

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 396 967**

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: **08.04.2003 E 03816604 (7)**

97 Fecha y número de publicación de la concesión europea: **03.10.2012 EP 1613767**

54 Título: **Método de definición del grado de diferenciación de un tumor**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
01.03.2013

73 Titular/es:

**F. HOFFMANN-LA ROCHE AG (100.0%)
GRENZACHERSTRASSE 124
4070 BASEL, CH**

72 Inventor/es:

**OKA, MASAOKI;
HAMAMOTO, YOSHIHIKO;
IIZUKA, NORIO;
OKABE, HISAFUMI y
HAMADA, KENJI**

74 Agente/Representante:

ARIAS SANZ, Juan

ES 2 396 967 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

Método de definición del grado de diferenciación de un tumor

5 **Campo técnico**

La presente invención se refiere a un método de definición del grado de diferenciación de un tumor. Más particularmente, la presente invención se refiere a un método de definición del grado de diferenciación de un tumor mediante la selección de genes y/o proteínas cuyo nivel de expresión se correlaciona con cada grado de diferenciación del carcinoma hepatocelular (CHC), midiendo de la expresión de los genes y/o proteínas de los tejidos tumorales humanos en cada grado de diferenciación.

También se divulga el uso de estos genes y/o proteínas para diagnosticar el grado de diferenciación del CHC y para el cribado de agentes anticancerosos para el tratamiento del CHC.

También se divulga un kit para realizar el método de la presente invención que comprende chips de ADN, chips de oligonucleótidos, chips de proteínas, péptidos, anticuerpos, sondas y cebadores que son necesarios para micromatrices de ADN, micromatrices de oligonucleótidos, matrices de proteínas, transferencia northern, hibridación *in situ*, ensayos de protección de RNasas, inmunotransferencia, ensayos ELISA, transcripción inversa reacción en cadena de la polimerasa (de aquí en adelante denominado RT-PCR) para examinar la expresión de los genes y/o proteínas cuyo nivel de expresión se correlaciona con el grado de diferenciación del tumor.

Antecedentes técnicos

El cáncer es la principal causa de muerte en el mundo. Particularmente, el carcinoma hepatocelular (CHC) es uno de los cánceres más comunes en el mundo, que representa un problema internacional principal de salud debido a su incidencia creciente en muchos países (Schafer, D.E. y Sorrell, M.F. Hepatocellular carcinoma, *Lancet* **353**, 1253-1257 (1999), Colombo, M. Hepatitis C virus and hepatocellular carcinoma, *Semin. Liver Dis.* **19**, 263-269 (1999), y Okuda, K. Hepatocellular carcinoma, *J. Hepatol.* **32**, 225-237 (2000)). La infección crónica por el virus de la hepatitis C (HCV) es uno de los principales factores de riesgo para el CHC así como la infección por el virus de la hepatitis B (HBV), consumo de alcohol y varios carcinógenos tal como aflatoxina B1 (Okuda, K. Hepatocellular carcinoma, *J. Hepatol.* **32**, 225-237 (2000)). Se han adoptado varias terapias para el tratamiento del CHC. Esas incluyen resección quirúrgica, radioterapia, quimioterapia y terapia biológica incluyendo terapia hormonal y génica. Sin embargo, ninguna de estas terapias pudo curar la enfermedad. Uno de los principales problemas del tratamiento del CHC es que las características de las células cancerosas cambian durante el desarrollo y evolución de la enfermedad. Particularmente, los cambios en el grado de diferenciación de las células tumorales son aparentes y frecuentes. Tales cambios alteran la capacidad de las células tumorales de invadir y metastatizar y también la sensibilidad de las células cancerosas a las diferentes terapias, produciendo resistencia a agentes anticancerosos. Si los cambios en las características de las células cancerosas se diagnostican y tratan de forma precisa, la terapia contra el cáncer será más eficaz.

Estudios previos sugirieron la implicación de genes supresores de tumores y oncogenes tales como los genes *p53*, *β-catenina* y *AXIN1* en la hepatocarcinogénesis (Okabe, H., Satoh, S., Kato, T., Kitahara, O., Yanagawa, R., Yamaoka, Y., Tsunoda, T., Furukawa, Y., y Nakamura, Y. Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: identification of genes involved in viral carcinogenesis and tumor progression, *Cancer Res.* **61**, 2129-2137 (2001)). También se ha sugerido que el desarrollo de CHC asociado al HCV se puede caracterizar por la evolución patológica desde el tumor temprano al avanzado, que se correlaciona con la dediferenciación de las células cancerosas (Kojiro, M. Pathological evolution of early hepatocellular carcinoma, *Oncology* **62**, 43-47 (2002)). Particularmente después de la introducción de las tecnologías de micromatrices de ADN en las ciencias médicas (Skena, M., Shalon, D., Davis, R.W., y Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray, *Science* **270**, 467-470 (1995), DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A., y Trent, J.M. Use of a cDNA microarray to analyse gene expression patterns in human cancer, *Nat. Genet.* **14**, 457-460 (1996)), muchos estudios mostraron patrones de expresión génica relacionados con algunos aspectos del CHC (Lau, W.Y., Lai, P.B., Leung, M.F., Leung, B.C., Wong, N., Chen, G., Leung, T.W., y Liew, C.T. Differential gene expression of hepatocellular carcinoma using cDNA microarray analysis, *Oncol. Res.* **12**, 59-69 (2000), Tackels-Horne, D., Goodman, M.D., Williams, A.J., Wilson, D.J., Eskandari, T., Vogt, L.M., Boland, J.F., Scherf, U., y Vockley, J.G. Identification of differentially expressed genes in hepatocellular carcinoma and metastatic liver tumors by oligonucleotide expression profiling, *Cancer* **92**, 395-405 (2001), Xu, L., Hui, L., Wang, S., Gong, J., Jin, Y., Wang, Y., Ji, Y., Wu, X., Han, Z., y Hu, G. Expression profiling suggested a regulatory role of liver-enriched transcription factors in human hepatocellular carcinoma, *Cancer Res.* **61**, 3176-3681 (2001), Xu, X.R., Huang, J., Xu, Z.G., Qian, B.Z. Zhu, Z.D., Yan, Q., Cai, T., Zhang, X., Xiao, H.S., Qu, J., Liu, F., Huang, Q.H., Cheng, Z.H., Li, N.G., Du, J.J., Hu, W., Shen, K.T., Lu, G., Fu, G., Zhong, M., Xu, S.H., Gu, W.Y., Huang, W., Zhao, X.T., Hu, G.X., Gu, J.R., Chen, Z., y Han, Z.G. Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding non-cancerous liver, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 15089-15094 (2001), Okabe, H., Satoh, S., Kato, T., Kitahara, O., Yanagawa, R., Yamaoka, Y., Tsunoda, T., Furukawa, Y., y Nakamura,

Y. Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: identification of genes involved in viral carcinogenesis and tumor progression, *Cancer Res.* **61**, 2129-2137 (2001), Shirota, Y., Kaneko, S., Honda, M., Kawai, H.F., y Kobayashi, K. Identification of differentially expressed genes in hepatocellular carcinoma with cDNA microarrays, *Hepatology* **33**, 832-840 (2001), Delpuech, O., Trabut, J.B., Carnot, F., Feuillard, J., Brechot, C., y Kremsdorf, D. Identification, using cDNA macroarray analysis, of distinct gene expression profiles associated with pathological and virological features of hepatocellular carcinoma, *Oncogene* **21**, 2926-2937 (2002), Iizuka, N., Oka, M., Yamada-Okabe, H., Mori, N., Tamesa, T., Okada, T., Takemoto, T., Tangoku, A., Hamada, K., Nakayama, H., Miyamoto, T., Uchimura, S., y Hamamoto, Y. Comparison of gene expression profiles between hepatitis B virus- and hepatitis C virus-infected hepatocellular carcinoma by oligonucleotide microarray data based on a supervised learning method, *Cancer Res.* **62**, 3939-3944 (2002), y Midorikawa, Y., Tsutsumi, S., Taniguchi, H., Ishii, M., Kobune, Y., Kodama, T., Makuuchi, M., y Aburatani, H. Identification of genes associated with dedifferentiation of hepatocellular carcinoma with expression profiling analysis, *Jpn. J. Cancer Res.* **93**, 636-643 (2002)). Entre ellos, dos estudios determinaron el perfil de expresión génica en CHC respecto a su desarrollo (Okabe, H., Satoh, S., Kato, T., Kitahara, O., Yanagawa, R., Yamaoka, Y., Tsunoda, T., Furukawa, Y., y Nakamura, Y. Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: identification of genes involved in viral carcinogenesis and tumor progression, *Cancer Res.* **61**, 2129-2137 (2001) y Midorikawa, Y., Tsutsumi, S., Taniguchi, H., Ishii, M., Kobune, Y., Kodama, T., Makuuchi, M., y Aburatani, H. Identification of genes associated with dedifferentiation of hepatocellular carcinoma with expression profiling analysis, *Jpn. J. Cancer Res.* **93**, 636-643 (2002)). Sin embargo, no se sabe nada sobre los genes y/o las proteínas que caracterizan y/o regulan cada grado de diferenciación del CHC durante el curso de la oncogénesis y el desarrollo de CHC asociado al HCV. Se pueden usar genes y/o proteínas que regulan el grado de diferenciación del CHC para diagnosticar el grado de diferenciación del CHC y para el cribado de agentes anticancerosos para el tratamiento del CHC que surge de la infección crónica de HCV.

Previamente, se ha divulgado un sistema de puntuación para la predicción de la reaparición de cáncer que implica el cociente de Fisher (documento WO 03/010337). Además, se ha descrito la comparación de los perfiles de expresión génica entre carcinoma hepatocelular infectado por virus de la hepatitis B y por virus de la hepatitis C mediante datos de micromatrices de oligonucleótidos en base a un método de aprendizaje supervisado que aplica el cociente de Fisher (Iizuka et al., *Cancer Research* 62: 3939-3944 (2002)). Además, los patrones de expresión génica en diferenciación hematopoyética se han interpretado previamente con mapas autoorganizados (Tamayo et al., *Proc. Natl. Acad. Sci. USA* 96: 29072912 (1999)). De forma similar, previamente se ha usado un mapa autoorganizado para el análisis de la expresión génica en (Törönen et al., *FEBS letters* 451: 142-146 (1999)).

En la presente invención, los inventores describen un método de diagnóstico del grado de diferenciación de un tumor y cribado de agentes anticancerosos para el tratamiento del mismo. Particularmente, los inventores describen un método de identificar 40 o más genes y/o proteínas cuya expresión se correlaciona con el grado de diferenciación del CHC, y el uso de estos genes y/o proteínas para diagnosticar el grado de diferenciación del CHC y para cribar agentes anticancerosos para el tratamiento del CHC en diferentes grados. Más particularmente, los inventores describen un método de predecir hígado no canceroso, hígado precanceroso y cada grado de diferenciación del CHC con 40 genes y/o proteínas.

