterior a la replicación hRAD18p Hs.21320 R59197
752	EIF4EBP2: **proteína de unión a factor 4E de inicio de la traducción eucariota 2 Hs.278712 H15159
753	**ARNm de *Homo sapiens*; ADNc DKFZp586M0723 (del clon DKFZp586M0723) Hs.27860 AA446650:
754	ORC3L: similar a complejo de reconocimiento de origen, subunidad 3 (homólogo de levaduras) Hs.74420 H99257
755	CDK7: cinasa dependiente de ciclinas 7 (homólogo de cinasa de activación de cdk MO15 de *Xenopus*) Hs.184298 AI311067
756	USP10: proteasa específica de ubiquitina 10 Hs.78829 AA455233
757	KIAA0733: proteína de unión a TAK1 2; proteína KIAA0733 Hs.109727 AA931658
758	R89286:
759	ALDH4: aldehído deshidrogenasa 4 (glutamato gamma-semialdehído deshidrogenasa; pirrolin-5-carboxilato deshidrogenasa) Hs.77448 AA181378
760	IDN3: proteína IDN3 Hs.225767 N62911
761	ESTs: Hs.50180 H48143
762	MIG2: inducible por mitógenos 2 Hs.75260 H29252
763	KIAA0856: proteína KIAA0856 Hs.13264 R12847
764	EST: Hs.47763 N54162
765	ARNm de *Homo sapiens*; ADNc DKFZp547C244 (del clon DKFZp547C244) Hs.9460 AA447553:
766	KIAA0855: golgina-67 Hs.182982 AA775625
767	ESTs: débilmente similar a nucleolina JH0148 - rata [*R. norvegicus*] Hs.30120 R54659
768	FLJ22313: proteína hipotética FLJ22313 Hs.30211 H52061
769	ESTs: Hs.71818 AI028074
770	KIAA0618: producto génico de KIAA0618 Hs.295112 AA455506
771	ESTs: Hs.59413 W93056
772	ESTs: Hs.165607 AA992090
773	UBAP: proteína asociada a ubiquitina Hs.75425 AA446016
774	HAN11: proteína de repeticiones de WD Hs.176600 AA725641
775	USP16: proteasa específica de ubiquitina 16 Hs.99819 AA489619
776	ESTs: Hs.67776 AA464963
777	SM-20: similar a proteína de músculo liso de rata SM-20 Hs.6523 H56028
778	CCNG2: ciclina G2 Hs.79069 AA489647
779	ARNm de *Homo sapiens*; ADNc DKFZp566P1124 (del clon DKFZp566P1124) Hs.321022 N62953:

780	FLJ20094: proteína hipotética FLJ20094 Hs.29700 N95490
781	LOC51174: delta-tubulina Hs.270847 W33133
782	ARNm de *Homo sapiens*; ADNc DKFZp434I1820 (del clon DKFZp434I1820); cds parcial Hs.14235 N52394:
783	FANCA: anemia de Fanconi, grupo de complementación A Hs.284153 AA644129
784	P5-1: ORF de región de MHC de clase 1 Hs.1845 T58146
785	DNA2L: similar a DNA2 (helicasa de replicación de ADN, levadura, homólogo) Hs.194665 AA974495 KIAA0083
786	LOC51578: **proteína de glándula suprarrenal AD-004 Hs.279586 AA150301
787	ESTs: Hs.326417 AA913304
788	CDKN2D: inhibidor de cinasa dependiente de ciclinas 2D (p19, inhibe CDK4) Hs.29656 R77517 p19-INK4D=inhibidor D de cinasa dependiente de ciclinas 4
789	FABP1: proteína de unión a ácidos grasos 1, hígado Hs.5241 AA682392
790	TERA: proteína TERA Hs.180780 AA906997
791	ESTs: Hs.145383 AI253072
792	SLC7A5: familia de portadores de soluto 7 (transportador de aminoácidos catiónicos, sistema y+), miembro 5 Hs.184601 AA419176
793	AXL: tirosina cinasa receptora AXL Hs.83341 HI 5336 axl=ufo=tirosina cinasa receptora
794	LOC57190: selenoproteína N Hs.8518 AA284276
795	ESTs: Hs.99037 AA443948
796	STCH: proteína de estrés 70, chaperona, asociada a microsomas, 60 kD Hs.288799 H85311
797	ESTs: Hs.88523 AA278591 UG desconocido Hs.88523 ESTs
798	ESD: **esterasa D/formilglutatión hidrolasa Hs.82193 AA250931
799	ESTs: Hs.122444 R31021
800	ESTs: Hs.283127 AI291262
801	KIAA0480: **producto génico de KIAA0480 Hs.92200 H91332
802	HP1-BP74: HP1-BP74 Hs.142442 AA598791
803	**ESTs: moderadamente similar a ALU1, ENTRADA DE ADVERTENCIA DE CONTAMINACIÓN DE SECUENCIA DE SUBFAMILIA J DE ALU HUMANA [*H. sapiens*] Hs.144662 AA987667
804	TTF2: factor de terminación de transcripción, ARN polimerasa II Hs.142157 AI023603
805	ESTs: Hs.13740 T70541
806	DJ37E16.5: proteína hipotética dJ37E16.5 Hs.5790 AA400021
807	CDH24: similar a cadherina 24 Hs.155912 AI732266
808	DJ465N24.2.1: **proteína hipotética dJ465N24.2.1 Hs.8084 AA932375
809	ESTs: débilmente similar a proteína S57447 HPBRII-7 [*H. sapiens*] Hs.16346 AA410490
810	ADNc de *Homo sapiens*: FLJ23285 fis, clon HEP09071 Hs.90424 AI005038:

811	KRAS2: homólogo de oncogén viral de sarcoma de rata de Kirsten v-Ki-ras2 2 Hs.184050 N95249
812	FLJ20038: proteína hipotética FLJ20038 Hs.72071 H96090
813	ESTs: débilmente similar a ALU4, ENTRADA DE ADVERTENCIA DE CONTAMINACIÓN DE SECUENCIA DE SUBFAMILIA SB2 DE ALU HUMANA [*H. sapiens*] Hs.28848 AA486607
814	H2AFN: familia de histonas H2A, miembro N Hs.134999 AI095013
815	RERE: repeticiones dipeptídicas de arginina-ácido glutámico (RE) Hs.194369 AA490249
816	USP1: proteasa específica de ubiquitina 1 Hs.35086 T55607
817	TIP47: proteína de selección de carga (proteína de unión a receptor de manosa-6-fosfato) Hs.140452 AA416787
818	KIAA0135: proteína KIAA0135 Hs.79337 AA427740 KIAA0135=relacionada con cinasa pim-1
819	ESTs: Hs.214410 T95273
820	PPP1R2: proteína fosfatasa 1, subunidad 2 reguladora (inhibidor) Hs.267819 N52605
821	ADNc de *Homo sapiens*: FLJ21210 fis, clon COL00479 Hs.325093 AA978323:
822	CSNK2A2: caseína cinasa 2, polipéptido principal alfa Hs.82201 AA054996
823	HSRTSBETA: proteína rTS beta Hs.180433 N66132
824	FLJ13110: proteína hipotética FLJ13110 Hs.7358 AA431233
825	ESTs: Hs.238797 N30704
826	FMN: oncogén de FYN relacionado cocn SRC, FGR, YES Hs.169370 N35086
827	RBM8A: proteína de motivos de unión a ARN 8A Hs.65648 AA448402
828	ESTs: Hs.21906 AA608546
829	ESTs: Hs.128081 AA971042
830	PP591: proteína hipotética PP591 Hs.118666 AA626336
831	N63866:
832	HM74: supuesto receptor de quimiocinas; proteína de unión a GTP Hs.137555 R02739
833	MID1: línea media 1 (síndrome de Opitz/BBB) Hs.27695 AA598640
834	KIAA1586: proteína KIAA1586 Hs.180663 AA938639
835	Secuencia de ARNm del clon CDABP0014 de *Homo sapiens* Hs.92679 AA443139:
836	HSU79274: proteína predicha por el clon 23733 Hs.150555 AA451900
837	AOC3: amina oxidasa, que contiene cobre 3 (proteína de adhesión vascular 1) Hs.198241 AA036974
838	AA548037:
839	FLJ10154: proteína hipotética FLJ10154 Hs.179972 AA457133
840	THBS1: trombospondina 1 Hs.87409 AA464532
841	DNAJB6: homólogo de DnaJ (Hsp40), subfamilia B, miembro 6 Hs.181195 AA431203
842	KIAA1547: proteína KIAA1547 Hs.31305 AI216623

843	GATA2: proteína de unión a GATA 2 Hs.760 R32405
844	ESTs: Hs.176950 R82522
845	KIAA1018: proteína KIAA1018 Hs.5400 AA156859
846	B4GALT1: **UDP-Gal:betaGlcNAc beta 1,4-galactosiltransferasa, polipéptido 1 Hs.198248 AA043795
847	HMGCR: 3-hidroxi-3-metilglutaril-coenzima A reductasa Hs.11899 AA779417
848	ESTs: débilmente similar a proteína 1819485A CENP-E [*H. sapiens*] Hs.167652 H94466
849	ESTs: Hs.294088 AA971073
850	KIAA1637: coactivador independiente de AF-2 (CIA); proteína KIAA1637 Hs.288140 AA918007
851	HSPC196: proteína hipotética Hs.239938 H66023
852	DR1: regulador por disminución de la transcripción 1, de unión a TBP (cofactor negativo 2) Hs.16697 AA132007
853	CG1I: supuesta proteína de interacción con ciclina G1 Hs.10028 AA486444
854	IGSF4: superfamilia de inmunoglobulinas, miembro 4 Hs.70337 AA487505
855	ESTs: Hs.179309 AA664350
856	HSPC163: proteína HSPC163 Hs.108854 AA053139
857	FLJ12788: proteína hipotética FLJ12788 Hs.20242 AI061317
858	FEM1B: homólogo b de FEM-1 (*C. elegans*) Hs.6048 H82273
859	FXR1: retraso mental asociado al cromosoma X frágil, homólogo autosómico 1 Hs.82712 N62761
860	NCOA3: coactivador del receptor nuclear 3 Hs.225977 AA156793
861	H2BFB: familia de histonas H2B, miembro B Hs.180779 N33927
862	ESTs: Hs.23830 AA460601
863	CDK7: cinasa dependiente de ciclinas 7 (homólogo de cinasa de activación de cdk MO15 de *Xenopus*) Hs.184298 AA031961 CAK=cdk7=NRTALRE=sdk=cinasa de activación de CDK
864	FLJ20259: proteína hipotética FLJ20259 Hs.9956 T55949
865	ADNc FLJ20678 fis de *Homo sapiens*, clon KAIA4163 Hs.143601 T95823:
866	RPS19: proteína ribosómica S19 Hs.298262 T72208
867	ARNm de *Homo sapiens*; ADNc DKFZp434M0420 (del clon DKFZp434M0420) Hs.326048 AA443976:
868	TP53: proteína tumoral p53 (síndrome de Li-Fraumeni) Hs.1846 R39356 p53
869	FBI1: inductor del VIH-1 de proteína de unión a transcritos cortos Hs.104640 R06252
870	GOT1: transaminasa glutámica-oxaloacética 1, soluble (aspartato aminotransferasa 1) Hs.597 H22855
871	FLJ21434: proteína hipotética FLJ21434 Hs.298503 AA680129
872	DNMT2: ADN (citosina-5-)-metiltransferasa 2 Hs.97681 R95731
873	ESTs: Hs.55272 W02785

874	H2BFQ: familia de histonas H2B, miembro Q Hs.2178 AA010223
875	NFIC: factor nuclear I/C (factor de transcripción de unión a CCAAT) Hs.184771 N20996
876	NPTX1: pentraxina neuronal 1 Hs.84154 H22445
877	TLOC1: proteína de translocación 1 Hs.8146 AA450205
878	MGC5302: proteína residente en el retículo endoplásmico 58; proteína hipotética MGC5302 Hs.44970 N39195
879	ACTR2: homólogo de ARP2 (proteína relacionada con actina 2, levadura) Hs.42915 AA032090
880	AI287555:
881	ABCA7: casete de unión a ATP, subfamilia A (ABC1), miembro 7 Hs.134514 AI668632
882	COL7A1: colágeno, tipo VII, alfa I (epidermólisis ampollosa, distrófica, dominante y recesiva) Hs.1640 AA598507
883	RFC2: factor de replicación C (activador 1) 2 (40 kD) Hs.139226 AA663472
884	FLJ22583: proteína hipotética FLJ22583 Hs.287700 AA135836
885	**ESTs: débilmente similar a ORF2 [*M. musculus*] Hs.172208 AI820570
886	ESTs: Hs.21667 R15709
887	RBBP4: proteína de unión a retinoblastoma 4 Hs.16003 AA705035
888	ARNm de *Homo sapiens*; ADNc DKFZp434J1027 (del clon DKFZp434J1027); cds parcial Hs.22908 R20166:
889	ESTs: Hs.166539 AI080987
890	NKTR: secuencia de reconocimiento de tumor de linfocitos naturales Hs.241493 AA279666 proteína de reconocimiento de tumor de NK=proteína relacionada con ciclofilina
891	MUC1: mucina 1, transmembrana Hs.89603 AA486365
892	AP4B1: complejo 4 de proteína relacionada con adaptador, subunidad beta 1 Hs.28298 AA481045
893	ESTs: Hs.94943 AA452165
894	MITF: factor de transcripción asociado a microftalmia Hs.166017 N66177
895	ESTs: Hs.183299 AA286914 UG desconocido Hs.183299 ESTs sc_id2032
896	BAG3: atanogén asociado a BCL2 3 Hs.15259 AI269958
897	INSR: receptor de insulina Hs.89695 AA001106
898	TRIP: proteína de interacción con TRAF Hs.21254 AA186426
899	EST: Hs.307975 R22182
900	**ADNc de *Homo sapiens*: FLJ23037 fis, clon LNG02036, altamente similar a HSU68019 ARNm de homólogo de proteína mad de *Homo sapiens* (hMAD-3) Hs.288261 W72201:
901	HLA-ADN: complejo principal de histocompatibilidad, clase II, DN alfa Hs.11135 AA702254 complejo principal de histocompatibilidad, clase II, DN alfa
902	FLJ10392: **proteína hipotética FLJ10392 Hs.20887 AI261305
903	MPHOSPH1: **fosfoproteína de fase M 1 Hs.240 N63752
904	STAG1: antígeno estromal 1 Hs.286148 R36160

905	USP1: proteasa específica de ubiquitina 1 Hs.35086 AA970066
906	ESTs: moderadamente similar a ALU4, ENTRADA DE ADVERTENCIA DE CONTAMINACIÓN DE SECUENCIA DE SUBFAMILIA SB2 DE ALU HUMANA [*H. sapiens*] Hs.181315 AA448251
907	PA26: proteína nuclear PA26 regulada por p53 Hs.14125 AA447661
908	ESTs: débilmente similar a proteína de dedos de zinc [*H. sapiens*] Hs.71243 N92478
909	SH3PX1: proteína que contiene dominios SH3 y PX SH3PX1 Hs.7905 R69163
910	**ADNc de *Homo sapiens*: FLJ22554 fis, clon HSI01092 Hs.93842 H58317:
911	RPS25: proteína ribosómica S25 Hs.113029 AA779404
912	ESTs: débilmente similar a región V-I de cadena kappa de Ig A49134 [*H. sapiens*] Hs.5890 N34799 fra-2=antígeno relacionado con fos 2
913	TXNRD1: tiorredoxina reductasa 1 Hs.13046 AA453335 tiorredoxina reductasa
914	**ESTs: Hs.184378 N77828
915	GCSH: proteína de sistema de escisión de glicina H (portador de aminometilo) Hs.77631 R71327
916	ADNc FLJ11904 fis de *Homo sapiens*, clon HEMBB1000048 Hs.285519 AA447098:
917	NCOA3: coactivador del receptor nuclear 3 Hs.225977 H51992 AIB1=amplificado en cáncer de mama=TRAM-1=RAC3=ACTR=CAGH16=nucl
918	FLJ20159: proteína hipotética FLJ20159 Hs.288809 R33122
919	IL7R: receptor de interleucina 7 Hs.237868 AA487121
920	RAB23: RAB23, miembro de familia de oncogenes RAS Hs.94769 AA134569
921	ESTs: Hs.132493 AA923168
922	ESTs: Hs.87507 AA236015
923	SHC1: proteína transformante de SHC (que contiene dominio de homología de Src 2) 1 Hs.81972 R52960
924	KIAA1321: proteína KIAA1321 Hs.24336 W37999
925	GL1: homólogo de oncogén asociado a glioma (proteína de dedos de zinc) Hs.2693 AI373071
926	ESTs: Hs.183299 AA291137 UG desconocido Hs.183299 ESTs sc_id2032
927	GPRK6: cinasa receptora acoplada a proteína G 6 Hs.76297 AA291284
928	ESTs: Hs.93704 AA702684
929	CAPS: calcifosina Hs.26685 AA858390
930	ADNc FLJ10976 fis de *Homo sapiens*, clon PLACE1001399 Hs.296323 R27711
931	C6: componente 6 del complemento Hs.1282 N59396
932	UBE2D3: enzima de conjugación con ubiquitina E2D 3 (homóloga a UBC4/5 de levadura) Hs.118797 AA465196
933	DDX8: polipéptido de 8 caja DEAD/H (Asp-Glu-Ala-Asp/His) (ARN helicasa) Hs.171872 AA465387 ARN helicasa (HRH1)
934	DKFZP434B168: proteína DKFZP434B168 Hs.48604 N62684
935	FLJ10512: proteína hipotética FLJ10512 Hs.93581 T39933

936	ARNm de *Homo sapiens*; ADNc DKFZp564F093 (del clon DKFZp564F093) Hs.18724 W87709:
937	F8A: asociado a factor coagulación VIII (transcrito intrónico) Hs.83363 AA463924
938	HSU53209: transformador-2 alfa (htra-2 alfa) Hs.24937 AA465172
939	UBQLN2: ubiquilina 2 Hs.4552 R43580
940	EIF2C2: factor de inicio de traducción eucariota 2C, 2 Hs.193053 N93082
941	ARNm de *Homo sapiens* para proteína FLJ00012, cds parcial Hs.21051 H17645
942	KIAA0841: proteína KIAA0841 Hs.7426 R20299
943	KCNAB2: subfamilia de canales de potasio dependientes de voltaje, relacionados con agitador, miembro beta 2 Hs.298184 H14383
944	KIAA1637: coactivador independiente de AF-2 (CIA); proteína KIAA1637 Hs.288140 AA521358
945	ESTs: Hs.27379 H17455
946	FLJ11323: proteína hipotética FLJ11323 Hs.25625 R49707
947	SSP29: proteína ácida rica en leucinas Hs.84264 AA489201
948	ESTs: Hs.69280 AA486011
949	ADAMTS1: similar a desintegrina y metaloproteasa (tipo reprolisina) con motivo de trombospondina tipo 1, 1 Hs.8230 AA057170
950	ESTs: Hs.43466 N23889
951	MLLT4: leucemia mieloide/linfoide o de linaje mixto (homólogo de Trithorax (*Drosophila*)); translocada, 4 Hs.100469 AA010818
952	ESTs: Hs.271034 AA406581
953	LMNB1: lámina B1 Hs.89497 AA983462
954	ADNc FLJ13547 fis de *Homo sapiens*, clon PLACE1007053 Hs.7984 AA629264:
955	PTMS: paratimosina Hs.171814 R10451
956	H2AFL: familia de histonas H2A, miembro 1 Hs.28777 AI268551
957	FLJ21603: proteína hipotética FLJ21603 Hs.129691 R72794
958	FLJ13287: proteína hipotética FLJ13287 Hs.53263 AA621725
959	CXCR4: quimiocina (motivo C-X-C), receptor 4 (fusina) Hs.89414 AA479357
960	INSM1: asociado a insulinoma 1 Hs.89584 R38640
961	FREQ: homólogo de frecuenina (*Drosophila*) Hs.301760 HI 6821
962	LOC58486: proteína similar a transposasa Buster1 derivada de transposones Hs.25726 AA868020
963	SMARCD1: regulador de la cromatina dependiente de actina, relacionado con SWI/SNF, asociado a matriz, subfamilia d, miembro 1 Hs.79335 H91691
964	ESTs: Hs.242998 T96522
965	INADL: proteína de dominio PDZ (similar a inaD de *Drosophila*) Hs.321197 AA005153

