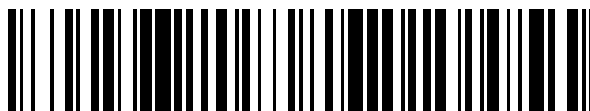


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 535 049**

51 Int. Cl.:

C12Q 1/68 (2006.01)

G01N 33/574 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **30.11.2006** **E 06819878 (7)**

97 Fecha y número de publicación de la concesión europea: **18.02.2015** **EP 1958109**

54 Título: **Procedimientos para la clasificación in vitro de carcinoma hepatocelular**

30 Prioridad:

30.11.2005 EP 05292533

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
04.05.2015

73 Titular/es:

**INSTITUT NATIONAL DE LA SANTE ET DE LA
RECHERCHE MEDICALE (INSERM) (100.0%)
101, RUE DE TOLBIAC
75013 PARIS, FR**

72 Inventor/es:

**ZUCMAN-ROSSI, JESSICA;
DE REYNIES, AURÉLIEN;
RICKMAN, DAVID y
BOYVAULT, SANDRINE**

74 Agente/Representante:

CURELL AGUILÁ, Mireia

ES 2 535 049 T3

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

DESCRIPCIÓN

Procedimientos para la clasificación *in vitro* de carcinoma hepatocelular.

5 La presente invención se refiere a procedimientos para la clasificación *in vitro* de carcinoma hepatocelular (HCC). Los presentes procedimientos están basados en la determinación del perfil de expresión de combinaciones de genes particulares.

10 El carcinoma hepatocelular (HCC) es uno de los tumores sólidos más frecuentes en todo el mundo y representa la tercera causa de mortalidad entre las muertes por cáncer (Bosch. F.X., *et al.* *Semin Liver Dis* 19, 271-85 (1999)). Su frecuencia es particularmente elevada en Asia y África debido a la elevada frecuencia de las infecciones por hepatitis vírica y a la exposición a Aflatoxina B1 (AFB1). Durante los últimos 10 años, la incidencia de HCC ha aumentado notablemente en el Reino Unido, Francia y los Estados-Unidos (Taylor-Robinson, S.D. *et al.* *Bmj* 319, 640 (1999); Deuffic, S. *et al.* *Lancet* 351, 214-5 (1998). El-Serag, H.B. & Mason, A.C. *N Engl J Med* 340, 745-50 (1999)). Este aumento está relacionado con el aumento de las infecciones por hepatitis C virales.

La cirrosis hepática de cualquier origen y los nódulos regenerativos displásicos han sido considerados durante mucho tiempo como precursores probables de HCC debido a su frecuente asociación con la ocurrencia de HCC (Edmondson, H.A. & Peters, R.L. *Semin Roentgenol* 18, 75-83 (1983); Thorgeirsson, S.S. & Grisham, J.W. *Nat Genet* 31, 339-46 (2002)). Del mismo modo que en otros tumores sólidos, un gran número de alteraciones genéticas se acumulan durante el proceso carcinogénico. Algunas de estas alteraciones genéticas son específicas de factores etiológicos de HCC, en particular de la infección por HBV que puede inducir inestabilidad cromosómica (Aoki, H., *et al.* *Proc Natl Acad Sci USA* 93, 7300-4 (1996)) o mutagénesis insercional (Brecht, C. *Gastroenterology* 127, S56-61 (2004)). Las demás alteraciones asociadas específicamente con factores de riesgo son la mutación R249S del gen TP53 en HCCs expuestos a Aflatoxina B1 (Bressac, B. *et al.* *Proc Natl Acad Sci USA* 87, 1973-7 (1990)), mutaciones de KRAS2 observadas en HCCs asociados a cloruro de vinilo (Weihrauch, M. *et al.* *Br J Cancer* 84, 982-9 (2001)) y mutaciones de TCF1 asociadas con adenomas hepatocelulares (Bluteau, O. *et al.* *Nat Genet* 32, 312-5 (2002)). De entre las alteraciones genéticas no relacionadas con factores de riesgo de HCC, estudios de alelotipos de microsatélites y de hibridación genómicas comparativa (CGH) han demostrado aberraciones cromosómicas recurrentes. Los brazos cromosómicos frecuentemente delecionados son el 17p, 8p, 16q, 16p, 4q, 9p, 13q, 1p y 6q mientras que las ganancias más frecuentes se encuentran en los cromosomas 1q, 7q, 8q y 17q (Boige, V. *et al.* *Cancer Res* 57, 1986-90 (1997); Wong, N. *et al.* *Clin Cancer Res* 6, 4000-9 (2000); Guan, X.Y. *et al.* *Genes Chromosomes Cancer* 29, 110-6 (2000)). El HCC es, por lo tanto, un grupo heterogéneo de tumores que difieren en los factores de riesgo y alteraciones genéticas.

35 Esto impone una necesidad de una clasificación de tumores HCC más precisa, fiable y fácil de llevar a cabo. De hecho, es por ejemplo muy difícil buscar terapias eficaces contra tumores muy heterogéneos. Por el contrario, una clasificación de HCC fiable permitiría estudiar cada subgrupo por separado y encontrar nuevas terapias dirigidas para cada subgrupo. Una prueba de clasificación fiable y fácil de llevar a cabo permitiría entonces elegir para cada paciente un tratamiento adaptado.

40 En particular, el pronóstico del HCC también es muy heterogéneo. Actualmente, el principal tratamiento para HCC es la eliminación quirúrgica del tumor de HCC, que puede o no estar seguida por quimioterapia adyuvante. La quimioterapia puede ser muy cansina y dolorosa para los pacientes pero es necesaria en caso de HCC con poco pronóstico. Por lo tanto, una clasificación y un procedimiento de pronóstico de tumores de HCC sería de gran ayuda para decidir si se debe o no administrar una terapia adyuvante a un paciente de HCC.

La heterogeneidad del HCC es bien conocida para el experto en la materia, y se han realizado bastantes esfuerzos para clasificar mejor y/o pronosticar los tumores HCC en la técnica anterior.

50 Por ejemplo, recientemente varios grupos han intentado clasificar los tumores de HCC mediante análisis de transcriptoma global. Algunos de ellos describen alteraciones de los perfiles de expresión significantes entre HCC derivados de HBV y de HCV respectivamente (Okabe *et al.* *Cancer Res.* 2001 Mar 1;61(5):2129-37; Iizuka *et al.* *Cancer Res.* 2002 Jul 15;62(14):3939-44).

55 Otros describen una clasificación de HCC en dos o tres subgrupos basada en varias características histológicas (Chung *et al.* *Mol Cells.* 2002 Dec 31;14(3):382-7; Chen *et al.* *Mol Bio Cell.* 2002 Jun;13(6):1929-39; WO 2004/090163) o en la probabilidad de supervivencia (Lee *et al.* *Hepatology* 2004 Sep;40(3):667-76).

60 Los inventores, en particular, han mostrado anteriormente que las alteraciones genéticas están de hecho estrechamente relacionadas con las características clínicas del HCC definiendo dos grupos de HCC (Legoix, P. *et al.* *Oncogene* 18, 4044-6 (1999); Laurent-Puig, P. *et al.* *Gastroenterology* 120, 1763-73 (2001)). El primer tipo de HCC se asoció no solamente con un elevado nivel de inestabilidad cromosómica y mutaciones de TP53 y AXIN1 frecuentes, sino que también se relacionó estrechamente con infecciones por HBV y con un mal pronóstico. A la inversa, de tumores HCC del segundo subgrupo son estables cromosómicamente, con una elevada incidencia de alteraciones que activan las β -cateninas y no están asociados con infecciones virales.

Con respecto al pronóstico, varios grupos han descrito genes implicados en invasión vascular (Chen *et al. Mol Bio Cell.* 2002 Jun;13(6):1929-39; Qin *et al. J Cancer Res Clin Oncol.* 2004 Sep;130(9):497-513) o metastasis (Qin *et al. J Cancer Res Clin Oncol.* 2004 Sep;130(9):497-513; Ye *et al. Nat Med.* 2003 Apr; 9(4):416-23).

- 5 Otros han identificado grupos de genes que permiten un pronóstico de recaída (Kurokawa *et al. J Hepatol.* 2004 Aug; 41(2):284-91; Iizuka *et al. Lancet.* 2003 Mar 15; 361(9261):923-9; WO 2005/017150) y/o de supervivencia (Lee *et al. Hepatology.* 2004 Sep; 40(3):667-76; WO 2005/017150).

- 10 Okabe *et al.* "Genome-wide Analysis of Gene Expression in Human Hepatocellular Carcinomas Using cDNA Microarray: Identification of Genes Involved in Viral Carcinogenesis and Tumor Progression", Cancer Research. Vol. 61, p. 2129-2137 (2001) divulga perfiles de expresión de genes que distinguen HCCs positivos para HVB y para HVC y diferentes fenotipos de tumores.

- 15 Sin embargo, una clasificación de los tumores HCC en dos o incluso tres subgrupos, basada solamente en características histológicas o genéticas, no es probable que refleje de forma precisa la gran heterogeneidad de HCC. Además, existe todavía una necesidad de una prueba de pronóstico simple y fácil de realizar.

- 20 Para seguir investigando correlaciones genotipo-fenotipo en HCC, identificando vías y/o procesos biológicos desregulados en dichos tumores heterogéneos y encontrar nuevos factores de pronóstico, los inventores han llevado a cabo de este modo un análisis exhaustivo a nivel clínico, genético y transcriptómico de una amplia serie de 123 tumores.

- 25 A pesar de que la mayoría de estudios previos sólo fueron capaces de subdividir los tumores de HCC en dos subgrupos, los inventores hallaron sorprendentemente que los tumores HCC se agrupaban de hecho en 6 subgrupos distintos, estrechamente asociados con varias alteraciones clínicas y genéticas. También determinaron un predictor de diagnóstico de 16 genes y un predictor de pertenencia a clase de 24 genes así como una firma de 5 genes que predice el pronóstico de un paciente independientemente del subgrupo de HCC y que obtiene mejores resultados que los marcadores de pronóstico clínicos comunes.

- 30 Más precisamente, los inventores han definido 6 subgrupos o clases de HCC distintos (de ahora en adelante denominados G1 a G6). Estos 6 subgrupos se definieron mediante un análisis no supervisado del análisis transcriptómico global de 57 HCC, 3 adenomas hepatocelulares y 5 muestras de tejidos no tumorales agrupados utilizando micromatrices Affymetrix HG-U133A GeneChip™. Los 6 subgrupos están muy asociados con factores clínicos y genéticos, tal como se muestra en la siguiente Tabla 1 y se resume en la Figura 1.

35

Tabla 1. Características clínicas y genéticas asociados con subgrupos de HCC

	Hibridaciones Affymetrix (57 HCC)	QRT-PCR	
		Set de validación (63 HCC)	Set completo (109 HCC)
G1			
HBV bajo número de copias	0,03	0,04	<10 ⁻⁵
AFP > 100 IU/ml	0,01	0,006	<10 ⁻⁴
origen africano	0,005	0,3*	0,004
hembra	0,06	0,05	0,002
Mutación de Axin1	0,1	0,009	0,001
16q LOH	0,05	0,04	0,001
G2			
HBV alto num. de copias	<10 ⁻⁴	0,07	0,004
hemocromatosis	1*	0,005*	0,03
invasión portal	0,6	0,05	0,01
Mutación de PIK3CA	0,009*	-	-
G3			
Mutación R249S de TP53	0,3*	0,002*	0,004
Metilación de CDKN2A	0,04	0,1	0,01
17p LOH	0,02	0,004	0,0002
5q LOH	0,02	0,02*	0,004
19p LOH	0,002	0,4*	0,004
21q LOH	0,001	0,05	<10 ⁻³
22q LOH	0,007	0,02*	<10 ⁻⁴
G1, G2, G3			
FAL	<10 ⁻³	<10 ⁻³	<10 ⁻⁵
4q LOH	<10 ⁻³	0,002	<10 ⁻⁵
16p LOH	0,005	0,05	<10 ⁻³

	Hibridaciones Affymetrix (57 HCC)	QRT-PCR	
		Set de validación (63 HCC)	Set completo (109 HCC)
Recaída temprana	0,005	1	0,3
Muerte temprana	0,05	0,7	0,4
G1, G2			
Edad < 63 años	0,03	0,08	0,001
13q LOH	0,08	0,0001	$<10^{-4}$
1p LOH	0,1	0,02	0,007
G2, G3			
Mutación de <i>TP53</i>	0,03	0,001	0,0001
G4			
Mutación de <i>TCF1</i> **	0,01*	—*	—*
no invasión vascular	0,2	0,03	0,01
G6			
Nódulos satélite	0,005	-	0,0005
G5, G6			
Metilación de <i>CDH1</i>	0,01	0,007	$<10^{-3}$
Mutación de <i>CTNNB1</i>	$<10^{-10}$	$<10^{-5}$	$<10^{-12}$

Los valores mostrados son los P-valores obtenidos a partir de unas pruebas exactas de Fisher basadas en la variable genética o clínica dada y (i) los grupos de clúster originales para las series de Affymetrix GeneChip, (ii) el grupo de clúster predicho (en base al predictor de 16 genes) para las series de QRT-PCR. * Igual o inferior a 5 muestras con esta característica en el set de tumores probado. ** incluyendo las 3 muestras de adenoma.

Como se describe más detalladamente en el Ejemplo 2, párrafo 2.2, los 6 subgrupos pueden estar definidos utilizando sus características principales tal y como se describe en la siguiente Tabla 2.

5 Tabla 2. Definición de los 6 subgrupos por la presencia (+) o ausencia (-) de características clínicas y genéticas principales

	G1	G2	G3	G4	G5	G6
Inestabilidad cromosómica	+	+	+	-	-	-
Recaída y muerte tempranas	+	+	+	-	-	-
Mutación de <i>TP53</i>	-	+	+	-	-	-
Infección HBV	+	+	-	-	-	-
Bajo número de copias	+	-	-	-	-	-
Alto número de copias	-	+	-	-	-	-
Mutación de <i>CTNNB1</i>	-	-	-	-	+	+
Nódulos satélite	-	-	-	-	-	+

10 Los procedimientos de clasificación según la invención permiten determinar fácilmente a cual de éstos 6 subgrupos HCC pertenece cualquier muestra de hígado de HCC.

La invención se refiere a un procedimiento de clasificación *in vitro* de un tumor de HCC en 6 subgrupos a partir de una muestra de hígado HCC de un sujeto que sufre HCC, que comprende:

- 15 a) determinar un perfil de expresión que comprende o que consiste en:
- a. la siguiente combinación de 16 genes: RAB1A, REG3A, NRAS, RAMP3, MERTK, PIR, EPHA1, LAMA3, G0S2, HN1, PAK2, AFP, CYP2C9, CDH2, HAMP, y SAE1, o
- 20 b. la siguiente combinación de 24 genes: ALDH1L1, CD24, CD74, CFHR3, CYP4F12, DNAJA3, DSCR1, EPHA1, EPHB4, FAAH, FGFR2, FLJ10159, GLT8D1, HAL, MATN2, MRPS7, PAK2, PLXNB1, RAB1A, RHOQ, SLC27A5, SLPI, SMARCE1, STRA13;
- 25 b) calcular a partir de dicho perfil de expresión 6 distancias de subgrupo; y
- c) clasificar dicho tumor HCC en el subgrupo para el cual la distancia es la más baja,

en el que los 6 subgrupos G1, G2, G3, G4, G5 y G6 están definidos por sus características clínicas y genéticas descritas en la Tabla 2.

30 La presente descripción también divulga un procedimiento de clasificación *in vitro* de un tumor HCC entre 6

subgrupos a partir de una muestra de hígado HCC de un sujeto que sufre HCC, que comprende:

- a) determinar un perfil de expresión que comprende o que consiste en una combinación de por lo menos 8, por lo menos 10, por lo menos 12, por lo menos 14 o por lo menos 16 genes seleccionados de entre el grupo que consiste en: RAB1A, REG3A, NRAS, RAMP3, MERTK, PIR, EPHA1, LAMA3, G0S2, HN1, PAK2, AFP, CYP2C9, CDH2, HAMP, SAE1, ADH6, DCN, FLJ10159, ALDH1L1, IGF1, LECT2, SLC38A1, SPARCL1, CTNNA2, GLUL, LEF1, MATN2, MME, PFN2, SPINT2, TBX3, y FGFR2;
- b) calcular a partir de dicho perfil de expresión 6 distancias de subgrupo; y
- c) clasificar dicho tumor HCC en el subgrupo para el cual la distancia de subgrupo es la más baja,

en el que los 6 subgrupos G1, G2, G3, G4, G5 y G6 están definidos por sus características clínicas y genéticas descritas en la Tabla 2.

Las características principales del primer set de genes implicados en la clasificación de HCC están descritos en la siguiente Tabla 3.

Tabla 3. Primer set de genes implicados en la clasificación de HCC.

Símbolo del gen *	Nombre del gen HUGO	ID del gen "Entrez Gene" **
RAB1A	RAB1A, miembro de la familia de oncogenes RAS	5861
REG3A	regenerador derivado de islote 3 alfa	5068
NRAS	homólogo de oncogen de neuroblastoma RAS viral (v-ras)	4893
RAMP3	Proteína modificadora de actividad receptor (calcitonina) 3	10268
MERTK	Tirosina quinasa de proto-oncogen c-mero	10461
PIR	pirina (proteína nuclear de unión a hierro)	8544
EPHA1	receptor de EPH A1	2041
LAMA3	laminina, alpha 3	3909
G0S2	Interrupor G0/G1 2	50486
HN1	Expresado hematológico y neurológico 1	51155
PAK2	Quinasa activada por p21 (CDKN1A) 2	5062
AFP	Fetoproteína alfa	174
CYP2C9	Citocromo P450, familia 2, subfamilia C, polipéptido 9	1559
CDH2	cadherina 2, tipo 1, N-cadherina (neuronal)	1000
HAMP	péptido antimicrobiano hepcidina	57817
SAE1	subunidad 1 de enzima activadora de SUMO-1	10055
ADH6	alcohol deshidrogenasa 6 (clase V)	130
DCN	decorina	1634
FLJ10159	Proteína hipotética FLJ10159	55084
ALDH1L1	familia aldehído deshidrogenasa 1, miembro L1	10840
IGF1	Factor de crecimiento tipo insulina 1 (somatomedina C)	3479
LECT2	quimiotaxina derivada de celular leucocitaria 2	3950
SLC38A1	Transportador de soluto familia 38, miembro 1	81539
SPARCL1	De tipo SPARC 1 (mast9, hevina)	8404
CTNNA2	catenina (proteína asociada a cadherina), alfa 2	1496
GLUL	Ligasa glutamato-amonio (glutamina sintetasa)	2752
LEF1	Factor de union a potenciador linfoide	51176
MATN2	matrilina 2	4147
MME	metalo-endopeptidasa de membranae (endopeptidasa neutral, encefalinasa, CALLA, CD10)	4311
PFN2	Profilina 2	5217
SPINT2	Inhibidor de serinpeptidasa, tipo Kunitz, 2	10653
TBX3	T-box 3 (síndrome ulnar mamario)	6926
FGFR2	Receptor de factor de crecimiento de fibroblasto 2 (quinasa expresada en bacteria, receptor de factor de crecimiento de queratinocito, disostosis craneofacial 1, síndrome de Crouzon, síndrome de Pfeiffer, síndrome de Jackson-Weiss)	2263

* Todos los símbolos y nombres de genes son según el Comité de Nomenclatura de Genes HUGO (disponible en <http://www.gene.ucl.ac.uk/nomenclature/>)

** Toda la información disponible en relación con los genes listados de la Tabla 3 puede recuperarse del portal

"Entrez Gene" (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=gene>) utilizando la ID de gen de "Entrez Gene" proporcionada en la Tabla 3.

Según la invención, una "clasificación" de tumores HCC pretende significar la determinación para cualquier tumor HCC, del "subgrupo" o "clase" HCC (estos dos términos "subgrupo" y "clase" se utilizarán indistintamente el uno por el otro a lo largo de la presente memoria) al cual pertenece, en la que los subgrupos están definidos por las características descritas en la Tabla 2.

También se divulga un procedimiento de clasificación *in vitro*, en el que el perfil de expresión comprende o consiste en por lo menos 8, por lo menos 10, por lo menos 12, por lo menos 14, o por lo menos 16 genes seleccionados de entre el grupo que consiste en: RAB1A, REG3A, NRAS, RAMP3, MERTK, PIR, EPHA1, LAMA3, GOS2, HN1, PAK2, AFP, CYP2C9, CDH2, HAMP, SAE1, ADH6, DCN, FLJ10159, ALDH1L1, IGF1, LECT2, SLC38A1, SPARCL1.

Además se divulga un procedimiento de clasificación *in vitro* en el que el perfil de expresión comprende o consiste en por lo menos 8, por lo menos 10, por lo menos 12, por lo menos 14, o por lo menos 16 genes seleccionados de entre el grupo que consiste en: RAB1A, REG3A, NRAS, RAMP3, MERTK, PIR, EPHA1, LAMA3, GOS2, HN1, PAK2, AFP, CYP2C9, CDH2, HAMP, y SAE1.

En una forma de realización preferida del procedimiento de clasificación *in vitro* según la presente invención, el perfil de expresión comprende o consiste en la combinación de 16 genes: RAB1A, REG3A, NRAS, RAMP3, MERTK, PIR, EPHA1, LAMA3, GOS2, HN1, PAK2, AFP, CYP2C9, CDH2, HAMP, y SAE1.