Divulgación de la invención

Compendio de la invención

La presente invención se refiere a las formas de realización caracterizadas en las reivindicaciones. Por tanto, se refiere a los siguientes puntos:

1. Un método in vitro de definir el grado de diferenciación de un tumor con genes y/o proteínas seleccionados mediante los análisis estadísticos basados en el nivel o patrón de expresión de los genes y/o las proteínas de tejidos tumorales humanos obtenibles de pacientes de cáncer, en donde los genes y/o las proteínas se seleccionan en orden descendente del cociente de Fisher y en donde el cociente de Fisher se determina sin el uso de una probabilidad anterior, en donde el cociente de Fisher para un gen j está dado por

$$F(j) = \frac{(\bar{\mu}_j(A) - \bar{\mu}_j(B))^2}{\sigma_j^2(A) + \sigma_j^2(B)}$$

en donde $\bar{\mu}_j(i)$ es la media de la muestra del nivel de expresión del gen j para las muestras en grado i , y $\sigma_j^2(i)$ es la varianza de la muestra del nivel de expresión del gen j para las muestra en grado i .

2. El método in vitro según el punto 1, en donde los tejidos humanos son tejidos hepáticos humanos.

3. El método in vitro según el punto 2, en donde el grado de diferenciación del tumor se selecciona del grupo que consiste en hígado no canceroso, hígado precanceroso, carcinoma hepatocelular (CHC) bien diferenciado, CHC moderadamente diferenciado y CHC mal diferenciado.
- 5 4. El método in vitro según el punto 3, en donde los genes y/o las proteínas se expresan diferencialmente entre hígado no canceroso e hígado precanceroso, hígado precanceroso y carcinoma hepatocelular (CHC) bien diferenciado, CHC bien diferenciado y CHC moderadamente diferenciado, o CHC moderadamente diferenciado y CHC mal diferenciado.
- 10 5. El método in vitro según cualquiera de los puntos 1 a 4, en donde el nivel o patrón de expresión de los genes y/o las proteínas se examina por medio de micromatriz de ADN, transcripción inversa reacción en cadena de la polimerasa o matriz de proteínas.
- 15 6. El método in vitro según el punto 5, en donde el número de genes y/o proteínas está entre 40 y 100.
7. El método in vitro según el punto 5, en donde el número de genes y/o proteínas está entre 35 y 45.
8. El método in vitro según el punto 7, en donde el número de genes y/o proteínas es 40.
- 20 9. Un método in vitro de definir el grado de diferenciación de un tumor, el método comprende los pasos de:
- (a) seleccionar genes y/o proteínas que tienen los mayores cocientes de Fisher en la comparación entre hígado no canceroso e hígado precanceroso, hígado precanceroso y carcinoma hepatocelular (CHC) bien diferenciado, CHC bien diferenciado y CHC moderadamente diferenciado, o CHC moderadamente diferenciado y CHC mal diferenciado; y
- 25 (b) definir el grado de diferenciación del tumor usando los genes y/o las proteínas, en donde el cociente de Fisher se determina sin el uso de una probabilidad anterior, en donde el cociente de Fisher para un gen j está dado por

30
$$F(j) = \frac{(\bar{\mu}_j(A) - \bar{\mu}_j(B))^2}{\sigma_j^2(A) + \sigma_j^2(B)}$$

donde $\bar{\mu}_j(i)$ es la media de la muestra del nivel de expresión del gen j para las muestras en grado i , y $\sigma_j^2(i)$ es la varianza de la muestra del nivel de expresión del gen j para las muestra en grado i .

- 35 10. Un método in vitro de definir el grado de diferenciación de un tumor, el método comprende los pasos de:
- (a) determinar el número de genes y/o proteínas para definir el grado de diferenciación del tumor,
- (b) seleccionar un número de genes y/o proteínas decididos en el paso (a) que tienen los mayores cocientes de Fisher en la comparación entre hígado no canceroso e hígado precanceroso, hígado precanceroso y carcinoma hepatocelular (CHC) bien diferenciado, CHC bien diferenciado y CHC moderadamente diferenciado, o CHC moderadamente diferenciado y CHC mal diferenciado;
- 40 (c) aplicar los datos de genes y/o proteínas seleccionados en el paso (b) a todas las muestras; y
- (d) definir el grado de diferenciación del tumor,
- en donde el cociente de Fisher se determina sin el uso de una probabilidad anterior, en donde el cociente de Fisher para un gen j está dado por
- 45

50
$$F(j) = \frac{(\bar{\mu}_j(A) - \bar{\mu}_j(B))^2}{\sigma_j^2(A) + \sigma_j^2(B)}$$

donde $\bar{\mu}_j(i)$ es la media de la muestra del nivel de expresión del gen j para las muestras en grado i , y $\sigma_j^2(i)$ es la varianza de la muestra del nivel de expresión del gen j para las muestra en grado i .

11. Un método in vitro de definir el grado de diferenciación de un tumor, el método comprende los pasos de:
- (a) determinar el número de genes y/o proteínas para definir el grado de diferenciación del tumor;
- 55 (b) seleccionar un número de genes y/o proteínas decididos en el paso (a) que tienen los mayores cocientes de Fisher en la comparación entre hígado no canceroso e hígado precanceroso, hígado precanceroso y carcinoma hepatocelular (CHC) bien diferenciado, CHC bien diferenciado y CHC moderadamente diferenciado, o CHC moderadamente diferenciado y CHC mal diferenciado;
- (c) aplicar los datos de genes y/o proteínas seleccionados en el paso (b) a todas las muestras;

- (d) diseñar un clasificador de distancia mínima con los datos de los genes y/o las proteínas seleccionados en el paso (b);
 (e) aplicar el clasificador de distancia mínima diseñado en el paso (d) a todas las muestras;
 (f) generar un mapa autoorganizado con los datos de todos los genes y/o proteínas seleccionados en el paso (b);
 (g) aplicar el mapa autoorganizado generado en paso (f) a todas las muestras; y
 (h) definir el grado de diferenciación del tumor,
 en donde el cociente de Fisher se determina sin el uso de una probabilidad anterior, en donde el cociente de Fisher para un gen j está dado por

$$F(j) = \frac{(\hat{\mu}_j(A) - \hat{\mu}_j(B))^2}{\hat{\sigma}_j^2(A) + \hat{\sigma}_j^2(B)}$$

donde $\hat{\mu}_j(i)$ es la media de la muestra del nivel de expresión del gen j para las muestras en el grado i , y $\hat{\sigma}_j^2(i)$ es la varianza de la muestra del nivel de expresión del gen j para las muestras en el grado i .

El carcinoma hepatocelular (CHC) es uno de los cánceres más comunes en el mundo. Sin embargo, no hay terapia que pueda curar la enfermedad. Esto se debe presumiblemente a los cambios secuenciales en las características de las células cancerosas durante el desarrollo y evolución de la enfermedad. Particularmente, la evolución del cáncer con frecuencia se asocia con los cambios del grado de diferenciación de las células tumorales. El diagnóstico y tratamiento de tales cambios de las células cancerosas hará la terapia del cáncer más eficaz. En la presente invención, se identifican genes cuya expresión se correlaciona con la oncogénesis y el desarrollo del CHC mediante una micromatriz de oligonucleótidos que representa aproximadamente 11.000 genes de 50 tejidos de CHC asociado con virus de la hepatitis C (HCV) y 11 tejidos hepáticos no tumorales (no cancerosos y precancerosos).

Los estadios de diferenciación se dividen en 5 grados.

El hígado no canceroso (L0) es el hígado que es histológicamente normal y es seronegativo tanto para el antígeno de superficie del virus de la hepatitis B como para anticuerpo de HCV. El hígado precanceroso (L1) es el hígado que está infectado por HCV y está histopatológicamente diagnosticado como hepatitis crónica o cirrosis hepática. El CHC bien diferenciado (G1) es el CHC que consiste en células cancerosas que se caracterizan por un aumento en la densidad celular con relaciones nuclear/citoplasma elevadas comparadas con hepatocitos normales pero muestran la morfología similar a hepatocitos normales. El CHC moderadamente diferenciado (G2) es el CHC que consiste en células cancerosas que son grandes e hiper cromáticas. Hay estructuras trabeculares o similares a glándulas en el nido de células cancerosas en grado G2. El CHC mal diferenciado (G3) es el CHC que consiste en las células cancerosas que son pleomórficas o multinucleadas. El tumor crece en masas sólidas o nidos celulares que carecen de organización arquitectural en el grado G3. Los tumores G1, G2 y G3 corresponden a los tipos I, II y III de la clasificación de Edmondson & Steiner, respectivamente (Edmondson, H.A. y Steiner, P.E. Primary carcinoma of the liver: a study of 100 cases among 48,900 necropsies, *Cancer* 7, 462-504 (1954)).

Se usa un método de aprendizaje supervisado seguido por una prueba de permutación aleatoria de los datos de la micromatriz de oligonucleótidos para seleccionar genes cuya expresión cambia significativamente durante la transición de hígado no canceroso sin infección de HCV (L0) a hígado precanceroso con infección de HCV (L1), de L1 a CHC bien diferenciado (G1), de G1 a CHC moderadamente diferenciado (G2), y de G2 a CHC mal diferenciado (G3). El mapa autoorganizado con los 40 genes seleccionados cuya expresión se altera significativamente en cada estadio de transición puede predecir correctamente el grado de diferenciación de los tejidos tumorales. Por tanto, estos genes se pueden usar para el diagnóstico del grado de diferenciación del CHC y para el cribado de agentes anticancerosos para el tratamiento del CHC en cada grado de diferenciación.

Descripción detallada de la invención

En la presente invención, se usan tejidos de carcinoma hepatocelular (CHC) y tejidos de hígado no tumoral (no canceroso y precanceroso) humanos. Los CHC con infección de HCV se usan para analizar los CHC. La presencia de infección por HCV y/o HBV se puede determinar mediante inmunoreactividad contra un anticuerpo anti-HCV y un anticuerpo anti-HVB o amplificando el genoma de HCV y/o HBV por PCR. El grado de diferenciación del CHC se puede determinar por examen histopatológico y los CHC se clasifican en CHC bien diferenciado (G1), CHC moderadamente diferenciado (G2) y CHC mal diferenciado (G3). Las muestras hepáticas no tumorales se pueden obtener de pacientes que se sometieron a resección hepática para tumores hepáticos benignos o metastásicos. Una muestra de hígado sin infección con HCV se clasifica como hígado no canceroso (L0) y esa con infección con HCV se clasifica como hígado precanceroso (L1). Después de extirpar los tejidos hepáticos durante la cirugía, es preferible que los tejidos se congelen inmediatamente en nitrógeno líquido o acetona que contenga nieve carbónica y se almacenen entre -70 y -80°C hasta su uso. Los tejidos se pueden o no embeber en un compuesto O.T.C. (Sakura-Seiki, Tokio, Japón, No. de catálogo 4583).