966	ESTs: débilmente similar a supuesta p150 [*H. sapiens*] Hs.37751 AA436174
967	MGC5338: proteína hipotética MGC5338 Hs.99598 H50550
968	W85890:
969	NUCKS: similar a caseína cinasa ubicua nuclear de rata 2 Hs.118064 AI053436
970	Secuencia de ARNm del clon 25110 de *Homo sapiens* Hs.27262 H18031:
971	AI333214:
972	GAS41: secuencia amplificada en glioma 41 Hs.4029 T62072
973	LOC51170: deshidrogenasa/reductasa de cadena corta retiniana retSDR2 Hs.12150 N79745
974	H2BFG: **familia de histonas H2B, miembro G Hs.182137 R98472
975	ABCC1: **casete de unión a ATP, subfamilia C (CFTR/MRP), miembro 1 Hs.89433 AA424804
976	EFNA1: efrina-A1 Hs.1624 AA857015
977	ARNm de Homo sapiens; ADNc DKFZp434A1014 (del clon DKFZp434A1014); cds parcial Hs.278531 H00596:
978	PPP2CA: proteína fosfatasa 2 (anteriormente 2A), subunidad catalítica, isoforma alfa Hs.91773 AA599092
979	ESTs: débilmente similar a producto de proteína sin nombre [*H. sapiens*] Hs.118338 W85843
980	ADNc FLJ11643 fis de *Homo sapiens*, clon HEMBA1004366 Hs.111496 AA598803:
981	ESTs: moderadamente similar a ALUE,!!!! ENTRADA DE ADVERTENCIA DE CLASE E DE ALU HUMANA!!! [*H. sapiens*] Hs.125407 AA878944
982	ESTs: moderadamente similar a ALU1, ENTRADA DE ADVERTENCIA DE CONTAMINACIÓN DE SECUENCIA DE SUBFAMILIA J DE ALU HUMANA [*H. sapiens*] Hs.144662 AI191290
983	KIAA0916: proteína KIAA0916 Hs.151411 R91388
984	CDC25A: ciclo de división celular 25A Hs.1634 R09062
985	PRIM2A: primasa, polipéptido 2A (58 kD) Hs.74519 R61073
986	DSP: desmoplaquina (DPI, DPII) Hs.74316 H90899
987	KIAA0101: producto génico de KIAA0101 Hs.81892 W68219
988	ESTs: débilmente similar a supuesta p150 [*H. sapiens*] Hs.268026 AA411454
989	ESTs: Hs.18140 T97707
990	H2AFL: familia de histonas H2A, miembro L Hs.28777 AA457566
991	ARNm de *Homo sapiens* para proteína KIAA1700, cds parcial Hs.20281 H00287:
992	STAG3: antígeno estromal 3 Hs.20132 AA453028
993	ZNF207: proteína de dedos de zinc 207 Hs.62112 N59119
994	BMP6: proteína morfogenética ósea 6 Hs.285671 AA424833
995	ESTs: moderadamente similar a sertolina [*R. norvegicus*] Hs.91192 H60690
996	LOC51064: **homólogo de subunidad 13 de glutatión S-transferasa Hs.279952 W88497

997	NUCKS: similar a caseína cinasa ubicua nuclear de rata 2 Hs.118064 AA927182
998	ESTs: débilmente similar a T00370, proteína hipotética KIAA0659 [*H. sapiens*] Hs.131899 W93155
999	FLJ13057: proteína hipotética FLJ13057 similar a GCL (*germ cell-less*) Hs.243122 R23254
1000	ESTs: Hs.144796 AI219737
1001	FLJ10511: proteína hipotética FLJ10511 Hs.106768 R25877
1002	DKFZP564A122: proteína DKFZP564A122 Hs.187991 N31577
1003	ODF2: fibra densa exterior de colas de espermatozoides 2 Hs.129055 AA400407
1004	AMY2A: amilasa, alfa 2A; pancreática Hs.278399 R64129
1005	**ESTs: débilmente similar a placofilina 2b [*H. sapiens*] Hs.12705 N91589
1006	CYP1B1: citocromo P450, subfamilia I (inducible por dioxina), polipéptido 1 (glaucoma 3, infantil primario) Hs.154654 AA029776
1007	CAPN7: calpaína 7 Hs.7145 N46420
1008	FLJ20069: proteína hipotética FLJ20069 Hs.273294 AA229966
1009	FLJ10618: proteína hipotética FLJ10618 Hs.42484 AA478847
1010	KIAA1637: **coactivador independiente de AF-2 (CIA); proteína KIAA1637 Hs.288140 AA452531
1011	FLJ20004: **proteína hipotética FLJ20004 Hs.17311 AA487895
1012	FLJ12892: proteína hipotética FLJ12892 Hs.17731 AA670363
1013	PLU-1: supuesto motivo de unión a ADN/cromatina Hs.143323 AA464869
1014	**ESTs: Hs.36828 AA418448
1015	KIAA0586: producto génico de KIAA0586 Hs.77724 AA905278
1016	MTHFD2: metilen-tetrahidrofolato deshidrogenasa (dependiente de NAD+), meteniltetrahidrofolato ciclohidrolasa Hs.154672 AA480994
1017	BRF1: **factor de respuesta a butirato 1 (factor de a respuesta EGF 1) Hs.85155 AA424743
1018	TFAP2A: factor de transcripción AP-2 alfa (proteína de unión a potenciador de activación 2 alfa) Hs.18387 R38044
1019	VIL2: vilina 2 (ezrina) Hs.155191 AA411440
1020	SDC1: sindecano 1 Hs.82109 AA074511
1021	RNTRE: relacionado con el extremo N-terminal de tre Hs.278526 AA281057
1022	HSPC207: proteína hipotética Hs.75798 H99997
1023	FLJ22376: proteína hipotética FLJ22376 Hs.29341 AI199155
1024	RNF10: proteína de dedos de Ring 10 Hs.5094 H73586
1025	PNN: pinina, proteína asociada a desmosomas Hs.44499 AA707321
1026	FLJ20516: proteína hipotética FLJ20516 Hs.70811 AA122393
1027	RPL13A: proteína ribosómica L13a Hs.119122 AI254200

1028	H2BFB: familia de histonas H2B, miembro B Hs.180779 AA885642
1029	OGT: N-acetilglucosamina (GlcNAc) transferasa unida a O (UDP-N-acetilglucosamina:polipéptido-N-acetilglucosaminiltransferasa) Hs.100293 R13317
1030	KIAA0155: producto génico de KIAA0155 Hs.173288 AA133684
1031	ILF2: factor de unión a potenciador de interleucina 2, 45 kD Hs.75117 H95638
1032	ARNm de *Homo sapiens*; ADNc DKFZp586I1518 (del clon DKFZp586I1518) Hs.21739 AA287917:
1033	PKNOX1: homeosecuencia 1 de PBX/knotted 1 Hs.158225 AI350546
1034	KMO: **quinurenina 3-monooxigenasa (quinurenina 3-hidroxilasa) Hs.107318 AA044326
1035	VCAM1: molécula de adhesión celular vascular 1 Hs.109225 H16591 CD106=VCAM-1
1036	N54811:
1037	KIAA0618: producto génico de KIAA0618 Hs.295112 H81940
1038	MAFG: familia de oncogenes de fibrosarcoma musculoaponeurótico v-maf (aviar), proteína G Hs.252229 N21609 MafG=factor de transcripción de cremallera de leucinas básica
1039	MATN2: matrilina 2 Hs.19368 AA071473
1040	HOXB4: homeosecuencia B4 Hs.126666 AA918749
1041	FLJ10466: proteína hipotética FLJ10466 Hs.121073 AA453607
1042	FLJ22557: proteína hipotética FLJ22557 Hs.106101 AA127879
1043	EST: Hs.149260 AI247680
1044	KIAA0677: producto génico de KIAA0677 Hs.155983 AA026751
1045	EST: Hs.104123 AA197344
1046	UCP4: proteína de desacoplamiento 4 Hs.40510 H94680
1047	EST: Hs.144224 N93807
1048	GATA2: proteína de unión a GATA 2 Hs.760 H00625 proteína de unión a GATA 2
1049	ESTs: Hs.14743 H61082
1050	EST: Hs.116174 AA626786
1051	ITGB3: integrina, beta 3 (glicoproteína de plaquetas IIIa, antígeno CD61) Hs.87149 AA666269
1052	FLJ23399: proteína hipotética FLJ23399 Hs.299883 R19895
1053	ESTs: Hs.21734 N72976
1054	FLJ20425: proteína hipotética FLJ20425 Hs.71040 AA424566
1055	CUL4A: culina 4A Hs.183874 AA598836
1056	PTP4A1: proteína tirosina fosfatasa de tipo IVA, miembro 1 Hs.227777 R61007 proteína tirosina fosfatasa PTPCAAX1 (hPTPCAAX1)
1057	ESTs: Hs.7913 N35592
1058	GRO1: oncogén de GRO1 (actividad estimulante de crecimiento de melanoma, alfa) Hs.789 W46900

1059	ESTs: moderadamente similar a convertasa NRD2 [*H. sapiens*] Hs.309734 1178796
1060	FLJ10826: proteína hipotética FLJ10826 Hs.24809 AA486738
1061	TOM34: translocasa de membrana mitocondrial externa 34 Hs.76927 AA457118
1062	H2AFL: familia de histonas H2A, miembro L Hs.28777 AA452933
1063	D10S170: **segmento de ADN, copia única, sonda pH4 (secuencia transformante, tiroide-1, Hs.315591 N35493
1064	SCYA2: citocina inducible pequeña A2 (proteína quimiotáctica de monocitos 1, homóloga a Sig-je de ratón) Hs.303649 T77816 MCP-1=MCAF=citocina inducible pequeña A2=JE=quimiocina
1065	FLJ10688: proteína hipotética FLJ10688 Hs.118793 AA465358
1066	PTD017: proteína PTD017Hs.274417 AA160498
1067	KIAA0026: gen relacionado con MORF X Hs.173714 AA676604
1068	BMP2: proteína morfogenética ósea 2 Hs.73853 AA489383
1069	MNT: proteína de unión a MAX Hs.25497 AA455508
1070	KIAA1170: proteína KIAA1170 Hs.268044 H80507
1071	CRYBA1: cristalina, beta A1 Hs.46275 AA487614
1072	KATNA1: subunidad A de katanina p60 (que contiene ATPasa) 1 Hs.289099 AA609740
1073	ADNc FLJ20796 fis de *Homo sapiens*, clon COL00301 Hs.113994 N53458:
1074	CEP4: proteína efectora de Cdc42 4; grupo de unión de GTPasas Rho 4 Hs.3903 W32509
1075	ESTs: Hs.117261 AA682521
1076	CYP1B1: citocromo P450, subfamilia I (inducible por dioxina), polipéptido 1 (glaucoma 3, infantil primario) Hs.154654 AA040872
1077	ALTE: elemento de transposón similar a Ac Hs.9933 AA630498
1078	RAD51: homólogo de RAD51 (*S. cerevisiae*) (homólogo de RecA de *E coli*) Hs.23044 AA873056
1079	MAN1A2: manosidasa, alfa, clase 1A, miembro 2 Hs.239114 R78501
1080	H53763.
1081	MET: proto-oncogén de met (receptor de factor de crecimiento de hepatocitos) Hs.285754 AA410591
1082	DYRK1A: cinasa regulada por fosforilación de tirosina (Y) de doble especificidad 1A Hs.75842 AA676749
1083	ARHGAP8: **proteína de activación de GTPasa Rho 8 Hs.102336 AA037410
1084	LMO4: único dominio de LIM 4 Hs.3844 H27986
1085	ADCY6: adenilato ciclasa 6 Hs.12373 AA148044
1086	EST: Hs.135448 AI078552
1087	NCOA3: coactivador del receptor nuclear 3 Hs.225977 W46433
1088	DNAJB4: homólogo de DnaJ (Hsp40), subfamilia B, miembro 4 Hs.41693 AA081471
1089	NAB1: proteína de unión a NGFI-A 1 (proteína de unión a ERG1 1) Hs.107474 AA486027

1090	ESTs: débilmente similar a proteína hipotética T08663 DKFZp547G0910.1 [*H. sapiens*] Hs.172084 N63646
1091	KIAA0735: producto génico de KIAA0735; homólogo 2B de proteína de vesículas sinápticas Hs.8071 R56082
1092	GNB1: proteína de unión a nucleótidos de guanina (proteína G), polipéptido beta 1 Hs.215595 AA487912
1093	ARNm de *Homo sapiens* para proteína KIAA1716, cds parcial Hs.21446 R49763:
1094	CINESINA: CADENA PESADA
1095	CCND1: ciclina D1 (PRAD1: adenomatosis paratiroidea 1) Hs.82932 AA487486 ciclina D1=BCL1=PRAD1=translocada en leucemia de células del manto
1096	ESTs: Hs.106129 R56716
1097	AA431931.
1098	PSEN1: presenilina 1 (enfermedad de Alzheimer 3) Hs.3260 AA403083
1099	ESTs: Hs.193804 AA010918
1100	DKFZp762P2111: proteína hipotética DKFZp762P2111 Hs.14217 AA429586
1101	KIAA1350: proteína KIAA1350 Hs.101799 W37627
1102	FLJ20847: proteína hipotética FLJ20847 Hs.13479 H16996
1103	HDCMA18P: proteína HDCMA18P Hs.278635 N64387
1104	FLJ12890: proteína hipotética FLJ12890 Hs.43299 N62475
1105	ESTs: Hs.127453 AA973625
1106	BAIAP2: proteína asociada a BAI1 2 Hs.7936 R60328
1107	ESTs: Hs.317584 AA191424
1108	DKFZP434J046: proteína DKFZP434J046 Hs.116244 AI024401
1109	ESTs: Hs.114055 AA701352
1110	ESTs: Hs.44380 N93122
1111	ESTs: Hs.20142 AA625570
1112	UBL3: similar a ubiquitina 3 Hs.173091 T82438
1113	H2AFL: familia de histonas H2A, miembro L Hs.28777 N50797
1114	SUCLG2: **succinato-CoA ligasa, de formación de GDP, subunidad beta Hs.247309 N68557
1115	ZWINT: interactor de ZW10 Hs.42650 AA706968
1116	FLJ10583: proteína hipotética FLJ10583 Hs.105633 R00425
1117	FLJ20552: proteína hipotética FLJ20552 Hs.69554 AA463982
1118	FADD: asociado a Fas (TNFRSF6) a través de dominio de muerte Hs.86131 AA430751 FADD=MORT
1119	SFRS7: factor de corte y empalme, rico en arginina/serina 7 (35 kD) Hs.184167 AA418813
1120	RAD54L: similar a RAD54 (*S. cerevisiae*) Hs.66718 AI372035
1121	MYLE: proteína MYLE Hs.11902 T68845

1122	LOC51334: proteína de células madre mesenquimatosas DSC54 Hs.157461 R63841
1123	PRIM2A: primasa, polipéptido 2A (58 kD) Hs.74519 AA434404
1124	KIAA0056: proteína KIAA0056 Hs.13421 AA430545
1125	ESTs: moderadamente similar a ALU7, ENTRADA DE ADVERTENCIA DE CONTAMINACIÓN DE SECUENCIA DE SUBFAMILIA SQ DE ALU HUMANA [*H. sapiens*] Hs.82590 N53024
1126	ESTs: Hs.17269 AA705050
1127	NSAP1: proteína asociada a NS1 1 Hs.155489 AA186327
1128	CEACAM5: molécula de adhesión celular relacionada con antígeno carcinoembrionario 5 Hs.220529 AA130547
1129	FLJ11021: proteína hipotética FLJ11021 similar a factor de corte y empalme, rico en arginina/serina 4 Hs.81648 AA291183 UG desconocido Hs.202583 ESTs, débilmente similar a rico en arginina
1130	FOSL1: antígeno similar a FOS 1 Hs.283565 T82817 fra-1=antígeno relacionado con fos 1
1131	U3-55K: proteína de 55 kDa asociada a U3 snoRNP Hs.153768 AA465355
1132	DNAJC6: homólogo de DnaJ (Hsp40), subfamilia B, miembro 6 Hs.44896 AA455940
1133	KIAA1382: transportador de aminoácidos 2 Hs.298275 R27255 similar a proteína transportadora
1134	PCAF: factor asociado a p300/CBP Hs.199061 N74637 P/CAF=factor asociado a p300/CBP
1135	ESTs: Hs.130460 AA927252
1136	ESTs: Hs.112570 AI014667
1137	FLJ10209: proteína hipotética FLJ10209 Hs.260150 AA454626
1138	ESTs: Hs.99014 AA485679
1139	ESTs: Hs.99621 AA464707
1140	ADNc FLJ11904 fis de *Homo sapiens*, clon HEMBB1000048 Hs.285519 N74617:
1141	AA928536:
1142	SQSTM1: **secuestosoma 1 Hs.182248 AA931964
1143	**ADNc FLJ13700 fis de *Homo sapiens*, clon PLACE2000216, altamente similar a CADENA BETA DE ESPECTRINA, CEREBRO Hs.324648 AA018591
1144	SLC22A3: familia 22 de portador de soluto 22 (transportador de monoaminas extraneuronales), miembro 3 Hs.81086 AA460012
1145	FLJ22557: proteína hipotética FLJ22557 Hs.106101 H00595
1146	FLJ20539: proteína hipotética FLJ20539 Hs.118552 R36152
1147	AA991624:
1148	TRAP150: proteína asociada a receptor de hormona tiroidea, subunidad de 150 kDa Hs.108319 W85832
1149	ESTs: Hs.221847 R91557
1150	TCFL1: similar a factor de transcripción 1 Hs.2430 AA443950

1151	ESTs: altamente similar a variante 1 de corte y empalme de oxitocinasa [*H. sapiens*] Hs.203271 AA487918
1152	PLAB: factor de diferenciación de próstata Hs.296638 AA450062
1153	RBM14: proteína de motivos de unión a ARN 14 Hs.1170 AA417283
1154	EGFL5: dominio similar a EGF, múltiple 5 Hs.5599 W67981
1155	H2AFO: familia de histonas H2A, miembro O Hs.795 AA047260
1156	ESTs: débilmente similar a leucotrieno B4 omega-hidroxilasa A46661 [*H. sapiens*] Hs.169001 N45556
1157	W78784:
1158	TOP3A: topoisomerasa (ADN) III alfa Hs.91175 N21546
1159	W73732: factor de células huésped 1=proteína de interacción con transactivador de VP16
1160	CYP1B1: citocromo P450, subfamilia I (inducible por dioxina), polipéptido 1 (glaucoma 3, infantil primario) Hs.154654 AA448157 citocromo P450 IB1 (inducible por dioxina)
1161	ESTs: Hs.135276 AI092102
1162	RHEB2: homólogo de Ras enriquecido en cerebro 2 Hs.279903 AA482117
1163	ESTs: altamente similar a EF-9 [*M. musculus*] Hs.8366 H94467
1164	POLA: polimerasa (dirigida por ADN), alfa Hs.267289 AA707650
1165	KIAA1008: proteína KIAA1008 Hs.323346 AA863115
1166	PIK3CD: fosfoinositida-3-cinasa, catalítica, polipéptido delta Hs.162808 AA281652
1167	T53625:
1168	**ARNm de *Homo sapiens*; ADNc DKFZp434A1114 (del clon DKFZp434A1114) Hs.326292 AA417274:
1169	ESTs: Hs.26744 H16988
1170	FLJ13912: proteína hipotética FLJ13912 Hs.47125 W74133
1171	ARNm de *Homo sapiens*; ADNc DKFZp762B195 (del clon DKFZp762B195) Hs.284158 AA625574:
1172	SSA2: antígeno de síndrome de Sjogren A2 (60 kD, autoantígeno de ribonucleoproteína SS-A/Ro) Hs.554 AA010351
1173	BK1048E9.5: proteína hipotética bK1048E9.5 Hs.6657 N68512
1174	TOP1: topoisomerasa (ADN) I Hs.317 AA232856 topoisomerasa I
1175	ESTs: Hs.15386 H18472
1176	KPNB1: carioferina (importina) beta 1 Hs.180446 AA121732
1177	MGC861: proteína hipotética MGC861 Hs.208912 N69694
1178	PMS2L8: **proteína 8 similar a proteína 2 de segregación posmeiótica aumentada Hs.323954 T62577
1179	TSC22: **proteína estimulada por factor de crecimiento transformante beta TSC-22 Hs.114360 R16390
1180	C8ORF1: marco de lectura abierto 1 del cromosoma 8 Hs 40539 AA278836