La presente descripción también divulga otro procedimiento de clasificación *in vitro* de un tumor de HCC entre 6 subgrupos a partir de muestras de hígado HCC de un sujeto que sufre HCC, que comprende:

- determinar un perfil de expresión que comprende o que consiste en una combinación de por lo menos 8, por lo menos 10, por lo menos 12, por lo menos 14, por lo menos 16, por lo menos 18, por lo menos 20, por lo menos 22 o por lo menos 24 genes seleccionados de entre el grupo que consiste en: ALDH1L1, CD24, CD74, CFHR3, CYP4F12, DNAJA3, DSCR1, EPHA1, EPHB4, FAAH, FGFR2, FLJ10159, GLT8D1, HAL, MATN2, MRPS7, PAK2, PLXNB1, RAB1A, RHOQ, SLC27A5, SLPI, SMARCE1, STRA13;
- calcular a partir de dicho perfil de expresión 6 distancias de subgrupo; y
- clasificar dicho tumor de HCC en el subgrupo para el que la distancia de subgrupo es la más baja,

en el que los 6 subgrupos G1, G2, G3, G4, G5 y G6 están definidos por sus características clínicas y genéticas descritas en la Tabla 2.

Las características principales de este segundo set de genes implicados en la clasificación de HCC están descritos en la siguiente Tabla 4.

Tabla 4. Segundo set de genes implicados en la clasificación de HCC

Símbolo del gen *	Nombre del gen HUGO *	ID del gen "Entrez Gene" **	Ubicación cromosómica	Otros alias
ALDH1L1	Formiltetrahidrofolato deshidrogenasa	10840	chr3q21.2	FTHFD
CD24	Antígeno CD24 (antígeno de clúster de células de carcinoma pulmonar 4 pequeño)	934	chr6q21	CD24A
CD74	Antígeno CD74	972	chr5q32	DHLA, HLADG, la-GAMMA, proteína 41
CFHR3 /// CFHR4	Relacionado con factor H de complemento 3 /// Relacionado con factor H de complemento 4	10878 /// 10877	chr1q32	CFHL3, DOWN16, FHR-3, FHR3, HLF4 /// CFHL4, FHR-4, FHR4, RP4-608O15.2
CYP4F12	citocromo P450, familia 4, subfamilia F, polipéptido 12	66002	chr19p13.1	F22329_1
DNAJA3	Homólogo DnaJ (Hsp40), subfamilia A, miembro 3	9093	chr16p13.3	FLJ45758, TID1, hTid-1
DSCR1	Gen 1 región crítica para síndrome de Down	1827	chr21q22.1-q22.2 21q22.12	ADAPT78, CSP1, DSC1, MCIP1, RCN1
EPHA1	EphA1	2041	chr7q34	EPH, EPHT, EPHT1
EPHB4	EphB4	2050	chr7q22	HTK, MYK1, TYRO11

Símbolo del gen *	Nombre del gen HUGO *	ID del gen "Entrez Gene" **	Ubicación cromosómica	Otros alias
FAAH	Hidrolasa de amida de ácidos grasos	2166	chr1p35-p34	MGC102823, MGC138146
FGFR2	receptor de factor de crecimiento de fibroblasto 2	2263	chr10q26	BEK, BFR-1, CD332, CEK3, CFD1, ECT1, JWS, K-SAM, KGFR, TK14, TK25
FLJ10159	Proteína hipotética FLJ10159	55084	chr6q21	
GLT8D1	Dominio glicosiltransferasa 8 que contiene 1, or glicosiltransferasa AD-017	55830	chr3p21.1	AD-017, DKFZp781020198, FLJ14611, MSTP139
HAL	histidina amonio liasa	3034	chr12q22-q24.1	HIS, HSTD, histidasa
MATN2	matrilina 2	4147	chr8q22	
MRPS7	Proteína ribosomal mitocondrial S7	51081	chr17q25	MRP-S, MRP-S7, RP-S7, RPMS7
PAK2	Quinasa activada por p21 (CDKN1A) 2	5062	chr3q29	PAK65, PAKgamma
PLXNB1	plexina B1	5364	chr3p21.31	KIAA0407, PLEXIN-B1, PLXN5, SEP
RAB1A	RAB1A, miembro familia oncogenes RAS	5861	chr2p14	DKFZP564B163, RAB1
RHOQ	Homólogo familia genes ras, miembro Q	23433	chr2p21	ARHQ, RASL7A, TC10, TC10A
SLC27A5	Familia de transportadores de soluto 27 (transportador de ácidos grasos), miembro 5	10998	chr19q13.43	ACSB, ACSVL6, FACVL3, FATP5, FLJ22987, VLACSR, VLCS-H2, VLCSH2
SLPI	Inhibidor de proteasa leucocitaria secretado (antileucoproteinasas)	6590	chr20q12	ALK1, ALP, BLPI, HUSI, HUSI-I, MPI, WAP4, WFDC4
SMARCE1	Regulador de cromatina relacionado con SWI/SNF, asociado a matriz, dependiente de actina, subfamilia e, miembro 1	6605	chr17q21.2	BAF57
STRA13	Estimulado por ácido retinoico 13	201254	chr17q25.3	E3, MGC14480

* Todos los símbolos y nombres de genes son según el Comité de Nomenclatura de Genes HUGO (disponible en <http://www.gene.ucl.ac.uk/nomenclature/>)

5 ** Toda la información disponible en relación con los genes listados de la Tabla 4 puede recuperarse del portal "Entrez Gene" (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=gene>) utilizando la ID de gen de "Entrez Gene" proporcionada en la Tabla 4.

10 En una forma de realización preferida del procedimiento de clasificación *in vitro* según la invención mencionado anteriormente en la presente memoria, el perfil de expresión comprende o consiste en la combinación de 24 genes siguiente: ALDH1L1, CD24, CD74, CFHR3, CYP4F12, DNAJA3, DSCR1, EPHA1, EPHB4, FAAH, FGFR2, FLJ10159, GLT8D1, HAL, MATN2, MRPS7, PAK2, PLXNB1, RAB1A, RHOQ, SLC27A5, SLPI, SMARCE1, STRA13.

15 Según la invención, una "muestra de hígado HCC" pretende significar cualquier muestra de hígado que comprende tejido de tumor de HCC. En una forma de realización preferida de un procedimiento de clasificación *in vitro* según la invención, la muestra de hígado HCC es una biopsia de hígado HCC o una extirpación quirúrgica de tumor HCC.

20 Por "determinar un perfil de expresión" se pretende decir la medición del nivel de expresión de un grupo de genes seleccionados. El nivel de expresión de cada gen puede ser determinado *in vitro* ya sea a nivel proteico o a nivel nucleico, utilizando cualquier tecnología conocida de la técnica.

25 Por ejemplo, a nivel proteico, la medición *in vitro* del nivel de expresión de una proteína en particular puede realizarse mediante cualquier procedimiento de cuantificación conocido por el experto en la materia, incluyendo pero no limitado a ELISA o análisis de espectrometría de masas. Estas tecnologías se adaptan fácilmente a cualquier muestra de HCC. De hecho, las proteínas de la muestra HCC pueden ser extraídas utilizando varias tecnologías bien conocidas por los expertos en la materia para medición por ELISA o por espectrometría de masas en solución. Alternativamente, el nivel de expresión de una proteína en una rodaja de tumor de HCC puede ser analizado utilizando espectroscopia de masas directamente sobre la rodaja de tejido.

En una forma de realización preferida de un procedimiento de clasificación *in vitro* según la invención, el perfil de expresión se determina *in vitro* a nivel nucleico. A nivel nucleico, la medición *in vitro* del nivel de expresión de un gen puede llevarse a cabo ya sea directamente sobre el ARN mensajero (ARNm), o sobre el ADN complementario retrotranscrito (ADNc). Cualquier procedimiento de medición del nivel de expresión puede ser utilizado, incluyendo pero no limitado a análisis con micromatrices, PCR cuantitativa, análisis Southern. En una forma de realización preferida de un procedimiento de clasificación *in vitro* según la invención, el perfil de expresión se determina utilizando una micromatriz. En otra forma de realización preferida de un procedimiento de clasificación *in vitro* según la invención, el perfil de expresión se determina utilizando PCR cuantitativa. En cualquier caso, el nivel de expresión de cualquier gene se normaliza preferentemente en comparación con el nivel de expresión de un gen de control interno, generalmente un gen constitutivo, incluyendo pero no limitado a RNA ribosomal (como por ejemplo ARN ribosomal 18S) o genes tales como la actina o el HRPT. Estas tecnologías son también fácilmente adaptables a cualquier muestra HCC. De hecho, varias tecnologías bien conocidas están disponibles para los expertos en la materia para extraer ARNm de una muestra de tejido y retrotranscribirlo ARNm a ADNc.

En una forma de realización preferida, cuando se utiliza un procedimiento de clasificación *in vitro* que involucra el primer set de genes (ver Tabla 3), el perfil de expresión se determina *in vitro* a nivel nucleico utilizando PCR cuantitativa. En otra forma de realización preferida, cuando se utiliza un procedimiento de clasificación que involucra el segundo set de genes (ver Tabla 4), el perfil de expresión se determina *in vitro* a nivel nucleico utilizando una micromatriz de ácidos nucleicos. En particular, puede utilizarse ventajosamente una micromatriz de Affymetrix U133A.

En cualquier procedimiento de clasificación *in vitro* de un tumor de HCC en los 6 subgrupos definidos por los inventores, para cada subgrupo se calcula una "distancia de subgrupo", que representa una distancia matemática entre dicho tumor de HCC y cada uno de los 6 subgrupos. Cuanto menor sea la distancia entre una muestra y un subgrupo, mayor es la probabilidad de que dicha muestra pertenezca a este subgrupo en particular.

Para una combinación de n genes ($n \geq 8$), el set de todos los tumores puede definirse como un set de n dimensiones en el que cada muestra de tumor puede ser caracterizada por n coordenadas que corresponden a los niveles de expresión de cada gen seleccionado en dicha muestra de tumor. Cada subgrupo o clase es un subconjunto del set de n dimensiones que puede definirse por un punto central y un porcentaje de variación aceptable alrededor de cada coordenada del punto central. Dependiendo de la tecnología utilizada para la determinación del perfil de expresión, pueden elegirse funciones matemáticas apropiadas permitiendo cada una calcular la distancia de cualquier muestra de tumor a uno de los 6 subgrupos o clases.

En particular, cuando el perfil de expresión se determina utilizando PCR cuantitativa, para una muestra_{*i*} de tumor de HCC dada, y un subgrupo o clase_{*k*} particular, la distancia de dicha muestra a dicha clase puede calcularse utilizando la siguiente fórmula (I):

$$Distancia (muestra_i, clase_k) = \sum_{t=1..n} \frac{(\Delta Ct(muestra_i, gen_t) - \mu(clase_k, gen_t))^2}{\sigma(gen_t)} \quad (I)$$

en la que

- n representa el número de genes en el perfil de expresión
- para cada gen_t , $\mu(clase_k, gen_t)$ y $\sigma(gen_t)$ son parámetros que dependen de la combinación de genes elegida y pueden ser calculados mediante optimización en un grupo de entreno de tumores de HCC, seguido de una validación en un grupo de prueba de tumores de HCC, tal como está descrito con más detalles en el Ejemplo 2 de la presente memoria.

Alternativamente, cuando el perfil de expresión se determina a nivel nucleico utilizando una micromatriz de ácidos nucleicos, para una muestra_{*i*} de tumor de HCC dada, y un subgrupo o clase_{*k*} particular, la distancia de dicha muestra_{*i*} a dicha clase_{*k*} puede ser calculada utilizando varias fórmulas derivadas de varios algoritmos bien conocidos por los expertos en la materia. Por ejemplo, dicha distancia de dicha muestra_{*i*} a dicha clase_{*k*} puede ser calculada utilizando la siguiente fórmula (II):

$$Distancia (muestra_i, clase_k) = \left(\frac{\sum_{t=n+1} (c(gen_t, clase_k))^2}{2} + 1,791759 \right) - \left(\sum_{t=n+1} \frac{(y(muestra_i, gene_t) - \mu(gen_t))}{\sigma(gen_t)} \times c(gen_t, clase_k) \right) \quad (II)$$

en la que

- n representa el número de genes en el perfil de expresión,

- para cada gen_t y $clase_k$, y $(muestra_i, gen_t)$ representa el valor de intensidad normalizado para el gen_t en la muestra i , y
- para cada gen_t y $clase_k$, $c(gen_t, clase_k)$, $\mu(gen_t)$ y $\sigma(gen_t)$ son parámetros que dependen de la combinación de genes elegida y pueden ser calculados mediante optimización en un grupo de entreno de tumores de HCC, seguido de una validación en un grupo de prueba de tumores de HCC, tal como está descrito en con más detalles en el Ejemplo 2 de la presente memoria.

La normalización puede realizarse utilizando cualquier procedimiento bien conocido, por ejemplo utilizando normalización RMA.

Para una muestra i dada, una vez se han calculado todas las distancias a todas las clases, la clase predicha para de muestra i se calcula según la siguiente fórmula (III):

$$clase\ predicha(muestra_i) = \arg \min_{k=1..6} (distancia(muestra_i, clase_k)) \quad (III)$$

que significa que la clase predicha para una muestra i dada es la clase para la que la distancia de la muestra i a la clase es la más baja.

En una forma de realización preferida de un procedimiento de clasificación *in vitro* de un tumor de HCC en 6 subgrupos definidos por los inventores utilizando el primer set de genes (ver Tabla 3), el perfil de expresión consiste en la siguiente combinación de genes: RAB1A, REG3A, NRAS, RAMP3, MERTK, PIR, EPHA1, LAMA3, GOS2, HN1, PAK2, AFP, CYP2C9, CDH2, HAMP, y SAE1, y se determina utilizando PCR cuantitativa, en la que cada distancia de una muestra i a una clase k se calcula utilizando la siguiente fórmula (IV):

$$Distancia(muestra_i, clase_k) = \sum_{t=1..16} \frac{(\Delta Ct(muestra_i, gen_t) - \mu(clase_k, gen_t))^2}{\sigma(gen_t)} \quad (IV)$$

en la que para cada gen_t y $clase_k$, los valores $\mu(clase_k, gen_t)$ y $\sigma(gen_t)$ están en un intervalo de 10%, preferentemente de 9%, 8%, 7%, 6%, 5%; o incluso 4%, 3%, 2% o 1% alrededor de los valores mostrados en la siguiente Tabla 5.

Tabla 5. Parámetros utilizados en la fórmula (IV) de distancia de PCR cuantitativa anterior para cada gen y para cada clase.

μ	clase 1	clase 2	clase 3	clase 4	clase 5	clase 6	σ
gen 1 (RAB1A)	-16,39	-16,04	-16,29	-17,15	-17,33	-16,95	0,23
gen 2 (PAP)	-28,75	-27,02	-23,48	-27,87	-19,23	-11,33	16,63
gen 3 (NRAS)	-16,92	-17,41	-16,25	-17,31	-16,96	-17,26	0,27
gen 4 (RAMP3)	-23,54	-23,12	-25,35	-22,36	-23,09	-23,06	1,23
gen 5 (MERTK)	-18,72	-18,43	-21,24	-18,29	-17,03	-16,16	7,23
gen 6 (PIR)	-18,44	-19,81	-16,73	-18,28	-17,09	-17,25	0,48
gen 7 (EPHA1)	-16,68	-16,51	-19,89	-17,04	-18,70	-21,98	1,57
gen 8 (LAMA3)	-20,58	-20,44	-20,19	-21,99	-18,77	-16,85	2,55
gen 9 (GOS2)	-14,82	-17,45	-18,18	-14,78	-17,99	-16,06	3,88
gen 10 (HN1)	-16,92	-17,16	-15,91	-17,88	-17,72	-17,93	0,54
gen 11 (PAK2)	-17,86	-16,56	-16,99	-18,14	-17,92	-17,97	0,58
gen 12 (AFP)	-16,68	-12,36	-26,80	-27,28	-25,97	-23,47	14,80
gen 13 (CYP2C9)	-18,27	-16,99	-16,26	-16,23	-13,27	-14,44	5,47
gen 14 (CDH2)	-15,20	-14,76	-18,91	-15,60	-15,48	-17,32	10,59

μ	clase 1	clase 2	clase 3	clase 4	clase 5	clase 6	σ
gen 15 (HAMP)	-19,53	-20,19	-21,32	-18,51	-25,06	-26,10	13,08
gen 16 (SAE1)	-17,37	-17,10	-16,79	-18,22	-17,72	-18,16	0,31

- 5 En una forma de realización preferida de un procedimiento de clasificación de un tumor de HCC en 6 subgrupos definidos por los inventores utilizando el segundo set de genes (ver Tabla 4), el perfil de expresión consiste en la siguiente combinación de genes: ALDH1L1, CD24, CD74, CFHR3, CYP4F12, DNAJA3, DSCR1, EPHA1, EPHB4, FAAH, FGFR2, FLJ10159, GLT8D1, HAL, MATN2, MRPS7, PAK2, PLXNB1, RAB1A, RHOQ, SLC27A5, SLPI, SMARCE1, STRA13, y se determina a nivel nucleico utilizando cuantitativo una micromatriz, en el que cada distancia de una muestra_i a una clase_k se calcula utilizando la siguiente fórmula (V):

$$Distancia(muestra_i, clase_k) = \left(\frac{\sum_{t=1,24} (c(gen_t, clase_k))^2}{2} + 1,791759 \right) - \left(\sum_{t=1,24} \frac{(y(muestra_i, gene_t) - \mu(gen_t))}{\sigma(gen_t)} \times c(gen_t, clase_k) \right) \quad (V)$$

- 10 en la que para cada gen_t y $clase_k$, los valores $c(gen_t, clase_k)$, $\mu(gen_t)$ y $\sigma(gen_t)$ están en un intervalo de 10%, preferentemente de 9%, 8%, 7%, 6%, 5%; o incluso 4%, 3%, 2% o 1% alrededor de los valores mostrados en la siguiente Tabla 6.

Tabla 6. Parámetros utilizados en la fórmula (V) de distancias de micromatriz anterior para cada gen y para cada clase.

Nº de gen	Símbolo del gen	C						μ	σ
		clase 1	clase 2	clase 3	clase 4	clase 5	clase 6		
1	MATN2	0,70615962	0,194820133	0	0,251226887	-0,43679693	-0,95672006	6,86694444	1,3591753
2	EPHB4	0,772905372	0	-0,23845281	0	0	-0,11554095	8,12722222	0,90851011
3	SLP	0,348202773	1,150217317	-0,30674955	0,772922269	-0,92749405	-1,03709919	9,19555556	1,58143433
4	FAAH	0	-0,05721942	-0,14546431	0,461062372	0	0	7,35527778	0,90656803
5	ALDH1L1	-0,90005519	-0,88021827	0	0,001883282	0,766706969	0,727033118	9,55166667	1,51233112
6	DNAJA3	-0,38971634	0	0	0	0,115122957	0	9,20444444	0,79106866
7	EPHA1	0,084999966	1,076607341	-0,39390753	0,336075517	-0,15411627	-0,94965903	8,34	1,18158329
8	CYP4F12	0	0	-0,50724032	0	0,411786437	0	8,61416667	0,89048235
9	CFHR4 // CFHR3	-0,3047186	0	-0,85911235	0,594911312	0,240167858	0	9,63111111	1,9450676
10	FGFR2	1,107271104	1,414413219	0	-0,33319989	-0,93870859	-0,78623447	6,45694444	1,51938352
11	CD24	0,877735471	0,272141136	0,391097168	-0,19493988	-0,09580985	-1,25022405	6,88527778	1,84942286
12	RAB1A	0	0,181176347	0,101889323	0	-0,13052075	0	10,9441667	0,75674426
13	PAK2	0,333394856	0	0,506491717	-0,14545539	-0,07557739	-0,39796907	7,58722222	1,0494811
14	STRA13	-0,61853503	-0,18082898	0,428679271	-0,14603227	0,113677426	0,403039586	9,56055556	0,91004758
15	CD74	-0,20143371	0	0	0	-0,40543025	0,741093354	10,9897222	1,28270148
16	SMARCE1	0,126430939	0	0,308479215	-0,17753221	0	-0,25610041	8,43638889	0,86973267
17	RHOQ	0	0,321593401	0,234570894	0	-0,196463	-0,09749623	8,02888889	0,97675881
18	DSCR1	-0,04133466	-0,10070514	-0,07191945	0,033538532	0	0,436538207	7,40194444	0,92638619
19	PLXNB1	0,550893643	0,055723559	-0,24782026	-0,15487024	0	-0,08389023	8,10722222	0,98619313
20	HAL	0,428036608	1,655082264	-0,24462171	0,139158147	-0,85676893	-1,12088638	7,73055556	1,21158219
21	MRPS7	-0,24232641	0	0,552001257	0	0	0	8,77777778	0,88623422
22	GLT8D1	-0,08380816	-0,21624193	-0,11283474	0	0,458734818	0,34625682	9,20027778	0,91869819
23	FLJ10159	1,275819214	0,257162026	0	-0,03229518	-0,27448613	-0,99690117	7,33638889	1,60341808
24	SLC27A5	-0,93094948	-0,2491716	-0,50142942	0,607483755	0,535897078	0,538169671	9,89388889	1,46675337

La presente descripción también divulga un procedimiento de pronóstico *in vitro* de supervivencia global y/o supervivencia sin recaída a partir de una muestra de hígado HCC de un sujeto que sufre HCC, que comprende:

- determinar un perfil de expresión que comprende una combinación de por lo menos 2, por lo menos 3, por lo menos 4 o por lo menos 5 genes seleccionados de entre un grupo que consiste en NRCAM, PIR, RAMP3, SLC21A2, TAF9, TNA, HN1, PSMD1, MRPS7, CDC20, ENO1, HLF, STRA13, RAGD, NRAS, ARFGEF2, RAB1A, G0S2, SMAD3, DNAJA3, HELO1, RHOQ, C14orf156, NPEPPS, PDCCD2, PHB, KIAA0090, IMP-3,

KPNA2, KIAA0268, UNQ6077, LOC440751, G6PD, STK6, TFRC, GLA, TRIP13, SPP1, AKR1C1, AKR1C2, GIMAP5, ADM, CCNB1, TKT, AGPS, RAN, NUDT9, HRASLS3, HLA-DQA1, NEU1, RARRES2, PAPOLA, ABCB6, BIRC5, FLJ20273, C14orf109, CHKA, TUBB2, HMGB3, TXNRD1, IFITM1, KIAA0992, MPPE1, KLRB1, CCL5, SYNE1, DNASE1L3, CYP2C18, PACSIN2, PON3, y PPP2R1B;

- b) calcular a partir de dicho perfil de expresión una puntuación de supervivencia global y/o una puntuación de supervivencia sin recaída; y
- c) comparar la puntuación de supervivencia global y/o la puntuación de supervivencia sin recaída obtenidas, cada una con un valor umbral, de manera que
- una puntuación de supervivencia global/de supervivencia sin recaída inferior a dicho valor umbral indica un buen pronóstico de supervivencia/supervivencia sin recaída, mientras que
 - una puntuación de supervivencia global/de supervivencia sin recaída superior o igual a dicho valor umbral indica un mal pronóstico de supervivencia/supervivencia sin recaída.