La expresión de genes y/o proteínas de tejidos de CHC y tejidos hepáticos no tumorales se puede analizar midiendo el nivel de ARN y/o proteínas. En la mayoría de los casos, el nivel del ARN y/o las proteínas se determina midiendo la fluorescencia de sustancias que incluyen fluoresceína y rodamina, quimioluminiscencia de luminol, radiactividad de materiales radioactivos incluyendo ^3H , ^{14}C , ^{35}S , ^{33}P , ^{32}P y ^{125}I , y densidad óptica. Por ejemplo, el nivel de expresión de ARN y/o proteínas se determina por métodos conocidos incluyendo micromatrices de ADN (Skena, M. et al. Quantitative monitoring of gene expression patterns with a complementary DNA microarray, *Science* **270**, 467-470 (1995) y Lipshutz, R.J. et al. High density synthetic oligonucleotide arrays, *Nat. Genet.* **21**, 20-24 (1999)), RT-PCR (Weis, J.H. et al. Detection of rare mRNAs via quantitative RT-PCR, *Trends Genet.* **8**, 263-264 (1992) y Bustin, S.A. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays, *J. Mol. Endocrinol.* **25**, 169-193 (2000)), transferencia northern e hibridación *in situ* (Parker, R.M. y Barnes, N.M. mRNA: detection *in situ* and northern hybridization, *Methods Mol. Biol.* **106**, 247-283 (1999)), ensayo de protección de RNasa (Hod, Y.A. Simplified ribonuclease protection assay, *BioTechniques* **13**, 852-854 (1992) y Saccomanno, C.F. et al. A faster ribonuclease protection assay, *BioTechniques* **13**, 846-850 (1992)), inmunotransferencia (Towbin, H. et al. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4350-4354 (1979) y Burnette, W.N. Western blotting: Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radioiodinated protein A, *Anal. Biochem.* **112**, 195-203 (1981)), ensayo ELISA (Engvall, E. y Perlman, P. Enzyme-linked immunosorbent assay (ELISA): Quantitative assay of immunoglobulin G, *Immunochemistry* **8**, 871-879 (1971)), y matrices de proteína (Merchant, M. y Weinberger, S.R. Review: Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry, *Electrophoresis* **21**, 1164-1177 (2000) y Paweletz, C.P. et al. Rapid protein display profiling of cancer progression directly from human tissue using a protein biochip, *Drug Dev. Res.* **49**, 34-42 (2000)).

Se seleccionan genes y/o proteínas que se expresan diferencialmente en cada grado de diferenciación de CHC e hígado no tumoral (no canceroso y precanceroso) comparando el nivel de expresión de los genes y/o proteínas entre los tejidos de CHC en cada grado de diferenciación y tejidos hepáticos no tumorales. Se identifican genes y/o proteínas que se expresan diferencialmente entre hígado no canceroso (L0) e hígado precanceroso que ha sido infectado por HCV (L1) comparando el nivel de expresión de cada gen y/o proteína entre tejidos hepáticos no cancerosos y tejidos hepáticos precancerosos. Se identifican genes y/o proteínas que se expresan diferencialmente entre hígado precanceroso y CHC bien diferenciado (G1) comparando el nivel de expresión de cada gen y/o proteína entre tejidos hepáticos precancerosos y tejidos de CHC bien diferenciado (CHC(G1)). Se identifican genes y/o proteínas que se expresan diferencialmente entre CHC bien diferenciado (G1) y CHC moderadamente diferenciado (G2) comparando el nivel de expresión de cada gen y/o proteína entre CHC(G1) y tejidos hepáticos de CHC moderadamente diferenciado (CHC(G2)). De forma similar, se identifican genes y/o proteínas que se expresan diferencialmente entre CHC moderadamente diferenciado (G2) y CHC mal diferenciado (G3) comparando el nivel de expresión de cada gen y/o proteína entre CHC(G2) y tejidos hepáticos de CHC mal diferenciado (CHC(G3)).

Las diferencias en el nivel de expresión de genes y/o proteínas de hígado no canceroso, hígado precanceroso, CHC bien diferenciado, CHC moderadamente diferenciado y CHC mal diferenciado se pueden analizar y detectar por métodos conocidos de análisis estadísticos. En todos los experimentos para comparar el nivel de expresión de genes y/o proteínas entre dos grados seleccionados de L0, L1, G1, G2 y G3, se toman los siguientes procedimientos.

En el primer paso, se seleccionan genes y/o proteínas con cierto nivel de expresión (por ejemplo, genes con un nivel de expresión mayor que 40 juzgado por las unidades arbitrarias por los resultados del chip de genes Affymetrix) en todas las muestras de CHC y en las muestras de hígado no canceroso e hígado precanceroso. Esta selección produce un cierto número de genes y/o proteínas. A continuación, se determina la capacidad discriminatoria de cada gen y/o proteína para discriminar L0 de L1, L1 de G1, G1 de G2, y G2 de G3 mediante el cociente de Fisher. El cociente de Fisher para un gen j está dado por

$$F(j) = \frac{(\hat{\mu}_j(A) - \hat{\mu}_j(B))^2}{\hat{\sigma}_j^2(A) + \hat{\sigma}_j^2(B)}$$

donde $\hat{\mu}_j(i)$ es la media de la muestra del nivel de expresión del gen j para las muestras en grado i , y $\hat{\sigma}_j^2(i)$ es la varianza de la muestra del nivel de expresión del gen j para la muestra en grado i .

En el segundo paso, los genes y/o proteínas seleccionados se clasifican en el orden de magnitud decreciente del cociente de Fisher. También se realiza una prueba de permutación aleatoria para determinar el número de genes y/o proteínas para definir el grado de diferenciación del CHC. En la prueba de permutación, las etiquetas de las muestras se permutan aleatoriamente entre dos grados que se van a comparar, y se computa de nuevo el cociente de Fisher para cada gen y/o proteína. Esta permutación al azar de las etiquetas de las muestras se repite 1.000 veces. Se asignan P_s a los cocientes de Fisher generados de los datos reales basados en la distribución de los cocientes de Fisher de los datos aleatorizados. A partir de la distribución de los cocientes de Fisher basados en los datos aleatorizados, se seleccionan los genes y/o proteínas que se determina que son estadísticamente significativos en dos grados por la prueba de permutación aleatoria. Más particularmente, se seleccionan los genes

y/o proteínas que tienen el valor de P menor de 0,005 mediante la prueba de permutación aleatoria entre dos grados. Entre estos genes y/o proteínas seleccionados, se seleccionan adicionalmente 40 genes y/o proteínas que tienen los mayores cocientes de Fisher en cada comparación entre hígado no canceroso (L0) e hígado precanceroso (L1), hígado precanceroso (L1) y CHC bien diferenciado (G1), CHC bien diferenciado (G1) y CHC moderadamente diferenciado (G2), y CHC moderadamente diferenciado (G2) y CHC mal diferenciado (G3).

La capacidad de los 40 genes y/o proteínas seleccionados para distinguir hígado no canceroso (L0) de hígado precanceroso (L1), hígado precanceroso (L1) de CHC bien diferenciado (G1), CHC bien diferenciado (G1) de CHC moderadamente diferenciado (G2), y CHC moderadamente diferenciado (G2) de CHC mal diferenciado (G3) se verifica mediante el clasificador de distancia mínima y el mapa autoorganizado (SOM).

El clasificador de distancia mínima se diseña usando los 40 genes y/o proteínas seleccionados en cada estadio de transición. El nivel de expresión de cada gen y/o proteína se normaliza para tener una media de cero y una varianza unidad usando todas las muestras de entrenamiento de dos grados. Después de medir la distancia euclidiana entre una muestra y cada vector media, la muestra se asigna al grado del vector media más cercano. El clasificador de distancia mínima que se crea con los 40 genes y/o proteínas seleccionados en cada estadio de transición también se usa para predecir el grado de diferenciación de las muestras de CHC cuyo grado de diferenciación no está determinado. Para diagnosticar el grado de diferenciación de los CHC, usando $\hat{\mu}_j(A)$ y $\hat{\mu}_j(B)$ previamente descritos, se obtiene la media de la muestra $\hat{\mu}_j$ de la mezcla que consiste en los grados A y B en un gen j mediante

$$\hat{\mu}_j = \frac{N_A}{N_A + N_B} \hat{\mu}_j(A) + \frac{N_B}{N_A + N_B} \hat{\mu}_j(B)$$

donde N_i es el número de muestras del grado i . A continuación, se obtiene la varianza de la muestra $\hat{\sigma}_j^2$ de la mezcla que consiste en los grados A y B en el gen j mediante

$$\hat{\sigma}_j^2 = \frac{1}{N_A + N_B - 1} \left[(N_A - 1) \hat{\sigma}_j^2(A) + (N_B - 1) \hat{\sigma}_j^2(B) + \frac{N_A N_B}{N_A + N_B} (\hat{\mu}_j(A) - \hat{\mu}_j(B))^2 \right]$$

Usando $\hat{\mu}_j$ y $\hat{\sigma}_j^2$, $\hat{\mu}$ y \hat{V} se definen mediante

$$\hat{\mu} = [\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_{40}]^T$$

$$\hat{V} = \begin{bmatrix} \frac{1}{\hat{\sigma}_1^2} & & & 0 \\ & \frac{1}{\hat{\sigma}_2^2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{\hat{\sigma}_{40}^2} \end{bmatrix}$$

A continuación, una muestra \mathbf{x} se normaliza mediante

$$\tilde{\mathbf{x}} = \hat{V}^T (\mathbf{x} - \hat{\mu})$$

donde $\tilde{\mathbf{x}}$ es la muestra normalizada. Usando las muestras normalizadas, se obtiene el vector media de la muestra para cada grado. En el clasificador de distancia mínima, el valor de puntuación se computa mediante

$$T_1(\tilde{\mathbf{x}}) = \|\tilde{\mathbf{x}} - \mu_{L0}\|^2 - \|\tilde{\mathbf{x}} - \tilde{\mu}_{L1}\|^2$$

$$T_2(\tilde{\mathbf{x}}) = \|\tilde{\mathbf{x}} - \mu_{L1}\|^2 - \|\tilde{\mathbf{x}} - \tilde{\mu}_{G1}\|^2$$

$$T_3(\tilde{\mathbf{x}}) = \|\tilde{\mathbf{x}} - \mu_{G1}\|^2 - \|\tilde{\mathbf{x}} - \tilde{\mu}_{G2}\|^2$$

$$T_4(\tilde{\mathbf{x}}) = \|\tilde{\mathbf{x}} - \mu_{G2}\|^2 - \|\tilde{\mathbf{x}} - \tilde{\mu}_{G3}\|^2$$