1181	ESTs: Hs.129165 AA989211
1182	DMTF: factor de transcripción similar a Myb de unión a ciclina D 1 Hs.5671 AA129860
1183	CDC7L1: similar a CDC7 (ciclo de división celular 7, *S. cerevisiae*, homólogo) 1 Hs.28853 N62245 cinasa relacionada con Cdc7
1184	LOC51700: citocromo b5 reductasa b5R.2 Hs.22142 AA425316
1185	FLNA: filamina A, alfa (proteína de unión a actina 280) Hs.195464 AA598978
1186	FLJ20257: proteína hipotética FLJ20257 Hs.178011 H78675
1187	ADNc FLJ13604 fis de *Homo sapiens*, clon PLACE1010401 Hs.23193 AA406599:
1188	ESTs: Hs.205227 R73480
1189	SCYB14: subfamilia B de citocinas inducibles pequeñas (Cys-X-Cys), miembro 14 (BRAK) Hs.24395 AA953842
1190	MAPK8IP2: **proteína 2 de interacción con proteína cinasa activada por mitógenos 8 Hs.80545 AA418293
1191	ZNF42: proteína de dedos de zinc 42 (sensible a ácido retinoico específico de mieloide) Hs.169832 AA932642
1192	ESTs: Hs.127054 AA862450
1193	NUDT4: motivo 4 de tipo Nudix (resto X unido a nucleósido difosfato) Hs.92381 AA425630
1194	ADNc FLJ10632 fis de *Homo sapiens*, clon NT2RP2005637 Hs.202596 H82421
1195	LOC51042: proteína de dedos de zinc Hs.102419 AA033532
1196	NUMA1: proteína de aparato mitótico nuclear 1 Hs.301512 AA679293
1197	ESTs: altamente similar a proteína I relacionada con kappa B A56429 [*H. sapiens*] Hs.144614 AA293771
1198	ESTs: Hs.127703 AA947258
1199	ADNc FLJ14214 fis de *Homo sapiens*, clon NT2RP3003576 Hs.321236 AA903913:
1200	NFKBIA: factor nuclear de potenciador de genes de polipéptido ligero kappa en inhibidor de células B, alfa Hs.81328 W55872 IkB alfa
1201	ESTs: Hs.120029 AA707598
1202	ESTs: moderadamente similar a cadena A, glucosamina-6-fosfato desaminasa isomerasa humana a 1,75 A [*H. sapiens*] Hs.21398 AA172012
1203	NFIA: factor nuclear I/A Hs.173933 AI912047
1204	RECQL4: similar a proteína RecQ 4 Hs.31442 AA620446
1205	*ESTs: débilmente similar a ALU1, ENTRADA DE ADVERTENCIA DE CONTAMINACIÓN DE SECUENCIA DE SUBFAMILIA J DE ALU HUMANA [*H. sapiens*] Hs.318894 R96212
1206	ADNc de Homo sapiens: FLJ21686 fis, clon COL09379 Hs.20787 R11371
1207	LOC57168: similar a aspartato beta hidroxilasa (ASPH) Hs.184390 H17272
1208	ESTs: Hs.26096 R54109

1209	ARNm de proteína relacionada con OSBP de *Homo sapiens* 6, cds completa Hs.318775 AA680281:
1210	APACD: proteína de unión a ATP asociada con la diferenciación celular Hs.153884 N80741
1211	VIM: **vimentina Hs.297753 AI668662
1212	ADNc FLJ13618 fis de *Homo sapiens*, clon PLACE1010925 Hs.17448 AA427980:
1213	NR3C1: subfamilia 3 de receptor nuclear, grupo C, miembro 1 Hs.75772 N30428 receptor de glucocorticoides
1214	ADNc de *Homo sapiens*: FLJ21814 fis, clon HEP01068 Hs.289008 R12808
1215	BRD7: proteína que contiene bromodominio 7 Hs.279762 AA488428
1216	MAP3K8: **proteína cinasa activada por mitógenos 8 Hs.248 W42450
1217	ESTs: Hs.23213 H29336
1218	ESTs: Hs.122444 AA939019
1219	TUSP: proteína de superfamilia de Tubby Hs.102237 H78234
1220	KIAA1117: proteína KIAA1117 Hs.278398 H01516
1221	Humano: ARNm del clon 137308, cds parcial Hs.322149 H91303
1222	ESTs: Hs.130214 AA456631
1223	RAB3A: RAB3A, miembro de la familia de oncogenes de RAS Hs.27744 H14230
1224	AA598795: proteína fosfatasa 2 (anteriormente 2A), subunidad reguladora B (P
1225	H2BFC: familia de histonas H2B, miembro C Hs.137594 AI340654
1226	CFLAR: regulador de apoptosis similar a CASP8 y FADD Hs.195175 N94588
1227	CD24: antígeno CD24 (antígeno de agrupación 4 de carcinoma de pulmón de células pequeñas) Hs.286124 H59915
1228	EST: Hs.48532 N62402
1229	CCRK: cinasa relacionada con el ciclo celular Hs.26322 H17616
1230	HECH: proteína similar a heterocromatina 1 Hs.278554 AI139106
1231	DKFZp547O146: proteína hipotética DKFZp547O146 Hs.91246 T80848
1232	ESTs: Hs.71574 AA135328
1233	HLXB9: homeosecuencia HB9 Hs.37035 AI459915
1234	AA600222:
1235	SPINK5: inhibidor de serina proteasa, tipo Kazal, 5 Hs.5476 W92134
1236	RNUT1: ARN, transportador de U 1 Hs.21577 AA447799
1237	ADNc de *Homo sapiens*: FLJ23013 fis, clon LNG00740 Hs.13075 AA464543:
1238	KIAA0063: producto génico de KIAA0063 Hs.3094 T82263
1239	DYRK2: cinasa regulada por fosforilación de tirosina (Y) de doble especificidad 2 Hs.173135 R63622
1240	R94947:
1241	ADNc FLJ14337 fis de *Homo sapiens*, clon PLACE4000494 Hs.180187 AA004903

1242	FLJ20624: proteína hipotética FLJ20624 Hs.52256 AA431909
1243	ESTs: Hs.43838 R38261
1244	FLJ23053: proteína hipotética FLJ23053 Hs.94037 R25654
1245	MGC11266: proteína hipotética MGC11266 Hs.293943 AA400456
1246	ESTs: moderadamente similar a ALU8, ENTRADA DE ADVERTENCIA DE CONTAMINACIÓN DE SECUENCIA DE SUBFAMILIA SX DE ALU HUMANA [*H. sapiens*] Hs.34174 AA126603
1247	PLAUR: activador de plasminógeno, receptor de urocinasa Hs.179657 AA147962
1248	TSG101: gen de susceptibilidad tumoral 101 Hs.118910 AA670215
1249	HCNGP: proteína reguladora transcripcional Hs.27299 AA457232
1250	KIAA0978: proteína KIAA0978 Hs.3686 AA857017
1251	ESTs: Hs.61708 AA033867
1252	ESTs: Hs.120734 AA827482
1253	ESTs: Hs.5909 AA972654
1254	CDH24: similar a cadherina 24 Hs.155912 AI668564
1255	CCND1: ciclina D1 (PRAD1: adenomatosis paratiroidea 1) Hs.82932 T77237
1256	ESTs: Hs.43148 AA284775
1257	ESTs: Hs.222566 T50982
1258	ESTs: Hs.194125 N52822
1259	EST: Hs.154621 AI138644
1260	MAN1A2: manosidasa, alfa, clase 1A, miembro 2 Hs.239114 R22905
1261	MAN2A2: manosidasa, alfa, clase 2A, miembro 2 Hs.295605 AA454175
1262	Secuencia de ADN humano del clon 967N21 en el cromosoma 20p12.3-13. Contiene el gen de CHGB para cromogranina B (secretogranina 1, SCG1), un pseudogén similar a parte de KIAA0172, el gen para una nueva proteína Hs.88959 W94690
1263	ESTs: altamente similar a CIKG, PROTEÍNA DE CANALES DE POTASIO DEPENDIENTES DE VOLTAJE HUMANA KV3.4 [*H. sapiens*] Hs.106486 H11376
1264	ARNm de HT023 de *Homo sapiens*, cds completa Hs.237225 AA169496:
1265	FLJ10339: **proteína hipotética FLJ10339 Hs.203963 H72354
1266	N66278:
1267	ESTs: Hs.6195 AA454745
1268	KIAA1404: proteína KIAA1404 Hs.200317 W72798
1269	PMAIP1: proteína inducida por forbol-12-miristato-13-acetato 1 Hs.96 AA458838 APR=gen de respuesta temprana inmediata=sensible a PMA derivado de ATL
1270	G3BP: proteína de unión a dominio SH3 de proteína de activación de Ras-GTPasa Hs.220689 AA598628
1271	ADNc de *Homo sapiens*: FLJ22807 fis, clon KAIA2887 Hs.261734 R26854:
1272	*Homo sapiens*, clon IMAGE:3535294, ARNm, cds parcial Hs.80449 T57359:

1273	CDC16: CDC16 (ciclo de división celular 16, *S. cerevisiae*, homólogo) Hs.1592 AA410559
1274	FGA: **fibrinógeno, polipéptido A alfa Hs.90765 AA026626
1275	ESTs: Hs.33446 N53560
1276	ADNc FLJ14175 fis de *Homo sapiens*, clon NT2RP2002979 Hs.288613 AA054704:
1277	ESTs: Hs.44243 AA011390
1278	Clon de ADNc de inserto de longitud completa de ARNm de *Homo sapiens* EUROIMAGE 42408 Hs.284123 R61732:
1279	ESTs: Hs.53455 AA454165
1280	FLJ11264: proteína hipotética FLJ11264 Hs.11260 AI219094
1281	MBD4: proteína de dominio de unión a metil-CpG 4 Hs.35947 AA010492
1282	FLJ11305: proteína hipotética FLJ11305 Hs.7049 N94612
1283	*Homo sapiens*, similar a producto génico de CG5057, clon MGC:5309, ARNm, cds completa Hs.13885 AA460004:
1284	ARHB: familia de genes de homólogo de ras, miembro B Hs.204354 H88963
1285	ITPR3: receptor de inositol 1,4,5-trifosfato, tipo 3 Hs.77515 AA865667
1286	HMG20B: grupo de alta movilidad 20B Hs.32317 AA775743
1287	ESTs: Hs.146276 AI214204
1288	PTPN9: proteína tirosina fosfatasa, tipo no receptora 9 Hs.147663 AA434420
1289	ARNm del clon FLB9213 PRO2474 de *Homo sapiens*, cds completa Hs.21321 AA486770:
1290	H21107:
1291	HSPC157: proteína HSPC157 Hs.279842 N20480
1292	ARNm de *Homo sapiens*; ADNc DKFZp564O2363 (del clon DKFZp564O2363) Hs.321403 AA406332:
1293	ESTs: Hs.150623 AA693532
1294	EST: Hs.188697 AA199733
1295	CLECSF2: lectina de tipo C (dependiente de calcio, dominio de reconocimiento de hidratos de carbono), miembro 2 de superfamilia (inducido por activación) Hs.85201 H11732 AICL=lectina de tipo C inducida por activación
1296	ITPR1: receptor de inositol 1,4,5-trifosfato, tipo 1 Hs.198443 AA035450
1297	CHML: similar a coroideremia (proteína de escolta de Rab 2) Hs.170129 R91881
1298	CDC42: ciclo de división celular 42 (proteína de unión a GTP, 25 kD) Hs.146409 AA668681
1299	FKBP5: **proteína de unión a FK506 5 Hs.7557 AA872767

Todas las publicaciones y solicitudes de patente mencionadas en la memoria descriptiva son indicativas del nivel de los expertos en la técnica a la que pertenece esta divulgación.

REIVINDICACIONES

1. Método para determinar la probabilidad de recidiva de cáncer o muerte específica de cáncer en un paciente que se somete a la prueba, que comprende:

(1) medir, en una muestra obtenida de dicho paciente que se somete a la prueba, los niveles de expresión de un panel de genes que comprende (i) al menos tres genes de prueba seleccionados del panel F y (ii) uno o más genes de prueba seleccionados de: KLK3, STX4, TAF5L, GTPB5, SIRT3, EIF3D, GTPBP5, LOC100126784, ABCG1, CYP1B1-AS1, CECR6, SLC45A3, ACPP, TRPM8 y MSMB;

(2) proporcionar una puntuación de expresión de prueba mediante (a) ponderar la expresión determinada de cada gen en dicho panel de genes con un coeficiente predefinido, y (b) combinar la expresión ponderada de cada gen en dicho panel de genes para proporcionar dicha puntuación de expresión de prueba, en el que dichos genes de prueba se ponderan para contribuir al menos el 25% en dicha puntuación de expresión de prueba;

(3) comparar dicha puntuación de expresión de prueba con una puntuación de expresión de referencia calculada a partir de niveles de expresión medidos del panel de genes en una población de referencia de pacientes; y

(4) diagnosticar que dicho paciente que se somete a la prueba tiene o bien (a) una probabilidad aumentada de recidiva de cáncer o muerte específica de cáncer si dicha puntuación de expresión de prueba supera una primera puntuación de expresión de referencia o bien (b) una probabilidad disminuida de recidiva de cáncer o muerte específica de cáncer si dicha puntuación de expresión de prueba no supera una segunda puntuación de expresión de referencia; o bien (c) ni probabilidad aumentada ni disminuida de recidiva de cáncer o muerte específica de cáncer si dicha puntuación de expresión de prueba supera dicha segunda puntuación de expresión de referencia pero no supera dicha primera puntuación de expresión de referencia,

en el que el cáncer comprende cáncer de próstata, cáncer de pulmón, cáncer de vejiga, cáncer de cerebro o cáncer de mama.

2. Método según la reivindicación 1, en el que dichos genes de prueba se ponderan para contribuir al menos el 30%, el 40%, el 50%, el 60%, el 70%, el 80%, el 90%, el 95%, el 96%, el 97%, el 98%, el 99% o el 100% de la ponderación total dada a la expresión de la totalidad de dichos genes de prueba en dicha puntuación de expresión de prueba.

3. Método según la reivindicación 1 ó 2, en el que dichos genes de prueba comprenden KLK3 y la expresión de KLK3 se incorpora en dicha puntuación de expresión de prueba de tal manera que una expresión de KLK3 disminuida aumenta dicha puntuación.

4. Método según una cualquiera de las reivindicaciones 1 a 3, en el que dichos genes de prueba comprenden uno o más genes seleccionados de KLK3, STX4, TAF5L, GTB5, SIRT3, EIF3D, GTPBP5, LOC100126784, ABCG1, CYPB1-AS1, CECR6, SLC45A3, ACPP, TRPM8 y MSMB, en el que el estado anómalo de uno o más de dichos genes indica una alta probabilidad de recidiva de cáncer o muerte específica de cáncer.

5. Método según una cualquiera de las reivindicaciones 1 a 4, en el que dicha etapa de medición comprende:

medir la cantidad de ARNm de panel en dicha muestra transcrito a partir de cada uno de (i) los al menos tres genes de prueba seleccionados del panel F y (ii) el uno o más genes seleccionados de: KLK3, STX4, TAF5L, GTPB5, SIRT3, EIF3D, GTPBP5, LOC100126784, ABCG1, CYP1B1-AS1, CECR6, SLC45A3, ACPP, TRPM8 y MSMB, o medir la cantidad de ADNc transcrito de manera inversa a partir de dicho ARNm de panel; y

medir la cantidad de ARNm de mantenimiento en dicha muestra transcrito a partir de uno o más genes de mantenimiento, o medir la cantidad de ADNc transcrito de manera inversa a partir de dicho ARNm de mantenimiento.

6. Método según una cualquiera de las reivindicaciones 1 a 5, en el que dichas puntuaciones de expresión de referencia primera y segunda son iguales.

7. Método según una cualquiera de las reivindicaciones 1 a 6, en el que un tercio de los pacientes con cáncer en una población de referencia tienen una puntuación de expresión que supera dicha primera puntuación de expresión de referencia y un tercio de los pacientes con cáncer en dicha población de referencia tienen una puntuación de expresión que no supera dicha segunda puntuación de expresión de referencia.

8. Método según cualquiera de las reivindicaciones 1 a 7, en el que la recidiva de cáncer se elige del grupo

que consiste en metástasis a distancia del cáncer primario; metástasis local del cáncer primario; recidiva del cáncer primario; progresión del cáncer primario; y desarrollo de enfermedad metastásica localmente avanzada.

9. Método para determinar la probabilidad de recidiva de cáncer o muerte específica de cáncer en un paciente con cáncer, que comprende:

(1) medir, en una muestra obtenida de dicho paciente, los niveles de expresión de un panel de genes que comprende (i) al menos tres genes de prueba seleccionados del panel F y (ii) uno o más genes de prueba seleccionados de: KLK3, STX4, TAF5L, GTPB5, SIRT3, EIF3D, GTPBP5, LOC100126784, ABCG1, CYP1B1-AS1, CECR6, SLC45A3, ACPP, TRPM8 y MSMB;

(2) proporcionar una puntuación de expresión de prueba mediante (1) ponderar la expresión determinada de cada gen en dicho panel de genes con un coeficiente predefinido, y (2) combinar la expresión ponderada para proporcionar dicha puntuación de expresión de prueba, en el que dichos genes de prueba se ponderan para contribuir al menos el 25% en dicha puntuación de expresión de prueba;

(3) proporcionar una puntuación de pronóstico de prueba combinando dicha puntuación de expresión de prueba con al menos una puntuación clínica de prueba que representa al menos una variable clínica;

(4) comparar dicha puntuación de pronóstico de prueba con una puntuación de pronóstico de referencia calculada a partir de niveles de expresión medidos del panel de genes en una población de referencia de pacientes; y

(5) diagnosticar que dicho paciente tiene o bien (a) una probabilidad aumentada de recidiva de cáncer o muerte específica de cáncer si dicha puntuación de pronóstico de prueba supera una primera puntuación de pronóstico de referencia o bien (b) probabilidad no aumentada de recidiva de cáncer o muerte específica de cáncer si dicha puntuación de pronóstico de prueba no supera una segunda puntuación de pronóstico de referencia,

en el que el cáncer comprende cáncer de próstata, cáncer de pulmón, cáncer de vejiga, cáncer de cerebro o cáncer de mama.

10. Método según la reivindicación 9, en el que dicha al menos una puntuación clínica incorpora al menos una variable clínica elegida del grupo que consiste en año de RP, márgenes quirúrgicos, extensión extracapsular, invasión de vesículas seminales, afectación ganglionar, puntuación primaria de Gleason, puntuación secundaria de Gleason o PSA preoperatorio.

11. Método según la reivindicación 9 ó 10, en el que dichas puntuaciones de pronóstico (i) incorporan (a) una primera puntuación clínica que representa la concentración de PSA preoperatorio, opcionalmente incorporada como una concentración numérica de ng/dl transformada mediante el algoritmo natural, sumando 1 para evitar valores nulos; y (b) una segunda puntuación clínica que representa la puntuación de Gleason, opcionalmente incorporada como una variable numérica continua o categorizada como < 7 (nivel de referencia), 7 o > 7; y/o (ii) se calculan según una fórmula que comprende los siguientes términos: (A x puntuación de expresión) + (B x puntuación clínica).

12. Método según la reivindicación 11(ii), en el que A = 0,58, dicha puntuación clínica es la puntuación de CAPRA y B = 0,41.

13. Fármaco antineoplásico para su uso en el tratamiento de cáncer en un paciente que se ha identificado que padece cáncer de próstata, cáncer de pulmón, cáncer de vejiga o cáncer de cerebro, y que se ha determinado que tiene un mal pronóstico mediante:

(1) determinar, en una muestra tumoral de un paciente al que se le ha diagnosticado cáncer de próstata, cáncer de pulmón, cáncer de vejiga o cáncer de cerebro, la expresión de un panel de genes en dicha muestra tumoral que incluye (i) al menos tres genes seleccionados del panel F y (ii) uno o más genes seleccionados de: KLK3, STX4, TAF5L, GTPB5, SIRT3, EIF3D, GTPBP5, LOC100126784, ABCG1, CYP1B1-AS1, CECR6, SLC45A3, ACPP, TRPM8 y MSMB; y

(2) proporcionar un valor de prueba mediante (1) ponderar la expresión determinada de cada uno de los genes de dicho panel de genes con un coeficiente predefinido, y (2) combinar la expresión ponderada para proporcionar dicho valor de prueba, en el que un nivel aumentado de expresión de dichos genes indica un mal pronóstico.

14. Kit de diagnóstico para pronosticar cáncer en un paciente al que se le ha diagnosticado cáncer de próstata, cáncer de pulmón, cáncer de vejiga o cáncer de cerebro, que comprende, en un recipiente

compartimentalizado:

(1) una pluralidad de pares de cebadores de PCR para la amplificación por PCR de al menos 5 genes de prueba que comprenden (i) al menos tres genes seleccionados del panel F y (ii) uno o más genes seleccionados de: KLK3, STX4, TAF5L, GTPB5, SIRT3, EIF3D, GTPBP5, LOC100126784, ABCG1, CYP1B1-AS1, CECR6, SLC45A3, ACPP, TRPM8 y MSMB, o una pluralidad de sondas para hibridarse con al menos 5 genes de prueba que comprenden (i) al menos tres genes seleccionados del panel F y (ii) uno o más genes seleccionados de: KLK3, STX4, TAF5L, GTPB5, SIRT3, EIF3D, GTPBP5, LOC100126784, ABCG1, CYP1B1-AS1, CECR6, SLC45A3, ACPP, TRPM8 y MSMB, en condiciones de hibridación rigurosas, en el que menos del 10%, el 30% o menos del 40% de la totalidad de dichos al menos 5 genes de prueba no se seleccionan del panel F; y

(2) uno o más pares de cebadores de PCR para la amplificación por PCR de al menos un gen de mantenimiento o una o más sondas para hibridarse con al menos un gen de mantenimiento.