Las características principales de los genes implicados en el pronóstico de HCC están descritas en la siguiente Tabla 7.

Tabla 7. Genes implicados en el pronóstico de HCC

Símbolo del gen*	Nombre del gen HUGO *	ID del gen "Entrez Gene" **
NRCAM	Molécula de adhesión de célula neuronal	4897
PIR	pirina (proteína nuclear de unión a hierro)	8544
RAMP3	proteína modificadora de actividad de receptor (calcitonina) 3	10268
SLCO2A1	familia transportadores de aniones orgánicos transportadores de solutos, miembro 2A1	6578
TAF9	TAF9 ARN polimerasa II, factor asociado a proteína de unión a TATA box (TBP), 32kDa	6880
CLEC3B	Familia dominios lectina tipo-C 3, miembro B	7123
HN1	expresado hematológico y neurológico 1	51155
PSMD1	proteosoma (prosoma, macropain) subunidad 26S, no-ATPasa, 1	5707
MRPS7	Proteína ribosomal mitocondrial S7	51081
CDC20	Homólogo ciclo división celular 20 (S. cerevisiae) CDC20	991
ENO1	enolasa 1, (alfa)	2023
HLF	factor de leucemia hepática	3131
STRA13	Homólogo estimulado por ácido retinoico 13 (ratón)	201254
RRAGD	D de unión a GTP asociado a ras	58528
NRAS	Homólogo oncogen viral RAS (v-ras) de neuroblastoma	4893
ARFGEF2	Factor ribosilación de ADP factor intercambio nucleótido guanina 2 (inhibido por brefeldina A)	10564
RAB1A	RAB1A, miembro familia oncogenes RAS	5861
G0S2	G0/G1 interruptor 2	50486
SMAD3	SMAD, homólogo madres contra DPP 3 (Drosophila)	4088
DNAJA3	Homólogo DnaJ (Hsp40), subfamilia A, miembro 3	9093
ELOVL5	miembro 5 familia ELOVL, elongación de cadena larga de ácidos grasos (FEN1/Elo2, tipo SUR4/Elo3, levadura)	60481
RHOQ	Familia de genes homólogos a ras, miembro Q	23433
C14orf156	cromosoma 14 patrón abierto de lectura 156	81892
NPEPPS	Amino peptidasa sensible a puromicina	9520
PDCD2	Muerte celular programada 2	5134
PHB	prohibitina	5245
KIAA0090	KIAA0090	23065
IMP-3	Proteína de unión a IGF-II ARNm 3	10643
KPNA2	carioferina alfa 2 (cohorte RAG 1, importina alfa 1)	3838
KIAA0268	Péptido reactivo a C219	348477
UNQ6077	AAAP6077	375056
LOC440751	similar a péptido reactivo C219	440751
G6PD	glucosa-6-fosfato deshidrogenasa	2539
STK6	serina/treonina quinasa 6	6790
TFRC	Receptor transferrina (p90, CD71)	7037
GLA	galactosidasa, alfa	2717

Símbolo del gen*	Nombre del gen HUGO *	ID del gen "Entrez Gene" **
TRIP13.	Interactuador receptor hormona tiroidea 13	9319
SPP1	fosfoproteína secretada 1 (osteopontina, sialoproteína de hueso I, activación temprana de linfocitos T 1)	6696
AKR1C1	Familia aldo-ceto reductasa 1, miembro C1 (dihidrodiol deshidrogenasa 1; 20-alfa (3-alfa)-hidroxiesteroide deshidrogenasa)	1645
AKR1C2	Familia aldo-ceto reductasa 1, miembro C2 (dihidrodiol deshidrogenasa 2; proteína de unión a ácido biliar; 3-alfa hidroxiesteroide deshidrogenasa, tipo III)	1646
GIMAP5	GTPasa, familia IMAP miembro 5	55340
ADM	adrenomedulina	133
CCNB1	Ciclina B1	891
TKT	transcetolasa (síndrome Wernicke-Korsakoff)	7086
AGPS	alquilglicerona fosfato sintasa	8540
RAN	RAN, miembro familia oncogenes RAS	5901
NUDT9	nudix (motivo de tipo residuo unido a nucleosidos difosfato X) 9	53343
HRASLS3	supresor de tipo HRAS 3	11145
HLA-DQA1	Complejo principal histocompatibilidad, clase II, DQ alfa 1	3117
NEU1	sialidasa 1 (lialidasa lisosomal)	4758
RARRES2	respondedor de receptor de ácido retinoico (inducido por tazaroteno) 2	5919
PAPOLA	poli(A) polimerasa alfa	10914
ABCB6	Casete de unión a ATP, subfamilia B (MDR/TAP), miembro 6	10058
BIRC5	IAP baculoviral que contiene repeticiones 5 (survivina)	332
FLJ20273	Proteína de unión a RNA	54502
C14orf109	Patrón abierto de lectura 109 cromosoma 14	26175
CHKA	colina quinasa alfa	1119
TUBB2	Tubulina, beta 2	7280
HMGB3	Caja grupo alta movilidad 3	3149
TXNRD1	Tioredoxina reductasa 1	7296
IFITM1	proteína transmembrana inducida por interferon 1 (9-27)	8519
KIAA0992	paladina	23022
MPPE1	Metalofosfoesterasa 1	65258
KLRB1	subfamilia B de receptores de tipo lecitina de célula asesina, miembro 1	3820
CCL5	Quimiocina (motivo C-C) ligando 5	6352
SYNE1	espectrina que contiene repeticiones, envuelta nuclear 1	23345
DNASE1L3	deoxiribonucleasa de tipo I 3	1776
CYP2C18	citocromo P450, familia 2, subfamilia C, polipéptido 18	1562
PACSIN2	Sustrato proteína quinasa C y caseína quinasa en neuronas 2	11252
PON3	paraoxonasa 3	5446
PPP2R1B	proteína fosfatasa 2 (anteriormente 2A), subunidad reguladora A (PR 65), isoforma beta	5519

* Todos los símbolos y nombres de genes son según el Comité de Nomenclatura de Genes HUGO (disponible en <http://www.gene.ucl.ac.uk/nomenclature/>)

- 5 ** Toda la información disponible en relación con los genes listados de la Tabla 7 puede recuperarse del portal "Entrez Gene" (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=gene>) utilizando la ID de gen de "Entrez Gene" proporcionada en la Tabla 7.

- 10 Preferentemente, en un procedimiento de pronóstico *in vitro* divulgado en la presente memoria, el perfil de expresión comprende y preferentemente consiste en una combinación de por lo menos 2, por lo menos 3, por lo menos 4, o por lo menos 5 genes seleccionados de entre un grupo que consiste en NRCAM, PIR, RAMP3, SLC21A2, TAF9, TNA, HN1, PSMD1, MRPS7, CDC20, ENO1, HLF, STRA13, RAGD, NRAS, ARFGF2, RAB1A, G0S2, SMAD3, DNAJA3, HELO1, y RHOQ.
- 15 En la presente descripción, un "pronóstico" de la evolución de HCC significa una predicción de la evolución futura de un tumor de HCC en particular en relación con un paciente que sufre de este tumor de HCC particular. Los procedimientos de la presente memoria permiten ambos, un pronóstico de supervivencia global y un pronóstico de supervivencia sin recaída.
- 20 Por "pronóstico de supervivencia global" se pretende decir pronóstico de supervivencia, con o sin recaída. Como se ha dicho anteriormente, el principal tratamiento hoy en día contra HCC es la extirpación quirúrgica del tumor. Como resultado, un "mal pronóstico de supervivencia global" se define como la ocurrencia de muerte en los 3 años

posteriores a la extirpación, mientras que un “buen pronóstico de supervivencia global” se define como la ausencia de muerte en los 5 años post-operación.

Por “pronóstico de supervivencia sin recaída” se pretende decir un pronóstico de supervivencia en ausencia de cualquier recaída. Un “mal pronóstico de supervivencia sin recaída” se define como la presencia de una recaída del tumor durante los dos años posteriores a la extirpación del hígado, mientras que un “buen pronóstico de supervivencia sin recaída” se define como la ausencia de recaída durante los 4 años post-operación.

Preferentemente, en un procedimiento de pronóstico *in vitro* de supervivencia global divulgado en la presente memoria, el perfil de expresión comprende y preferentemente consiste en una combinación de genes seleccionada de entre:

- TAF9, PIR, NRCAM y RAMP3,
- TAF9, NRCAM, SLC21A2 y PSMD1,
- TAF9, NRCAM, RAMP3 y PSMD1,
- TAF9, NRCAM, NRAS, RAMP3 y PSMD1, o
- TAF, NRCAM, RAMP3, PSMD1 y ARFGEF2.

Más preferentemente, en un procedimiento de pronóstico *in vitro* de supervivencia global divulgado en la presente memoria, el perfil de expresión comprende y preferentemente consiste en la siguiente combinación de genes:

- TAF9, NRCAM, RAMP3, PSMD1 y ARFGEF2.

Alternativamente, en un procedimiento de pronóstico *in vitro* de supervivencia sin recaída divulgado en la presente memoria, el perfil de expresión comprende y preferentemente consiste en una combinación de genes seleccionada de entre:

- TAF9 y G0S2,
- TAF9, NRCAM y RAMP3,
- TAF9, G0S2 y RAMP3,
- TAF9, NRCAM, DNAJA3 y RAMP3, o
- TAF9, NRCAM, G0S2, DNAJA3 y RAMP3.

Más preferentemente, en un procedimiento de pronóstico *in vitro* de supervivencia sin recaída divulgado en la presente memoria, el perfil de expresión comprende y preferentemente consiste en la siguiente combinación de genes:

- TAF9, NRCAM y RAMP3

Puede hacerse referencia a una combinación particular de genes con el término predictor.

Para un procedimiento de pronóstico *in vitro* de supervivencia global y/o de supervivencia sin recaída divulgado en la presente memoria, la “muestra de hígado HCC” puede ser cualquier muestra de hígado que comprende tejido de tumor de HCC. Preferentemente, en un procedimiento de pronóstico *in vitro* de supervivencia global y/o de supervivencia sin recaída divulgado en la presente memoria, la muestra de hígado HCC es una biopsia de hígado HCC o una extirpación quirúrgica de un tumor de HCC.

Al igual que para los procedimientos de clasificación *in vitro*, en un procedimiento de pronóstico *in vitro* de supervivencia global y/o de supervivencia sin recaída divulgado en la presente memoria, el nivel de expresión de cada gen puede ser determinado *in vitro* tanto a nivel proteico como a nivel nucleico, utilizando cualquier tecnología conocida de la técnica, en particular cualquier tecnología descrita anteriormente.

Preferentemente, en un procedimiento de pronóstico *in vitro* de supervivencia global y/o de supervivencia sin recaída divulgado en la presente memoria, el perfil de expresión se determina *in vitro* a nivel nucleico. Preferentemente, el perfil de expresión se determina utilizando una micromatriz. También preferentemente, el perfil de expresión se determina utilizando PCR cuantitativa. En cualquier caso, el nivel de expresión de cualquier gen es preferentemente normalizado en comparación con el nivel de expresión de un gen de control interno, generalmente un gen constitutivo, incluyendo pero no limitado a ARN ribosomal (tal como por ejemplo ARN ribosomal 18S) o genes como la actina o el HPRT.

En cualquier procedimiento de pronóstico *in vitro* de supervivencia global y/o supervivencia sin recaída divulgados en la presente memoria, se calcula una “puntuación” de supervivencia global y/o supervivencia sin recaída. Para una combinación de n genes ($n \geq 2$) seleccionada, una “puntuación” es una función logística que tiene en cuenta los n niveles de expresión de cada gen seleccionado en dicha muestra de tumor, balanceado por parámetros que dependen de la combinación de genes elegida y que pueden ser calculados mediante optimización sobre un grupo de entrenamiento de tumores de HCC, seguido de una validación sobre un grupo de prueba de tumores de HCC, tal

y como está explicado con más detalle en el Ejemplo 3 de la presente memoria.

En función de la tecnología utilizada para la determinación del perfil de expresión, puede determinarse una función de puntuación apropiada con parámetros adecuados.

En particular, cuando el perfil de expresión se determina utilizando PCR cuantitativa, para una muestra_i dada, se puede calcular una "puntuación" de supervivencia global o de supervivencia sin recaída utilizando la siguiente fórmula:

$$Puntuación(muestra_i) = \sum_{t=1..n} \beta(gen_t) \cdot (2^{-\Delta Ct(muestra_i, gen_t)} - \mu(gen_t))$$

en la que

- n representa el número de genes en el perfil de expresión,
- para cada gen_t , $\beta(gen_t)$ y $\mu(gen_t)$ son parámetros que dependen de la combinación de genes elegida y pueden ser calculados mediante optimización sobre un grupo de entrenamiento de tumores de HCC, seguido de una validación sobre un grupo de prueba de tumores de HCC, tal y como está explicado con más detalle en el Ejemplo 3 de la presente memoria.

En cualquier procedimiento de pronóstico *in vitro* de supervivencia global o de supervivencia sin recaída divulgado en la presente memoria, la puntuación o puntuaciones de supervivencia global y/o de supervivencia si recaída obtenidas se comparan entonces con por lo menos un valor umbral, que determina si el pronóstico es malo o bueno.

Para una combinación de genes dada en el perfil de expresión, dicho valor umbral puede ser determinado utilizando el mismo procedimiento que para los parámetros $\beta(gen_t)$ y $\mu(gen_t)$, es decir, mediante optimización sobre un grupo de entrenamiento de tumores de HCC, seguido de una validación sobre un grupo de prueba de tumores de HCC, tal y como está explicado con más detalle en el Ejemplo 3 de la presente memoria.

Para un valor umbral dado, el pronóstico de la muestra será:

- un mal pronóstico: si su puntuación es superior o igual a dicho valor umbral, o
- un buen pronóstico: si su puntuación es estrictamente inferior a dicho valor umbral.

Preferentemente, en un procedimiento de pronóstico *in vitro* de supervivencia global divulgado en la presente memoria, el perfil de expresión consiste en una combinación de genes de la siguiente Tabla 8 y se determina utilizando PCR cuantitativa y la siguiente fórmula:

$$Puntuación supervivencia global(muestra_i) = \sum_{t=1..n} \beta(gen_t) \cdot (2^{-\Delta Ct(muestra_i, gen_t)} - \mu(gen_t))$$

en la que:

- n representa el número de genes de la combinación
- t representa el número de cada gen en la combinación mostrada en la siguiente Tabla 8, y
- el valor de cada uno de los coeficientes $\beta(gen_t)$ y $\mu(gen_t)$ está en un intervalo de 10%, preferentemente 9%, 8%, 7%, 6%, 5%; o incluso 4%, 3%, 2% o 1% alrededor de los valores mostrados en la Tabla 8.

Tabla 8. Combinación de genes (predictor) preferida y parámetros para la determinación de la puntuación global.

Predictor 1				
Número de gen	Símbolo de gen	μ	β	Valor umbral
1	TAF9	7,28	0,129	-0,393
2	NRCAM	1,59	0,252	
3	RAMP3	0,14	-6,133	
4	PSMD1	4,66	0,024	
5	ARFGEF2	3,66	-0,025	

Preferentemente, el valor umbral utilizado para el pronóstico de supervivencia global está en un intervalo de 10%, preferentemente 9%, 8%, 7%, 6%, 5%; o incluso 4%, 3%, 2% o 1% alrededor de los valores mostrados en la Tabla 8.

Preferentemente, en un procedimiento de pronóstico *in vitro* de supervivencia sin recaída divulgado en la presente memoria, el perfil de expresión consiste en una combinación de genes de la siguiente Tabla 9 y se determina utilizando PCR cuantitativa y la siguiente fórmula:

$$\text{Puntuación supervivencia sin recaída (muestra}_i\text{)} = \sum_{t=1..n} \beta(\text{gen}_t) \cdot (2^{-\Delta\text{Ct}(\text{muestra}_i, \text{gen}_t)} - \mu(\text{gen}_t))$$

en la que

- n representa el número de genes en la combinación
- t representa el número de cada gen en la combinación mostrada en la siguiente Tabla 9, y
- el valor de cada uno de los coeficientes $\beta(\text{gen}_t)$ y $\mu(\text{gen}_t)$ está en un intervalo de 10%, preferentemente 9%, 8%, 7%, 6%, 5%; o incluso 4%, 3%, 2% o 1% alrededor de los valores mostrados en la Tabla 9.

Tabla 9. Combinación de genes (predictor) preferida y parámetros para la determinación de la puntuación global.

Predictor 1				
Número de gen	Símbolo de gen	μ	β	Valor umbral
1	TAF9	7,28	0,127	-0,461
2	NRCAM	1,59	0,196	
3	RAMP3	0,14	-3,886	

Preferentemente, el valor umbral utilizado para el pronóstico está en un intervalo de 10%, preferentemente 9%, 8%, 7%, 6%, 5%; o incluso 4%, 3%, 2% o 1% alrededor de los mostrados en la Tabla 9.

Preferentemente, los valores de $\Delta\text{Ct}(\text{muestra}_i, \text{gen}_t)$ se calculan en relación a ARN ribosomal 18S (R18S).

La presente descripción divulga además un procedimiento para el diagnóstico/determinación *in vitro* respecto a si es recomendable la terapia adyuvante a partir de una muestra de hígado HCC de un sujeto que sufre HCC, que comprende:

- a) determinar un pronóstico de supervivencia global y/o de supervivencia sin recaída según cualquier procedimiento divulgado en la presente memoria y
- b) determinar si es recomendable la terapia adyuvante a partir de dicho pronóstico, de manera que:
 - en presencia de un mal pronóstico, se recomienda la terapia adyuvante, mientras que
 - en ausencia de un mal pronóstico, no se recomienda la terapia adyuvante.

Por "terapia adyuvante" se pretende decir una terapia antitumoral adicional que puede ser administrada a un sujeto que sufre HCC después de la extirpación quirúrgica del tumor de HCC. Las terapias adyuvantes pueden incluir, sin estar limitadas a, quimioterapia y radioterapia.

La presente invención también se refiere a un kit para la clasificación *in vitro* de un tumor de HCC en 6 subgrupos G1, G2, G3, G4, G5 y G6 a partir de una muestra de hígado HCC de un sujeto que sufre HCC, que consiste en reactivos para la determinación de un perfil de expresión que consiste en:

- (i) la siguiente combinación de 16 genes: RAB1A, REG3A, NRAS, RAMP3, MERTK, PIR, EPHA1, LAMA3, G0S2, HN1, PAK2, AFP, CYP2C9, CDH2, HAMP, y SAE1, o
- (ii) la siguiente combinación de 24 genes: ALDH1L1, CD24, CD74, CFHR3, CYP4F12, DNAJA3, DSCR1, EPHA1, EPHB4, FAAH, FGFR2, FLJ10159, GLT8D1, HAL, MATN2, MRPS7, PAK2, PLXNB1, RAB1A, RHOQ, SLC27A5, SLPI, SMARCE1, STRA13,

en el que los 6 subgrupos G1, G2, G3, G4, G5 y G6 están definidos por sus características clínicas y genéticas descritas en la Tabla 2.

La presente descripción también divulga un kit para la clasificación *in vitro* de un tumor HCC en 6 subgrupos a partir de una muestra de hígado HCC de un sujeto que sufre HCC, que comprende reactivos para la determinación *in vitro* de un perfil de expresión que comprende o que consiste en una combinación de por lo menos 8, por lo menos 10, por lo menos 12, por lo menos 14, o por lo menos 16 genes seleccionados de entre el grupo que consiste en: RAB1A, REG3A, NRAS, RAMP3, MERTK, PIR, EPHA1, LAMA3, G0S2, HN1, PAK2, AFP, CYP2C9, CDH2, HAMP, SAE1, ADH6, DCN, FLJ10159, ALDH1L1, IGF1, LECT2, SLC38A1, SPARCL1, CTNNA2, GLUL, LEF1, MATN2,

MME, PFN2, SPINT2, TBX3, y FGFR2. Estos genes corresponden al primer set de genes (ver Tabla 3) identificados como útiles para clasificar tumores HCC en los subgrupos G1 a G6 tal como están definidos por sus características clínicas y genéticas de la Tabla 2.

Un kit para la clasificación *in vitro* de un tumor de HCC divulgada en la presente memoria comprende los reactivos para la determinación *in vitro* de un perfil de expresión que comprende o que consiste en una combinación de por lo menos 8, por lo menos 10, por lo menos 12, por lo menos 14, o por lo menos 16 genes seleccionados de entre el grupo que consiste en: RAB1A, REG3A, NRAS, RAMP3, MERTK, PIR, EPHA1, LAMA3, G0S2, HN1, PAK2, AFP, CYP2C9, CDH2, HAMP, SAE1, ADH6, DCN, FLJ10159, ALDH1L1, IGF1, LECT2, SLC38A1, SPARCL1.

Más preferentemente, un kit para la clasificación *in vitro* de un tumor de HCC divulgado en la presente memoria comprende los reactivos para la determinación *in vitro* de un perfil de expresión que comprende o que consiste en una combinación de por lo menos 8, por lo menos 10, por lo menos 12, por lo menos 14, o por lo menos 16 genes seleccionados de entre el grupo que consiste en: RAB1A, REG3A, NRAS, RAMP3, MERTK, PIR, EPHA1, LAMA3, G0S2, HN1, PAK2, AFP, CYP2C9, CDH2, HAMP, y SAE1.