Usando cuatro clasificadores de distancia mínima, se puede diagnosticar el grado de diferenciación de los CHC como sigue:

- 5 (i) Una muestra normalizada \bar{x} se clasifica en el grado L0 si $T_1(\bar{x}) < 0, T_2(\bar{x}) < 0, T_3(\bar{x}) < 0$ y $T_4(\bar{x}) < 0$
- (ii) Una muestra normalizada \bar{x} se clasifica en el grado L1 si $T_1(\bar{x}) > 0, T_2(\bar{x}) < 0, T_3(\bar{x}) < 0$ y $T_4(\bar{x}) < 0$
- (iii) Una muestra normalizada \bar{x} se clasifica en el grado G1 si $T_1(\bar{x}) > 0, T_2(\bar{x}) > 0, T_3(\bar{x}) < 0$ y $T_4(\bar{x}) < 0$
- 10 (iv) Una muestra normalizada \bar{x} se clasifica en el grado G2 si $T_1(\bar{x}) > 0, T_2(\bar{x}) > 0, T_3(\bar{x}) > 0$ y $T_4(\bar{x}) < 0$
- (v) Una muestra normalizada \bar{x} se clasifica en el grado G3 si $T_1(\bar{x}) > 0, T_2(\bar{x}) > 0, T_3(\bar{x}) > 0$ y $T_4(\bar{x}) > 0$

15 El SOM es un algoritmo de red neuronal ampliamente usado para agrupamiento y es bien conocido como una herramienta eficaz para la visualización de datos multidimensionales (Tamayo, P. et al. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation, Proc. Natl. Acad. Sci. U.S.A. 96, 2907-2912 (1999) y Sultan, M. et al. Binary tree-structured vector quantization approach to clustering and visualizing microarray data, Bioinformatics Supl 1, S111-S119 (2002)). El SOM con los 40 genes y/o proteínas seleccionados se lleva a cabo según el método de MATLAB R13 con la caja de herramientas de SOM disponible en el sitio web, <http://www.cis.hut.fi/projects/somtoolbox/> (Kohonen, 2001).

25 Cada conjunto de cuarenta genes y/o proteínas cuya expresión se altera significativamente durante la transición de hígado no canceroso (L0) a hígado precanceroso (L1), de hígado precanceroso (L1) a CHC bien diferenciado (G1), de CHC bien diferenciado (G1) a CHC moderadamente diferenciado (G2), y de CHC moderadamente diferenciado (G2) a CHC mal diferenciado (G3) se usa para diagnosticar el grado de hepatocarcinogénesis del CHC, y también para el cribado de agentes anticancerosos que se usan para el tratamiento del CHC en cada grado.

30 Cada conjunto de cuarenta genes y/o proteínas cuya expresión se altera significativamente durante la transición de hígado no canceroso (L0) a hígado precanceroso (L1), de hígado precanceroso (L1) a CHC bien diferenciado (G1), de CHC bien diferenciado (G1) a CHC moderadamente diferenciado (G2), y de CHC moderadamente diferenciado (G2) a CHC mal diferenciado (G3) se expresa en bacterias, células eucariotas y sistemas sin células. Los agentes que afectan la expresión y/o función de los genes y/o proteínas se criban siguiendo la expresión y/o función. También se generan anticuerpos monoclonales contra las proteínas y se usan para el tratamiento del CHC en diferentes grados. Como anticuerpos monoclonales, se pueden generar anticuerpos monoclonales de ratón enteros, anticuerpos humanizados, anticuerpos quiméricos, anticuerpos de cadena sencilla, anticuerpos de cadena sencilla divalentes y/o anticuerpos biespecíficos contra las proteínas purificadas y se usan para el diagnóstico del grado del CHC y el tratamiento del mismo.

40 También se crea un kit para examinar la expresión de los genes y/o proteínas. El kit consiste en los componentes incluyendo reactivos para la extracción de ADN, enzimas para la síntesis del ADNc y ARNc, chips de ADN, chips de oligonucleótidos, chips de proteínas, sondas y cebadores para los genes, fragmentos de ADN de genes control y anticuerpos contra las proteínas. Los componentes del kit están fácilmente disponibles del mercado.

45 **Breve descripción de las figuras**

La figura 1 ilustra las imágenes en color de la expresión de 152 genes cuya expresión estaba significativamente alterada durante la transición de L0 a L1 (a), 191 genes cuya expresión estaba significativamente alterada durante la transición de L1 a G1 (b), 54 genes cuya expresión estaba significativamente alterada durante la transición de G1 a G2 (c) y 40 genes cuya expresión estaba significativamente alterada durante la transición de G2 a G3 (d). Los paneles e, f, g y h ilustran la expresión de los 40 genes seleccionados en cada estadio de transición en todas las muestras. Se muestra la expresión de los 40 genes seleccionados cuya expresión estaba significativamente alterada durante la transición de L0 a L1 (e), de L1 a G1 (f), de G1 a G2 (g) y de G2 a G3 (h). Los 40 genes seleccionados en cada estadio de transición distinguen muestras antes y después de la transición. Los genes se muestran en orden descendiente de cociente de Fisher y se indican por los números de acceso a GenBank.

El nombre de cada muestra se indica en la parte superior de cada foto (e-h); NL-64, NL-65, NL-66, NL-67, NL-68, NL-69, IL-49, IL-58, IL-59, IL-60, IL-62, G1-26T, G1-42T, G1-85T, G1-86T, G1-87T, G1-147T, G1-165T, G2-1T, G2-2T, G2-6T, G2-8T, G2-10T, G2-12T, G2-16T, G2-18T, G2-20T, G2-22T, G2-23T, G2-27T, G2-28T, G2-29T, G2-31T, G2-34T, G2-37T, G2-43T, G2-45T, G2-46T, G2-49T, G2-58T, G2-59T, G2-60T, G2-62T, G2-89T, G2-90T, G2-105T, G2-151T, G2-155T, G2-161T, G2-162T, G2-163T, G2-171T, G2-182T, G3-19T, G3-21T, G3-25T, G3-35T, G3-80T, G3-81T, G3-107T, G3-174T, desde la izquierda.

El nombre de cada gen se indica a la derecha de la foto. En el caso del panel e, M18533, AF035316, AL049942, L27479, "Fibronectina, ajuste alt. 1", U19765, X55503, AL046394, AB007886, AL050139, AF012086, A1539439,

M19828, U92315, D76444, X02761, AF001891, AI400326, AI362017, L13977, D32053, AF038962, AL008726, J03909, Z69043, AL080080, M63138, L09159, AF017115, M13560, M36035, U47101, U81554, M21186, D32129, AL022723, M83664, U50523, M81757, AF102803, desde arriba. En el caso del panel f, M93221, AF079221, V01512, D88587, U12022, AF055376, R93527, R92331, U83460, AF052113, H68340, M10943, M13485, U75744, X02544, M93311, Z24725, U22961, M62403, M35878, U84011 AF055030, L13977, D13891, M63175, AB023157, U20982, M14058, AL049650, U61232, AI991040, U64444, D63997, X55503, AL080181, X76228, AB018330, D76444, U70660, U10323, desde arriba. En el caso del panel g, M87434, M12963, AI625844, M97936, Z99129, L07633, D50312, U07364, AA883502, M97935, AF061258, AB007447, M97935, W28281, M97935, Y00281, D28118, AF104913, AA675900, L27706, D32050, M63573, AF014398, X70944, U70671, AA447263, AB014569, M23115, D38521, X00351, L11672, X82834, AB007963, U76247, X68560, AB015344, AB018327, AF004430, D14697, AB028449 desde arriba. En el caso panel h, AA976838, Z11793, AB002311, Y18004, AL031230, AF002697, AB014596, U49897, AF070570, M80482 AI263099, U22961, Z24725, U77594, L34081, M88458, U68723, X92098, D10040, AB023194, AF001903, X96752, AB006202, M75106, Y12711, D14662, S87759, Z48199, AF088219, AA453183, D31767, AB000095, AB006782, M21186, AB002312, U44772, AI541308, Z49107, U77735, M38449, desde arriba.

La figura 2 ilustra la validación de los 40 genes seleccionados en cada estadio de transición para distinguir el grado de diferenciación de CHC.

En cada transición, de L0 a L1 (a), de L1 a G1 (b), de G1 a G2 (c) y de G2 a G3 (d), se construyó el clasificador de distancia mínima con las muestras en dos grados de diferenciación consecutivos como se indica mediante la barra roja (muestras de entrenamiento) y se aplicó a las muestras en los grados de diferenciación restantes como se indica mediante la barra negra (muestras de prueba). El clasificador resultante clasificó las muestras de prueba con una precisión del 92% (a), el 98% (b), el 84% (c) y el 100% (d).

La figura 3 ilustra el resultados del análisis mediante el algoritmo del mapa autoorganizado (SOM) de los genes cuya expresión cambió durante la transición de hígado no canceroso (L0) a hígado precanceroso (L1), de hígado precanceroso (L1) a CHC bien diferenciado (G1), de CHC bien diferenciado (G1) a CHC moderadamente diferenciado (G2), y de CHC moderadamente diferenciado (G2) a CHC mal diferenciado (G3).

La figura 3a ilustra grupos de las muestras (tabla 1). Cada célula en la rejilla SOM corresponde a un grupo. Los vectores de células vecinas habitualmente se localizan próximos entre sí.

(m, n), índice de la célula localizada en la fila m y la columna n. NL-XX, muestras de hígado no canceroso sin infección por HCV (L0); IL-XX, muestras de hígado precanceroso infectado por HCV (L1); G1-XXT, muestras de CHC bien diferenciado (G1); G2-XXT, muestras de CHC moderadamente diferenciado; G3-XXT, muestras de hígado moderadamente diferenciado (G3).

El mapa muestra que las muestras claramente formaron una curva sigmoidea en el orden de L0, L1, G1, G2 y G3. Las muestras de G2 sin implicación vascular (letras azules) están situadas cerca de las muestras de G1 y las muestras de G2 con implicación vascular (letras rojas) están situadas cerca de las muestras de G3.

La figura 3b ilustra la distancia entre los grupos vecinos.