15. Uso de

(1) una pluralidad de (i) pares de cebadores de PCR adecuados para la amplificación por PCR de genes que comprenden (i) al menos cuatro genes seleccionados del panel F y (ii) uno o más genes seleccionados de: KLK3, STX4, TAF5L, GTPB5, SIRT3, EIF3D, GTPBP5, LOC100126784, ABCG1, CYP1B1-AS1, CECR6, SLC45A3, ACPP, TRPM8 y MSMB, o (ii) sondas para hibridarse con genes que comprenden (i) al menos cuatro genes seleccionados del panel F y (ii) uno o más genes seleccionados de: KLK3, STX4, TAF5L, GTPB5, SIRT3, EIF3D, GTPBP5, LOC100126784, ABCG1, CYP1B1-AS1, CECR6, SLC45A3, ACPP, TRPM8 y MSMB, en condiciones de hibridación rigurosas; y

(2) uno o más (i) pares de cebadores de PCR adecuados para la amplificación por PCR de al menos un gen de mantenimiento o (ii) sondas para hibridarse con al menos un gen de mantenimiento en condiciones de hibridación rigurosas,

para la fabricación de un producto de diagnóstico para determinar la expresión de dichos genes de prueba en una muestra tumoral de un paciente al que se le ha diagnosticado cáncer de próstata, cáncer de pulmón, cáncer de vejiga o cáncer de cerebro, para predecir el pronóstico de cáncer, en el que un nivel aumentado de dicha expresión indica un mal pronóstico o una probabilidad aumentada de recidiva de cáncer en el paciente.

16. Método según cualquiera de las reivindicaciones 1 a 12, en el que la expresión de KLK3 se incorpora en dicha puntuación de expresión de prueba de tal manera que una expresión de KLK3 disminuida se correlaciona con una puntuación de expresión de prueba que da lugar a un diagnóstico de probabilidad aumentada de recidiva de cáncer o muerte específica de cáncer.

17. Método según la reivindicación 16, en el que dicha puntuación de expresión de prueba incorpora el negativo del valor numérico de expresión de KLK3 de tal manera que una mayor puntuación de expresión de prueba da lugar a un diagnóstico de probabilidad aumentada de recidiva de cáncer o muerte específica de cáncer.

18. Método implementado por ordenador para evaluar la clasificación de riesgo de cáncer de próstata según la AUA de un paciente, que comprende

(1) obtener la clasificación de riesgo de cáncer de próstata según la AUA de dicho paciente;

(2) proporcionar un valor de prueba mediante

(a) ponderar la expresión determinada de cada uno de una pluralidad de genes de prueba seleccionados de un panel de biomarcadores que comprende (i) al menos tres genes de prueba seleccionados del panel F y (ii) uno o más genes de prueba seleccionados de: KLK3, STX4, TAF5L, GTPB5, SIRT3, EIF3D, GTPBP5, LOC100126784, ABCG1, CYP1B1-AS1, CECR6, SLC45A3, ACPP, TRPM8 y MSMB con un coeficiente predefinido; y

(b) combinar la expresión ponderada para proporcionar el valor de prueba, en el que la ponderación combinada dada a los genes seleccionados del panel F es al menos el 40% de la ponderación total dada a la expresión de dicha pluralidad de genes de prueba; y

(3) (a) reclasificar que dicho paciente tiene un riesgo mayor que el indicado por la clasificación según la AUA si dicho valor de prueba es mayor que la mediana de valores de prueba para todos los pacientes en una población de referencia que tiene la misma clasificación según la AUA; o

(4) (b) reclasificar que dicho paciente tiene un riesgo menor que el indicado por la clasificación según la

AUA si dicho valor de prueba es menor que la mediana de valores de prueba para todos los pacientes en una población de referencia que tiene la misma clasificación según la AUA; o

(5) (c) confirmar la clasificación según la AUA de dicho paciente si dicho valor de prueba es sustancialmente igual que la mediana de valores de prueba para todos los pacientes en una población de referencia que tiene la misma clasificación según la AUA.

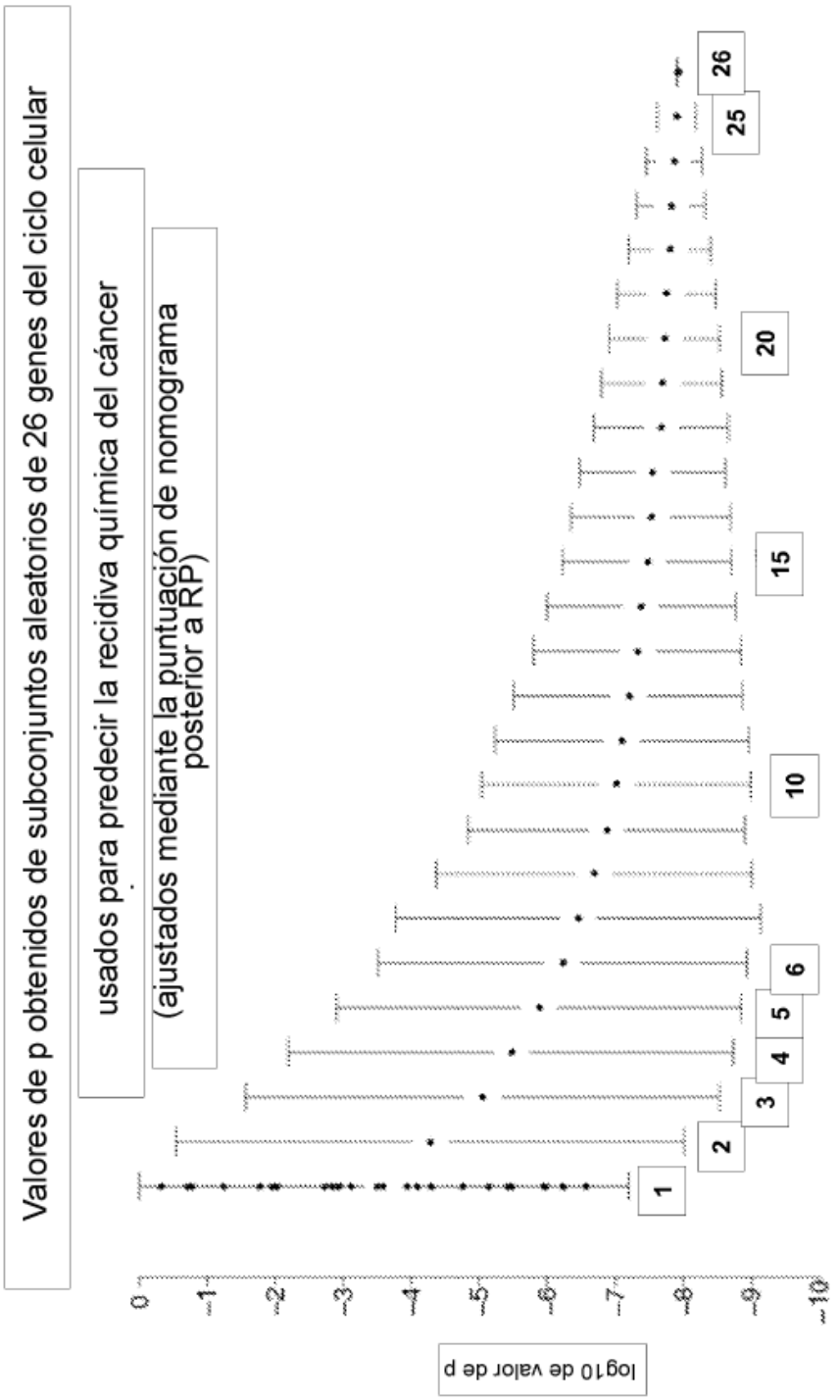

Valores de p obtenidos de subconjuntos aleatorios de 26 genes del ciclo celular
usados para predecir la recidiva química del cáncer
(ajustados mediante la puntuación de nomograma posterior a RP)
log10 de valor de p
0
−1
−2
−3
−4
−5
−6
−7
−8
−9
−10
1
2
3
4
5
6
10
15
20
25
26
Análisis de RP de Cox multivariante del tiempo hasta la recidiva tras la cirugía de próstata
100 muestras FFPE
Expresión de ARNm mediante TLDA
1000 simulaciones aleatorias extraídas de 26 CCG

Figura 1

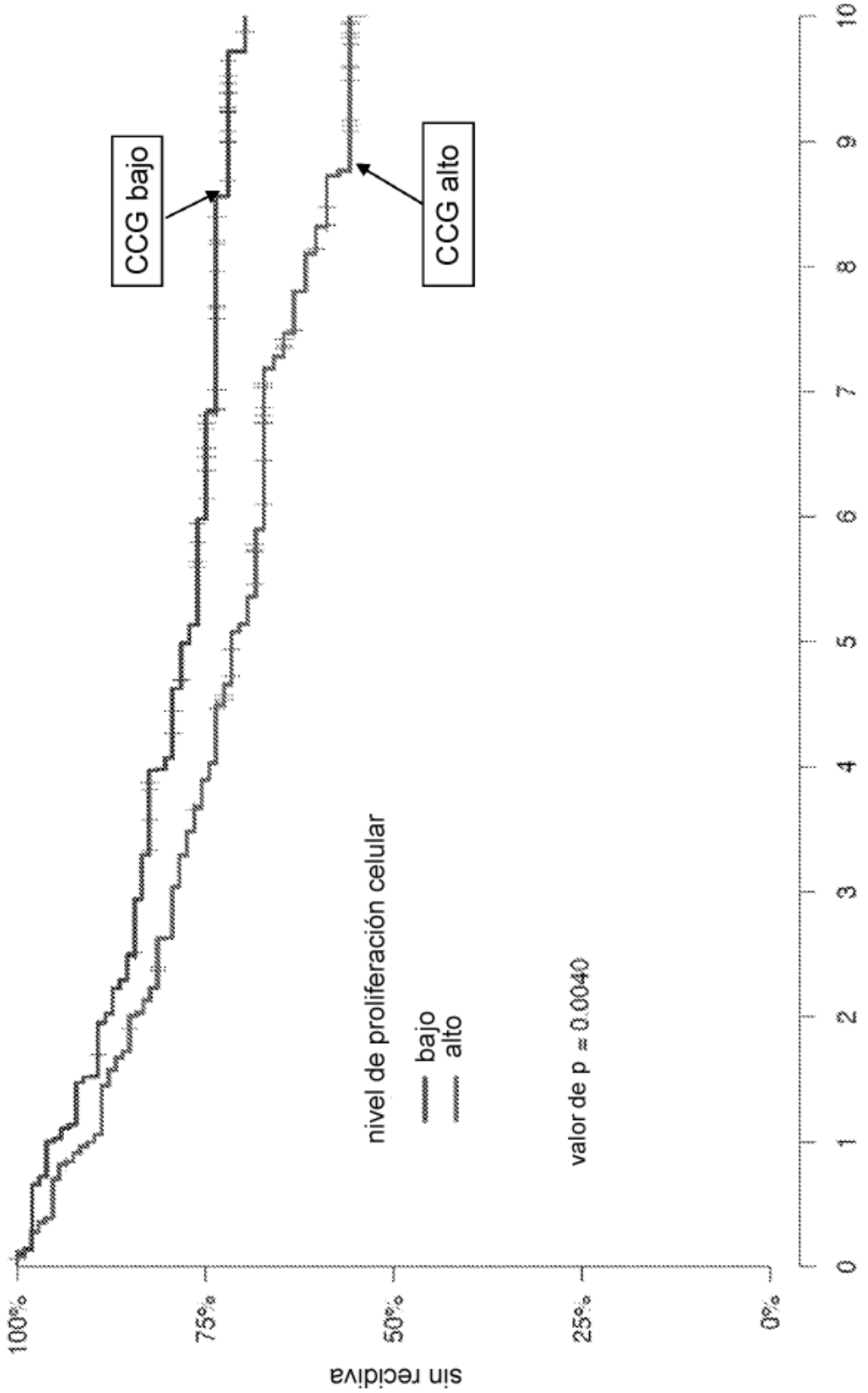

CCG bajo
CCG alto
nivel de proliferación celular
bajo
alto
valor de p = 0.0040
sin recidiva
100%
75%
50%
25%
0%
0
1
2
3
4
5
6
7
8
9
10
años hasta la recidiva

Figura 2

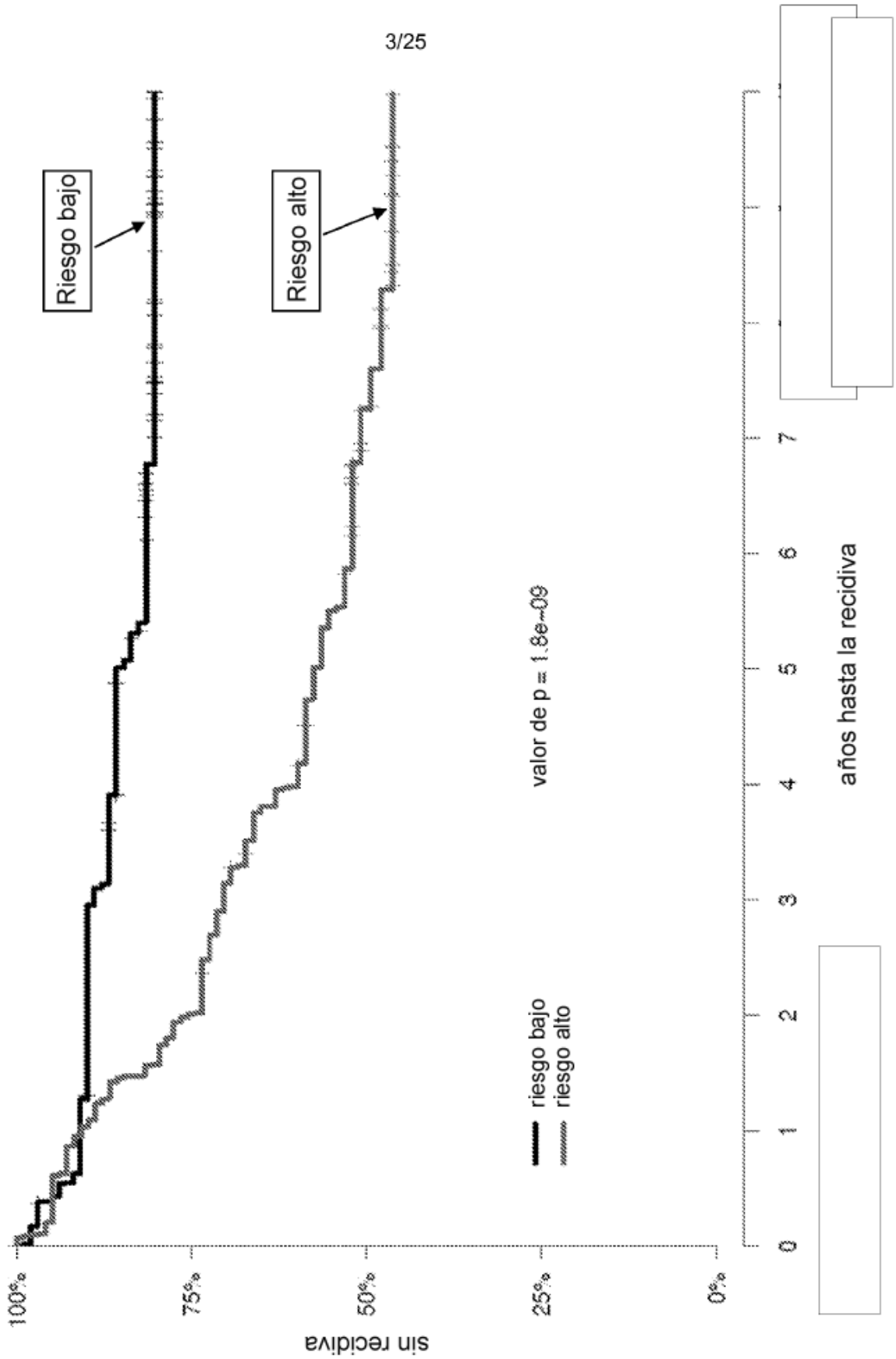

3/25
Riesgo bajo
Riesgo alto
sin recidiva
100%
75%
50%
25%
0%
valor de p = 1.8e-09
riesgo bajo
riesgo alto
0
1
2
3
4
5
6
7
años hasta la recidiva

Figura 3

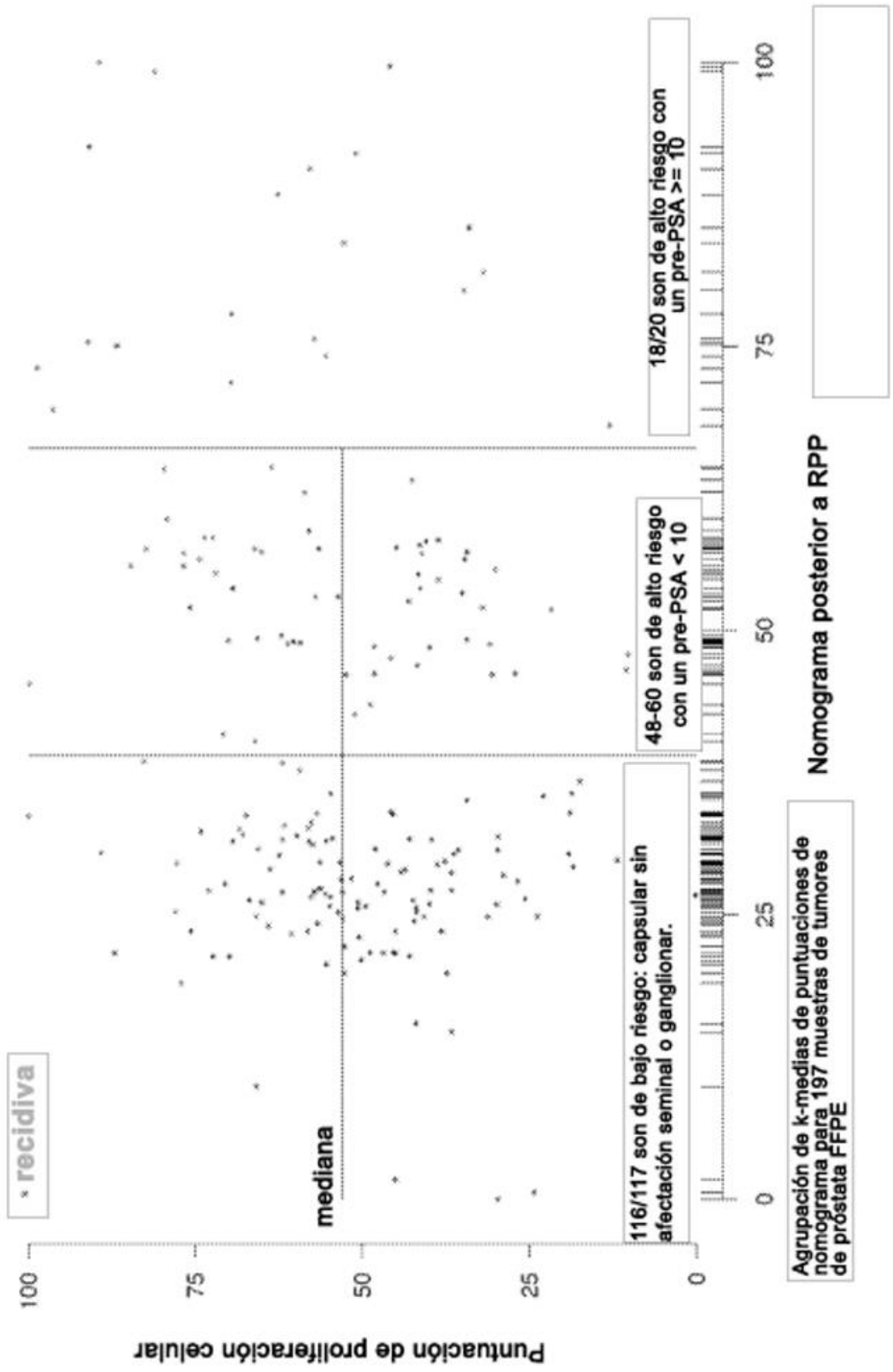

Puntuación de proliferación celular
100
75
50
25
0
× recidiva
mediana
116/117 son de bajo riesgo: capsular sin afectación seminal o ganglionar.
48-60 son de alto riesgo con un pre-PSA < 10
18/20 son de alto riesgo con un pre-PSA >= 10
0
25
50
75
100
Nomograma posterior a RPP
Agrupación de k-medias de puntuaciones de nomograma para 197 muestras de tumores de próstata FFPE