La presente descripción divulga además un kit para la clasificación *in vitro* de un tumor de HCC en 6 subgrupos a partir de una muestra de hígado de HCC de un sujeto que sufre HCC, que comprende los reactivos para la determinación *in vitro* de un perfil de expresión que comprende o que consiste en una combinación de por lo menos 8, por lo menos 10, por lo menos 12, por lo menos 14, por lo menos 16, por lo menos 18, por lo menos 20, por lo menos 22 o por lo menos 24 genes seleccionados de entre el grupo que consiste en: ALDH1L1, CD24, CD74, CFHR3, CYP4F12, DNAJA3, DSCR1, EPHA1, EPHB4, FAAH, FGFR2, FLJ10159, GLT8D1, HAL, MATN2, MRPS7, PAK2, PLXNB1, RAB1A, RHOQ, SLC27A5, SLPI, SMARCE1, STRA13. Estos genes corresponden al segundo set de genes (ver Tabla 4) identificados como útiles para clasificar tumores de HCC en subgrupos G1 a G6 tal como están definidos por sus características clínicas y genéticas de la Tabla 2.

La presente descripción divulga un kit para el pronóstico *in vitro* de supervivencia global y/o de supervivencia sin recaída a partir de una muestra de hígado HCC de un sujeto que sufre HCC, que comprende los reactivos para la determinación *in vitro* de un perfil de expresión que comprende o que consiste en una combinación de por lo menos 2, por lo menos 3, por lo menos 4 o por lo menos 5 genes seleccionados de entre el grupo que consiste en NRCAM, PIR, RAMP3, SLC21A2, TAF9, TNA, HN1, PSMD1, MRPS7, CDC20, ENO1, HLF, STRA13, RAGD, NRAS, ARFGEF2, RAB1A, G0S2, SMAD3, DNAJA3, HELO1, RHOQ, C14orf156, NPEPPS, PDCD2, PHB, KIAA0090, IMP3, KPNA2, KIAA0268, UNQ6077, LOC440751, G6PD, STK6, TFRC, GLA, TRIP13, SPP1, AKR1C1, AKR1C2, GIMAP5, ADM, CCNB1, TKT, AGPS, RAN, NUDT9, HRASLS3, HLA-DQA1, NEU1, RARRES2, PAPOLA, ABCB6, BIRC5, FLJ20273, C14orf109, CHKA, TUBB2, HMGB3, TXNRD1, IFITM1, KIAA0992, MPPE1, KLRB1, CCL5, SYNE1, DNASE1L3, CYP2C18, PACSIN2, PON3, y PPP2R1B.

Preferentemente, el kit divulgado en la presente memoria para el pronóstico *in vitro* de supervivencia global y/o de supervivencia sin recaída a partir de una muestra de hígado HCC de un sujeto que sufre HCC, comprende reactivos para la determinación *in vitro* de un perfil de expresión que comprende o que consiste en una combinación de por lo menos 2, por lo menos 3, por lo menos 4 o por lo menos 5 genes seleccionados de entre el grupo que consiste en NRCAM, PIR, RAMP3, SLC21A2, TAF9, TNA, HN1, PSMD1, MRPS7, CDC20, ENO1, HLF, STRA13, RAGD, NRAS, ARFGEF2, RAB1A, G0S2, SMAD3, DNAJA3, HELO1, y RHOQ.

Más preferentemente, el kit divulgado en la presente memoria para el pronóstico *in vitro* de supervivencia global a partir de una muestra de hígado HCC de un sujeto que sufre HCC, comprende reactivos para la determinación *in vitro* de un perfil de expresión que comprende o que consiste en una de las combinaciones de genes siguientes:

- TAF9, PIR, NRCAM y RAMP3,
- TAF9, NRCAM, SLC21A2 y PSMD1,
- TAF9, NRCAM, RAMP3 y PSMD1,
- TAF9, NRCAM, NRAS, RAMP3 y PSMD1, o
- TAF9, NRCAM, RAMP3, PSMD1 y ARFGEF2.

Más preferentemente, el kit divulgado en la presente memoria para el pronóstico *in vitro* de supervivencia global a partir de una muestra de hígado HCC de un sujeto que sufre HCC, comprende reactivos para la determinación *in vitro* de un perfil de expresión que comprende o que consiste en la siguiente combinación de genes:

- TAF9, NRCAM, RAMP3, PSMD1 y ARFGEF2.

Preferentemente, el kit divulgado en la presente memoria para el pronóstico *in vitro* de supervivencia sin recaída a partir de una muestra de hígado HCC de un sujeto que sufre HCC comprende reactivos para la determinación *in vitro* de un perfil de expresión que comprende o que consiste en una de las combinaciones de genes siguientes:

- TAF9 y G0S2,
- TAF9, NRCAM y RAMP3,

- TAF9, G0S2 y RAMP3,
- TAF9, NRCAM, DNAJA3 y RAMP3, o
- TAF9, NRCAM, G0S2, DNAJA3, y RAMP3.

5 Más preferentemente, el kit divulgado en la presente memoria para el pronóstico *in vitro* de supervivencia sin recaída a partir de una muestra de hígado HCC de un sujeto que sufre HCC comprende reactivos para la determinación *in vitro* de un perfil de expresión que comprende o que consiste en la siguiente combinación de genes:

- TAF9, NRCAM y RAMP3.

10 En un kit para el pronóstico *in vitro* de supervivencia global y/o de supervivencia sin recaída divulgado en la presente memoria, los reactivos que se proporcionan pueden permitir el pronóstico de sólo supervivencia global o de sólo supervivencia sin recaída, o pueden permitir el pronóstico de ambos supervivencia global y supervivencia sin recaída.

15 En cualquier kit según la invención, los reactivos para la determinación de un perfil de expresión pueden incluir cualquier reactivo útil para la determinación de un nivel de expresión de un gen. Dicha determinación del nivel de expresión puede realizarse a nivel proteico o a nivel nucleico.

20 Reactivos adecuados para la determinación de un perfil de expresión a nivel proteico incluyen, sin estar limitados a, anticuerpos y fragmentos de anticuerpos, reactivos para análisis por espectrometría de masas, y micromatrices de proteínas.

25 A la inversa, reactivos adecuados para la determinación de un perfil de expresión a nivel nucleico incluyen, sin estar limitados a, cebadores de amplificación, sondas nucleicas y micromatrices de ácidos nucleicos.

30 En particular, en un kit para la clasificación de HCC que comprende reactivos para determinar un perfil de expresión que implica genes del primer set de genes útiles para la clasificación de HCC (ver tabla 3), dicho kit puede ventajosamente comprender cebadores de amplificación, y opcionalmente sondas nucleicas útiles para el análisis por PCR cuantitativa de la expresión génica. Dicho kit también puede contener opcionalmente otros reactivos de PCR cuantitativa útiles.

35 De forma alternativa, en un kit para la clasificación de HCC que comprende reactivos para determinar un perfil de expresión que involucra genes del segundo set de genes útiles para la clasificación de HCC (ver Tabla 4), dicho kit puede ventajosamente comprender una micromatriz de ácidos nucleicos, y opcionalmente otros reactivos útiles para el análisis con micromatriz de la expresión génica.

40 Adicionalmente, en cualquier kit según la invención, dichos reactivos pueden estar provistos de instrucciones para llevar a cabo un procedimiento de clasificación *in vitro* según la invención. Por ejemplo, dichas instrucciones pueden:

- 1) permitir al usuario por si mismo llevar a cabo la clasificación o pronóstico, por ejemplo proporcionando las fórmulas necesarias y varios valores de parámetros, o
- 2) instruir al usuario para entrar sus datos de expresión en un software dedicado que puede proporcionarse en el kit o puede, por ejemplo, ser accesible por Internet.

En este caso, los reactivos e instrucciones pueden proporcionarse juntos en el mismo paquete o de forma separada en dos paquetes distintos.

50 La presente descripción divulga además un procedimiento de tratamiento de un sujeto que sufre de HCC, que comprende:

- a) determinar la supervivencia global y/o supervivencia sin recaída a partir de una muestra de hígado HCCC de dicho sujeto según un procedimiento divulgado en la presente memoria, y
- b) administrar a dicho sujeto una terapia adyuvante en presencia de un mal pronóstico, mientras que no administrar dicha terapia adyuvante en ausencia de un mal pronóstico.

60 La presente descripción también divulga un procedimiento de criba *in vitro* de compuestos útiles para el tratamiento de uno de los 6 subgrupos de HCC divulgados en la presente memoria, que comprende:

- a) proporcionar muestras de tumor de HCC,
- b) clasificar dichas muestras de tumor de HCC según un procedimiento de la invención, y
- c) probar la habilidad de dichos compuestos de inhibir el crecimiento *in vitro* de las muestras de tumor de HCC

que han sido clasificadas en dicho subgrupo de HCC:

La presente descripción divulga además un procedimiento de tratamiento de un sujeto que sufre HCC, que comprende:

- a) clasificar dicha muestra de tumor de HCC de un sujeto en uno de los 6 subgrupos divulgados en la presente memoria, y
- b) administrar a dicho sujeto un tratamiento terapéutico orientado al subgrupo de HCC al cual pertenece la muestra de tumor de HCC.

Habiendo descrito de forma general la invención, un entendimiento adicional de las características y ventajas de la invención puede obtenerse mediante referencia a ciertos ejemplos específicos 1-2 y figuras 1-4 que se proporcionan en la presente memoria con propósitos de ilustración.

El Ejemplo 3 y la figura 5 ilustran contenido que no se reivindica en la presente solicitud.

Descripción de las figuras

Figura 1. Esquema de los diferentes subgrupos de HCC definidos por análisis transcriptómico con sus clínica, genética y vías relacionadas.

G1 a G6 indica los subgrupos de HCC de tumores definidos por análisis transcriptómico. Las líneas verticales indican características asociadas significativas (ver Tabla 1). LOH, pérdida de homocigosis; Hemocrom, hemocromatosis; AFP, alfa-fetoproteína, HBV, virus hepatitis B. Las líneas sólidas y punteadas subrayando palabras principalmente indican genes sobre- y subexpresados, respectivamente, en esa categoría funcional particular.

Figura 2. Agrupamiento jerárquico no supervisado.

El dendograma mostrado se obtuvo en base al perfil de expresión de 6.176 sets de sondas de datos de Affymetrix provenientes de 57 tumores de HCC, 3 adenomas y 5 cuando positivo y negativo respectivamente. En caso de infección por HBV, las cajas grises indican un bajo número de copias de ADN viral. FAL indica Pérdida Alélica Fraccional (negro indica tumores que contienen la delección de más de 5 brazos cromosómicos (FAL > 0,128)). Otras abreviaciones son: CTNNB1, gen β -catenina; Mut, mutación; met, metilación; nódulos sat., nódulos satélite a menos de 1 cm del tumor principal; AFP, fetoproteína alfa; CDH1, gen de E-cadherina; inv. portal, invasión portal.

Figura 3: Caracterización de un número seleccionado de genes específicos del subgrupo de HCC G1 utilizando QRT-PCR.

- a. Validación de genes específicamente sobreexpresados en el subgrupo G1 de HCC predicho. Se analizaron *IGF2* (Factor de crecimiento de insulina 2), *AFP* (feto proteína alfa), *SOX9* (región Y-box9 determinada por sexo), *MYH4* (Cadena pesada de miosina IIb), *PEG1* y *PEG3* (expresado paternalmente 1 y 3) en 109 HCC. Los diagramas de cajas y bigotes muestran (extendiéndose desde el percentil 25 al percentil 75 con una línea en la mediana (percentil 50) el rango de valores de expresión log10 relativos (tumor versus media de 21 no tumores (T/NT)) obtenidos para el gen indicado en cada uno de los 6 subgrupos predichos (G1 a G6), en 21 muestras no tumorales (NT) y en 19 muestras de hígados fetales (FL). Los bigotes se extienden por encima y por debajo de la caja para mostrar los valores superior e inferior. Los P-valores a partir de pruebas ANOVA que comparan los valores de expresión en los diferentes subgrupos de HCC están indicados debajo del símbolo del gen.
- b. Validación de genes sobreexpresados en tumores mutados en *PIK3CA* (*PIK3CA* mut) comparados con 107 HCC no mutados (*PIK3CA* NM) para *EEF1A2* (factor de elongación de la traducción eucariótico 1 alfa) y *PRSS7* (precursor enteroquinasa). Los P-valores resultantes de una prueba T comparando muestras mutadas y no mutadas se muestran debajo del símbolo del gen.

Figura 4: Caracterización de tumores de HCC que lleva a los subgrupos G5 y G6.

- a. Validación de genes especialmente sobreexpresados en los subgrupos G5 y G6 de HCC predichos utilizando QRT-PCR. Los diagramas de cajas y bigotes que representan el rango de valores de expresión log10 relativos (tumor versus media no tumores (T/NT)) obtenidos para *GLUL* (glutamina sintasa), *TBX3* (factor de transcripción *TBX3*), *MME* (metaloendopeptidasa de membrana, *CD10*), *LAMA3* (cadena alfa-3 de laminina 5), *SPARCL1* (hevína), *MERTK* (tirosina quinasa de proto-oncogen c-mero), *PAP* (proteína asociada a pancreatitis), *EPHB2* (receptor de efrina B2), *LEF1* (factor de unión a intensificador linfóide 1) y *CDH1* (E-cadherina) analizados en 109 muestras de HCC como se describe en la Figura 3.
- b. Inmunotinción de β -catenina en casos representativos de HCC mutado para la β -catenina y que lleva a G5 y

G6. En el caso HCC303 (G5), debe notarse un bajo número de núcleos teñidos y una tinción intensa de la membrana plasmática (flechas blancas). En el caso HCC305 (G6), el citoplasma y los núcleos de los hepatocitos están intensamente teñidos (flechas negras) sin señal en la membrana plasmática.

- 5 c. Expresión de proteína E-cadherina en HCC de G6 utilizando western blot (paneles superiores) comparado con el nivel de expresión de ARNm (grupo G5 y G6) utilizando QRT-PCR (panel inferior).

Figura 5: Predictores de supervivencia

- 10 a. Se muestran los resultados para el mejor predictor de supervivencia global. La curva ROC muestra la especificidad y sensibilidad para diferentes umbrales de puntuación. Los círculos corresponden al set de entrenamiento (n = 42) y las cruces al set de validación (n=53). Los círculos y cruces encuadrados indican la sensibilidad y especificidad obtenidas para el umbral elegido (-0,393). El umbral se eligió con el fin de tener una tasa de éxito máxima y un P-valor mínimo en base a una prueba exacta de Fisher para la tabla de contingencia "clase predicha/clase verdadera" de las muestras del set de entrenamiento. La curva de puntuación muestra las puntuaciones obtenidas para el set de entrenamiento (curva superior) y el set de validación (curva inferior) a partir de la fórmula de puntuación de supervivencia global descrita en el Ejemplo 3.2 con los parámetros de la Tabla 11, mientras que la línea punteada indica la puntuación umbral elegida. Los trazos horizontales representan pacientes vivos mientras que los trazos verticales representan pacientes muertos. Las curvas de supervivencia para el set de entrenamiento (líneas punteadas) y el set de validación (líneas sólidas) están estratificadas mediante la puntuación umbral indicada.
- 15
- 20
- b. Se muestran los resultados para el mejor predictor de supervivencia libre de enfermedad. Se utiliza el mismo código de representación que en la Figura 5A. La **curva de puntuación** muestra las puntuaciones obtenidas a partir de la fórmula de puntuación libre de enfermedad descrita en el Ejemplo 3.2 con los parámetros de la Tabla 14. Los trazos horizontales representan pacientes libres de enfermedad mientras que los trazos verticales representan pacientes no libres de enfermedad.
- 25

Ejemplos

Ejemplo 1: Estrategia de análisis transcripcional de tumores de HCC

Tumores y muestras, datos clínicos

- 35 Una serie de 120 carcinomas hepatocelulares y 3 adenomas hepatocelulares con sus tejidos no tumorales correspondientes se recogieron de 123 pacientes tratados por extirpación de hígado en tres departamentos quirúrgicos franceses entre 1992 y 1999. Para todos los casos incluidos en este estudio, estuvieron disponible los datos clínicos y el seguimiento completos. Todos estos tumores se caracterizaron clínicamente tal y como se ha descrito anteriormente (Laurent-Puig, P. *et al. Gastroenterology* 120, 1763-73 (2001)). La proporción por sexo (M:F)
- 40 fue de 1:4 y la media de edad de los pacientes fue de 60 años (edad mediana = 63 años, rango entre 18 y 85 años). Los pacientes nacieron en Francia (92 casos), África subsahariana (11 casos), área mediterránea (7 casos), las Antillas (4 casos) y Asia (4 casos). Los factores de riesgo para HCC de virus hepatitis B, virus hepatitis C, abuso de alcohol y hemocromatosis ocurrieron en 36, 30, 40 y 6% de los tumores, respectivamente. En 25 casos los HCC se desarrollaron en ausencia de factores de riesgo conocidos y en 16 casos se encontraron por lo menos dos factores de riesgo. El grado histológico de diferenciación del tumor se asignó de acuerdo con el sistema de graduación de Edmonson y Steiner, grado I (7%), II (49%), III (39%) y IV (4%). En 103 casos el nivel preoperatorio de α -fetoproteína en suero estuvo disponible y para 37 pacientes por encima de 100 IU/ml. Se registró invasión vascular microscópica y/o microscópica en 37% de los casos. Se registraron tumores satélites, definidos por nódulos localizados a menos de 1 cm del tumor principal en el 41% de los casos. Se evaluó la supervivencia global y la supervivencia libre de enfermedad en 99 pacientes con una extirpación completa R0 después de eliminar los pacientes tratados con un trasplante de hígado o que murieron en un periodo de 3 meses post-operatorios. Para minimizar el efecto de la ocurrencia de un segundo tumor no relacionado en cirrosis, no se tuvieron en cuenta los datos de supervivencia después de 5 años. El seguimiento medio en toda la serie fue de 38 meses (rango de 3-60 meses) y fue de 49 meses para pacientes todavía vivos. Se construyeron dos variables de pronóstico cuantitativas:
- 55 (1) "recaída temprana" si o no se definió por la presencia de una recaída del tumor durante los dos años después de la extirpación del hígado y la ausencia de recaída durante los 4 años post-operatorios; (2) "muerte temprana" si o no se definió mediante la ocurrencia de muerte durante los 3 años después de la extirpación de hígado y la ausencia de muerte durante los 5 años post-operatorios. Para el análisis de Affymetrix, se utilizaron 5 conjuntos de 3 tejidos de hígado no tumorales correspondientes con los tumores analizados incluyendo cirrosis alcohólica (conjunto 1), hígado alcohólico no cirrótico (conjunto 2), hígado no cirrótico HBV (conjunto 3), cirrosis HCV (conjunto 4) y cirrosis HBV (conjunto 5). En los experimentos de QRT-PCR, se analizaron individualmente estos 15 ARNs no tumorales y 6 ARNs de hígado no tumoral normal adicionales. También se recogieron 19 muestras de hígado fetal humano en diferentes etapas del embarazo (desde 11 hasta 29 semanas de embarazo). El estudio fue aprobado por el Comité Ético local (CCPPRB Paris Saint-Louis) y se obtuvo el consentimiento informado de acuerdo con la legislación francesa.
- 60
- 65

Análisis del transcriptoma básico

Se realizaron análisis con micromatriz utilizando como material de partida 5 µg de ARN total de cada muestra y 20 µg de ARNc por hibridación (GeneChip Fluidics Station 400) de matrices de Affymetrix GeneChip™ HG-U133A. Se generaron imágenes de cada matriz utilizando HP GeneArray 2500 y se analizaron siguiendo los protocolos del fabricante (http://www.affymetrix.com/support/technical/manual/expression_manual.affx). Excepto cuando se indique, todos los análisis de transcriptoma se llevaron a cabo utilizando ya sea un surtido de paquetes del software de sistema R (v 1.9.0) incluyendo aquellos de Bioconductor (v 1.1.1) o código de R original. Los paquetes de R y las versiones están indicadas cuando es necesario. Los datos de características crudos de 65 micromatrices Affymetrix GeneChip™ HG-U133A se normalizaron y los valores resumen de expresión de intensidad log2 para cada set de sondas se calcularon utilizando una media multimatriz robusta (paquete Affy V1.4.32). Sets de sonda correspondientes a genes control o que tienen una anotación “_x_” se enmascararon resultando en un total de 19.787 sets de sondas disponibles para posteriores análisis.