(m, n), índice de la célula localizada en la fila m y la columna n. El color de las células indica la distancia entre los grupos vecinos; un color rojo significa una distancia larga. Las células rojas en el área superior claramente muestran que las muestras de hígado no tumoral (no canceroso y precanceroso) y las muestras de CHC están relativamente separadas en todos los 40 genes seleccionados.

La tabla 1 ilustra los grupos de muestras con perfil de L0, L1, G1, G2 y G3 como se muestran en la figura 3a.

La tabla 2 ilustra los factores clinicopatológicos del CHC usado en la presente invención.

La tabla 3 ilustra los 40 mejores genes diferenciadores en L0 y L1.

La tabla 4 ilustra los 40 mejores genes diferenciadores en L1 y G1.

La tabla 5 ilustra los 40 mejores genes diferenciadores en G1 y G2.

La tabla 6 ilustra los 40 mejores genes diferenciadores en G2 y G3.

Mejor manera de llevar a cabo la invención

Los siguientes ejemplos únicamente ilustran el método preferido para la identificación y el uso de genes y/o proteínas que se expresan diferencialmente en hígado no canceroso, hígado precanceroso, CHC bien diferenciado, CHC moderadamente diferenciado y CHC mal diferenciado.

A continuación en el presente documento, la presente invención se describirá específicamente usando ejemplos, sin embargo, no se debe interpretar como que está limitada a los mismos.

5 **Ejemplo 1. Preparación de tejidos humanos**

Cincuenta pacientes se sometieron a tratamiento quirúrgico para CHC en el Hospital de la Universidad de Yamaguchi entre mayo de 1997 y agosto de 2000. Se obtuvo consentimiento informado escrito de todos los pacientes antes de la cirugía. El protocolo del estudio fue aprobado por el comité de ética para el uso de sujetos humanos de la Facultad de Medicina de la Universidad de Yamaguchi. Los 50 pacientes eran seropositivos para anticuerpo de HCV (HCVAb) y seronegativos para antígeno de superficie del virus de la hepatitis B (HBsAg). Se hizo un diagnóstico histopatológico del CHC en todos los casos después de la cirugía. Este examen histopatológico mostró que siete pacientes tenían CHC bien diferenciado (G1), 35 tenían CHC moderadamente diferenciado (G2) y los ocho restantes tenían CHC mal diferenciado (G3). Se determinaron factores clinicopatológicos según la clasificación TNM de la Unión Internacional contra el Cáncer. Se usaron la prueba exacta de Fisher, la prueba de la t de Student y la prueba U de Mann-Whitney para elucidar diferencias en las características clinicopatológicas entre los 3 grados de CHC G1, G2 y G3. $P < 0,05$ se consideró significativa.

Se obtuvieron seis muestras de hígado no canceroso de seis pacientes que se sometieron a resección hepática para tumores hepáticos benignos o metastásicos, y se confirmó que tenían hígados histopatológicamente normales. Todos eran seronegativos tanto para HBsAg como para HCVAb. También se prepararon cinco muestras de hígado infectado por HCV de áreas no tumorales de cinco pacientes con CHC. Las cinco muestras de hígado se diagnosticaron histopatológicamente como hepatitis crónica o cirrosis hepática. Se obtuvo el consentimiento informado por escrito de todos los pacientes antes de la cirugía.

25 **Ejemplo 2. Características clinicopatológicas de los CHC**

Los exámenes histológicos mostraron que, entre los 50 CHC asociados con HCV inscritos en este estudio, siete eran CHC bien diferenciado (G1), 35 eran CHC moderadamente diferenciado (G2) y los ocho restantes eran CHC mal diferenciado (G3) (tabla 2). El tamaño del tumor de los CHC G2 y G3 era significativamente mayor que el del CHC G1 ($p=0,0007$ y $p=0,028$, respectivamente, mediante la prueba U de Mann-Whitney). La incidencia de implicación vascular en los CHC G2 y G3 era significativamente mayor que en CHC G1 ($p=0,038$ mediante la prueba exacta de Fisher). En paralelo a la desdiferenciación de G1 a G3, el estadio del tumor era más avanzado ($p=0,066$ mediante la prueba exacta de Fisher). Por tanto, cada tipo de los CHC G1, G2 y G3 inscritos en este estudio mostró características correspondientes a desdiferenciación, es decir, tamaño tumoral, potencial metastásico y estadio tumoral, según los propuestos por Kojiro (Kojiro, M. Pathological evolution of early hepatocellular carcinoma, *Oncology* **62**, 43-47 (2002)).

40 **Ejemplo 3. Extracción del ARN de tejidos**

Se resuspendieron trozos de los tejidos (aproximadamente 125 mm^3) en TRIZOL (Life Technologies, Gaithersburg, EE UU, No. de catálogo 15596-018) o Sepasol-RNAI (Nacalai tesque, Kioto, Japón, no. de catálogo 306-55) y se homogenizaron dos veces con un Polytron (Kinematica, Littau, Suiza) (5 segundos a velocidad máxima). Después de la adición de cloroformo, los homogenizados de tejido se centrifugaron a $15.000 \times g$ durante 10 minutos y se recogieron las fases acuosas, que contenían el ARN. El ARN celular total se precipitó con alcohol isopropílico, se lavó una vez con etanol al 70% y se resuspendió en agua tratada con DEPC (Life Technologies, Gaithersburg, EE UU, No. de catálogo 10813-012). Después de tratar con 1,5 unidades de DNasa I (Life Technologies, Gaithersburg, EE UU, No. de catálogo 18068-015), el ARN se volvió a extraer con TRIZOL/cloroformo, se precipitó con etanol y se disolvió en agua tratada con DEPC. Después de ello, se eliminaron los nucleótidos de pequeño peso molecular usando el kit RNeasy Mini (QIAGEN, Hilden, Alemania, No. de catálogo 74104) según el manual de instrucciones del fabricante. La calidad del ARN total se juzgó a partir de la relación de ARN ribosómico de 28S y 18S después de electroforesis en gel de agarosa. El ARN total purificado se almacenó a -80°C en una solución de etanol al 70% hasta su uso.

55 **Ejemplo 4. Síntesis de ADNc y sondas de ARNc marcadas**

Se sintetizó el ADNc usando el sistema inverso SuperScript Choice (Life Technologies, Gaithersburg, EE UU, No. de catálogo 18090-019) según el manual de instrucciones del fabricante. Se hibridaron cinco microgramos del ARN total purificado con cebadores de oligo-dT (Sawady Technology, Tokio, Japón) que contenían secuencias para el promotor de T7 y 200 unidades de transcriptasa inversa SuperScriptII y se incubó a 42°C durante 1 hora. El ADNc resultante se extrajo con fenol/cloroformo y se purificó con Phase Lock Gel™ Light (Eppendorf, Hamburgo, Alemania, No. de catálogo 0032 005.101).

También se sintetizó el ARNc usando el kit MEGAscript T7 (Ambion, Austin, EE UU, No. de catálogo 1334) y ADNc como molde según las instrucciones del fabricante. Se incubaron aproximadamente $5 \mu\text{g}$ del ADNc con $2 \mu\text{l}$ de mezcla de enzima que contenía polimerasa de T7, $7,5 \text{ mM}$ de cada uno de adenosina trifosfato (ATP) y guanosina

trifosfato (GTP), 5,625 de cada uno de citidina trifosfato (CTP) y uridina trifosfato (UTP), y 1,875 de cada uno de Bio-11-CTP y Bio-16-UTP (ENZO diagnostics, Farmingdale, EE UU, No. de catálogo 42818 y 42814, respectivamente) a 37°C durante 6 horas. Se eliminaron los mononucleótidos y los oligonucleótidos cortos por cromatografía en columna en una columna de CHROMA SPIN +STE-100 (CLONTECH, Palo Alto, EE UU, No. de catálogo K1302-2), y el ARNc en los eluatos se sedimentó añadiendo etanol. La calidad del ARNc se juzgó a partir de la longitud del ARNc después de electroforesis en gel de agarosa. El ARNc purificado se almacenó a -80°C en solución de etanol al 70% hasta su uso.

Ejemplo 5. Análisis de expresión génica de CHC en diferente grado de diferenciación

Se examinó la expresión génica de tumores primarios humanos de pacientes de glioma mediante micromatrices de oligonucleótidos de alta densidad (matriz U95A, Affymetrix, Santa Clara, EE UU, No. de catálogo 510137) (Lipshutz, R.L. et al. High density synthetic oligonucleotide arrays, *Nat. Genet.* **21**, 20-24 (1999)). Para la hibridación con los oligonucleótidos en los chips, el ARNc se fragmentó a 95°C durante 35 minutos en un tampón que contenía Tris 40 mM (Sigma, St. Louis, EE UU, No. de catálogo T1503)-ácido acético (Wako, Osaka, Japón, No. de catálogo 017-00256) (pH 8,1), acetato de potasio 100 mM (Wako, Osaka, Japón, No. de catálogo 160-03175) y acetato de magnesio 30 mM (Wako, Osaka, Japón, No. de catálogo 130-00095). La hibridación se realizó en 200 µl de un tampón que contenía ácido 2-(N-morfolino) etanosulfónico (MES) (Sigma, St. Louis, EE UU, No. de catálogo M-3885) 0,1 M (pH 6,7), NaCl 1 M (Nacalai tesque, Kioto, Japón, No. de catálogo 313-20), polioxileno (10) octilfeniléter al 0,01% (Wako, Osaka, Japón, No. de catálogo 168-11805), 20 µg de ADN de esperma de arenque (Promega, Madison, EE UU, No. de catálogo D181B), 100 µg de seroalbúmina bovina acetilada (Sigma, St. Louis, EE UU, No. de catálogo B-8894), 10 µg del ARNc fragmentado y oligonucleótidos biotinilados-control, biotina-5'-CTGAACGGTAGCATCTTGAC-3' (Sawady technology, Tokio, Japón) a 45°C durante 12 horas. Después de lavar los chips con un tampón que contenía MES 0.01 M (pH 6,7), NaCl 0,1 M y polioxileno (10) octilfeniléter al 0,001%, los chips se incubaron con anticuerpo anti-estreptavidina biotinilado (Funakoshi, Tokio, Japón, No. de catálogo BA0500) y se tñeron con estreptavidina R-ficoeritrina (Molecular Probes, Eugene, EE UU, No. de catálogo S-866) para aumentar las señales de hibridación como se describe en el manual de instrucciones (Affymetrix, Santa Clara, EE UU) y se calcularon los niveles de expresión de cada ADNc y la fiabilidad (llamada presente/ausente) con los softwares Affymetrix GeneChip ver. 3.3 y Affymetrix Microarray Suite ver. 4.0. A partir de estos experimentos, se determinó la expresión de aproximadamente 11.000 genes en los tumores primarios humanos de pacientes de glioma.