Figura 4

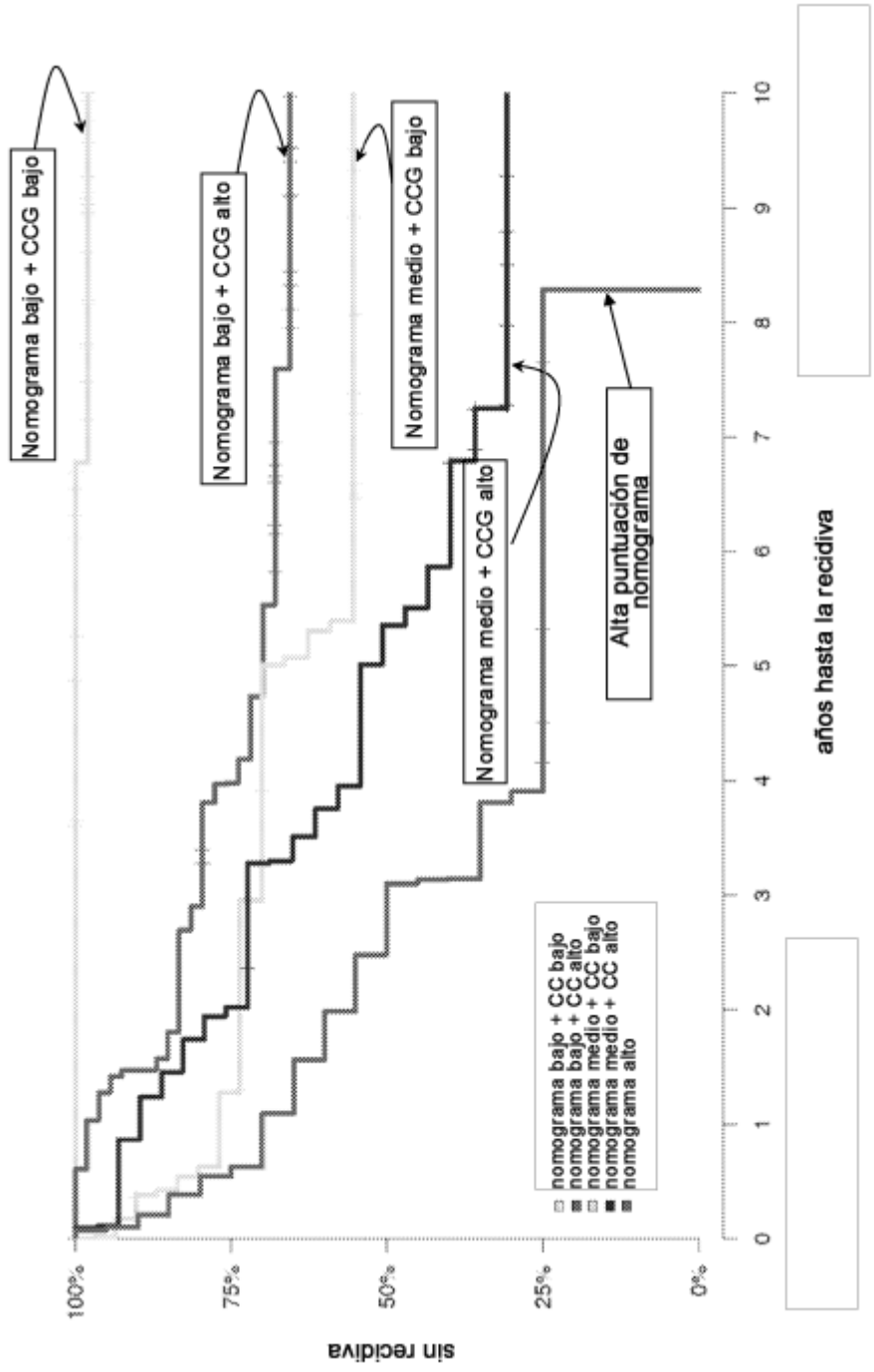

Nomograma bajo + CCG bajo
Nomograma bajo + CCG alto
Nomograma medio + CCG bajo
Nomograma medio + CCG alto
Alta puntuación de nomograma
nomograma bajo + CC bajo
nomograma bajo + CC alto
nomograma medio + CC bajo
nomograma medio + CC alto
nomograma alto
100%
75%
50%
25%
0%
sin recidiva
0
1
2
3
4
5
6
7
8
9
10
años hasta la recidiva

Figura 5

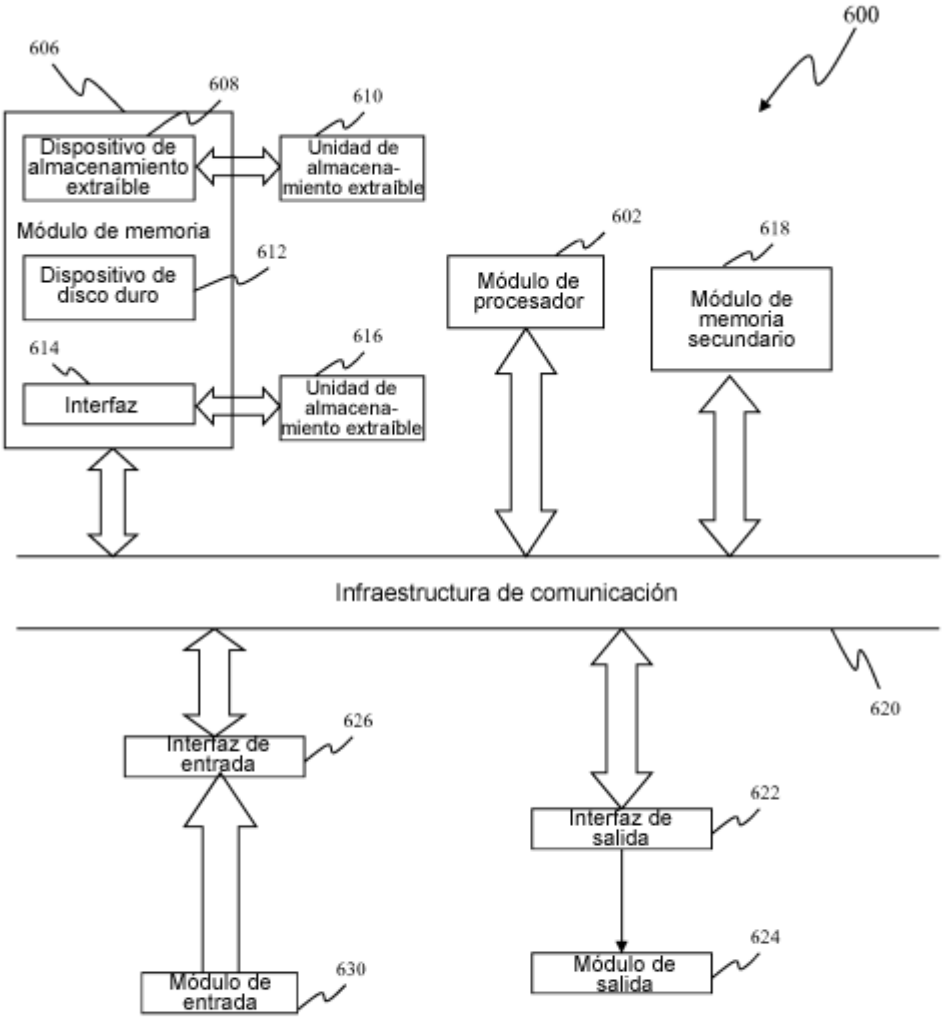

600
606
608
610
Dispositivo de almacenamiento extraíble
Unidad de almacena-miento extraíble
Módulo de memoria
612
Dispositivo de disco duro
602
Módulo de procesador
618
Módulo de memoria secundario
614
616
Interfaz
Unidad de almacena-miento extraíble
Infraestructura de comunicación
620
626
Interfaz de entrada
622
Interfaz de salida
624
Módulo de salida
630
Módulo de entrada

Figura 6

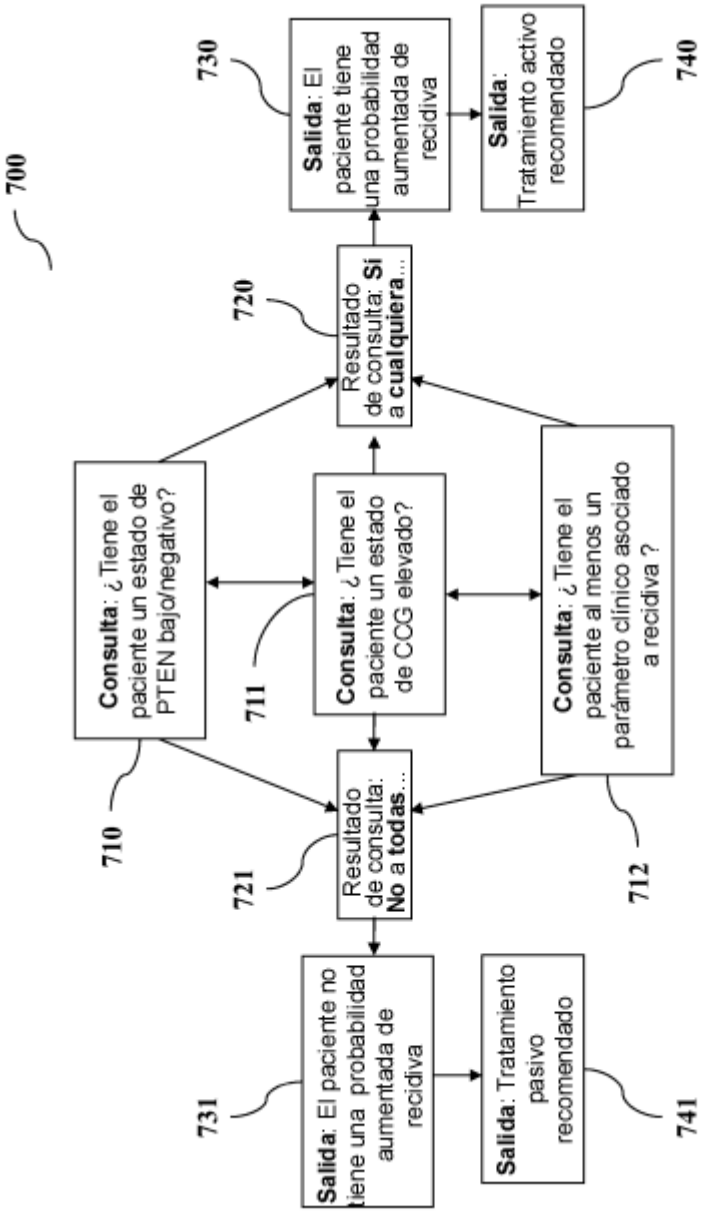

700
Consulta: ¿Tiene el paciente un estado de PTEN bajo/negativo?
710
Consulta: ¿Tiene el paciente un estado de CCG elevado?
711
Consulta: ¿Tiene el paciente al menos un parámetro clínico asociado a recidiva ?
712
Resultado de consulta: Si a cualquiera...
720
Resultado de consulta: No a todas...
721
Salida: El paciente tiene una probabilidad aumentada de recidiva
730
Salida: El paciente no tiene una probabilidad aumentada de recidiva
731
Salida: Tratamiento activo recomendado
740
Salida: Tratamiento pasivo recomendado
741

Figura 7

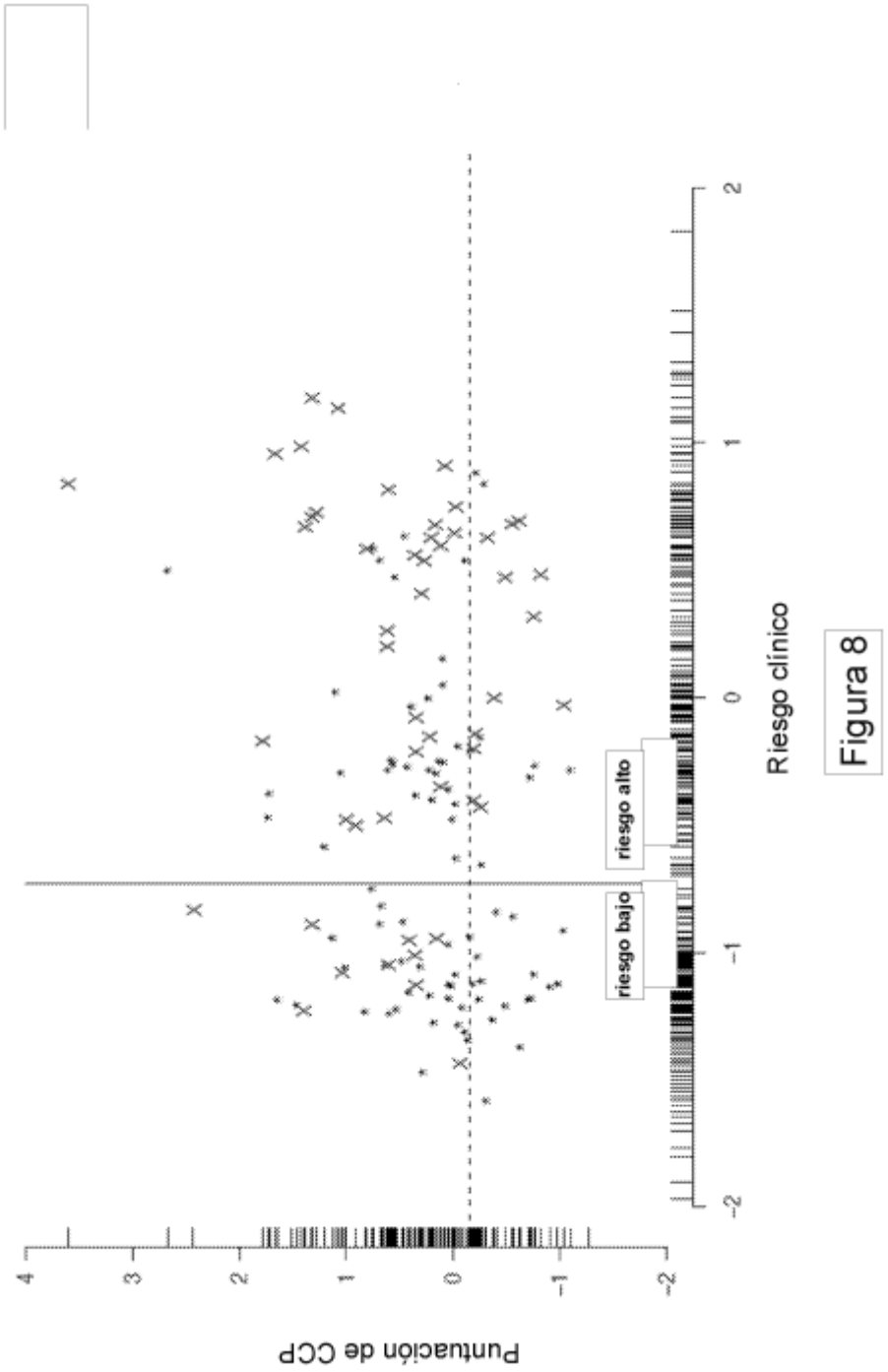

Puntuación de CCP
4
3
2
1
0
−1
−2
riesgo bajo
riesgo alto
Riesgo clínico
−2
−1
0
1
2

Figura 8

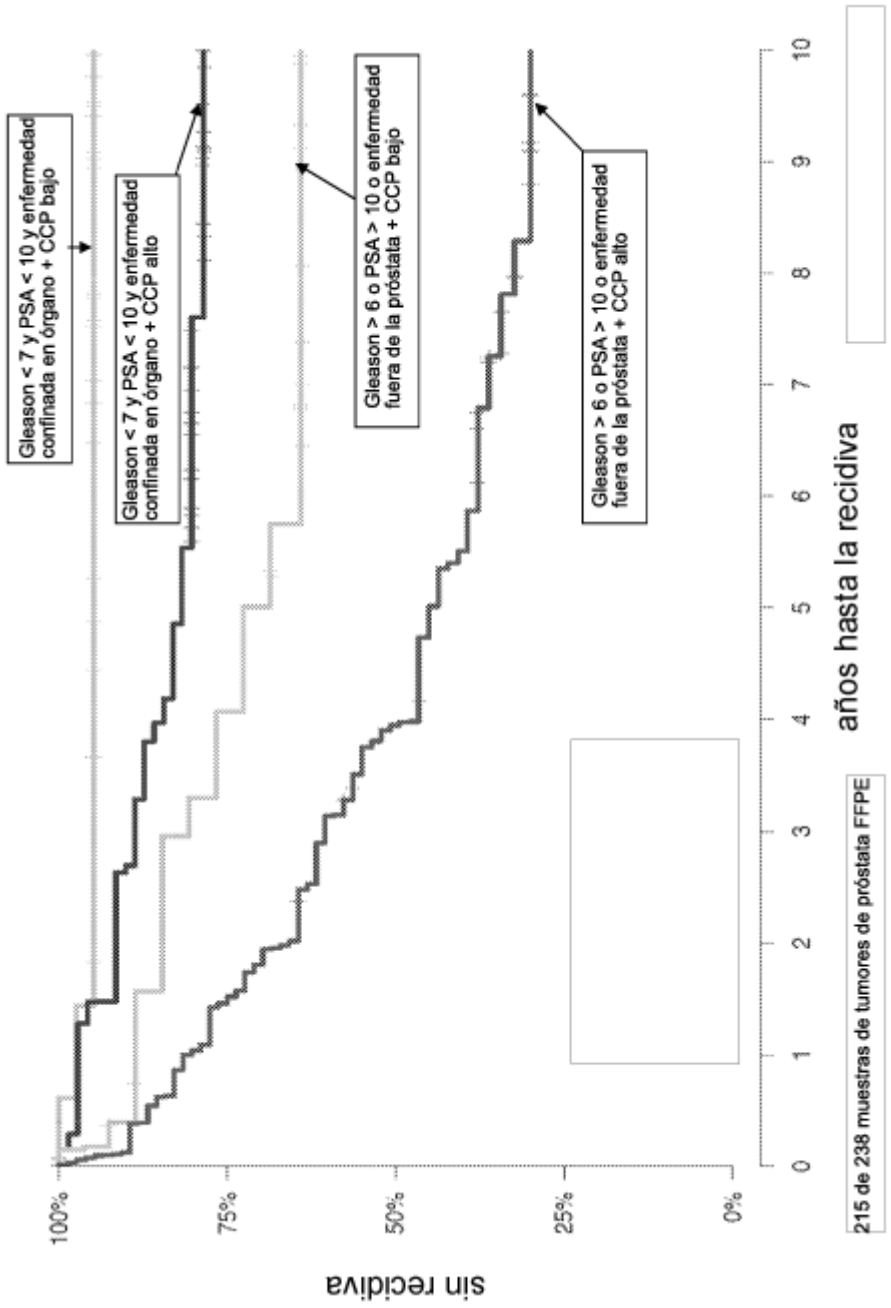

Gleason < 7 y PSA < 10 y enfermedad confinada en órgano + CCP bajo
Gleason < 7 y PSA < 10 y enfermedad confinada en órgano + CCP alto
Gleason > 6 o PSA > 10 o enfermedad fuera de la próstata + CCP bajo
Gleason > 6 o PSA > 10 o enfermedad fuera de la próstata + CCP alto
sin recidiva
100%
75%
50%
25%
0%
0
1
2
3
4
5
6
7
8
9
10
años hasta la recidiva
215 de 238 muestras de tumores de próstata FFPE

Figura 9

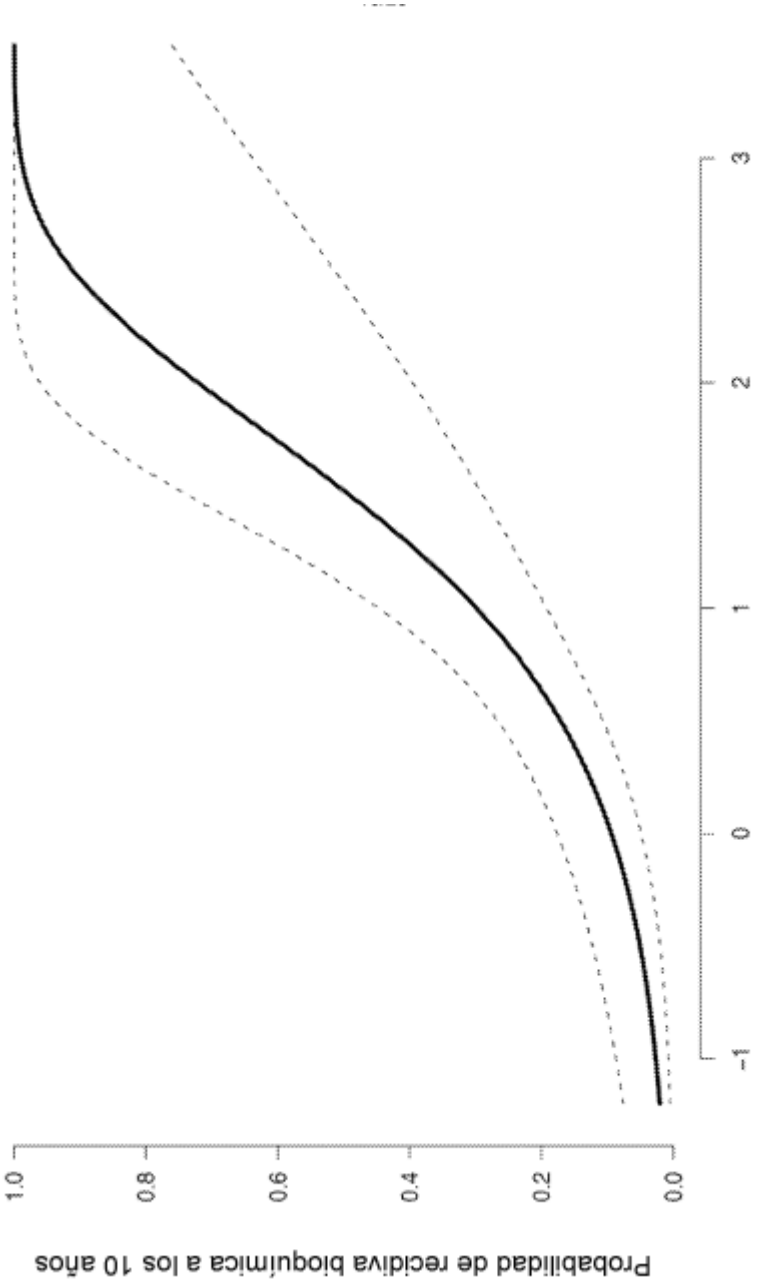

Probabilidad de recidiva bioquímica a los 10 años
1.0
0.8
0.6
0.4
0.2
0.0
−1
0
1
2
3
Puntuación de CCP