Ejemplo 2: Clasificación de tumores de HCC

2.1 Material y métodos

Mutaciones génicas, desequilibrio cromosómico, cuantificación del genoma HBV y metilación de ADN

Para todas las muestras, los ADNs de tumor y de no tumor se diseccionaron y se almacenaron a -80°C hasta que tuvo lugar la extracción de ADN y ARN utilizando los kits de extracción Qiaquick y Rneasy, respectivamente (Qiagen). Los ADNs se cuantificaron mediante fluorometría (Fluoroskan Thermo Lab-system), el ARN se cuantificó utilizando un espectrofotómetro a 260 nm (Nanodrop). La calidad del ADN y el ARN se controló mediante electroforesis en gel seguida de tinción con bromuro de etidio y bioanalizador Agilent 2100. Los ARNs calificaron para los experimentos de Affymetrix si la proporción 28S/18S era más de 1,5 y para análisis de PCR cuantitativa si la proporción 28S/18S era más de 1. Se buscaron mutaciones génicas en los exones 2 a 11 de *TP53*, exones 2 a 4 de *CTNNB1* que codifica para la β-catenina, exones 1 a 10 de *AXIN1*, exones 1 a 20 de *PIK3CA*, mediante secuenciación directa de los ADNs de tumor utilizando un secuenciador 3100 Applied Biosystems. Las pérdidas alélicas y desequilibrios cromosómicos se investigaron mediante el genotipado de 400 marcadores del panel de microsatélites LMS2 (Applied Biosystems) como se ha descrito anteriormente (Laurent-Puig, P. *et al. Gastroenterology* 120, 1763-73 (2001)). Para todas las muestras relacionadas con una infección HBV, ya sea por resultados serológicos o por amplificación de ADN viral (Laurent-Puig, P. *et al. Gastroenterology* 120, 1763-73 (2001)), se cuantificaron copias de HBS y HBX de ADN en los ADN tumoral y no tumorales utilizando el método *Sybr green* (Applied Biosystems). Las secuencias de ADN de HBS y HBX se determinaron en todos los tumores para asegurar que los cebadores utilizados para la cuantificación se habían elegido en regiones fuera de polimorfismos o mutaciones virales. La cuantificación de ADN viral se relacionó con la amplificación por PCR del cromosoma 22. La amplificación por PCR se midió a 95, 97 y 94% para HBS, HBX y referencia de cromosoma 22, respectivamente. Las muestras de ADN de tumor y de no tumor también se cuantificaron cuidadosamente utilizando fluorimetría con Hoetsch y las concentraciones se verificaron mediante electroforesis en gel de agarosa. Se definió un bajo número de copias de ADN viral en los tumores mediante una proporción HBX/referencia y HBS/referencia inferior a 0,5 (media: 0,01, rango: 0,002-0,5, error estándar: 0,14). Se definió un alto número de copias de ADN viral en los tumores mediante una proporción HBX/referencia y HBS/referencia superior a 1,5 (media: 25, rango: 1,6-212, error estándar: 46). No se encontró ningún valor entre 0,5 y 1,6. Se investigó la metilación del ADN en los promotores de *CDH1* y *CDKN2A* utilizando ADN bisulfito y una amplificación específica de metilación tal y como se ha descrito anteriormente (Lee, S. *et al. Am J Pathol* 163, 1371-8 (2003); Zochbauer-Muller, S. *et al. Cancer Res* 61, 249-55 (2001)).

Análisis por RT-PCR cuantitativa

Para los análisis de RT-PCR cuantitativa, 3 µg de ARN total se retrotranscribieron utilizando el Archive kit de alta capacidad y hexámeros aleatorios (Applied Biosystems). Para cada muestra y cada gen probado, se analizó 1 µl de ADNc, correspondiente a 2 ng de ARN retrotranscrito, mediante análisis de PCR TaqMan, en duplicado, utilizando la matriz de Baja Densidad TaqMan® y el Sistema ABI PRISM® 7900HT (Applied Biosystems). La calidad del ADNc se evaluó utilizando una cuantificación de R18S mediante PCR a tiempo real (Coeficiente de variación 7% para la serie entera). La cantidad relativa del ARNm probado en las muestras se determinó utilizando el método $2^{-\Delta\Delta CT}$ en el que $\Delta\Delta CT = (CT_{PROBADA} - CT_{R18S})_{muestra} - (CT_{PROBADA} - CT_{R18S})_{calibrador}$ (Livak, K.J. & Schmittgen, T.D. *Methods* 25, 402-8 (2001)). Brevemente, los resultados de expresión de un gen fueron normalizados con respecto a un control interno 18S ribosomal y en relación a un calibrador, que consiste en el nivel de expresión medio del gen correspondiente en muestras de no tumor normalizadas con respecto a un control interno 18S ribosomal. Los valores mostrados en las tablas y gráficos expresan la proporción n-veces de la expresión génica en una muestra probada comparada con la media de tejidos no tumorales.

Western-blot e inmunocitoquímica

Se homogeneizaron tejidos congelados con un Dounce en 500 µl de tampón de lisis RIPA (Santa Cruz) enfriado en

hielo y se determinó la concentración de proteína mediante el kit de ensayo de proteína BCA (Pierce). Se realizó un análisis de inmunoblot utilizando 50 µg de proteínas migradas en un gel de SDS poliacrilamida 6%, un anticuerpo E-cadherina policlonal (SC-7870, 1:500, Santa Cruz), un anticuerpo secundario conjugado con peroxidasa (1:2000, Santa Cruz) y quimioluminiscencia mejorada (ECL, Pierce). La inmunotinción se realizó sobre secciones de 5 µm de muestras de hígado fijadas en formalina y embebidas en parafina, utilizando anti-β-catenina (1:400, BD biosciences 610153), anti-raton biotinilado (1:200, Vector Laboratories BA-2000), Vectastain ABC Elite standard kit (Vector Laboratories PK-6100), DaB kit (Vector laboratories SK-4100). Antes de la inmunotinción, la peroxidasa endógena se bloqueó y se realizó la recuperación de antígeno con tampón citrato 0,001 M pH 7 en una olla a presión (Biogenex).

Clasificación basada en el análisis de transcriptoma

La clasificación de las 65 muestras se basó en una serie de 24 análisis de clústeres jerárquicos, obtenidos a partir de 8 subsets de datos y 3 métodos de enlace diferentes (media, completo y de Ward), utilizando una correlación 1-Pearson como métrica de distancia (paquete clase V 2.2-2). Los 8 subsets de datos correspondieron a 8 selecciones no-supervisadas de los perfiles de expresión más variables. Los criterios para esta selección fueron la diferencia significativa de la varianza para un set sonda dado en comparación con la varianza mediana ($P < 0,01$), así como un coeficiente "robusto" diferente coeficientes de umbrales de variación (rCVs, calculados dividiendo la desviación estándar por la media de los valores de intensidad log2 de un set sonda para $n-2$ muestras eliminando los valores más bajo y la más alto). Se seleccionaron entre 99 y 6.712 sets sonda (percentiles de rCV 99,5 y 60). La estabilidad de los 24 dendogramas iniciales se evaluó mediante comparación de cada uno con los resultados de clúster obtenidos antes de perturbar/re-muestrear (100 iteraciones para cada uno, ver información suplementaria para detalles sobre la estabilidad de la puntuación). El modelo también se probó utilizando una aproximación de agrupamiento de K-medias, con diferentes números de clústeres iniciales ($k = 7..15$). Utilizando el mejor resultado de 200 para cada k (es decir, el resultado con la mayor distancia entre los k grupos), las muestras se agruparon de forma consistente en los 6 subgrupos de HCC o como subsets de los mismos.

Asociación de subgrupos de HCC y variables clínicas/genéticas

Se realizaron pruebas exactas de Fisher con el fin de determinar la significación de la asociación entre diferentes variables clínicas y genéticas y los 6 subgrupos de tumores de HCC. Se probaron los resultados de clústeres obtenidos a partir de todas las combinaciones de métodos de enlace y listas de genes. Además, se recodificaron variables con múltiples modalidades (por ejemplo, HBT) en valores binomiales y se probó cada combinación. Para el predictor global utilizando resultados de QRT-PCR (ver Tabla 1), se utilizaron 1.000 permutaciones aleatorias de etiquetas de clase para corregir los P-valores originales lo que significa que un P de 0 en la tabla equivale a $P < 0,001$.

Construcción de un predictor global

Predictor de datos de Affymetrix

Para construir un predictor multiclase, las 65 muestras se dividieron en un set de entrenamiento ($n=36$, 6 para cada grupo de clúster, seleccionados de forma aleatoria), y en un set de prueba ($n= 24$ tumores más 5 muestras no tumorales). Se utilizó una estrategia de aprendizaje en 6 etapas: (1) selección supervisada del gen utilizando pruebas F y en base a las muestra de entrenamiento ($n = 376$ sets sonda); (2) filtrado de los set de sondas del gen en base a submuestreo, a niveles de intensidad globales y a símbolos de genes HUGO redundantes y a control de la tasa de descubrimiento falso ($n = 258$); (3) subselecciones aleatorias de 8-25 genes, segmentados (o no) mediante "bins" de clústeres de genes; (4) aprendizaje de regla utilizando 5 algoritmos de predicción; (5) selección de regla en base a la tasa de éxito en predecir el set de prueba; (6) validación de regla utilizando datos de RT-PCR y pruebas exactas de Fisher para evaluar la asociación de factores clínicos y genéticos con los grupos predichos.

De forma más precisa, las diferentes etapas se llevaron a cabo como se explica a continuación:

1. Selección de gen y 2. Filtrado de set sonda del gen: utilizando S1, se realizó una prueba F utilizando una prueba de permutación multivariable en base a 1.000 permutaciones de etiquetas de muestra para corregir para pruebas múltiples (BRB ArrayTools3.0.beta2). Esta prueba resultó en 1.041 sets de sondas que contenían menos de 10 descubrimientos falsos. Se realizó la misma prueba sobre una subselección de 18 muestras en S1 (3 casos por grupo seleccionados de forma aleatoria) y se encontraron 515 sets de sondas significativos (aquí se utilizó un p-valor $< 0,001$ como criterio umbral debido al bajo número de muestras por clase que hace que la prueba de permutación no sea fiable). La intersección de estas dos listas y el filtrado de los sets de sondas que tuvieron una intensidad máxima menor de 100 unidad en los 6 grupos de clúster de HCC resultaron en 258 sets de sondas.

3. Subselecciones aleatorias de genes: (i) A partir de la lista de 225 sets de sondas se generaron de forma aleatoria 1000 sublistas de un número k de sets de sondas ($k = 8..25$) por lista (total de 18.000 sublistas). (ii) A partir de la lista de 225 sets de sondas, se generaron 1.000 sublistas de un número k -de sets de sondas ($k=15, 30$, total de 2.000 sublistas), eligiendo la misma proporción de genes de bins de clúster de genes individuales. Los bins de clústeres de genes se construyeron en base a un agrupamiento por enlace completo de los 225 sets de sondas

(utilizando un coeficiente 1-Pearson como métrica de distancia), y después cortando el dendograma para obtener 15 nodos de clúster ($n = 4$ a 49 sets de sondas por bin de clúster); (iii) también se generaron 1.000 sublistas de k sets de sondas ($k = 8, 16, 24$, un total de 3.000 sublistas) representando igualmente los bins de clúster de genes derivados a partir de un agrupamiento de enlace medio, (coeficiente 1-Pearson) y cortando el dendograma en 20 nodos, fusionando nodos de clúster pequeños (representados por menos de 10 sets de sondas o una correlación mayor de 0,3 con el nodo adyacente más cercano) resultando en un total de 8 bins de clústeres de genes ($n = 9 - 87$ sets de sondas por bin de clúster).

4. Aprendizaje de regla: Los datos de expresión del set S1, restringidos a las 23.000 sublistas de sets de sondas, sirvieron para entrenar 5 algoritmos de predicción (SVM (e1071, v1.4-1), PAM (pamr, 1.14.2), k-NN (class, v7.2-2), DQDA y DLDA (sma, v0.5.14)), resultando en 115.000 predictores.

5. Selección de regla: Para cada algoritmo, en combinación con cada valor de k ((i) $k = 8..25$; (ii) $k=15, 30$; (iii) $k = 8, 16, 24$), se seleccionó la mejor sublista de entre 1.000 en base a la tasa de éxito del predictor correspondiente (entrenado con un set de entrenamiento) para clasificar sets de muestra de validación (en caso de igualdad se conservó la primera sublista encontrada); esta selección resultó en 115 predictores. De entre estos, 7 predictores relacionados con 6 sublistas distintas, obtuvieron tasas de éxito superiores al 93% al clasificar muestras del set de validación. Se seleccionaron las 24 sublistas de genes relacionadas con 2 predictores (algoritmos SVM y PAM) de entre estas 6 sublistas. Se observó que esta sublista tuvo una tasa de éxito de 97% utilizando Centroides Encogidos Más Cercanos (PAM) en predecir pertenencias verdaderas de las 36 muestras en el set de entrenamiento (100% para SVM) y 93,1% de las 29 muestras en el grupo de validación.

6. Validación de regla: Se obtuvieron datos de RT-PCR para 23 de los 24 genes en la sublista previamente seleccionada (no hubo cebadores disponibles para CD24), sobre una serie de 109 tumores, incluyendo 46 muestras analizadas previamente con micromatrices Affymetrix HG-U133A GeneChip™ (28 en el set de entrenamiento original (S1a), y 18 HCC en el set de validación original (S2a)). Utilizando datos de ΔCt (con 18S como gen control) para el set S1a, se entrenaron 5 algoritmos de predicción (SVM, PAM, k-NN, DQDA y DLDA): aplicado al set S2a, el predictor derivado de SVM tuvo una tasa de éxito de 81% (100% para S1a).

Predictor de datos de PCR

Dado el éxito parcial del predictor de 24 genes transferido de los datos Affymetrix a los datos de PCR, se investigó un nuevo predictor, empezando el proceso desde una lista inicial de 103 genes de entre los 140 genes analizados por QRT-PCR. Estos 103 genes correspondieron a las pruebas estadísticas supervisadas comparando los diferentes grupos de clústeres, utilizando todas las 65 muestras en el set de datos de Affymetrix. Se siguió entonces la misma estrategia de aprendizaje descrita para los datos de Affymetrix: subselecciones aleatorias de genes; aprendizaje de regla; selección de regla; validación de regla. De forma aleatoria, se generaron 500 sublistas de un número k de sets de sondas ($k = 5..16$) por lista (total de 6.000 sublistas). Utilizando los datos de ΔCt (gen control 18S), se entrenaron 5 algoritmos de predicción (SVM, PAM, k-NN, DQDA, DLDA) sobre el set S1 y se obtuvieron 30.000 predictores. Para cada algoritmo, en combinación con cada valor de k , se seleccionó la mejor sublista de entre las 500, en base a la tasa de éxito del predictor correspondiente (entrenado con S1) para clasificar el set de prueba S2 (en caso de igualdad se conservó la primera sublista encontrada); esta selección resultó en 60 predictores. De entre estos, 3 predictores dieron la tasa de éxito más altas para clasificar el set de prueba S2, uno de los cuales, obtuvo una tasa de éxito (de S2) superior a 88% para 3 algoritmos diferentes, y se consideró por lo tanto como el mejor. Este predictor se obtuvo con el algoritmo DLDA y predijo el set S1 con una tasa de éxito de 100% y el set S2 con una tasa de éxito de 94%. Como validación del predictor seleccionado, la relevancia de las clases predichas para el set S3 se evaluó vía el P-valor de una prueba exacta de Fisher midiendo el nivel de asociación entre las clases predichas (1, 2, 3 vs 4, 5, 6) y FAL ($P=8,5 \cdot 10^{-4}$) así como entre las clases predichas (4 vs 5, 6) y la mutación CTNNB1 ($P = 5 \cdot 10^{-5}$).

Determinación de los genes expresados diferencialmente en un subgrupo específico de HCC y análisis GO subsiguientes

Todas las pruebas t y F univariantes se realizaron utilizando BRB ArrayTools (v3.2 b5) sobre los valores de intensidad transformados $\log_2$ para los 19.787 sets de sondas. Se designó un nivel de significación nominal de cada prueba univariable de $P < 0,001$ así como un 90% confianza de menos de 10 descubrimientos falsos en base a una prueba multivariable utilizando 1.000 permutaciones. Todas las pruebas t inter-grupo se realizaron para identificar genes que se hallaron diferencialmente expresados entre un subgrupo dado (o una combinación de subgrupos) y las muestras restantes (Gx versus Gno-x) así como entre 5 muestras no tumorales agrupadas (Gx versus no tumoral). Los genes que se hallaron mediante los dos tipos de pruebas para un subgrupo dado (y no en comparación con cualquier otro grupo) y que a su vez fueron significativos ($P < 0,001$, menos de 10 descubrimientos falsos como descrito) en un análisis ANOVA (prueba F anteriormente descrita) se consideraron genes específicos de un subgrupo de HCC (o de una combinación de subgrupos).

Evaluación de estabilidad de la clasificación

Para la perturbación, se añadió ruido Gaussiano aleatorio ($\mu = 0$, $\sigma = 1,5 \times$ varianza mediana calculada a partir del set de datos) a un set de datos dado. Cada dendrograma se dividió en k grupos ($k = 2..18$) y la proporción de pares de muestras retenidas en cada grupo comparado con el dendrograma inicial se utilizó como puntuación de estabilidad (rango de puntuación de 0 a 1, en el que una puntuación de 1 significa que la perturbación (o remuestreo) no tuvo ningún efecto sobre la pertenencia del grupo de clúster).

2.2 Resultados

El análisis de transcriptoma no supervisado define clústeres de tumores estrechamente relacionados con alteraciones clínicas y genéticas

Se analizaron cincuenta y siete HCC, 3 adenomas hepatocelulares y 5 muestras de tejidos no tumorales agrupados utilizando matrices Affymetrix HG-U133A GeneChip™. En base a un análisis no supervisado se ha desarrollado un modelo robusto de clasificación de tumores de HCC en 6 subgrupos (Figura 2) cada uno de los cuales está estrechamente asociado con factores clínicos y genéticos en base a pruebas exactas de Fisher (ver Tablas 1 y 2 arriba). En base a los análisis realizados, las 60 muestras de tumores se subdividen en 2 grupos principales cada uno estando dividido en 3 subgrupos más pequeños (denominados G1 a G6). Esta clasificación probó ser extremadamente robusta cuando se confrontó con pruebas de perturbación/remuestreo (las puntuaciones de reproducibilidad medias para cada análisis de clúster fueron de por lo menos 0,9 para los dos grupos principales y los 6 subgrupos) así como consistente con un análisis de clúster de k -medias iterativo (ver Materiales y Métodos). Además la topología de la división de la muestra se conservó a lo largo de las diferentes listas de genes y métodos de enlace de clústeres. Los dos grupos principales corresponden con muestras cromosómicamente inestables (G1, G2 y G3) y estables (G4, G5 y G6) ya que G1 a G3 mostraron una mayor pérdida alélica fraccional (FAL) significativa que G4 a G6 ($P = 0,05$, Tabla 2). Los diferentes subgrupos se caracterizaron por mutaciones de *TP53* (G2 y G3), una infección HBV (G1 y G2), con un bajo número de copias de ADN HBV (G1) y mutaciones del gen *CTNNB1* (G5 y G6). La presencia de nódulos cancerosos localizados a menos de 1 centímetro del tumor primario se asoció con G6 ($P = 0,04$, Tabla 2), indicando el gran potencial de invasión local de estos tumores. Los 5 conjuntos de muestras de tejidos de hígado no tumoral se agruparon juntos estrechamente y se encontró que un amplio grupo heterogéneo (G4) que contiene 20 tumores, 4 de los cuales, en el mismo clúster pequeño, presentaba mutaciones de *TCF1* (3 adenomas y un HCC).

Identificación de 2 predictores de la clasificación de 6 grupos

Dada la relevancia clínica de los subgrupos y el potencial de diagnóstico de esta clasificación, el objetivo de los inventores fue identificar un predictor de clase más adaptado a un ambiente clínico utilizando la tecnología de transcriptasa inversa cuantitativa (QRT-PCR), más eficiente en tiempo y coste. Con el fin de buscar genes que pueden predecir la pertenencia a una clase de los 6 subgrupos se construyó, en primer lugar, un predictor utilizando los datos de Affymetrix (ver Material y métodos y siguiente Tabla 10). Este análisis identificó un primer predictor de 24 genes (ALDH1L1, CD24, CD74, CFHR3, CYP4F12, DNAJA3, DSCR1, EPHA1, EPHB4, FAAH, FGFR2, FLJ10159, GLT8D1, HAL, MATN2, MRPS7, PAK2, PLXNB1, RAB1A, RHOQ, SLC27A5, SLP1, SMARCE1, STRA13) resultando en una alta tasa de éxito de predicción de clase utilizando los datos de Affymetrix (93,1%) pero resultó ser menos satisfactorio utilizando datos de QRT-PCR (81%).

Tabla 10. Lista de los 24 genes que fueron identificados como el predictor de HCC global para las muestras de Affymetrix utilizando datos de Affymetrix. Se incluyen el símbolo del gen HUGO, el estadístico F a partir de una ANOVA entre las 6 clases de muestras y la media geométrica de valores de intensidad no log asociada por subgrupo de HCC (G1-G6).

Símbolo de gen HUGO	Estadístico F	G1	G2	G3	G4	G5	G6
ALDH1L1	13,1	221	226	766	993	2213	2123
CD24	10,7	511	235	274	66	75	17
CD74	5,7	1343	2008	2481	1896	1121	4973
CFHR3///CFHR4	5,8	368	1000	174	2527	1567	980
CYP4F12	8,4	341	347	243	461	595	459
DNAJA3	6,8	412	521	621	662	727	658
DSCR1	5,5	139	134	136	205	170	265
EPHA1	16,2	432	972	189	530	230	120
EPHB4	9,3	538	264	204	261	288	220
FAAH	5,6	140	134	127	258	169	186
FGFR2	20,8	373	515	71	47	25	29
FLJ10159	11,9	896	289	168	116	89	40
GLT8D1	8,5	471	433	463	542	933	868
HAL	27,7	380	1065	139	298	83	66

Símbolo de gen HUGO	Estadístico F	G1	G2	G3	G4	G5	G6
MATN2	10,4	291	180	114	190	60	37
MRPS7	6,5	321	389	725	403	438	448
PAK2	7,6	297	199	337	143	150	119
PLXNB1	6,9	482	343	194	207	304	217
RAB1A	6,5	1960	2490	2388	1820	1601	1722
RHOQ	5,2	220	389	366	259	191	204
SLC27A5	12,2	282	564	437	2309	2147	2152
SLPI	18,0	1148	2766	313	1829	159	141
SMARCE1	7,3	439	395	489	265	303	253
STRA13	13,7	433	570	1170	583	959	1151

De este modo, se realizaron una serie de pruebas supervisadas utilizando datos de Affymetrix y anotaciones clínicas y genéticas relevantes (es decir, el estatus mutacional de los genes *TP53*, *CTNNB1* y *AXIN1*, presencia y título de HBV, recaída temprana y supervivencia global). Se elaboró una lista de 140 genes que mostraron ser significativos en una o más de estas pruebas supervisadas. Todos menos 5 de estos genes fueron validados por QRT-PCR en 109 tumores de HCC (incluyendo 46 de los 57 HCC analizados utilizando micromatrices Affymetrix y un set de validación de 63 HCC) y 21 tejidos de hígado no tumoral. Se encontró una elevada correlación entre los datos de Affymetrix y los datos de QRT-PCR en 135 de los 140 genes seleccionados (coeficiente de correlación de mediana rho de Spearman de 0,84 utilizando valores de ΔCt). Utilizando los datos de QRT-PCR, se probaron múltiples sublistas de un subset de 103 genes (dentro los 135 probados) con el fin de identificar el mejor predictor global de los 6 subgrupos de HCC. Para este propósito, los 46 HCCs analizados con micromatrices Affymetrix se dividieron en un set de entrenamiento (n=28) y un set de prueba (n=18). Se demostró que todos los genes en la Tabla 3 previamente descrita fueron significativos en una o más de estas pruebas supervisadas, y la mayoría de ellos estuvieron presentes en por lo menos uno o dos predictores de clasificaciones buenos. La mejor tasa de éxito en predecir la pertenencia verdadera a una clase del set de entrenamiento (100%) y del set de prueba (94,4 %) se obtuvo con los valores de Ct de 16 genes (*RAB1A*, *PAP*, *NRAS*, *RAMP3*, *MERTK*, *PIR*, *EPHA1*, *LAMA3*, *G0S2*, *HN1*, *PAK2*, *AFP*, *CYP2C9*, *CDH2*, *HAMP*, *SAE1*) utilizando el algoritmo de predicción DLDA.