Ejemplo 6. Análisis estadístico de los datos de micromatrices de oligonucleótidos

Se seleccionaron genes con diferencias medias mayores de 40 (unidades arbitrarias en Affymetrix) en todas las 50 muestras de CHC y las 11 muestras hepáticas no tumorales (no cancerosas y precancerosas). Este procedimiento dio 3.559 genes de aproximadamente 11.000. A continuación, se determinó el cociente de Fisher (Iizuka, N., Oka, M., Yamada-Okabe, H., Mori, N., Tamesa, T., Okada, T., Takemoto, T., Tangoku, A., Hamada, K., Nakayama, H., Miyamoto, T., Uchimura, S., y Hamamoto, Y. Comparison of gene expression profiles between hepatitis B virus- and hepatitis C virus-infected hepatocellular carcinoma by oligonucleotide microarray data based on a supervised learning method, *Cancer Res.* **62**, 3939-3944 (2002) y Luo, J., Duggan, D.J., Chen, Y., Sauvageot, J., Ewing, C.M., Bittner, M.L., Trent, J.M., y Isaacs, W.B. Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling, *Cancer Res.* **61**, 4683-4688 (2001)) para evaluar estos genes como diferenciadores de L0 de L1, L1 de G1, G1 de G2, y G2 de G3. Los 3.559 genes anteriores se clasificaron en orden de magnitud decreciente del cociente de Fisher. También se realizó una prueba de permutación aleatoria para determinar el número de genes para definir el grado de diferenciación de CHC. La prueba de permutación aleatoria se llevó a cabo como se ha descrito previamente Fisher (Iizuka, N., Oka, M., Yamada-Okabe, H., Mori, N., Tamesa, T., Okada, T., Takemoto, T., Tangoku, A., Hamada, K., Nakayama, H., Miyamoto, T., Uchimura, S., y Hamamoto, Y. Comparison of gene expression profiles between hepatitis B virus- and hepatitis C virus-infected hepatocellular carcinoma by oligonucleotide microarray data based on a supervised learning method, *Cancer Res.* **62**, 3939-3944 (2002) y Luo, J., Duggan, D.J., Chen, Y., Sauvageot, J., Ewing, C.M., Bittner, M.L., Trent, J.M., y Isaacs, W.B. Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling, *Cancer Res.* **61**, 4683-4688 (2001)). En la prueba, las etiquetas de muestras se permutaron aleatoriamente entre dos grados que se van a considerar, y el cociente de Fisher para cada gen se computó otra vez. Esta permutación aleatoria de las etiquetas de las muestras se repitió 1.000 veces. Los cocientes de Fisher generados de los datos reales se asignaron después Ps basados en la distribución de los cocientes de Fisher de datos aleatorizados. De la distribución de los cocientes de Fisher basados en los datos aleatorizados, se seleccionaron todos los genes que pudieran pasar la prueba de permutación aleatoria ($P < 0,005$). Este procedimiento se realizó en todos los experimentos para la comparación de dos grados. Como resultado, 152 genes con los cocientes de Fisher mayores de 4,90 fueron diferenciadores estadísticamente significativos entre L0 y L1. Asimismo, se identificaron 191 genes con los cocientes de Fisher mayores de 4,08 para diferenciar L1 de G1, 54 genes con los cocientes de Fisher mayores de 1,52 para diferenciar G1 de G2, y 40 genes con los cocientes de Fisher mayores de 1,34 para diferenciar G2 de G3.

Ejemplo 7. Selección de genes cuya expresión se correlaciona con el grado de diferenciación de CHC

5 Con los datos de matrices de oligonucleótidos, se analizaron los cambios en la expresión génica durante la oncogénesis, es decir, de hígado no canceroso (L0) a hígado precanceroso infectado con HCV (L1) y de L1 a CHC bien diferenciado (G1), y durante la desdiferenciación de CHC (G1 a G2 y G2 a G3). El método de aprendizaje supervisado seguido por una prueba de permutación aleatoria identificó 152 genes cuyo nivel de expresión estaba significativamente cambiado durante la transición de L0 a L1. Entre los 152 genes, 67 estaban aumentados y 85 estaban disminuidos durante esta transición. De la misma manera, se identificaron 191 genes cuyo nivel de expresión estaba significativamente cambiado durante la transición de L1 a CHC G1, Entre los 191 genes, 95 estaban aumentados y 96 estaban disminuidos. Cincuenta y cuatro genes parecían estar diferencialmente expresados entre CHC G1 y G2, y entre ellos la expresión de 36 genes estaba aumentada y la de 18 genes estaba disminuida durante la transición de G1 a G2. Cuarenta genes resultaron estar diferencialmente expresados entre CHC G2 y G3, y entre ellos la expresión de 10 genes estaba aumentada y la de 30 genes estaba disminuida durante la transición de G2 a G3.

Para examinar el rendimiento de los genes seleccionados en cada grado en la oncogénesis y el desarrollo de CHC, los inventores aplicaron datos de estos genes a todas las muestras. Como resultado, casi todos de estos genes seleccionados en cada estadio de transición se colocaron en la transición L0-L1, la transición L1-G1, la transición G1-G2 y la transición G2-G3. Por ejemplo, los 191 genes que discriminan L1 de CHC G1 podrían distinguir claramente hígados no tumorales (L0 y L1) de CHC (G1, G2 y G3) (figura 1). Estos resultados indican que el nivel alterado de los genes seleccionados desempeña papeles esenciales en la determinación de cada grado de la patogénesis del CHC.

Ejemplo 8. Genes cuya expresión cambió durante la transición de hígado no canceroso (L0) a hígado precanceroso (L1)

La expresión de la mayoría de los genes relacionados con la respuesta inmune, genes relacionados con el metabolismo, genes relacionados con el transporte, genes relacionados con la proteólisis y genes relacionados con la oncogénesis aumentó y la de los genes relacionados con la transcripción disminuyó durante la transición de L0 a L1 (tabla 3).

Los genes relacionados con la respuesta inmune incluyen la familia de MHC de clase I (HLA-A, -C, -E y -F), la familia de MHC de clase II (HLA-DPB1 y HLA-DRA), CD74, NK4, LILRB1, FCGR3B e IFI30. El aumento de un gen inducible de interferón (IFN) tal como IFI30 puede representar la defensa del huésped contra la infección vírica; sin embargo, se debe indicar que varios genes relacionados con IFN disminuyeron durante la desdiferenciación de G1 a G2 como se menciona en la siguiente sección (véase el ejemplo 10).

Los genes relacionados con el metabolismo incluyen KARS, ALDOA, ASAH, MPI y GAPD. Los niveles aumentados de KARS y ALDOA aumentan la biosíntesis de proteínas y la glicólisis, respectivamente. El aumento de ASAH, MPI y GAPD aumenta la biosíntesis de ácidos grasos, manosa y gliceraldehído, respectivamente.

Los genes relacionados con transporte incluyen VDAC3, SSR4, BZRP y ATOX1. SSR4 es responsable para el transporte eficaz de polipéptidos recién sintetizados. ATOX1 es un transportador de cobre y un aumento en su expresión produce la activación de varias rutas metabólicas, porque muchas enzimas requieren el ion cobre como cofactor de actividad enzimática.

Los genes relacionados con proteólisis incluyen CST3 y CTSD. CST3 está implicado en la formación vascular. El nivel aumentado en suero de la proteína CTSD se observó en pacientes cirróticos que pueden desarrollar nódulos hepáticos precancerosos (Leto, G., Tumminello, F.M., Pizzolanti, G., Montalto, G., Soresi, M., Ruggeri, I., y Gebbia, N. Cathepsin D serum mass concentrations in patients with hepatocellular carcinoma and/or liver cirrhosis, *Eur. J. Clin. Chem. Clin. Biochem.* **34**, 555-560 (1996)).

Los genes relacionados con oncogénesis incluyen MBD2, RPS19, RPS3, RPS15 y RPS12. La metilación de ADN es un cambio epigenético común en muchas neoplasias malignas, por tanto, los patrones de metilación del ADN se determinan por los procesos enzimáticos de metilación y desmetilación. El aumento de MBD2, que inhibe la transcripción de ADN metilado, desempeña un papel importante en la disminución de genes supresores de tumores que tienen ADN metilado en sus regiones promotoras.

La disminución de un gen relacionado con transcripción, RB1CC1, se observó durante la transición de L0 a L1. La proteína RB1CC1 es un regulador principal del gen supresor de tumores RB1, por tanto niveles disminuidos de RB1CC1 pueden fomentar la oncogénesis a través de la actividad disminuida de la proteína RB1.

Por tanto, el hígado precanceroso infectado con HCV se caracteriza por la expresión alterada de estos genes, lo que sugiere que el inicio de la hepatocarcinogénesis se produce durante la infección con HCV. Entre los genes cuya

expresión cambia durante la transición de L0 a L1, los implicados en proteólisis y oncogénesis pueden servir como dianas moleculares para la quimioprevención de CHC asociado a HCV.

Ejemplo 9. Genes cuya expresión cambió durante la transición de hígado precanceroso (L1) a CHC bien diferenciado (G1)

Los genes cuya expresión estaba alterada durante la transición de L1 a G1 incluyen la mayoría de los genes relacionados con oncogénesis, genes relacionados con transducción de señales, genes relacionados con transcripción, genes relacionados con transporte, genes relacionados con detoxificación y genes relacionados con la respuesta inmune (Tabla 4).

Los genes relacionados con la oncogénesis tales como BNIP3L, FOS, MAF e IGFBP3 que pueden inducir apoptosis de algunas células cancerosas e IGFBP4 que actúa como un inhibidor de la proliferación celular inducida por IGF estaban disminuidos durante la transición, lo que indica que la disminución de estos genes también es importante para el fomento de la hepatocarcinogénesis. Un artículo previo también mostró la expresión disminuida de IGFBP3 e IGFBP4 en CHC comparada con hígado no tumoral (Okabe, H., Satoh, S., Kato, T., Kitahara, O., Yanagawa, R., Yamaoka, Y., Tsunoda, T., Furukawa, Y., y Nakamura, Y. Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: identification of genes involved in viral carcinogenesis and tumor progression, *Cancer Res.* **61**, 2129-2137 (2001) y Delpuech, O., Trabut, J.B., Carnot, F., Feuillard, J., Brechot, C., y Kremsdorf, D. Identification, using cDNA macroarray analysis, of distinct gene expression profiles associated with pathological and virological features of hepatocellular carcinoma, *Oncogene* **21**, 2926-2937 (2002)). Los datos de la presente invención proporcionan percepciones adicionales de que la disminución de estos dos genes ya se ha producido en CHC bien diferenciado. MAF funciona como un regulador para la diferenciación celular. BNIP3L induce apoptosis celular a través de la inhibición de la actividad de BCL2. En algunos casos, la expresión de FOS parece estar asociada con muerte celular apoptótica. Por tanto, la disminución de estos cinco genes es probable que desencadene la transformación de los hepatocitos después de la infección crónica con HCV.