Figura 10

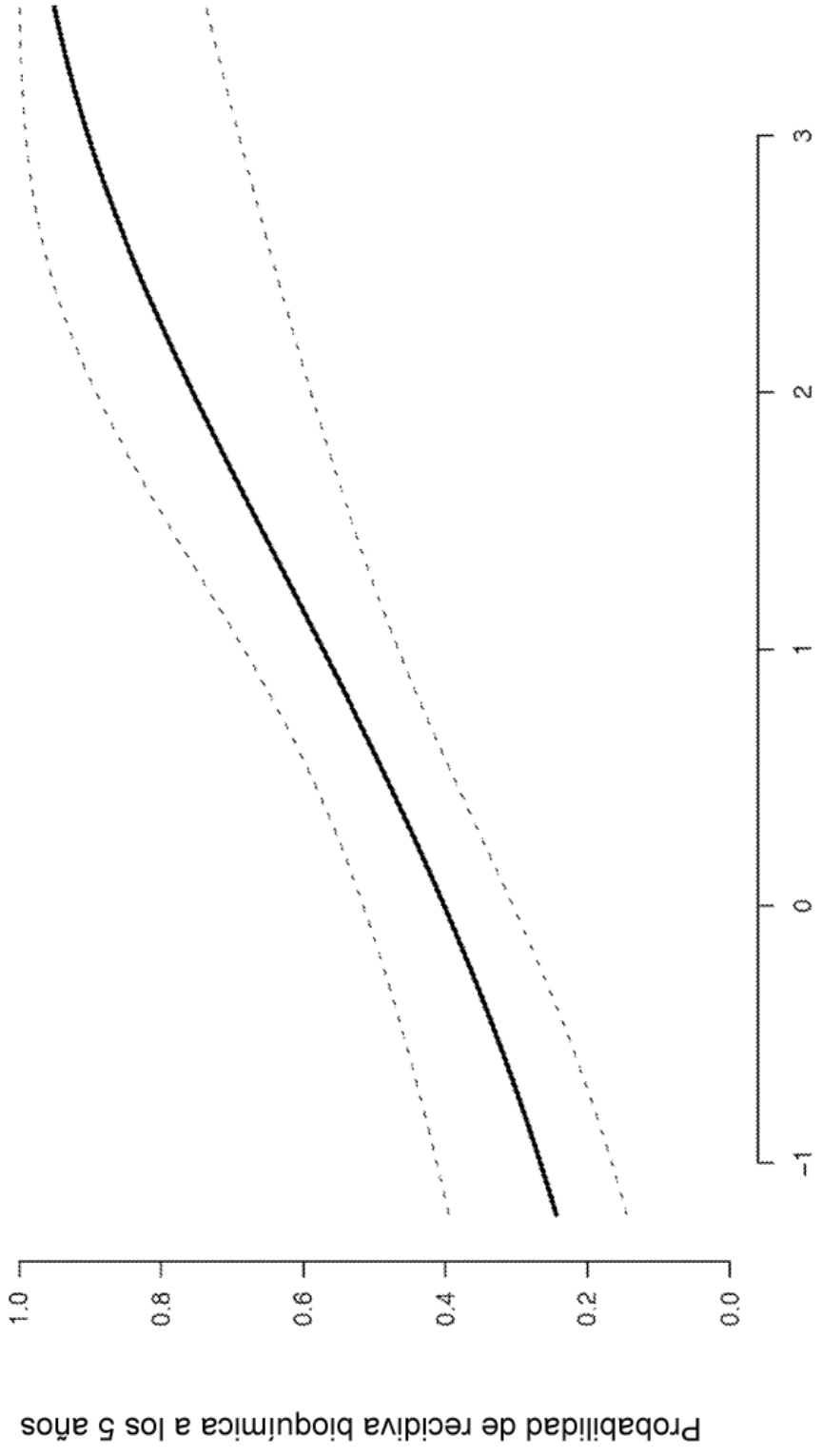

Probabilidad de recidiva bioquímica a los 5 años
1.0
0.8
0.6
0.4
0.2
0.0
−1
0
1
2
3
Puntuación de CCP

Figura 11

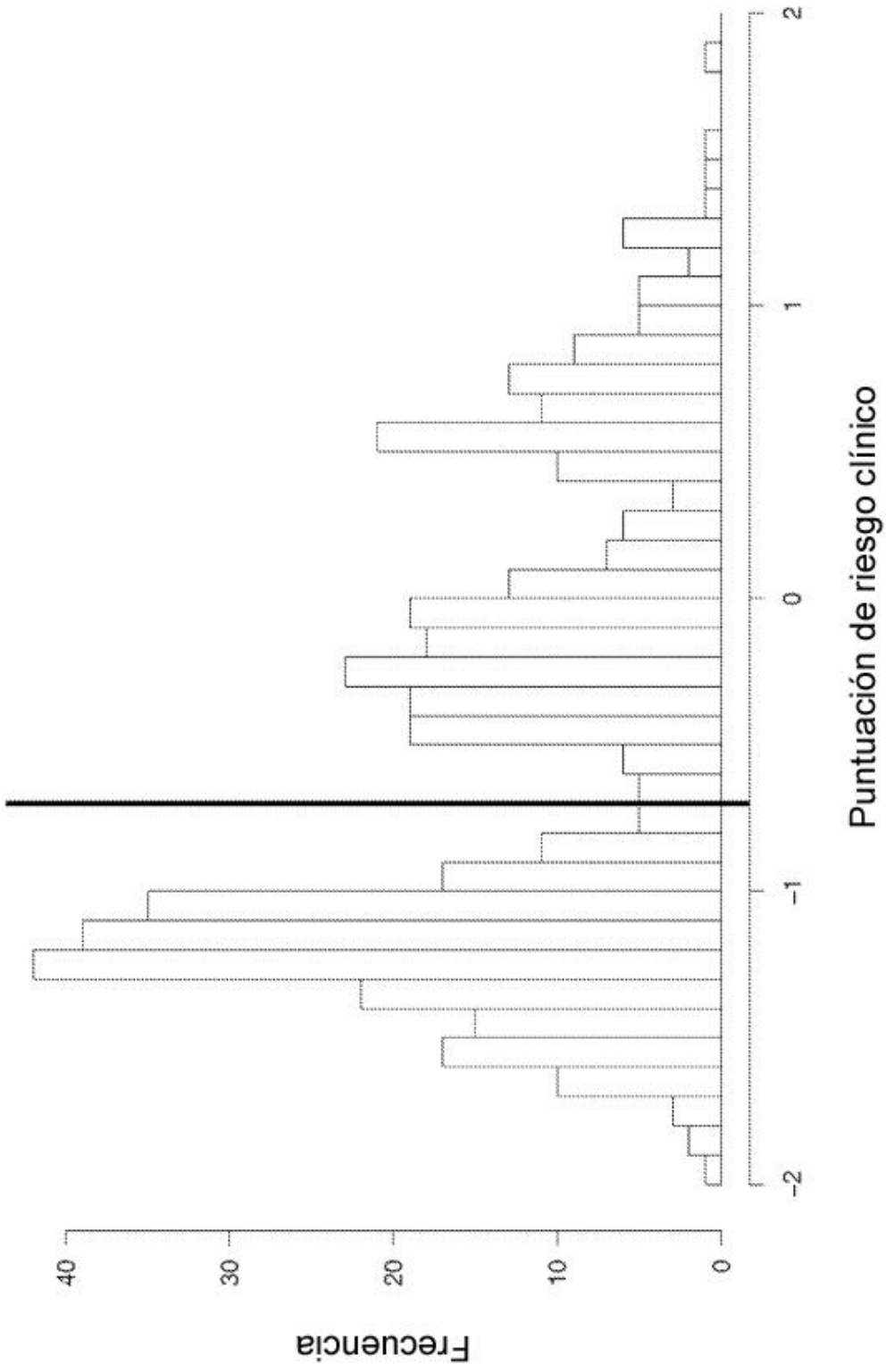


Figura 12

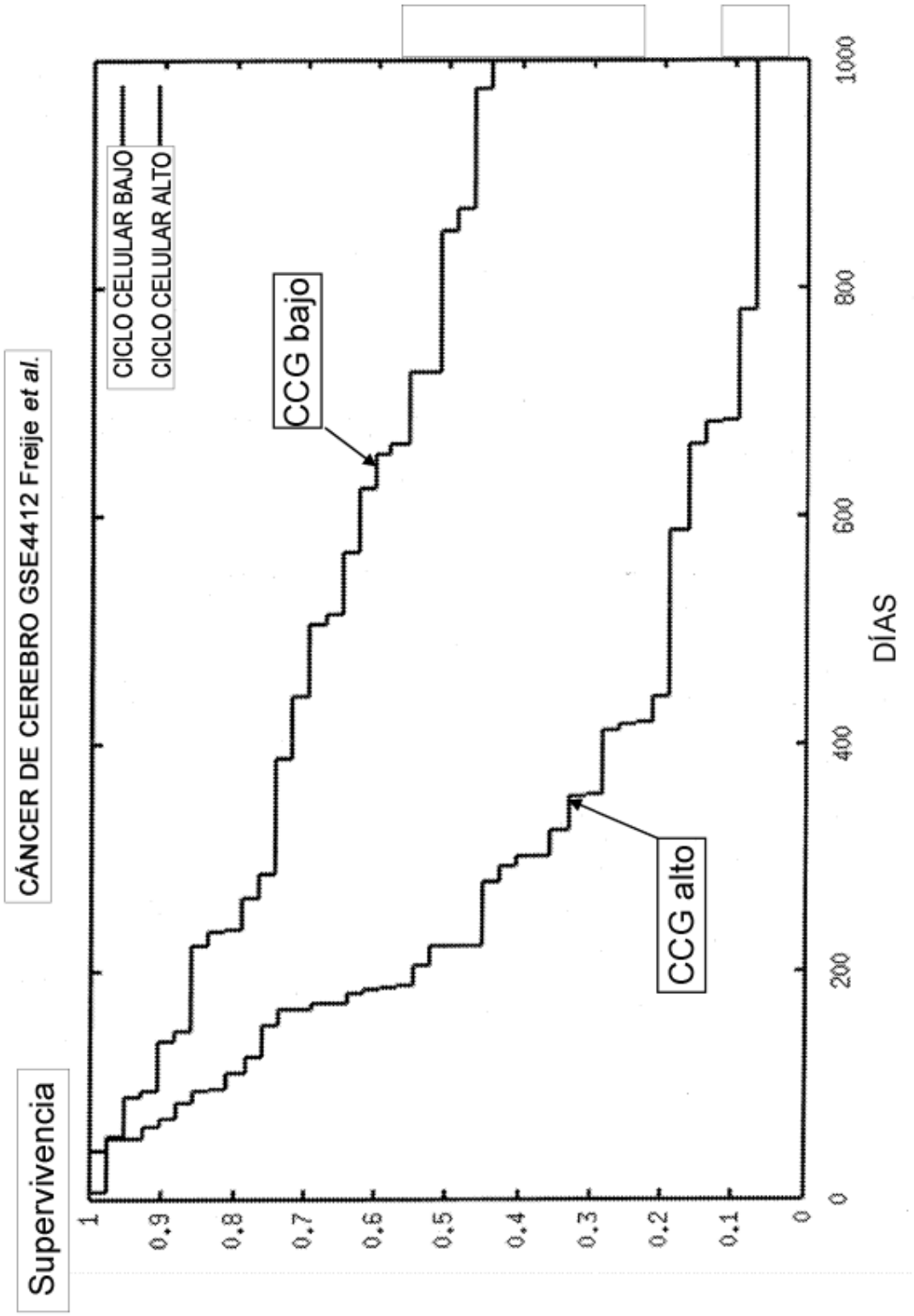

CÁNCER DE CEREBRO GSE4412 Freije et al.
CICLO CELULAR BAJO
CICLO CELULAR ALTO
CCG bajo
CCG alto
Supervivencia
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
0
200
400
600
800
1000
DÍAS

Figura 13

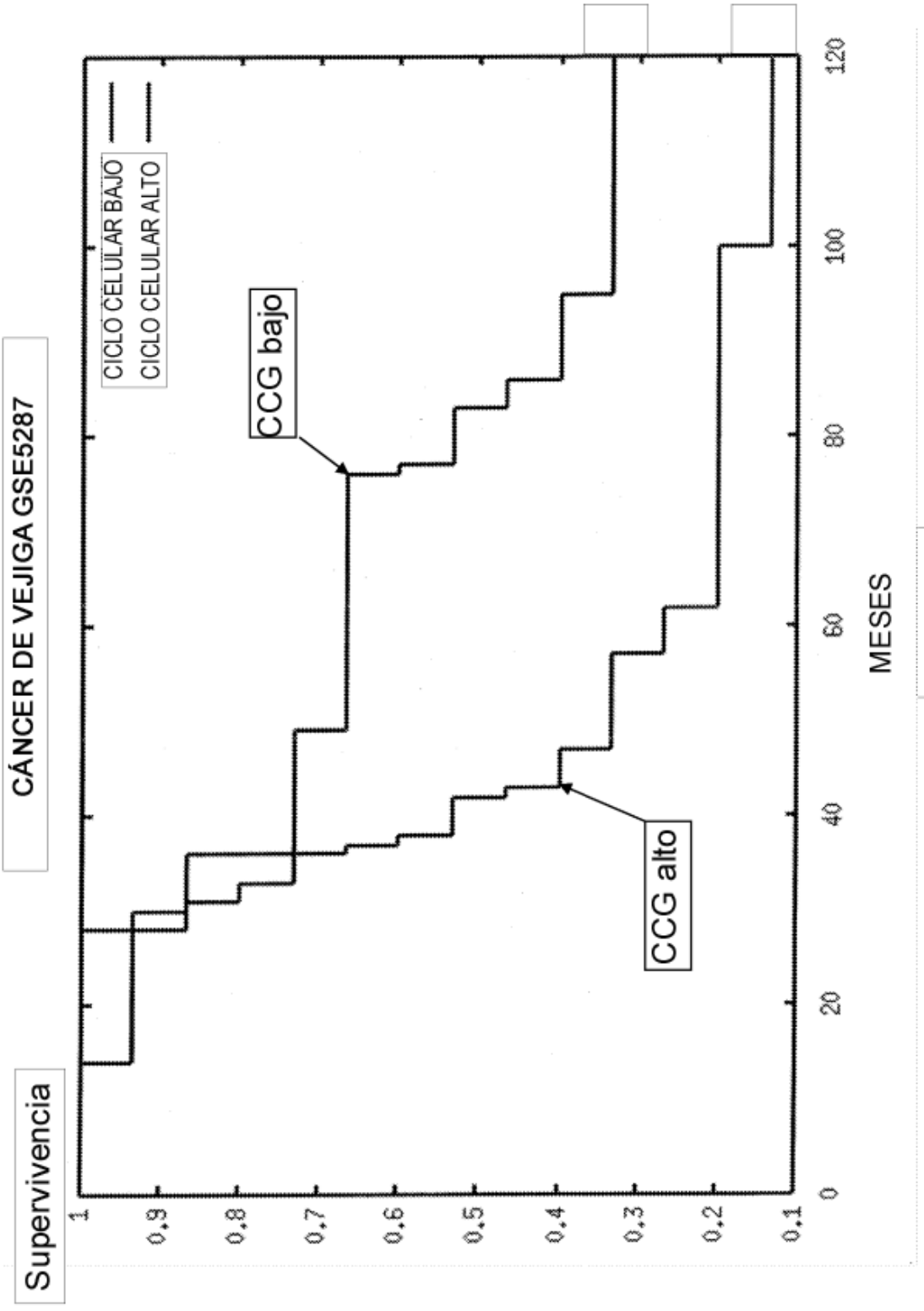

CÁNCER DE VEJIGA GSE5287
CICLO CELULAR BAJO
CICLO CELULAR ALTO
CCG bajo
CCG alto
Supervivencia
MESES
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
20
40
60
80
100
120

Figura 14

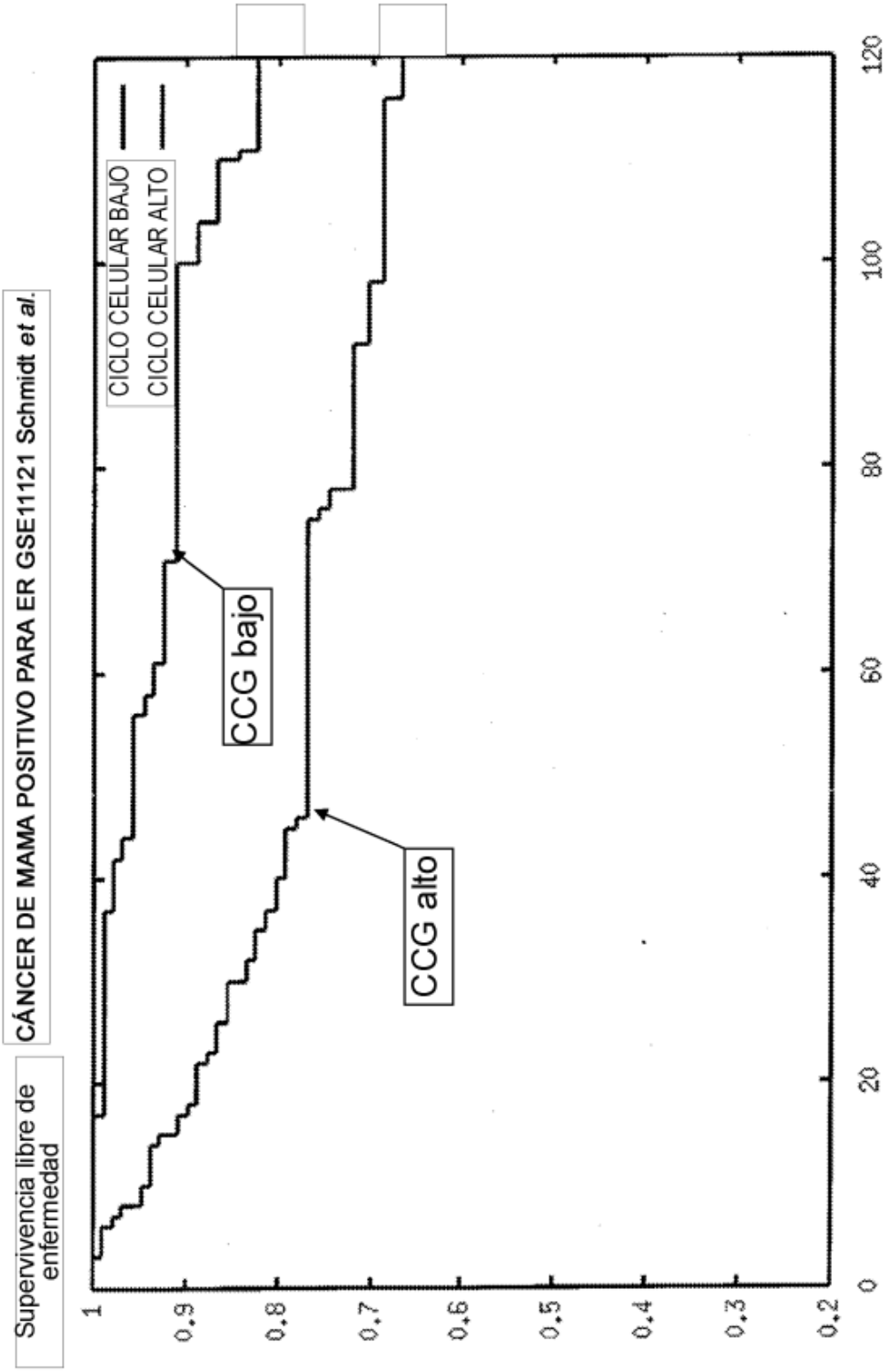

CÁNCER DE MAMA POSITIVO PARA ER GSE11121 Schmidt et al.
Supervivencia libre de enfermedad
CICLO CELULAR BAJO
CICLO CELULAR ALTO
CCG bajo
CCG alto
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0
20
40
60
80
100
120
MESES