El mejor predictor sigue las siguientes fórmulas:

$$clase\ predicha\ (muestra_i) = \arg \min_{k=1..6} (distancia(muestra_i, clase_k))$$

en la que

$$distancia\ (muestra_i, clase_k) = \sum_{t=1..16} \frac{(\Delta Ct(muestra_i, gen_t) - \mu(clase_k, gen_t))^2}{\sigma(gen_t)}$$

y en la que los parámetros $\mu(clase_k, gen_t)$ y $\sigma(gen_t)$ son aquellos listados en la Tabla 5 previamente descrita.

De este modo, después de haber calculado la distancia entre la muestra dada y la representación centroe de cada clase, la nueva muestra se asigna a la clase más cercana.

Esta firma se utilizó entonces para dividir las 63 muestras del set de tumores de validación en 6 subgrupos. Como se ha observado en el primer set de tumores analizados en el experimento de Affymetrix, se hallaron, utilizando pruebas exactas de Fisher, asociaciones significativas entre FAL, TP53, infección HBV y mutación del gen CTNNB1 y los diferentes subgrupos predichos, así como con aquellos que utilizan la serie completa de 109 tumores de HCC (Tabla 1).

Identificación de vías de señalización claves y categorías funcionales de genes implicados en cada subgrupo HCC

Para identificar vías clave afectadas en los diferentes subgrupos de HCC se identificaron 1.560 genes específicamente desreglados en uno o más subgrupos de HCC en base a los resultados de un "análisis de prueba t de todo el grupo combinado con ANOVA. Para todas las listas de genes específicos de subgrupos de HCC, se investigó la asociación de genes de vías conocidas. Se observó un enriquecimiento de genes del ciclo celular /proliferación/ metabolismo de ADN específicamente sobreexpresados en los subgrupos G1 a G3, correspondientes a las muestras cromosómicamente inestables ($P < 0,01$). Un elevado número de genes específicamente sobreexpresados se observaron para el subgrupo G1 (relacionados con infección HBV con bajo número de copias de ADN, mutaciones *AXIN1*, una edad más joven, niveles elevados de AFP en suero y orígenes frecuentes en África, Tablas 1 y 2). Entre ellos se encontraron genes que codifican para proteínas expresadas durante el desarrollo: cadena pesada de miosina IIb, *MYH4*, factores de transcripción *SOX9* y *SOX4* y genes con impronta parental: factor de crecimiento de tipo insulina (*IGF2*), gene expresado paternamente 1, 3 y 10 (*PEG1*, *PEG3* y *PEG10*), fetoproteína-alfa (AFP) y sarcoglicano epsilon (*SGCE*). La expresión diferencial de todos estos genes fue

validada utilizando QRT-PCR en 109 tumores (Figura 3a). Los genes marcado probados estuvieron altamente sobreexpresados en hígados fetales normales (Figura 3a). El ARNm *H19* también estuvo sobreexpresado, no sólo en muestras G1 sino también en muestras fetales, correlacionando con IGF2 en estos dos grupos ($R^2 = 0,4$ y $0,6$, respectivamente).

Los tumores del subgrupo G2 (relacionado con infección HBV con alto número de copias de ADN viral, invasión local y vascular frecuente y mutaciones de *TP53*) estuvieron asociados de forma significativa con la sobreexpresión de genes del ciclo celular/proliferación/metabolismo de ADN ($P < 0,01$), un enriquecimiento que se observó igualmente en G3 (relacionado con mutaciones de *TP53* y metilación de promotor de *CDKN2A*) y todas las muestras cromosómicamente inestables ($P < 0,007$). También se identificó una sobrerepresentación significativa de genes sobreexpresados implicados en fosforilación de proteína ($P < 0,009$). De manera interesante, se identificaron mutaciones en el gen *PIK3CA* de las cuales se ha predicho que resultan en la activación de la vía de fosfatidilinositol 3-quinasa (PI3K)-AKT en dos tumores pertenecientes a G2. Estas dos muestras estuvieron estrechamente asociadas en el análisis de agrupamiento no supervisado (Figura 2). Se identificaron 38 genes específicamente sobreexpresados en las muestras mutadas en *PIK3CA* cuando se compararon con los otros tumores en los grupos G1 a G3. Entre estos genes, se validó utilizando QRT-PCR la sobreexpresión de dos genes que codifican para el factor de elongación de proteína *EEF1A2* y la enteroquinasa *PRSS7*, que están específicamente sobreexpresados en los tumores mutados en *PIK3CA* ($P = 0,001$, Figura 3b). Además, el análisis GO demostró un enriquecimiento de genes de comunicación celular en los tumores mutados en *PIK3CA* ($P = 0,07$).

En G5 (*CTNNB1* mutado, no nódulos distantes), se observó un enriquecimiento de genes subexpresados involucrados en estrés y respuesta inmune como *IFI16*, *IL4R*, *IFI44*, *STAT1*, *IL10RA*, *CTSS* y *HLA-DPA1/B1* ($P < 0,002$). Los subgrupos de HCC G5 y G6 contienen 23 y 11 tumores mutados en *CTNNB1* en 70 y 100% de los casos, respectivamente. En una investigación sobre posibles genes diana de β -catenina, se halló una lista de 280 genes sobreexpresados de forma significativa en G5 y G6. Además de *GPR49* y *GLUL*, dos genes diana de β -catenina en el hígado conocidos (Cadoret, A. *et al. Oncogene* 21, 8293-301 (2002); Yamamoto, Y. *et al. Hepatology* 37, 528-33 (2003)), se confirmó la sobreexpresión de 7 genes diana de β -catenina putativos utilizando QRT-PCR. Estos genes incluyen: *EPHB2*, un receptor de tirosina quinasa; *MME*, la encefalinasa CD10; *MERTK*, un oncogén tirosina quinasa; *LAMA3*, que codifica para la cadena alfa 3 de la laminina 5; *PAP/HIP*, que codifica por una proteína asociada a pancreatitis; *SPARCL1* que codifica por la hevína que está asociada con la matriz extracelular; y el factor de transcripción *TBX3* (Figura 4a). Se observó un nivel de expresión significativamente más elevado de todos estos genes diana de β -catenina putativos en G6 en comparación con G5, incluso después de excluir las muestras sin la mutación *CTNNB1*. También se demostró que la β -catenina estaba más sobreexpresada en tumores G6, en comparación con tumores G5, con una pérdida de señal en la membrana plasmática y una fuerte localización en el citoplasma y el núcleo (Figura 4b). De forma consistente con esta observación, se halló en G6 una sobreexpresión de *LEF1*, un factor de transcripción que interactúa con la β -catenina para activar genes diana sensibles a Wnt. Mientras que ambos subgrupos G5 y G6 estuvieron asociados con LOH del cromosoma 8p, no se identificó ninguna otra delección cromosómica específica de G6. Sin embargo, se halló una subexpresión de *CDH1* (que codifica para E-cadherina) en el subgrupo G6 (en experimentos de Affymetrix y de QRT-PCR, Figura 4a) que puede explicar la invasión local de estos HCC como se muestra por la casi-constante presencia de nódulos satélite encontrados alrededor del tumor principal (Figura 4c y Tablas 1 y 2). Se demostró que el nivel de subregulación del ARNm de *CDH1* está altamente relacionado con la expresión subregulada de la proteína E-cadherina en G6, consistente con el elevado nivel de metilación de promotor de *CDH1* en estos tumores (datos no mostrados).

2.3 Conclusión

Utilizando una aproximación en todo el genoma no supervisada, los inventores obtuvieron una clasificación robusta de HCC en 6 subgrupos principales que reflejan la gran diversidad natural de estos tumores (Bosch, F. X. *et al. Gastroenterology* 127, S5-S16 (2004), El-Serag, H.B. *Gastroenterology* 127, S27-34 ("004)). Además esta clasificación pudo ser reproducida utilizando solamente 16 genes analizados mediante QRT-PCR y, de forma más importante, fue confirmada en un set de tumores independiente.

Esta clasificación va en acorde con los análisis de HCC previamente publicados (Lee, J.S. *et al. Hepatology* 40, 667-76 (2004); Breuhahn, K. *et al. Cancer Res* 64, 6058-64 (2004); Chen, X. *et al. Mol Biol Cell* 13, 1929-39 (2002)) que han descrito dos grupos principales de tumores relacionados con estabilidad cromosómica (Correspondiendo con los metagrupos G1-G3 y G4-G6). Sin embargo, el presente análisis ha extendido y refinado esta clasificación.

En síntesis, los inventores creen que la elucidación de la clasificación multifaceta de HCC solo ha sido posible en esta solicitud, en comparación con las clasificaciones previamente publicadas en dos grupos, debido a que (1) la serie de tumores estudiados tratados quirúrgicamente en Francia incluyó los diferentes factores de riesgo principales de HCC, es decir, infección por HBV y HCV, abuso de alcohol y hemocromatosis y (2) el gran número de anotaciones clínicas, histopatológicas y genéticas disponibles para la población muestral estudiada. De hecho, el principal determinante clínico de pertenencia a clase es la infección HVB mientras que otros determinantes principales son alteraciones genéticas y epigenéticas incluyendo inestabilidad cromosómica, mutaciones de *TP53* y *CTNNB1*, metilación de *CDKN2A* y *CDH1* y la impronta parental (ver Figura 1).

Centrándose en la historia natural de HCC; resulta que los tumores relacionados con HBV que definen los subgrupos G1 y G2 son claramente molecularmente distintos de otras etiologías. Los tumores relacionados con infección HCV y abuso de alcohol están dispersados en los subgrupos G3 a G6. La presente clasificación transcriptómica ha permitido la identificación de nuevas entidades de tumores. El subgrupo G1 incluye tumores relacionados con HBV de pacientes más jóvenes (con respecto a los otros HCCs HBV), frecuentemente de África, con una proporción de sexos igual, un bajo número de copias de ADN viral, mutaciones AXIN1 frecuentes, ausencia de mutación de TP53 y una sobreexpresión de genes normalmente con impronta parental. Estos resultados sugieren que la infección HBV a una edad temprana lleva a un tipo específico de HCC que demuestra características inmaduras con una impronta genética parental anormal, posiblemente debido a la persistencia de hepatocitos fetales o la dediferenciación de hepatocitos adultos. Dicha diversidad en los tumores puede estar relacionada con las poblaciones de alto riesgo halladas mediante estudios epidemiológicos (Brecht, C. *Gastroenterology* 127, S56-61 (2004); Yu, M.C. & Yuan, J.M. *Gastroenterology* 127, S72-8 (2004)).

El subgrupo G6, con un 100% de incidencia de la mutación CTNNB1, un elevado nivel de activación de la vía Wnt patológica (superior que en G5) e inactivación de E-cadherina (Kozyraki, R. *et al. Gastroenterology* 110, 1137-49 (1996)) es consistente con el gran potencial invasivo de estos tumores ya que es conocido que la inactivación de E-cadherina participa en el proceso de invasión celular (Behrens, J. *et al. J Cell Biol* 108, 2435-47 (1989)).

A parte de estos amplios subgrupos de tumores, el presente análisis transcriptómico también ha sugerido subgrupos homogéneos de tumores relacionados con alteraciones genéticas raras como mutaciones de *TCF1* o *PIK3CA* (Bluteau, O. *et al. Nat Genet* 32, 312-5 (2002); Lee, J.W. *et al. Oncogene* 24, 1477-80 (2005)). Todavía han quedado por identificar nuevas alteraciones genéticas estructurales de otros pequeños subgrupos homogéneos de tumores y en cambio esto puede ser una potente herramienta para descubrir nuevas dianas terapéuticas.

Ejemplo 3: Pronóstico de tumores de HCC

3.1 Material y métodos

Análisis por RT-PCR cuantitativa

El análisis por RT-PCR cuantitativa se realizó tal como se ha descrito en la sección Material y Métodos del Ejemplo 2.

Construcción del predictor de pronóstico

En base a los valores $2^{-\Delta Ct}$ ($\Delta Ct = Ct_{\text{PROBADA}} - Ct_{\text{R18S}}$) para 135 genes de la serie de 42 muestras analizadas con Affymetrix GeneChips, se identificaron los mejores 16 genes (rango logP máximo $\leq 10^{-2}$) asociados con el estado de pronóstico (Supervivencia Global en 60 meses) utilizando un modelo Cox univariable (paquete survival V2.15). Utilizando las mismas 42 muestras, se seleccionaron las mejores combinaciones de 5 genes o menos de entre estos 16 genes (rango logP máximo $< 10^{-5}$) utilizando un modelo Cox multivariable de entre todas las combinaciones posibles. Una segunda serie de 53 HCC independientes se utilizó entonces para validar estos modelos (rango logP máximo $< 10^{-3}$), conservando 42 de ellas. La robustez de cada modelo se evaluó a continuación con la siguiente aproximación de remuestreo: se obtuvieron 1.000 muestreos dividiendo 1.000 veces, aleatoriamente, la serie entera de 95 tumores en 2 grupos de 47 y 48 muestras cada uno (equilibrando el número de muertes entre ambos grupos); a continuación, utilizando cada una de las 42 listas de genes, para ambos grupos de cada muestreo, se construyeron modelos Cox multivariables y se almacenaron el valor de rango log P-calculado a partir de ambos modelos. Se conservó la combinación de genes que condujo al rango logP mediano más bajo en ambos grupos (de entre estos 1.000 remuestreos) y se derivó entonces un predictor a partir de esta combinación.

3.2 Resultados

Identificación y validación de genes que predicen pronóstico

Aunque útil para el diagnóstico, la firma de clasificación de 16 genes (ver Ejemplo 2) no fue suficiente para predecir el pronóstico de HCC ya que las pruebas de rango log resultaron en p valores altos ($P = 0,2$ y $0,1$) probando ya sea los dos grupos de tumores principales (G1 a G3 vs G4 a G6) o los 6 subgrupos individuales, respectivamente. Como resultado, se construyó un predictor de pronóstico específico tal y como se ha descrito en la sección de Material y Métodos.

Globalmente, los genes que se encontraron útiles para el pronóstico de la supervivencia global y/o la supervivencia sin recaída fueron los que están incluidos en la lista mostrada en la Tabla 7 descrita anteriormente.

De forma más precisa, se determinó que los mejores 16 genes asociados con el estado de pronóstico de supervivencia global fueron: *NRCAM*, *PIR*, *RAMP3*, *SLC21A2*, *TAF9*, *TNA*, *HN1*, *PSMD1*, *MRPS7*, *CDC20*, *ENO1*, *HLF*, *STRA13*, *RAGD*, *NRAS*, *ARFGF2*.

Después de probar todas las combinaciones posibles de 5 genes o menos de entre estos 16 genes como se ha descrito en Material y Métodos, los mejores 5 modelos predijeron la supervivencia global utilizando un análisis Cox multivariable con un $P < 10^{-8}$. Finalmente, la combinación más útil para predecir la mala superación es la asociación de 5 genes: un bajo nivel de *RAMP3* combinado con un alto nivel de *TAF9*, *NRCAM*, *PSMD1* y *ARFGEF2*.

El mejor predictor de supervivencia global sigue las siguientes fórmulas:

$$\text{Puntuación supervivencia global}(\text{muestra}_i) = \sum_{t=1..n} \beta(\text{gen}_t) \cdot (2^{-\Delta Ct(\text{muestra}_i, \text{gen}_t)} - \mu(\text{gen}_t))$$

en la que los diferentes parámetros $\beta(\text{gen}_t)$ y $\mu(\text{gen}_t)$ son los que están incluidos en la lista de la siguiente Tabla 11.

Tabla 11. Parámetros que deben utilizarse en la fórmula anterior para el mejor de los 5 mejores predictores de supervivencia global.

Supervivencia global	μ	β
gen 1 (<i>TAF9</i>)	7,28	0,129
gen 2 (<i>NRCAM</i>)	1,59	0,252
gen 3 (<i>RAMP3</i>)	0,14	-6,133
gen 4 (<i>PSMD1</i>)	4,66	0,024
gen 5 (<i>ARFGEF2</i>)	3,66	-0,025

Los resultados para el mejor predictor de supervivencia global en términos de curva ROC, curva de puntuación y curvas de supervivencia se muestran en la Figura 5a, mientras que las estadísticas relacionadas con dicho mejor predictor de supervivencia global están incluidas en la lista de la siguiente Tabla 12.

Tabla 12. Estadísticas relacionadas con el predictor de supervivencia global (para los sets de entrenamiento y de validación)

	Supervivencia global	
	Set de entrenamiento	Set de validación
Área bajo la curva	0,88	0,67
Especificidad	72,8%	70%
Sensibilidad	88,8%	73,6%
P de prueba exacta de Fisher	$4 \cdot 10^{-5}$	$9 \cdot 10^{-5}$
Tasa de éxito	80,9%	79,2%

En el set de validación de 53 HCC, esta combinación de genes predijo la recaída temprana en 79% de los casos (70% para (+), 89% para (-)); y las muertes tempranas fueron correctamente predichas en 81% de los casos (73% para (+); 92% para (-)). Entre las características clínicas y morfológicas, el grado de Edmondson y la invasión vascular estuvieron asociados de forma significativa con un pronóstico pobre (rango log $P < 0,04$ y $0,0002$ respectivamente). Se realizó un modelo Cox multivariable incluyendo estas dos variables además del mejor predictor de supervivencia global (ver siguiente Tabla 13). Este modelo muestra que nuestra combinación de genes es una variable de pronóstico independiente.

Tabla 13. Proporción de riesgo (HR), P-valor rango log (P) e Intervalo de confianza (CI) 95% del modelo Cox multivariable obtenido (para supervivencia global) utilizando las siguientes variables predictivas: (i) atribución binaria (umbral de puntuación superior e inferior) para el mejor predictor de supervivencia global, (ii) invasión vascular y (iii) grado de Edmondson (se unieron los grados I y II, ya que solo estaban disponibles 7 casos para el grado I).

	HR	P	95% CI
Gen predictor	7,8	0,00001	3,1-19,7
Invasión vascular	2,6	0,02	1,2-5,8
Grado de Edmondson III	0,5	0,09	0,2-1,1
Grado de Edmondson IV	2,8	0,14	0,7-10,6

La misma estrategia se aplicó para encontrar combinaciones de genes que predicen la supervivencia libre de enfermedad. De forma interesante, entre los mejores 16 genes (*TAF9*, *NRCAM*, *ENO1*, *RAB1A*, *ARFGEF2*, *G0S2*, *PSMD1*, *MRPS7*, *RAGD*, *HN1*, *PIR*, *SMAD3*, *DNAJA3*, *HELO1*, *RAMP3*, *RHOQ*), diez fueron identificados anteriormente como mejores predictores de supervivencia global utilizando un modelo Cox univariable. Finalmente, los 3 genes incluidos en el mejor predictor de supervivencia libre de enfermedad se incluyeron también en el mejor predictor de supervivencia global.

El mejor predictor de supervivencia sin recaída (o libre de enfermedad) sigue las fórmulas siguientes:

$$\text{Puntuación libre de enfermedad (muestra}_i) = \sum_{t=1..n} \beta(\text{gen}_t) \cdot (2^{-\Delta Ct(\text{muestra}_i, \text{gen}_t)} - \mu(\text{gen}_t))$$

en la que los diferentes parámetros $\beta(\text{gen}_t)$ y $\mu(\text{gen}_t)$ son los que están incluidos en la lista de la siguiente Tabla 14.

Tabla 14. Parámetros que deben ser usados en la fórmula anterior para el mejor de los mejores 5 predictores de supervivencia libre de enfermedad.

Supervivencia libre de enfermedad	μ	β
gen 1 (<i>TAF9</i>)	7,28	0,127
gen 2 (<i>NRCAM</i>)	1,59	0,196
gen 3 (<i>RAMP3</i>)	0,14	-3,886

Los resultados para el mejor predictor de supervivencia global en términos de curva ROC, curva de puntuación y curvas de supervivencia se muestran en la Figura 5b, mientras que las estadísticas relacionadas con este mejor predictor de supervivencia global están incluidas en la lista de la Tabla 15 a continuación.

Tabla 15. Estadísticas relacionadas con el predictor de supervivencia libre de enfermedad.