Los genes relacionados con la transducción de señales tales como CAMKK2, GMFB, RALBP1, CDITP, ZNF259 y RAC1 y genes relacionados con transcripción tales como DRAP1, ILF2, BMI1 y PMF1 estaban aumentados durante la transición de L1 a G1. Otros genes relacionados con la transducción de señales tales como CALM1, RAB14, TYROBP y MAP2K1 estaban disminuidos durante esta transición. La disminución de TYROBP en CHC G1 puede reflejar la respuesta inmune disminuida. La alteración de la expresión de genes implicados en varias rutas de transducción de señales puede reflejar un verdadero retrato en CHC bien diferenciado que surge de hígado precanceroso infectado con HCV.

Los genes relacionados con transporte tales como TBCE, ATP6V1E, ATOX1 y SEC61G estaban aumentados y esos tales como SLC31A1 y DDX19 estaban disminuidos durante la transición de L1 a G1. ATOX1 que es un transportador intracelular de cobre estaba aumentado durante la transición de L0 a L1 y aumentó más durante la transición de L1 a G1. Puesto que el cobre excesivo es tóxico o incluso mortal para los hepatocitos, la expresión distinta de los genes ATOX1 altera las concentraciones iónicas de cobre intracelular, por tanto fomenta el daño al ADN y la lesión celular. De hecho, un estudio reciente mostró el efecto preventivo de agentes quelantes de cobre sobre el desarrollo tumoral en el modelo de xenoinjerto murino de CHC (Yoshii, J., Yoshiji, H., Kurivama, S., Ikenaka, Y., Noguchi, R., Okuda, H., Tsujinoue, H., Nakatani, T., Kishida, H., Nakae, D., Gomez, D.E., De Lorenzo, M.S., Tejera, A.M., y Fukui, H. The copper-chelating agent, trientine, suppresses tumor development and angiogenesis in the murine hepatocellular carcinoma cells, *Int. J. Cancer.* **94**, 768-773 (2001)).

El daño al ADN y la lesión celular pueden aumentar por la disminución de un gen antioxidante CAT y genes relacionados con la detoxificación tales como MT1H, MT1E, MT1F, MT1B, MT3 y UGT2B7, lo que fomenta la desdiferenciación de CHC.

Usando un anticuerpo anti-receptor 1 de ácido hialurónico, Carreira y col. mostraron que el número de vasos linfáticos era menor en CHC que en tejidos hepáticos no tumorales tal como cirrosis hepática (Mouta Carreira, C., Nasser, S.M., di Tomaso, E., Padera, T.P., Boucher, Y., Tomarev, S.I., y Jain, R.K. LYVE-1 is not restricted to the lymph vessels: expression in normal liver blood sinusoids and down-regulation in human liver cancer and cirrhosis, *Cancer Res.* **61**, 8079-8084 (2001)). En la presente invención, la expresión de genes relacionados con la respuesta inmune tales como ORM1, C1R, C6, IL4R, C8B y C1S estaba disminuida durante la transición de L1 a G1, lo que indica que se producen cambios en el microambiente en CHC durante la transición de L1 a G1. Como se ha descrito previamente, muchos genes que codifican componentes del complemento estaban disminuidos durante esta transición (Okabe, H., Satoh, S., Kato, T., Kitahara, O., Yanagawa, R., Yamaoka, Y., Tsunoda, T., Furukawa, Y., y Nakamura, Y. Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: identification of genes involved in viral carcinogenesis and tumor progression, *Cancer Res.* **61**, 2129-2137 (2001) y Iizuka, N., Oka, M., Yamada-Okabe, H., Mori, N., Tamesa, T., Okada, T., Takemoto, T., Tangoku, A., Hamada, K., Nakayama, H., Miyamoto, T., Uchimura, S., y Hamamoto, Y. Comparison of gene expression profiles between hepatitis B virus- and hepatitis C virus-infected hepatocellular carcinoma by oligonucleotide microarray data based on a supervised learning method, *Cancer Res.* **62**, 3939-3944 (2002)).

Ejemplo 10. Genes cuya expresión cambió durante la transición de CHC bien diferenciado (G1) a CHC moderadamente diferenciado (G2)

5 Los genes cuya expresión estaba alterada durante la transición de G1 a G2 incluyen genes relacionados con IFN, genes relacionados con la estructura y la motilidad celulares, genes relacionados con la transcripción y genes supresores de tumores (tabla 5).

10 Durante la transición de G1 a G2, los cambios genéticos más prominentes parecen ser la disminución de genes relacionados con IFN tales como OAS2, STAT1, PSME1, ISGF3G y PSMB9. También se observaron cambios genéticos similares en células de cáncer de próstata (Shou, J., Soriano, R., Hayward, S.W., Cunha, G.R., Williams, P.M., y Gao, W.Q. Expression profiling of a human cell line model of prostatic cancer reveals a direct involvement of interferon signaling in prostate tumor progression, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2830-2835 (2002)). El IFN actúa no solo como un agente antiviral sino también como un agente anticanceroso; sin embargo, ciertos tipos de CHC no responden a IFN. La disminución de los genes relacionados con IFN puede atenuar la respuesta de las células tumorales a IFN, lo que sugiere que la resistencia del CHC a IFN se explota durante la transición de G1 a G2. Entre los genes relacionados con IFN, STAT1 apareció cuatro veces en nuestra lista de diferenciadores de G1 a G2 (tabla 5). A diferencia de otros genes de la misma familia, STAT1 funciona como un supresor de tumores (Bromberg, J.F. Activation of STAT proteins and growth control, *Bioessays* **23**, 161-169 (2001)). De forma interesante, el tratamiento con IFN aumenta la expresión de STAT1 en hepatocitos así como muchos genes relacionados con IFN (Radaeva, S., Jaruga, B., Hong, F., Kim, W.H., Fan, S., Cai, H., Strom, S., Liu, Y., El-Assal, O., y Gao, B. Interferon-alpha activates multiple STAT signals and down-regulates c-Met in primary human hepatocytes, *Gastroenterology* **122**, 1020-1034 (2002)). El aumento de STAT1 en líneas celulares de CHC se observó durante la diferenciación inducida por butirato de sodio (Hung, W.C. y Chuang, L.Y. Sodium butyrate enhances STAT 1 expression in PLC/PRF/5 hepatoma cells and augments their responsiveness to interferon-alpha, *Br. J. Cancer* **80**, 705-710 (1999)). Los hechos de que STAT1 es una diana transcripcional del efecto apoptótico de IGFBP3 independiente de IGF (Spagnoli, A., Torello, M., Nagalla, S.R., Horton, W.A., Pattee, P., Hwa, V., Chiarelli, F., Roberts, C. T. Jr., y Rosenfeld, R.G. Identification of STAT-1 as a molecular target of IGFBP-3 in the process of chondrogenesis, *J. Biol. Chem.* **277**, 18860-18867 (2002)) y que IGFBP3 disminuye durante la transición de L1 a G1 sugieren fuertemente que la expresión disminuida de STAT1 durante la transición de CHC G1 a G2 facilita la desdiferenciación adicional de CHC.

35 El gen relacionado con transcripción TRIM16 que está implicado en una variedad de procesos biológicos incluyendo crecimiento, diferenciación y patogénesis celular, y el gen supresor de tumores TPD52L2 que fomenta la proliferación celular también estaban aumentados durante la transición de G1 a G2. El aumento de estos genes en CHC G2 puede fomentar el crecimiento y la invasión de células tumorales.

Ejemplo 11. Genes cuya expresión cambió durante la transición de CHC moderadamente diferenciado (G2) a CHC mal diferenciado (G3)

40 Los genes cuya expresión estaba alterada durante la transición de G2 a G3 incluyen genes relacionados con proteólisis y genes relacionados con metabolismo y generación de energía (tabla 6).

45 SPINT1 y LGALS9 resultaron estar aumentados durante la transición de G2 a G3. SPINT1 está implicado en la regulación de la activación proteolítica del factor de crecimiento de hepatocitos (HGF) en tejidos lesionados. Previamente, Nagata y col. mostraron que la transducción de SPINT1 antisentido (HAI-1) inhibía el crecimiento de células de hepatoma humano, lo que sugiere que SPINT1 desempeña un papel importante en la evolución de CHC (Nagata, K., Hirono, S., Ido, A., Kataoka, H., Moriuchi, A., Shimomura, T., Hori, T., Hayashi, K., Koono, M., Kitamura, N., y Tsubouchi, H. Expression of hepatocyte growth factor activator and hepatocyte growth factor activator inhibitor type 1 in human hepatocellular carcinoma, *Biochem. Biophys. Res. Commun.* **289**, 205-211 (2001)). LGALS9 pertenece a una familia de lectinas que está implicada en adhesión celular, regulación del crecimiento celular, inflamación, inmunomodulación, apoptosis y metástasis. Se piensa que varias galectinas están relacionadas con la adhesión de células cancerosas (Ohannesian, D.W., Lotan, D., Thomas, P., Jessup, J.M., Fukuda, M., Gabius, H.J., y Lotan, R. Carcinoembryonic antigen and other glycoconjugates act as ligands for galectin-3 in human colon carcinoma cells, *Cancer Res.* **55**, 2191-2199 (1995)).

60 BNIP3, un gen relacionado con BCL-2, estaba disminuido durante la transición de G2 a G3. BNIP3 comparte el 56% de identidad de secuencia de aminoácidos con BNIP3L. Como se ha mencionado anteriormente, la expresión de BNIP3L estaba disminuida durante la transición de L1 a G1. Puesto que BCL2 funciona como un factor antiapoptótico, la disminución de BNIP3L y BNIP3 fomenta la oncogénesis, facilitando la desdiferenciación de las células tumorales.

65 Muchos genes relacionados con metabolismo y generación de energía también estaban disminuidos durante esta transición. Además, la expresión de PGRMC1 que codifica una proteína rica en el hígado que se une a progesterona y RARRES2 también estaba disminuida durante la transición de G2 a G3. La expresión disminuida de RARRES2 puede ser causante de mala respuesta de CHC G3 a ácidos retinoicos.