Figura 15

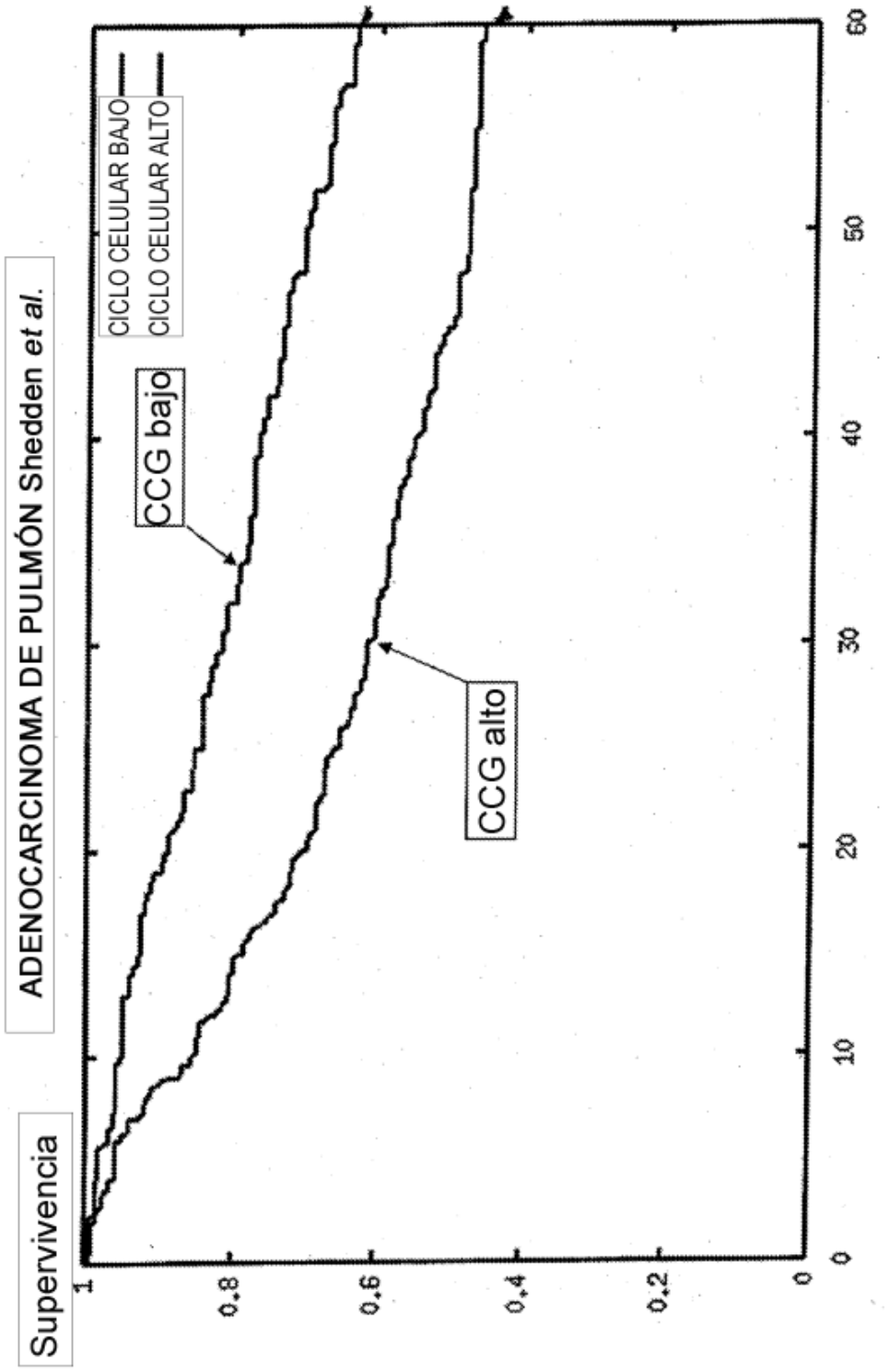

ADENOCARCINOMA DE PULMÓN Shedden et al.
Supervivencia
CICLO CELULAR BAJO
CICLO CELULAR ALTO
CCG bajo
CCG alto
1
0,8
0,6
0,4
0,2
0
0
10
20
30
40
50
60
MESES

Figura 16

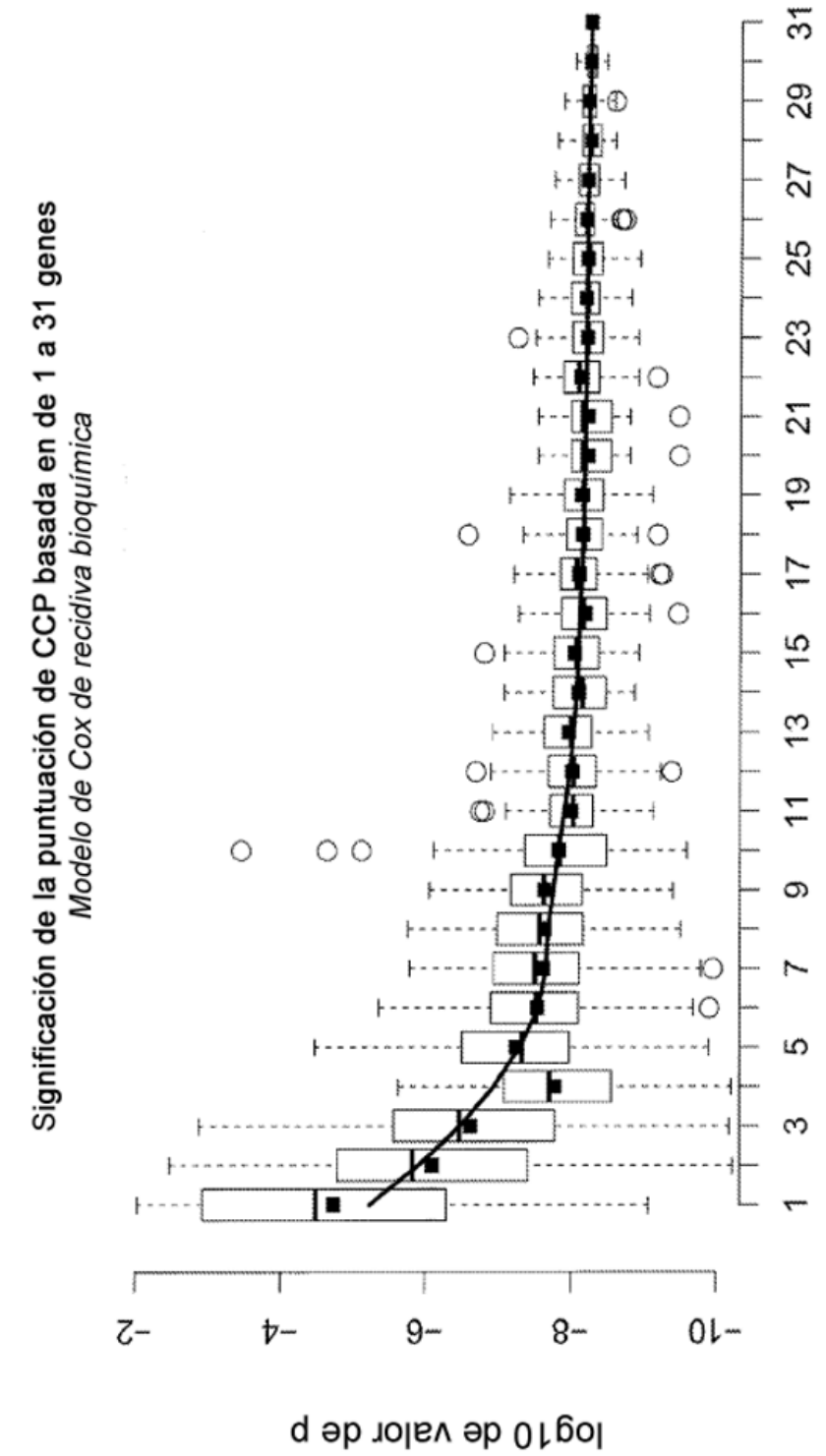

Significación de la puntuación de CCP basada en de 1 a 31 genes
Modelo de Cox de recidiva bioquímica
log10 de valor de p
−2
−4
−6
−8
−10
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
31
número de genes de CCP
Para cada n, 100 combinaciones aleatorias de tamaño n se extrajeron de los 31 genes.

Figura 17

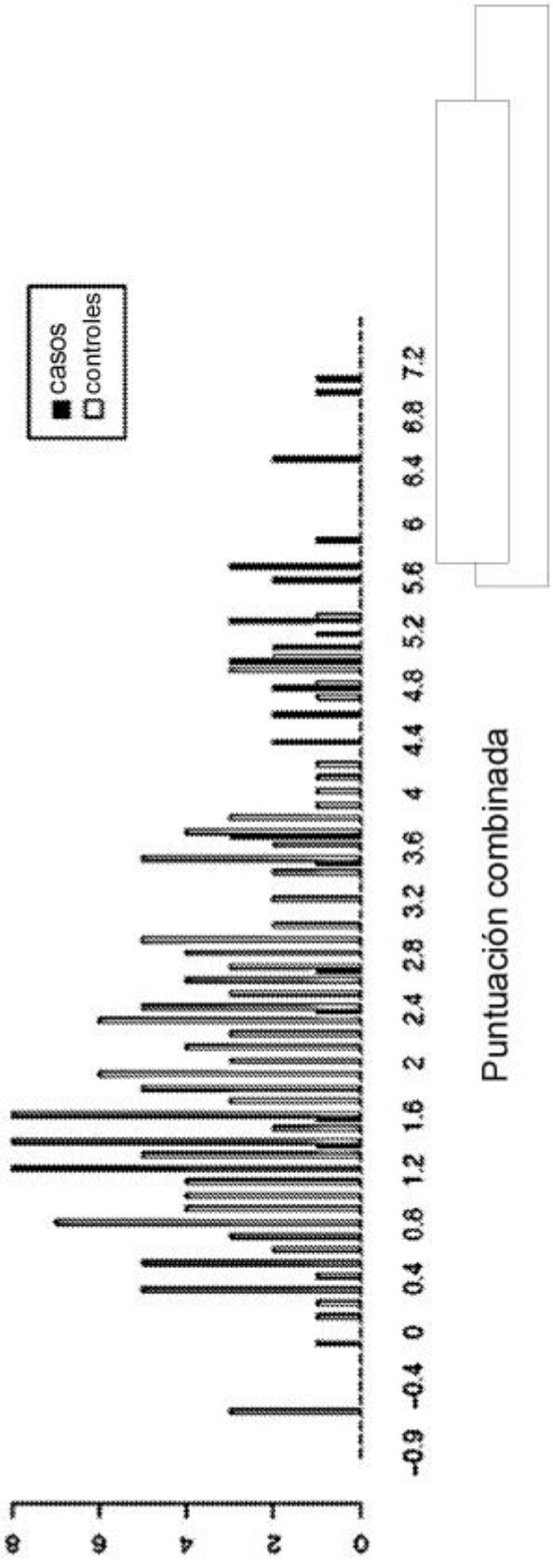

casos
controles
-0.9 -0.4 0 0.4 0.8 1.2 1.6 2 2.4 2.8 3.2 3.6 4 4.4 4.8 5.2 5.6 6 6.4 6.8 7.2
Puntuación combinada

Figura 18

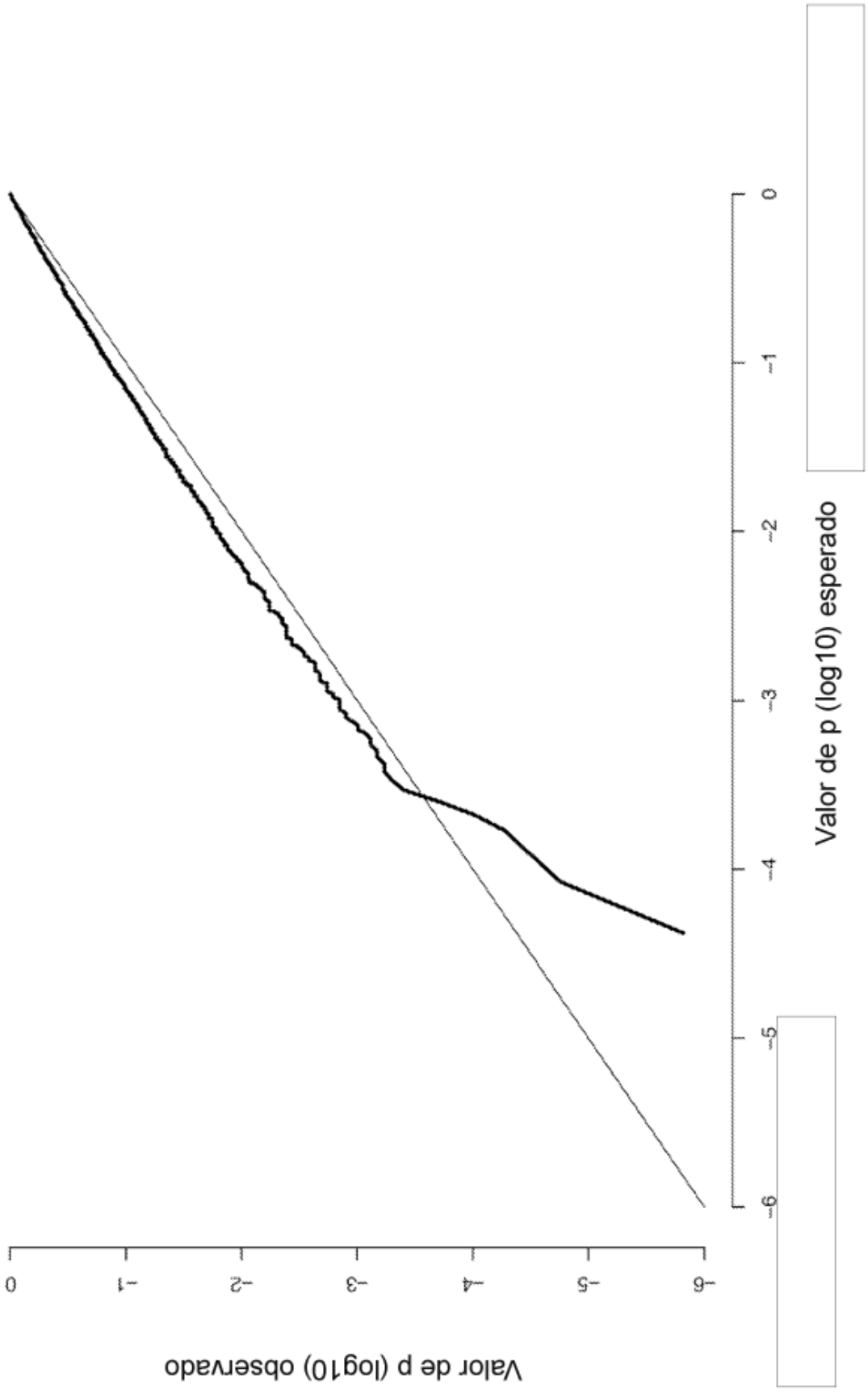

Valor de p (log10) observado
0
-1
-2
-3
-4
-5
-6
Valor de p (log10) esperado
0
-1
-2
-3
-4
-5
-6

Figura 19

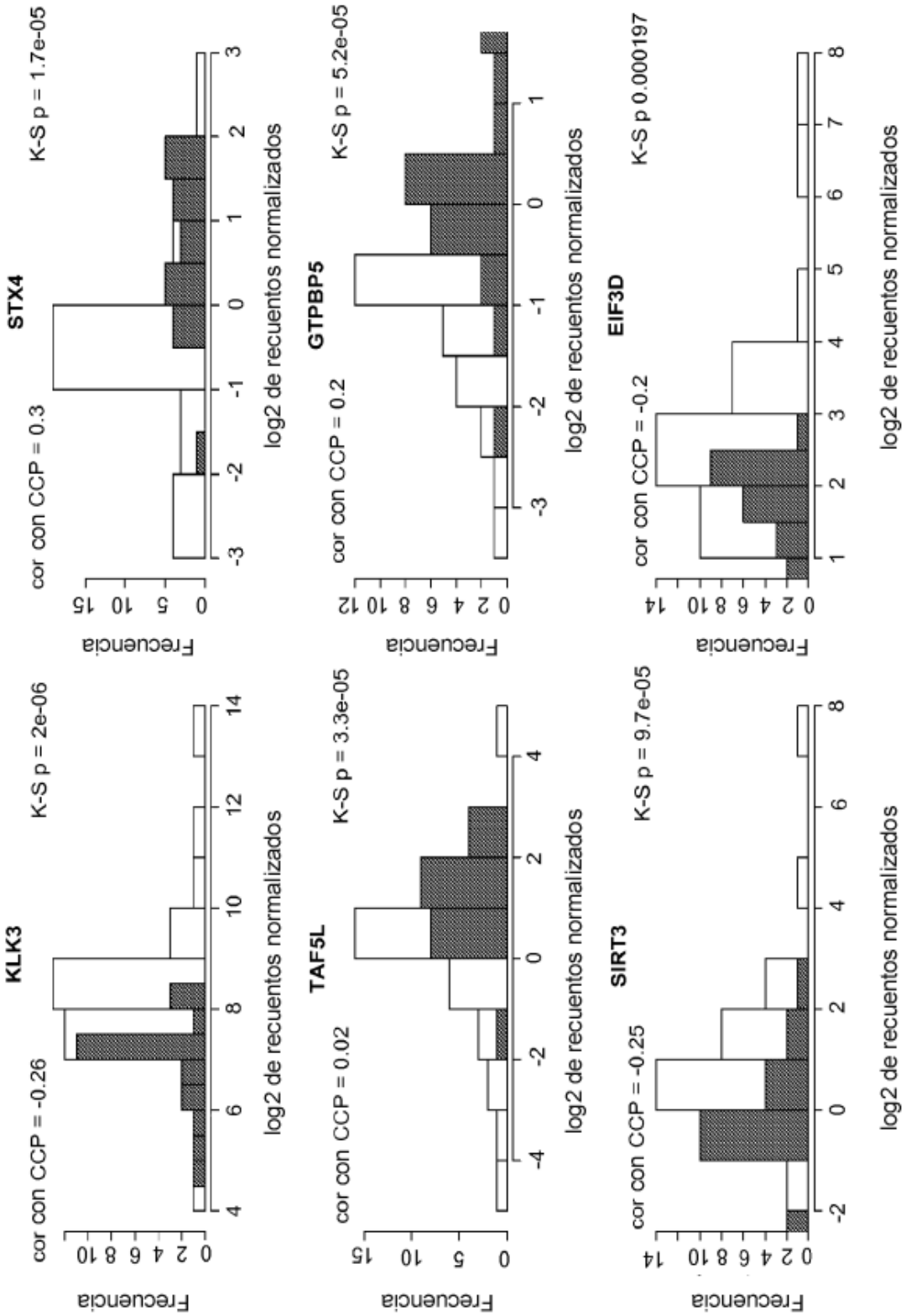

KLK3
cor con CCP = -0.26
K-S p = 2e-06
STX4
cor con CCP = 0.3
K-S p = 1.7e-05
TAF5L
cor con CCP = 0.02
K-S p = 3.3e-05
GTPBP5
cor con CCP = 0.2
K-S p = 5.2e-05
SIRT3
cor con CCP = -0.25
K-S p = 9.7e-05
EIF3D
cor con CCP = -0.2
K-S p 0.000197
Frecuencia
log2 de recuentos normalizados