	Supervivencia libre de enfermedad	
	Set de entrenamiento	Set de validación
Área debajo de la curva	0,86	0,84
Especificidad	83,4%	84,7%
Sensibilidad	74%	78,5%
P de prueba exacta de Fisher	$4 \cdot 10^{-3}$	$6 \cdot 10^{-6}$
Tasa de éxito	73,8%	81,1%
Rango Log P	$3 \cdot 10^{-4}$	$7 \cdot 10^{-6}$

3.3 Conclusión

La elucidación de la clasificación transcriptómica es de particular interés para aplicaciones clínicas. En particular, parece ser que los HCC pertenecientes a los grupos G1 a G3 estuvieron ligeramente relacionados con una recaída temprana y una muerte temprana en comparación con los HCC de G4 a G6, mostrando que la clasificación y el pronóstico están de alguna manera relacionados.

Sin embargo, los inventores hallaron que utilizar un pequeño subconjunto específico de aproximadamente 5 genes era mejor que utilizar la firma de 16 genes de clasificación global en predecir el pronóstico de pacientes tratados mediante extirpación quirúrgica completa. En contraste con los análisis transcriptómicos publicados previos, el funcionamiento del predictor de supervivencia determinado se verificó en un segundo set de tumores independientes incluyendo todos los factores de riesgo etiológicos y dichas validaciones se realizaron utilizando QRT-PCR en lugar de los datos de hibridación (Lee, J.S. *et al. Hepatology* 40, 66776 ("004); Ye, Q.H. *et al. Nat Med* 9, 416-23 (2003); Iizuka, N. *et al. Lancet* 361, 923-9 (2003); Kurokawa, Y. *et al. J Hepatol* 41, 284-91 (2004)).

Los genes identificados en las presentes aplicaciones como útiles para predecir la supervivencia nunca antes habían sido encontrados asociados con el pronóstico del paciente y pueden estar implicados en procesos celulares generales como degradación en proteosoma de proteínas (*PSMD1*, ver Yokota, K. *et al. Mol Biol Cell* 7, 853-70 (1996)), la iniciación de la transcripción de ARN (*TAF9*, ver Michel, B., Komarnitsky, P. & Buratowski, S. *Mol Cell* 2, 663-73 (1998)) y proliferación celular (*NRCAM*, ver Sehgal, A., *et al. Anticancer Res* 19, 4947-53 (1999); y *ARFGEF2*, ver Sheen, V.L. *et al. Nat Genet* 36, 69-76 (2004)).

De forma interesante, las mejores combinaciones de genes que predicen la supervivencia global así como la supervivencia libre de enfermedad (es decir, la supervivencia sin recaída) fueron muy similares en este estudio, demostrando que los predictores determinados reflejan de forma precisa la progresión del tumor independientemente de una enfermedad hepática no relacionada con un tumor.

Sería también muy interesante evaluar estos predictores en pacientes tratados con trasplante de hígado o radiofrecuencia con el fin de estimar la potencial utilidad de estos marcadores en la elección terapéutica.

En conclusión, el presente análisis transcriptómico global se ha llevado a cabo y se ha validado utilizando una amplia serie de tumores altamente anotados. Este análisis ha establecido una clasificación robusta que refleja la diversidad natural de los HCCs humanos, las alteraciones genéticas estructurales y desregulaciones epigenéticas acumuladas durante la progresión del tumor. La amplia diversidad de tumores de HCC tiene implicaciones clínicas y la presente clasificación a resultado en herramientas de pronóstico no sólo para pacientes tratados quirúrgicamente sino también para identificar además pacientes que van a beneficiarse de terapias dirigidas.

Bibliografía

1. Bosch, F.X., *et al.* Semin Liver Dis 19, 271-85 (1999)
- 5 2. Taylor-Robinson, S.D. *et al.* Bmj 319, 640 (1999);
3. Deuffic, S. *et al.* Lancet 351, 214-5 (1998).
4. El-Serag, H.B. & Mason, A.C. N Engl J Med 340, 745-50 (1999)
5. Edmondson, H.A. & Peters, R.L. Semin Roentgenol 18, 75-83 (1983);
6. Thorgeirsson, S.S. & Grisham, J.W. Nat Genet 31, 339-46 (2002)).
- 10 7. Aoki, H., *et al.* Proc Natl Acad Sci U S A 93, 7300-4 (1996)
8. Brechot, C. Gastroenterology 127, S56-61 (2004)
9. Bressac, B. *et al.* Proc Natl Acad Sci U S A 87, 1973-7 (1990)
10. Weihrauch, M. *et al.* Br J Cancer 84, 982-9 (2001)
11. Bluteau, O. *et al.* Nat. Genet 32, 3.12-5 (2002)
- 15 12. Boige, V. *et al.* Cancer Res 57, 1986-90 (1997);
13. Wong, N. *et al.* Clin Cancer Res 6, 4000-9 (2000);
14. Guan, X.Y. *et al.* Genes Chromosomes Cancer 29, 110-6 (2000)
15. Okabe *et al.* Cancer Res. 2001 Mar 1;61(5):2129-37;
16. Iizuka *et al.* Cancer Res. 2002 Jul 15;62(14):3939-44
- 20 17. Chung *et al.* Mol Cells. 2002 Dec 31;14(3):382-7 ;
18. Chen *et al.* Mol Biol Cell. 2002 Jun; 13(6):1929-39;
19. WO 2004/090163
20. Lee *et al.* Hepatology. 2004 Sep;40(3):667-76
21. Legoix, P. *et al.* Oncogene 18, 4044-6 (1999);
- 25 22. Laurent-Puig, P. *et al.* Gastroenterology 120, 1763-73 (2001)
23. Qin *et al.* J Cancer Res Clin Oncol. 2004 Sep;130(9):497-513;
24. Ye *et al.* Nat Med. 2003 Apr ;9(4) :416-23
25. Kurokawa *et al.* J Hepatol. 2004 Aug;41(2):284-91 ;
26. Iizuka *et al.* Lancet. 2003 Mar 15 ;361(9361) :923-9;
- 30 27. WO 2005/017150
28. Lee, S. *et al.* Am J Pathol 163, 1371-8 (2003);
29. Zochbauer-Muller, S. *et al.* Cancer Res 61, 249-55 (2001)
30. Livak, K.J. & Schmittgen, T.D. Methods 25, 402-8 (2001)
31. Cadoret, A. *et al.* Oncogene 21, 8293-301 (2002);
- 35 32. Yamamoto, Y. *et al.* Hepatology 37, 528-33 (2003)
33. Bosch, F.X., *et al.* Gastroenterology 127, S5-S16 (2004);
34. El-Serag, H.B. Gastroenterology 127, S27-34 (2004)
35. Breuhahn, K. *et al.* Cancer Res 64, 6058-64 (2004);
36. Yu, M.C. & Yuan, J.M. Gastroenterology 127, S72-8 (2004)
- 40 37. Kozyraki, R. *et al.* Gastroenterology 110, 1137-49 (1996)
38. Behrens, J., *et al.* J Cell Biol 108, 2435-47 (1989)
39. Lee, J.W. *et al.* Oncogene 24, 1477-80 (2005)
40. Yokota, K. *et al.* Mol Biol Cell 7, 853-70 (1996)
41. Michel, B., Komarnitsky, P. & Buratowski, S. Mol Cell 2, 663-73 (1998)
- 45 42. Sehgal, A., *et al.* Anticancer Res 19, 4947-53 (1999);
43. Sheen, V.L. *et al.* Nat Genet 36, 69-76 (2004)).

REIVINDICACIONES

1. Procedimiento de clasificación *in vitro* de un tumor de HCC en 6 subgrupos a partir de una muestra de hígado HCC de un sujeto que sufre HCC, que comprende:

a) determinar un perfil de expresión que comprende o que consiste en:

(i) la siguiente combinación de 16 genes: RAB1A, REG3A, NRAS, RAMP3, MERTK, PIR, EPHA1, LAMA3, G0S2, HN1, PAK2, AFP, CYP2C9, CDH2, HAMP, y SAE1, o

(ii) la siguiente combinación de 24 genes: ALDH1L1, CD24, CD74, CFHR3, CYP4F12, DNAJA3, DSCR1, EPHA1, EPHB4, FAAH, FGFR2, FLJ10159, GLT8D1, HAL, MATN2, MRPS7, PAK2, PLXNB1, RAB1A, RHOQ, SLC27A5, SLPI, SMARCE1, STRA13;

b) calcular a partir de dicho perfil de expresión 6 distancias de subgrupo; y

c) clasificar dicho tumor HCC en el subgrupo para el cual la distancia de subgrupo es la más baja,

en el que los 6 subgrupos G1, G2, G3, G4, G5 y G6 están definidos por la presencia (+) o ausencia (-) de características clínicas y genéticas descritas en la siguiente Tabla 2:

	G1	G2	G3	G4	G5	G6
Inestabilidad cromosómica	+	+	+	-	-	-
Recaída temprana y muerte	+	+	+	-	-	-
Mutación TP53	-	+	+	-	-	-
Infección HBV	+	+	-	-	-	-
Bajo número de copias	+	-	-	-	-	-
Alto número de copias	-	+	-	-	-	-
Mutación CTNNB1	-	-	-	-	+	+
Nódulos satélite	-	-	-	-	-	+

2. Procedimiento según la reivindicación 1, en el que el perfil de expresión se determina a nivel de ácido nucleico.

3. Procedimiento según las reivindicaciones 1 o 2, en el que el perfil de expresión consiste en la siguiente combinación de 16 genes: RAB1A, REG3A, NRAS, RAMP3, MERTK, PIR, EPHA1, LAMA3, G0S2, HN1, PAK2, AFP, CYP2C9, CDH2, HAMP, y SAE1.

4. Procedimiento según cualquiera de las reivindicaciones 1 a 3, en el que el perfil de expresión se determina a nivel de ácido nucleico utilizando PCR cuantitativa.

5. Procedimiento según la reivindicación 4, en el que el perfil de expresión consiste en la siguiente combinación de 16 genes: RAB1A, REG3A, NRAS, RAMP3, MERTK, PIR, EPHA1, LAMA3, G0S2, HN1, PAK2, AFP, CYP2C9, CDH2, HAMP, y SAE1, y en el que cada distancia de una muestra_i a una clase_k se calcula utilizando la siguiente fórmula:

$$\text{Distancia}(\text{muestra}_i, \text{clase}_k) = \sum_{t=1..16} \frac{(\Delta Ct(\text{muestra}_i, \text{gen}_t) - \mu(\text{clase}_k, \text{gen}_t))^2}{\sigma(\text{gen}_t)}$$

en la que para cada gen_t y clase_k, los valores de $\mu(\text{clase}_k, \text{gen}_t)$ y $\sigma(\text{gen}_t)$ están en un intervalo de 10% alrededor de los valores que se muestran en la siguiente Tabla 5.

μ	clase 1	clase 2	clase 3	clase 4	clase 5	clase 6	σ
gen 1 (RAB1A)	-16,39	-16,04	-16,29	-17,15	-17,33	-16,95	0,23
gen 2 (PAP)	-28,75	-27,02	-23,48	-27,87	-19,23	-11,33	16,63
gen 3 (NRAS)	-16,92	-17,41	-16,25	-17,31	-16,96	-17,26	0,27
gen 4 (RAMP3)	-23,54	-23,12	-25,34	-22,36	-23,09	-23,06	1,23
gen 5 (MERTK)	-18,72	-18,43	-21,24	-18,29	-17,03	-16,16	7,23
gen 6 (PIR)	-18,44	-19,81	-16,73	-18,28	-17,09	-17,25	0,48
gen 7	-16,68	-16,51	-19,89	-17,04	-18,70	-21,98	1,57

μ	clase 1	clase 2	clase 3	clase 4	clase 5	clase 6	σ
(EPA1)							
gen 8 (LAMA3)	-20,58	-20,44	-20,19	-21,99	-18,77	-16,85	2,55
gen 9 (G0S2)	-14,82	-17,45	-18,18	-14,78	-17,99	-16,06	3,88
gen 10 (HN1)	-16,92	-17,16	-15,91	-17,88	-17,72	-17,93	0,54
gen 11 (PAK2)	-17,86	-16,56	-16,99	-18,14	-17,92	-17,97	0,58
gen 12 (AFP)	-16,68	-12,36	-26,80	-27,28	-25,97	-23,47	14,80
gen 13 (CYP2C9)	-18,27	-16,99	-16,26	-16,23	-13,27	-14,44	5,47
gen 14 (CDH2)	-15,20	-14,76	-18,91	-15,60	-15,48	-17,32	10,59
gen 15 (HAMP)	-19,53	-20,19	-21,32	-18,51	-25,06	-26,10	13,08
gen 16 (SAE1)	-17,37	-17,10	-16,79	-18,22	-17,72	-18,16	0,31

6. Procedimiento según las reivindicaciones 1 o 2, en el que el perfil de expresión consiste en la siguiente combinación de 24 genes: ALDH1L1, CD24, CD74, CFHR3, CYP4F12, DNAJA3, DSCR1, EPA1, EPHB4, FAAH, FGFR2, FLJ10159, GLT8D1, HAL, MATN2, MRPS7, PAK2, PLXNB1, RAB1A, RHOQ, SLC27A5, SLPI, SMARCE1, STRA13.

7. Procedimiento según la reivindicación 6, en el que el perfil de expresión se determina a nivel de ácido nucleico utilizando una micromatriz de ácidos nucleicos.

8. Procedimiento según la reivindicación 7, en el que el perfil de expresión consiste en la siguiente combinación de 24 genes: ALDH1L1, CD24, CD74, CFHR3, CYP4F12, DNAJA3, DSCR1, EPA1, EPHB4, FAAH, FGFR2, FLJ10159, GLT8D1, HAL, MATN2, MRPS7, PAK2, PLXNB1, RAB1A, RHOQ, SLC27A5, SLPI, SMARCE1, STRA13, y en el que cada distancia de una muestra_i a una clase_k se calcula utilizando la siguiente fórmula:

$$\text{Distancia}(muestra_i, clase_k) = \left(\frac{\sum_{t=1..24} (c(gen_t, clase_k))^2}{2} + 1,791759 \right) - \left(\sum_{t=1..24} \frac{(y(muestra_i, gen_t) - \mu(gen_t))}{\sigma(gen_t)} \times c(gen_t, clase_k) \right)$$

(V)

en la que para cada gen_t y $clase_k$, los valores $c(gen_t, clase_k)$, $\mu(gen_t)$ y $\sigma(gen_t)$ están en un intervalo de 10% alrededor de los valores que se muestran en la siguiente Tabla 6:

Nº de gen	Símbolo del gen	C						μ	σ
		clase 1	clase 2	clase 3	clase 4	clase 5	clase 6		
1	MATN2	0,70615962	0,194820133	0	0,251226887	-0,43679693	-0,95672006	6,86694444	1,3591753
2	EPHB4	0,772905372	0	-0,23845281	0	0	-0,11554095	8,12722222	0,90851011
3	SLP	0,348202773	1,150217317	-0,30674955	0,772922695	-0,92749405	-1,03709919	9,19555556	1,58143433
4	FAAH	0	-0,05721942	-0,14546431	0,461062372	0	0	7,35527778	0,90656803
5	ALDH1L1	-0,90005519	-0,88021827	0	0,001883282	0,766706969	0,727033118	9,55166667	1,51233112
6	DNAJA3	-0,38971634	0	0	0	0,115122957	0	9,20444444	0,79106868
7	EPHA1	0,084999966	1,076607341	-0,39390753	0,336075517	-0,15411627	-0,94965903	8,34	1,18158325
8	CYP4F12	0	0	-0,50724032	0	0,411786437	0	8,61416667	0,89048235
9	CFHR4 // CFHR3	-0,3047186	0	-0,85911235	0,594911312	0,240167858	0	9,63111111	1,9450676
10	FGFR2	1,107271104	1,414413219	0	-0,33319985	-0,93870859	-0,78623447	6,45694444	1,51938352
11	CD24	0,877735471	0,272141138	0,391097168	-0,19493988	-0,09580985	-1,25022405	6,88527778	1,84942286
12	RAB1A	0	0,181176347	0,101889323	0	-0,13052075	0	10,9441667	0,75674426
13	PAK2	0,333384856	0	0,506491717	-0,14545539	-0,07557739	-0,39796907	7,58722222	1,04944811
14	STRA13	-0,61853503	-0,18082898	0,428679271	-0,14603227	0,113677426	0,403039586	9,56055556	0,91004758
15	CD74	-0,20143371	0	0	0	-0,40543025	0,741093354	10,9897222	1,28270148
16	SMARCE1	0,126430939	0	0,308479215	-0,17753221	0	-0,25610041	8,43638889	0,86973267
17	RHOA	0	0,321593401	0,234570894	0	-0,196463	-0,09749623	8,02888889	0,97675881
18	DSCR1	-0,04133466	-0,10070514	-0,07191945	0,033538532	0	0,436538207	7,40194444	0,92638619
19	PLXNB1	0,550893643	0,055723559	-0,24782026	-0,15487024	0	-0,08389023	8,10722222	0,98619313
20	HAL	0,428036608	1,655082264	-0,24462171	0,139158147	-0,85676893	-1,12088638	7,73055556	1,21158219
21	MRPS7	-0,24232641	0	0,552001257	0	0	0	8,77777778	0,88623422
22	GLT8D1	-0,08380816	-0,21624193	-0,11283474	0	0,458734816	0,34625682	9,20027778	0,91869819
23	FLJ10159	1,275819214	0,257167026	0	-0,03229518	-0,27448613	-0,99690117	7,33638889	1,60341806
24	SLC27A5	-0,93094948	-0,2491716	-0,50142942	0,607483755	0,535897078	0,538169671	9,89388889	1,46675337

9. Procedimiento según cualquiera de las reivindicaciones 1 a 8, en el que la muestra de hígado HCC es una biopsia de hígado HCC o una extirpación quirúrgica de tumor de HCC.

5 10. Kit para la clasificación *in vitro* de un tumor de HCC en 6 subgrupos G1, G2, G3, G4, G5 y G6 a partir de una muestra de hígado HCC de un sujeto que sufre HCC, que consiste en reactivos para la determinación de un perfil de expresión que consiste en:

10 (i) la siguiente combinación de 16 genes: RAB1A, REG3A, NRAS, RAMP3, MERTK, PIR, EPHA1, LAMA3, GOS2, HN1, PAK2, AFP, CYP2C9, CDH2, HAMP, y SAE1, o

- (ii) la siguiente combinación de 24 genes: ALDH1L1, CD24, CD74, CFHR3, CYP4F12, DNAJA3, DSCR1, EPHA1, EPHB4, FAAH, FGFR2, FLJ10159, GLT8D1, HAL, MATN2, MRPS7, PAK2, PLXNB1, RAB1A, RHOQ, SLC27A5, SLPI, SMARCE1, STRA13,

5 en el que los 6 subgrupos G1, G2, G3, G4, G5 y G6 están definidos por la presencia (+) o ausencia (-) de características clínicas y genéticas descritas en la siguiente Tabla 2:

	G1	G2	G3	G4	G5	G6
Inestabilidad cromosómica	+	+	+	-	-	-
Recaída temprana y muerte	+	+	+	-	-	-
Mutación TP53	-	+	+	-	-	-
Infección HBV	+	+	-	-	-	-
Bajo número de copias	+	-	-	-	-	-
Alto número de copias	-	+	-	-	-	-
Mutación CTNNB1	-	-	-	-	+	+
Nódulos satélite	-	-	-	-	-	+

- 10 11. Kit según la reivindicación 10, en el que dichos reactivos se seleccionan de entre cebadores de amplificación, sondas nucleicas y micromatrices de ácidos nucleicos.
12. Kit según la reivindicación 10, que consiste en reactivos para la determinación de un perfil de expresión que consiste en la siguiente combinación de 16 genes: RAB1A, REG3A, NRAS, RAMP3, MERTK, PIR, EPHA1, LAMA3, G0S2, HN1, PAK2, AFP, CYP2C9, CDH2, HAMP, y SAE1, en el que dichos reactivos comprenden cebadores de amplificación, y opcionalmente sondas nucleicas útiles para un análisis por PCR cuantitativa de dicha expresión de genes.
- 15 13. Kit según la reivindicación 10, que consiste en reactivos para la determinación de un perfil de expresión que consiste en la siguiente combinación de 24 genes: ALDH1L1, CD24, CD74, CFHR3, CYP4F12, DNAJA3, DSCR1, EPHA1, EPHB4, FAAH, FGFR2, FLJ10159, GLT8D1, HAL, MATN2, MRPS7, PAK2, PLXNB1, RAB1A, RHOQ, SLC27A5, SLPI, SMARCE1, STRA13, en el que dichos reactivos comprenden una micromatriz de ácidos nucleicos.
- 20 14. Kit según cualquiera de las reivindicaciones 10 a 13, que comprende además instrucciones para realizar el procedimiento según cualquiera de las reivindicaciones 1 a 9.
- 25

Vías biológicas	Características clínicas	Metilación génica	Mutación génica	LOH cromosómica			
<u>Ciclo celular,</u> <u>unión a ácido</u> <u>nucleico</u>	<u>Genes de desarrollo</u> <u>y de impronta</u>	Mujer África, joven AFP alta suero	Núm. copias bajo	G1	<i>AXIN1</i>	16q	Inestabilidad cromosómica
	<u>Fosforilación</u> <u>Actividad quinasa</u>	Jóven Hemocrom Invasión portal	HBV Núm. copias alto	G2	<i>PIK3CA</i>	13q, 1p	
	<u>Importación/</u> <u>Exportación</u> <u>nuclear</u>			G3	<i>CDKN2A</i>	<i>TP53</i> 17p, 5q 19p, 21q 22q	
 <u>Estrés y respuesta</u> <u>inmune</u> <u>Adhesión celular</u>		No invasión vascular		G4	<i>TGF1</i>		Estabilidad cromosómica
				G5			
				G6	<i>CDH1</i>	<i>CTNNB1</i> 8p	



Figura 1

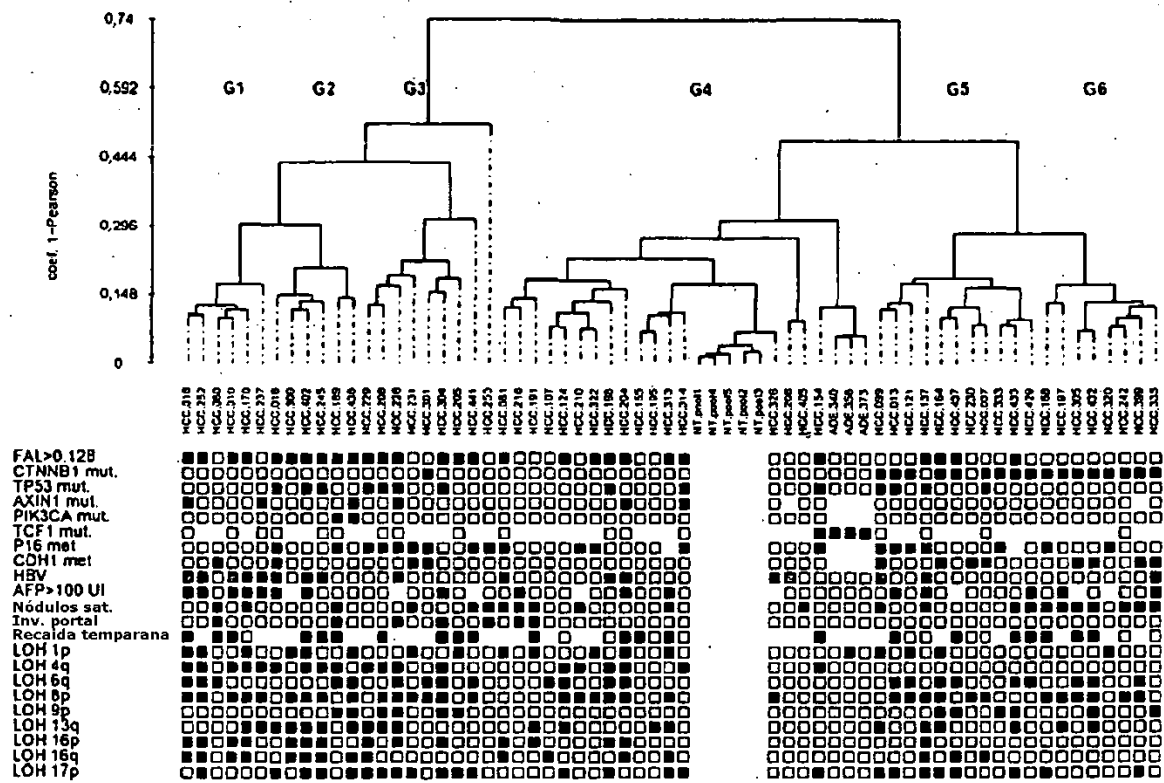


Figura 2

a

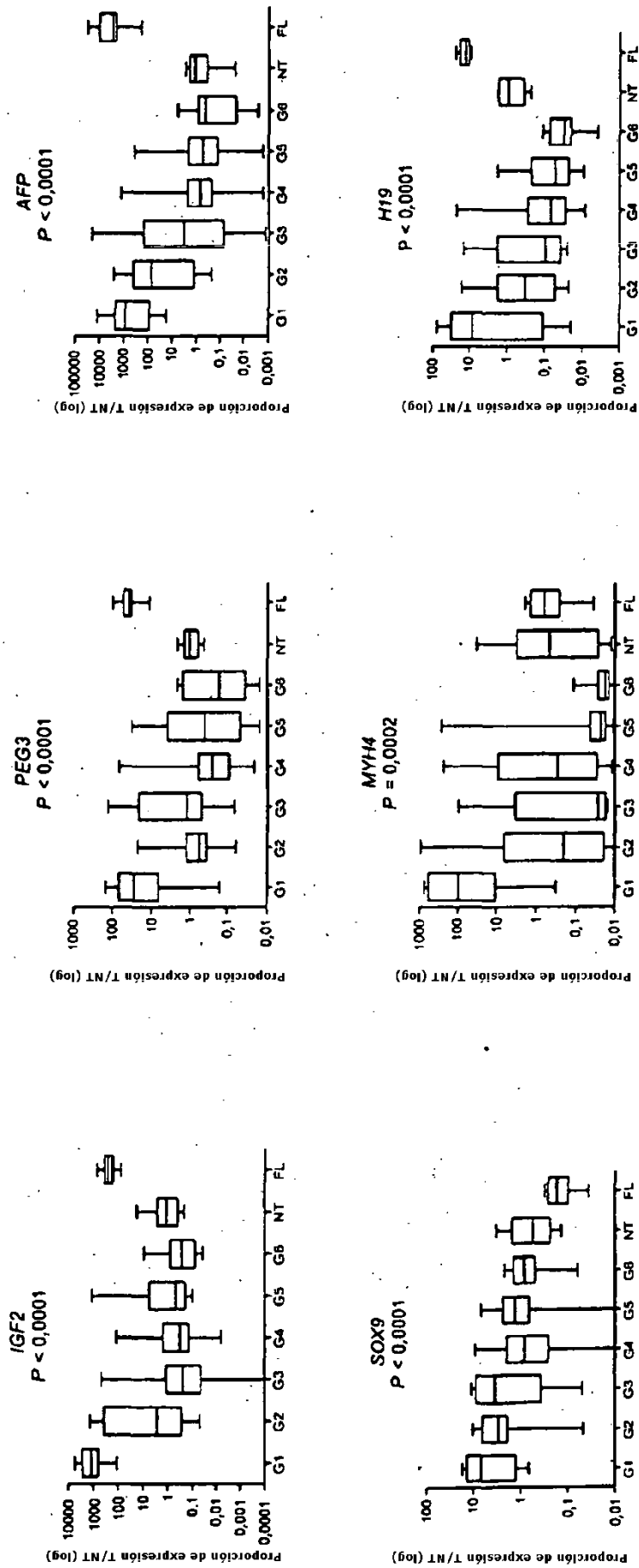


Figura 3a

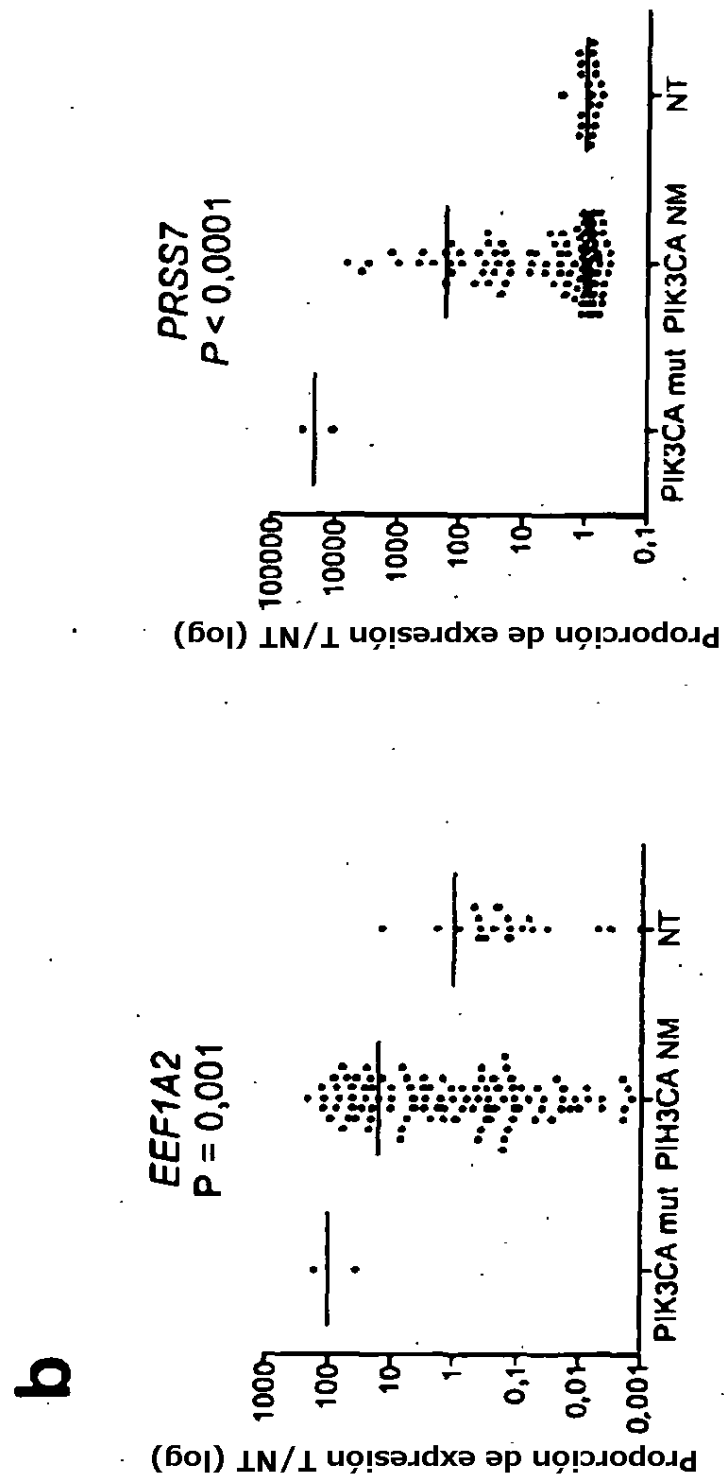


Figura 3b

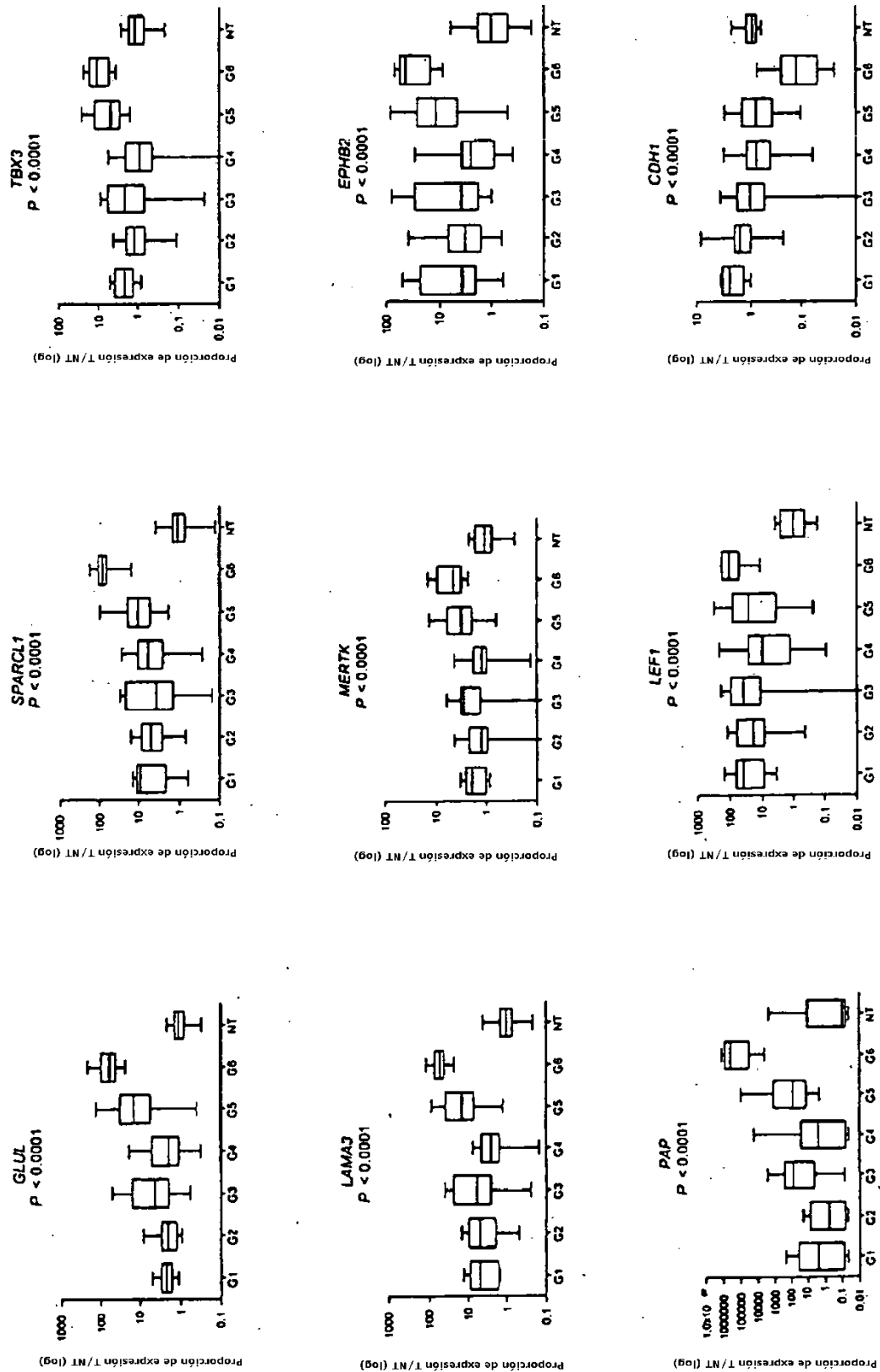
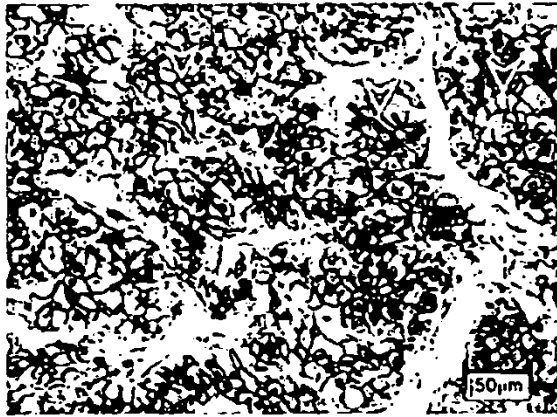
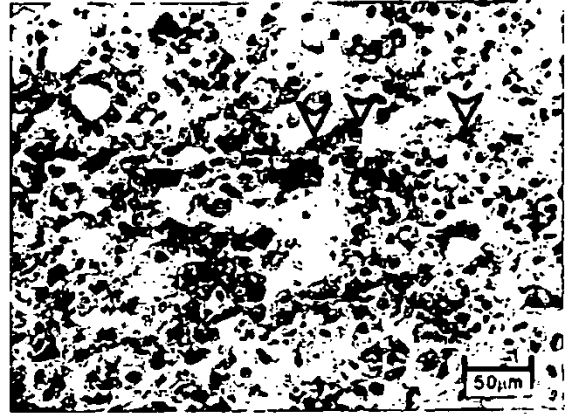


Figura 4a



HCC 303 G5



HCC 305 G6

Figura 4b

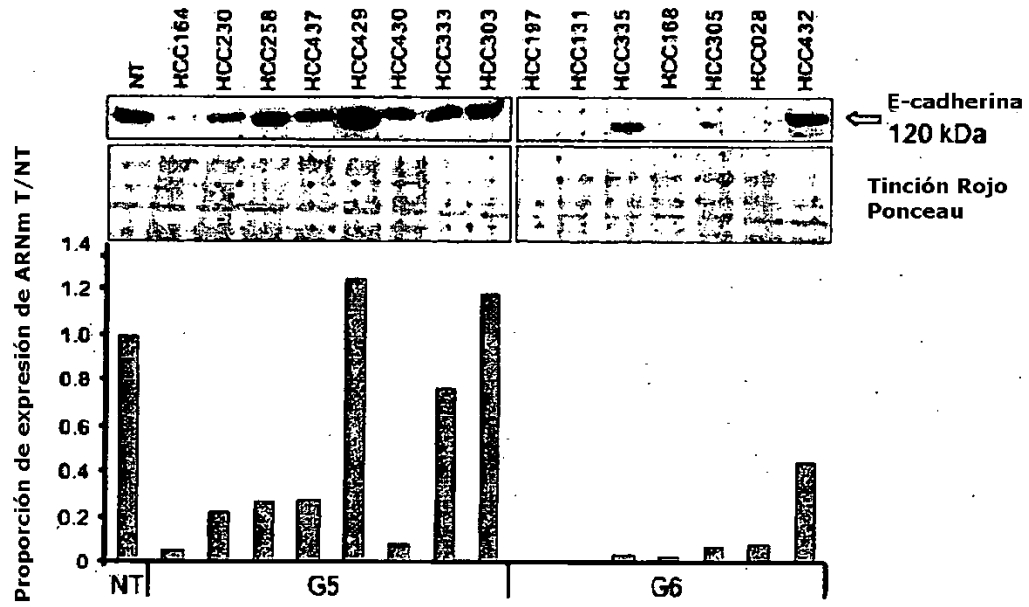


Figura 4c

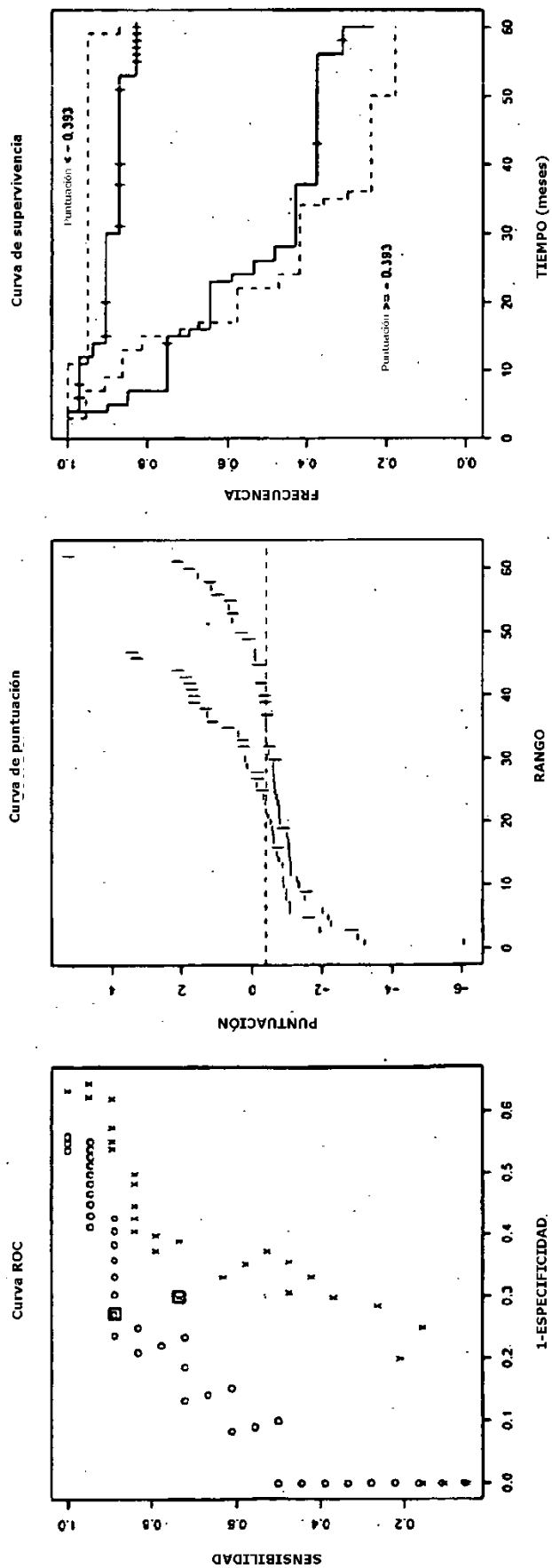


Figura 5a

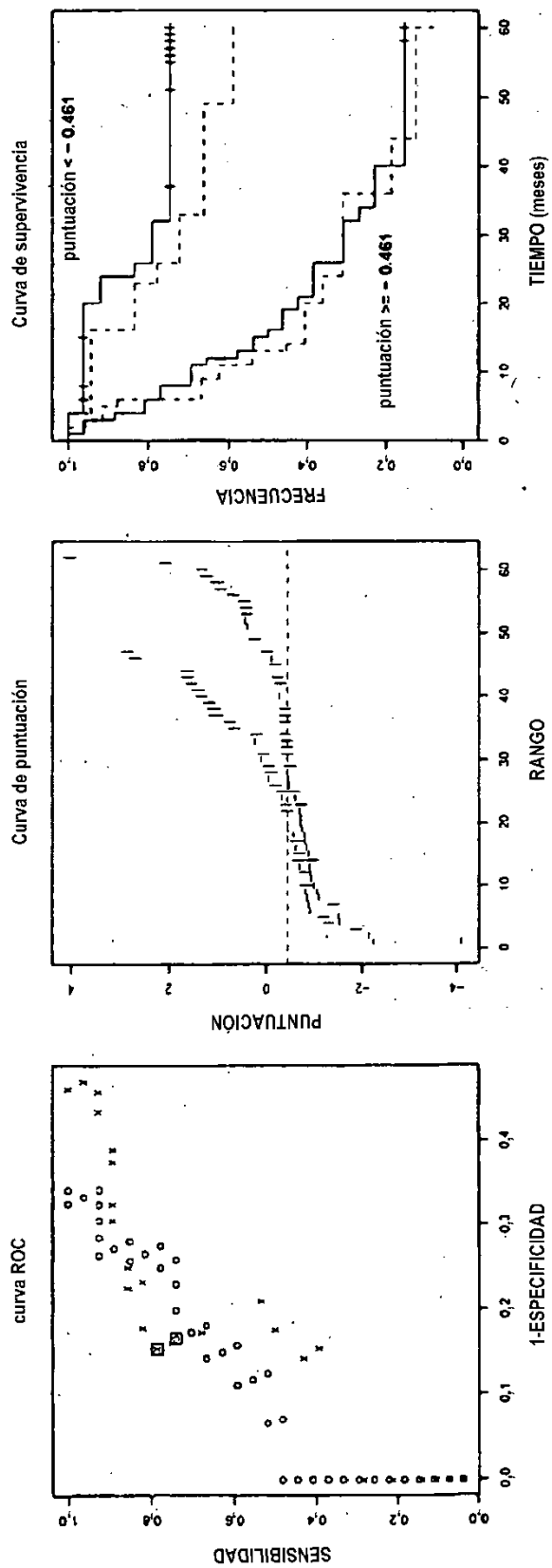


Figura sb