Figura 20

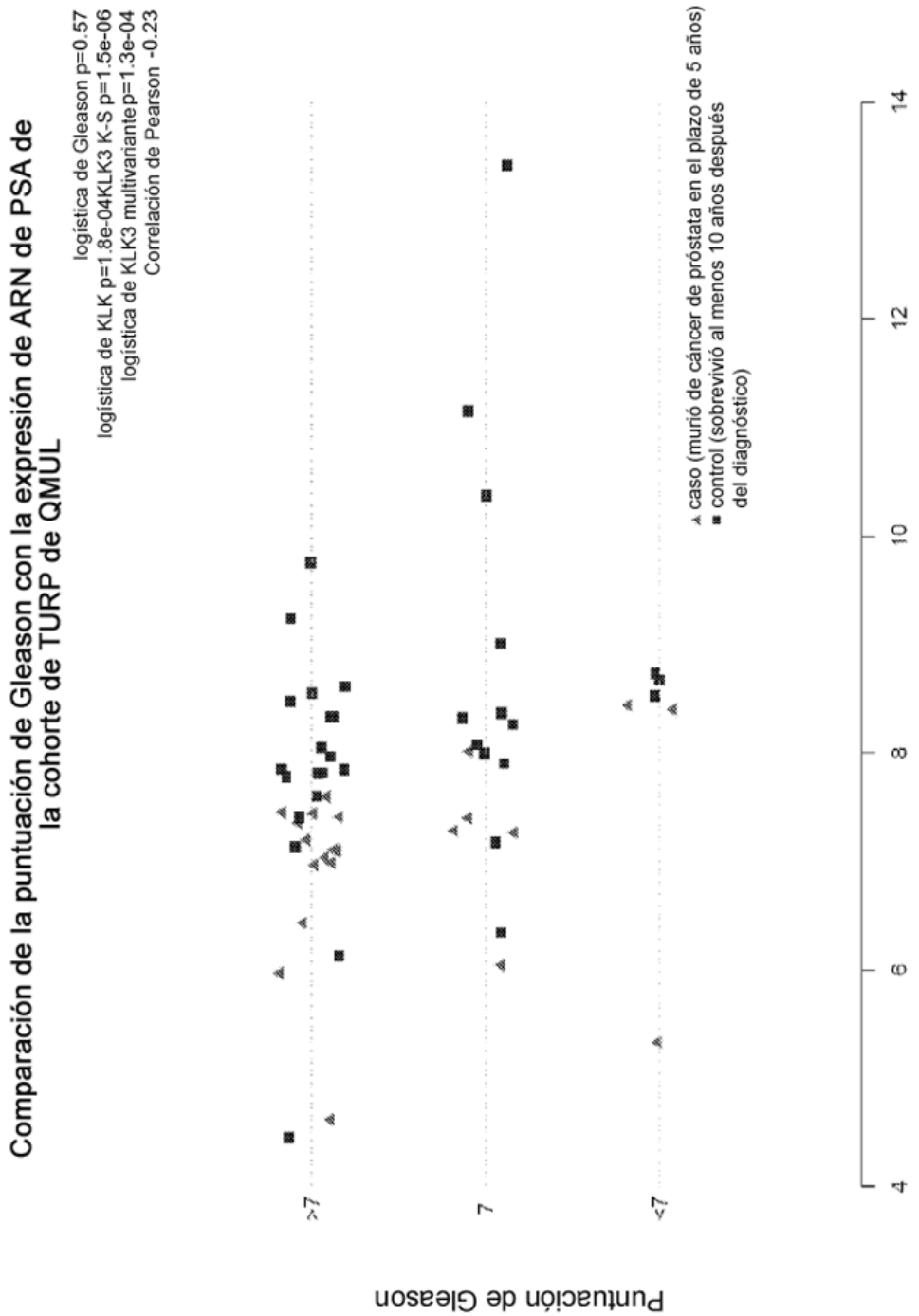

Comparación de la puntuación de Gleason con la expresión de ARN de PSA de la cohorte de TURP de QMUL
logística de Gleason p=0.57
logística de KLK p=1.8e-04KLK3 K-S p=1.5e-06
logística de KLK3 multivariante p=1.3e-04
Correlación de Pearson -0.23
caso (murió de cáncer de próstata en el plazo de 5 años)
control (sobrevivió al menos 10 años después del diagnóstico)
Puntuación de Gleason
>7
7
<7
4
6
8
10
12
14
log2 de recuentos normalizados de KLK3

Figura 21

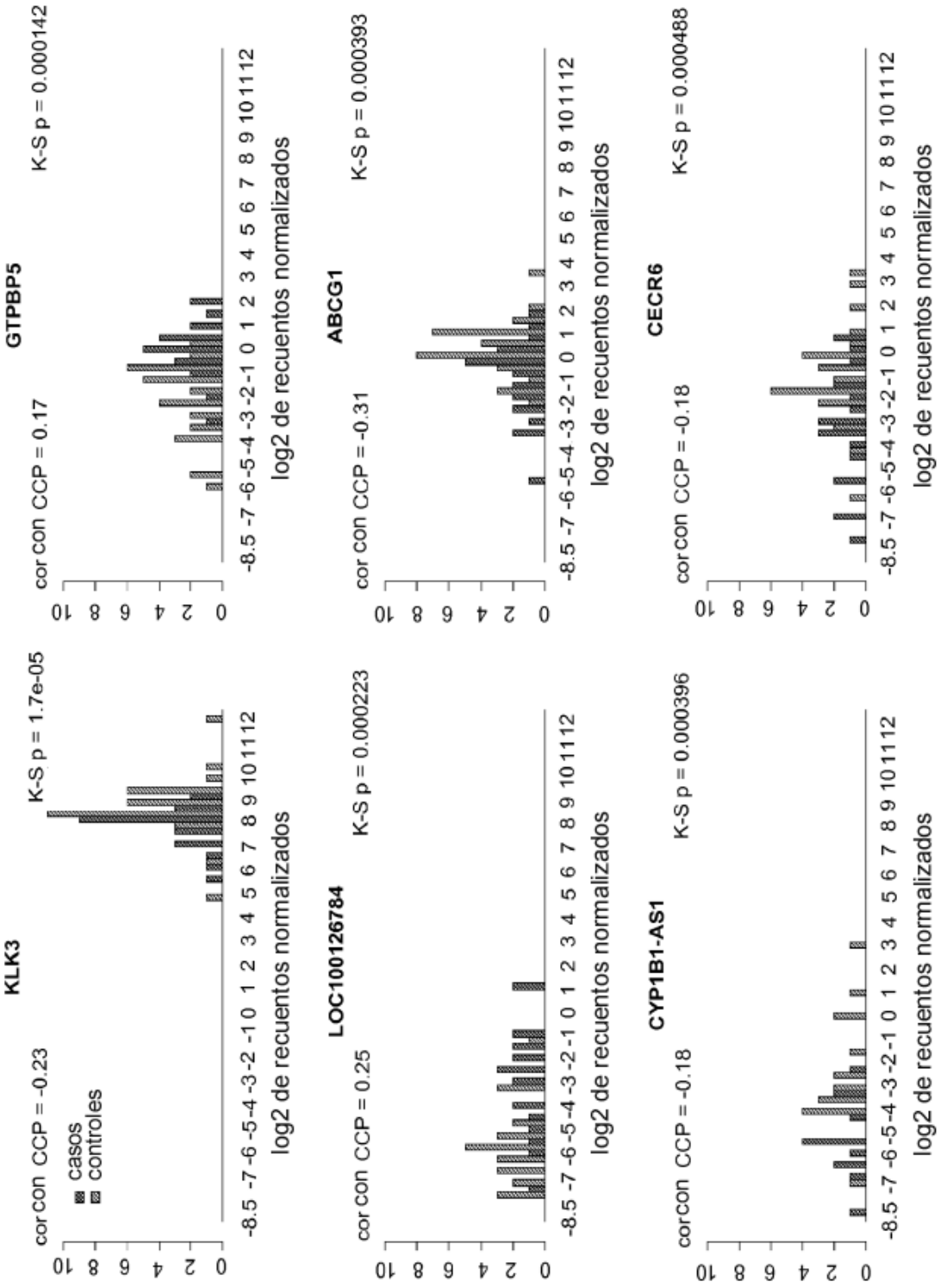

KLK3
K-S p = 1.7e-05
cor con CCP = -0.23
casos
controles
log2 de recuentos normalizados
GTPBP5
K-S p = 0.000142
cor con CCP = 0.17
log2 de recuentos normalizados
LOC100126784
K-S p = 0.000223
cor con CCP = 0.25
log2 de recuentos normalizados
ABCG1
K-S p = 0.000393
cor con CCP = -0.31
log2 de recuentos normalizados
CYP1B1-AS1
K-S p = 0.000396
cor con CCP = -0.18
log2 de recuentos normalizados
CECR6
K-S p = 0.000488
cor con CCP = -0.18
log2 de recuentos normalizados

Figura 22

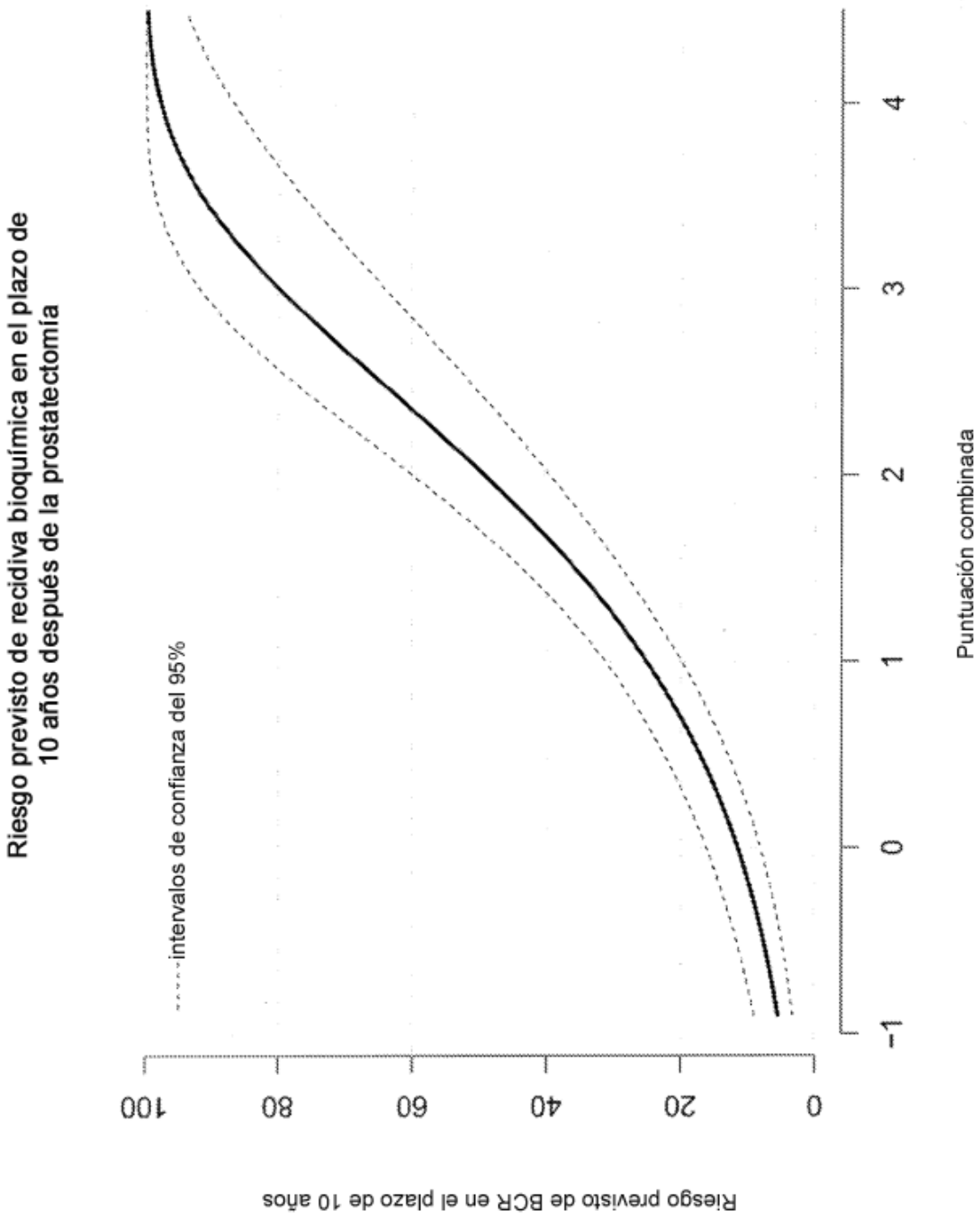

Riesgo previsto de recidiva bioquímica en el plazo de 10 años después de la prostatectomía
intervalos de confianza del 95%
Riesgo previsto de BCR en el plazo de 10 años
0
20
40
60
80
100
−1
0
1
2
3
4
Puntuación combinada

Figura 23

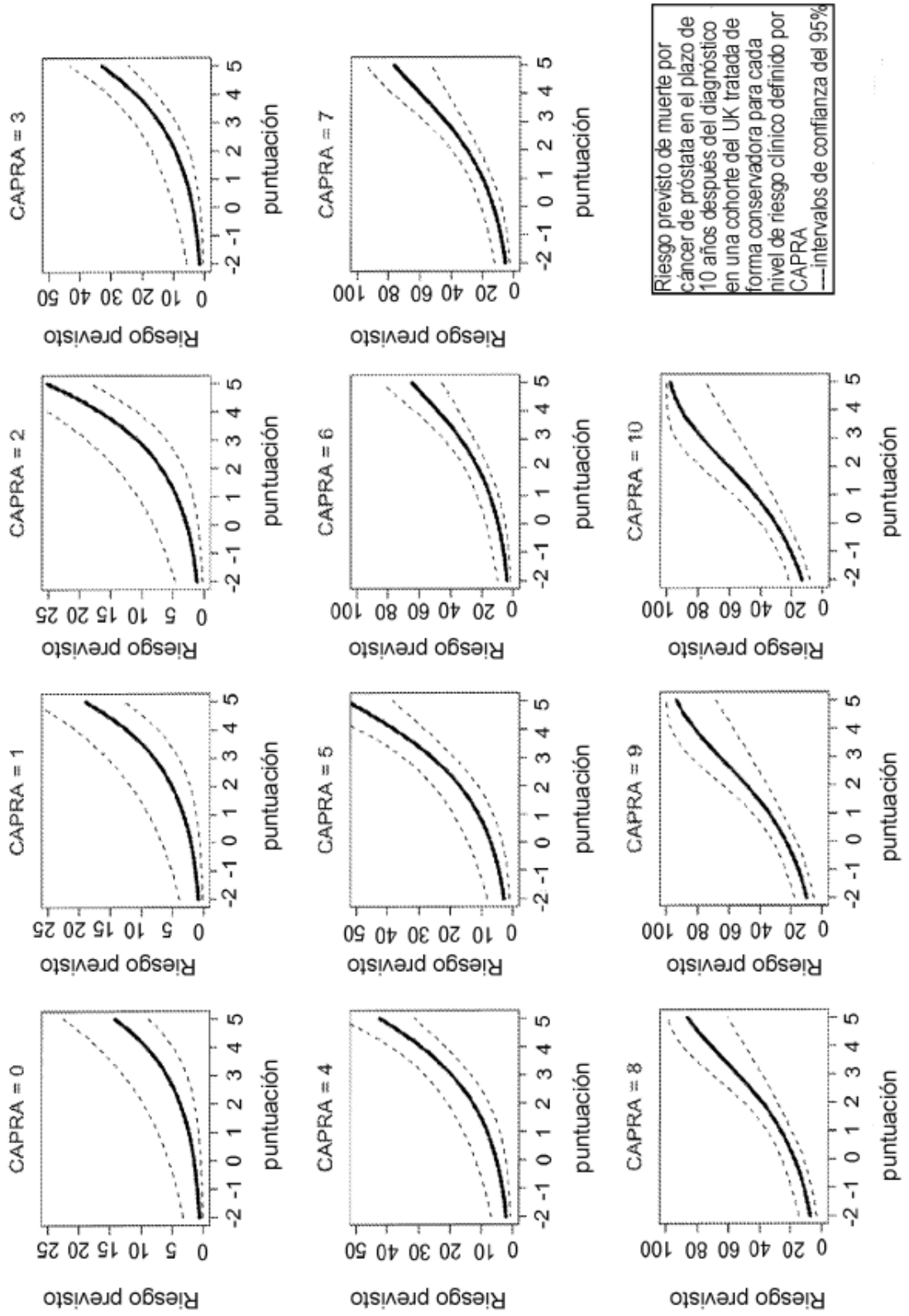

CAPRA = 0
CAPRA = 1
CAPRA = 2
CAPRA = 3
CAPRA = 4
CAPRA = 5
CAPRA = 6
CAPRA = 7
CAPRA = 8
CAPRA = 9
CAPRA = 10
Riesgo previsto
puntuación
Riesgo previsto de muerte por cáncer de próstata en el plazo de 10 años después del diagnóstico en una cohorte del UK tratada de forma conservadora para cada nivel de riesgo clínico definido por CAPRA
—intervalos de confianza del 95%

Figura 24

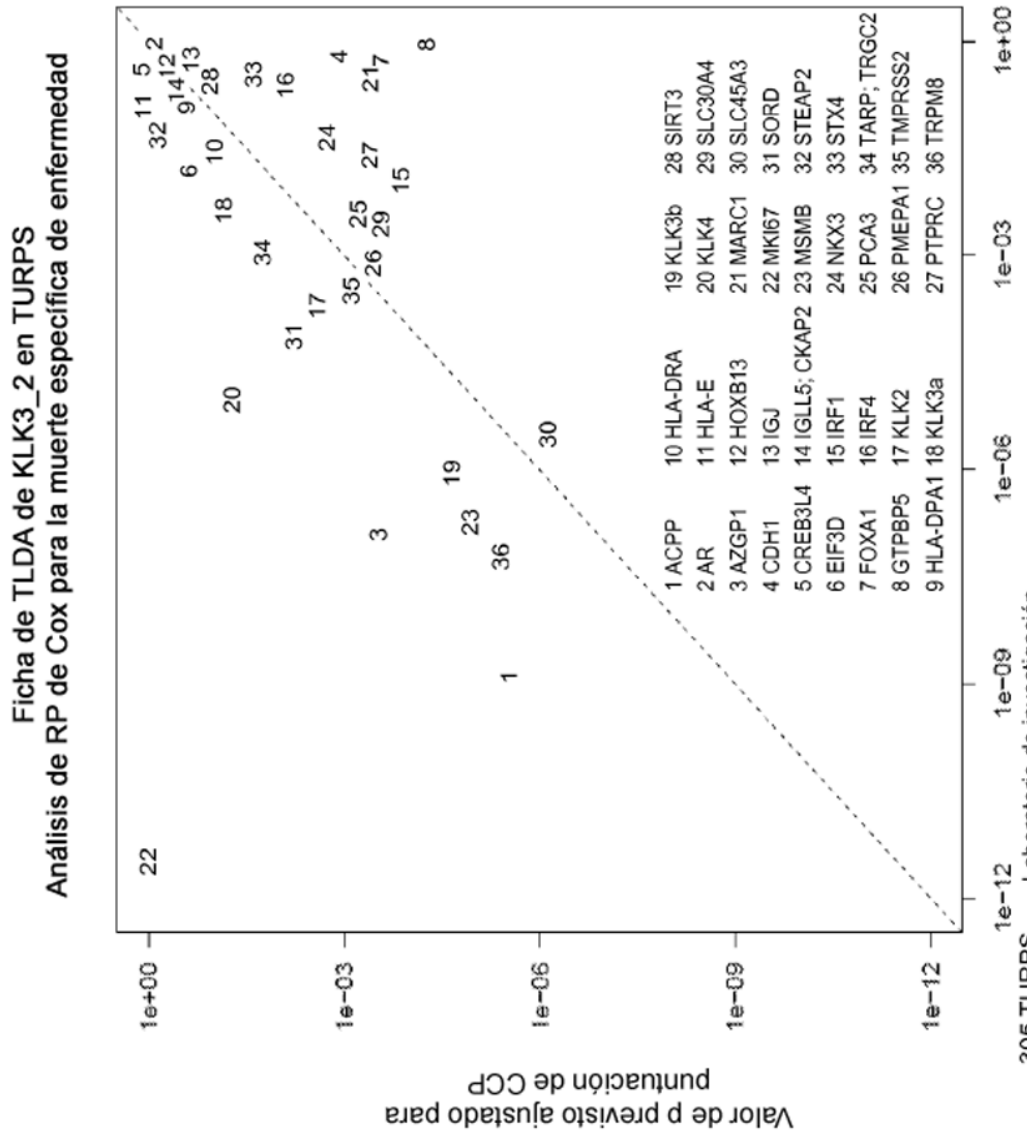

Ficha de TLDA de KLK3_2 en TURPS
Análisis de RP de Cox para la muerte específica de enfermedad
Valor de p previsto ajustado para puntuación de CCP
1e+00
1e−03
1e−06
1e−09
1e−12
1 ACPP
2 AR
3 AZGP1
4 CDH1
5 CREB3L4
6 EIF3D
7 FOXA1
8 GTPBP5
9 HLA-DPA1
10 HLA-DRA
11 HLA-E
12 HOXB13
13 IGJ
14 IGLL5; CKAP2
15 IRF1
16 IRF4
17 KLK2
18 KLK3a
19 KLK3b
20 KLK4
21 MARC1
22 MKI67
23 MSMB
24 NKX3
25 PCA3
26 PMEPA1
27 PTPRC
28 SIRT3
29 SLC30A4
30 SLC45A3
31 SORD
32 STEAP2
33 STX4
34 TARP; TRGC2
35 TMPRSS2
36 TRPM8
305 TURPS Laboratorio de investigación
Sin falta de PSA, Gleason, gradoEnfermedad excluir muestra con id 146887 (expresión muy baja)
Valor de p univariante

Figura 25