Ejemplo 12. Imagen en color de la expresión de los genes seleccionados en cada estadio de transición

La expresión de los 152 genes cuya expresión estaba significativamente alterada durante la transición de L0 a L1 (figura 1a), 191 genes cuya expresión estaba significativamente alterada durante la transición de L1 a G1 (figura 1b), 54 genes cuya expresión estaba significativamente alterada durante la transición de G1 a G2 (figura 1c), y 40 genes cuya expresión estaba significativamente alterada durante la transición de G2 a G3 (figura 1d) se muestran mediante imagen en color. Estos genes claramente distinguían las muestras en los dos grados de diferenciación consecutivos. La figura 1e-h indica la expresión de los 40 genes seleccionados en cada estadio de transición en todas las muestras. La expresión de los 40 genes seleccionados cuya expresión estaba significativamente alterada durante la transición de L0 a L1 (figura 1e), de L1 a G1 (figura 1f), de g1 a G2 (figura 1g) y de G2 a G3 (figura 1h) también se muestra mediante imagen en color. Los 40 genes seleccionados en cada estadio de transición distinguieron muestras antes y después de la transición.

Ejemplo 13. Validación de los 40 genes seleccionados en cada estadio de transición para distinguir el grado de diferenciación de CHC

Para validar la actuación diferenciadora de los 40 genes seleccionados en cada estadio de transición, se creó el clasificador de distancia mínima con los 40 genes seleccionados en cada estadio de transición. En cada transición, se construyó el clasificador de distancia mínima con las muestras en dos grados de diferenciación consecutivos como se indica mediante la barra roja (muestras de aprendizaje) y se aplicó a las muestras en los grados de diferenciación restantes como se indica mediante la barra negra (muestras de prueba). El clasificador resultante clasificó las muestras de prueba con una precisión del 92% (figura 2a), el 98% (figura 2b), el 84% (figura 2c) y el 100% (figura 2d).

Ejemplo 14. Análisis mediante el algoritmo de mapa autoorganizado (SOM) de los genes cuya expresión cambió durante la transición de hígado no canceroso (L0) a hígado precanceroso (L1), de hígado precanceroso (L1) a CHC bien diferenciado (G1), de CHC bien diferenciado (G1) a CHC moderadamente diferenciado (G2), y de CHC moderadamente diferenciado (G2) a CHC mal diferenciado (G3)

Se analizó la expresión de los genes cuya expresión era estadísticamente significativamente diferente entre hígado no canceroso (L0) e hígado precanceroso (L1), hígado precanceroso (L1) y CHC bien diferenciado (G1), CHC bien diferenciado (G1) y CHC moderadamente diferenciado (G2), CHC moderadamente diferenciado (G2) y CHC mal diferenciado (G3) según el método de MATLAB R13 con la caja de herramientas de SOM disponible en el sitio web, <http://www.cis.hut.fi/projects/somtoolbox/> (Kohonen, 2001). Se usaron 40 genes en cada comparación entre hígado no canceroso (L0) e hígado precanceroso (L1), hígado precanceroso (L1) y CHC bien diferenciado (G1), CHC bien diferenciado (G1) y CHC moderadamente diferenciado (G2), CHC moderadamente diferenciado (G2) y CHC mal diferenciado (G3). Los vectores de células vecinas se colocaron cerca uno del otro en el espacio génico 155-dimensional (figura 3a), donde (m, n) indicaba la célula localizada en la fila m y la columna n, NL-XX indicaba muestras de hígado no canceroso sin infección con HCV (L0), IL-XX indicaba muestras de hígado precanceroso infectadas con HCV (L1), G1-XXT indicaba muestras de CHC bien diferenciado (G1), G2-XXT indicaba muestras de CHC moderadamente diferenciado (G2), G3-XXT indicaba muestras de CHC mal diferenciado (G3). El mapa mostró que las muestras claramente formaron una curva sigmoidea en el orden de L0, L1, G1, G2 y G3. Las muestras de G2 sin implicación vascular (letras azules) se situaron cerca de las muestras de G1 y las muestras de G2 con implicación vascular (letras rojas) se situaron cerca de las muestras de G3 (figura 3a). Las muestras de G2 sin invasión venosa se situaron cerca de las muestras de G1 y las muestras de G2 con invasión venosa se situaron cerca de las muestras de G3. Por tanto, el SOM clasificó las muestras de G2 en dos subtipos, es decir, tumor con invasión venosa y ese sin invasión venosa, en la corriente del grado de desdiferenciación. Cuando la distancia entre los grupos vecinos se mostró mediante colores donde el rojo indicaba distancia larga, las células rojas en el área superior claramente demostraron que las muestras de hígado no tumoral (no canceroso y precanceroso) y de CHC estaban relativamente separadas en el espacio génico 155-dimensional (figura 3b).

Aplicabilidad industrial

El carcinoma hepatocelular (CHC) es uno de los cánceres más comunes en el mundo. Sin embargo, no hay terapia que pueda curar la enfermedad. Este se debe presumiblemente a los cambios secuenciales en las características de las células cancerosas durante el desarrollo y la evolución de la enfermedad. Particularmente, la evolución del cáncer se asocia con frecuencia con los cambios en el grado de diferenciación de las células tumorales. El diagnóstico y tratamiento de tales cambios de las células cancerosas hará la terapia contra el cáncer más eficaz. En la presente invención, se identifican genes cuya expresión se correlaciona con oncogénesis y desarrollo de CHC. Se usa un método de aprendizaje supervisado seguido por una prueba de permutación aleatoria para seleccionar genes cuya expresión cambia significativamente durante la transición de hígado no canceroso sin infección con HCV (L0) a hígado precanceroso con infección por HCV (L1), de L1 a CHC bien diferenciado (G1), de G1 a CHC moderadamente diferenciado (G2) y de G2 a CHC mal diferenciado (G3). El clasificador de distancia mínima y el mapa autoorganizado (SOM) con los 40 genes seleccionados cuya expresión está significativamente alterada en cada estadio de transición pueden predecir correctamente el grado de diferenciación de los tejidos tumorales. Por

ES 2 396 967 T3

tanto, estos genes se pueden usar para el diagnóstico del grado de diferenciación de CHC y para el cribado de agentes anticancerosos para el tratamiento de los CHC en cada grado de diferenciación.

5

Tabla 1. Grupos de muestras con perfiles a L0, L1, G1, G2 y G3.

Célula	Muestra
(1,1)	IL-49, IL-58, IL-59, IL-60, IL-62
(1,2)	
(1,3)	NL-64, NL-65, NL-68, NL-69
(1,4)	NL-66, NL-67
(1,5)	
(2,1)	
(2,2)	G2-34T
(2,3)	
(2,4)	
(2,5)	G2-16T, G2-29T, G2-45T G2-2T
(3,1)	G1-85T, G1-87T
(3,2)	
(3,3)	G1-42T G2-22T
(3,4)	
(3,5)	
(4,1)	G1-86T G2-105T
(4,2)	G1-26T
(4,3)	
(4,4)	G2-8T, G2-27T
(4,5)	G2-151T
(5,1)	G1-147T, G1-165T
(5,2)	
(5,3)	G2-60T
(5,4)	G2-18T
(5,5)	G2-31T G2-20T, G2-59T
(6,1)	G3-21T
(6,2)	G3-80T
(6,3)	G2-1T, G2-163T G2-161T
(6,4)	G2-28T, G2-155T
(6,5)	G2-90T
(7,1)	G3-107T
(7,2)	G3-25T
(7,3)	G2-46T, G2-62T, G2-171T G2-162T
(7,4)	
(7,5)	G2-37T G2-6T, G2-58T
(8,1)	G3-35T, G3-81T, G3-174T
(8,2)	G2-49T G2-23T
(8,3)	G2-12T G2-10T G3-19T
(8,4)	G2-89T
(8,5)	G2-43T, G2-182T

Tabla 2. Características clinicopatológicas por grupos de estudio

Factores	Bien (G1)	Moderadamente (G2)	Mal (G3)	valor de P
Sexo				N.S.
Hombre	4	24	6	
Mujer	3	11	2	
Edad (años)	65,3 ± 2,6	65,4 ± 1,2	67,2 ± 3,3	N.S.
Lesión primaria				N.S.
Tumor único	6	15	2	
Tumores múltiples	1	20	6	
Tamaño del tumor (cm)	2,0 ± 0,3	5,0 ± 0,5	6,0 ± 2,5	p=0,0007 (G1 frente a G2) p=0,028 (G1 frente a G3)
Estadio*				p=0,066
I	6	10	2	
II	1	17	3	
III/IV	0	8	3	
Invasión venosa*				p=0,038
(-)	7	21	3	
(+)	0	14	5	
Hígado no tumoral				N.S.
Normal o hepatitis crónica	2	15	2	
Cirrosis hepática	5	20	6	

*, La diferenciación tumoral, el estadio y la invasión venosa se determinaron en la base a la clasificación TNM de la UICC.

Se usaron la prueba exacta de Fisher, la prueba de la t de Student y la prueba U de Mann-Whitney para elucidar las diferencias en los fondos entre cada grado de diferenciación.

N.S., no significativo.

Tabla 3. 40 mejores genes diferenciadores en L0 y L1

Dieciocho genes disminuidos en L1 en comparación con L0

Cociente de Fisher	Número de GB	Descripción	Símbolo	Locus	Función
50,45	M18533	distrofina	DMD	Xp21.2	citoesqueleto
23,02	AF035316	homólogo a la cadena beta de la tubulina		6p24.3	desconocida
20,65	AL049942	proteína con dedo de zinc 337	ZNF337	20p11.1	desconocida
18,34	L27479	gen de la región de la ataxia de Friedreich X123	X123	9q13-q21	desconocida
16,63	Fibronectina, Ajuste alt. 1	fibronectina (ajuste alt. 1)			matriz extracelular
16,13	U19765	proteína con dedo de zinc 9	ZNF9	3q21	transcripción/proteína de unión a ácido nucleico retrovírico
14,91	X55503	metalotioneína IV	MTIV	16q13	detoxificación
13,71	AL046394	proteína de unión a poli (rC) 3	PCBP3	21q22.3	proteína de unión a ARN/control postranscripcional
12,56	AB007886	producto del gen KIAA0426	KIAA0426	6p22.2-p21.3	desconocida
12,41	AL050139	proteína hipotética FLJ13910	FLJ13910	2p11.1	desconocida
12,37	AF012086	similar a proteína 2 de unión a RAN 1	RANBP2L1	2q12.3	transducción de señales/proteína pequeña de unión a GTP
11,66	AI539439	proteína de unión a calcio S100 A2	S100A2	1q21	estímulos extracelulares y respuestas extracelulares
11,24	M19828	apolipoproteína B	APOB	2p24-p23	metabolismo lipídico
10,59	U92315	familia sulfotransferasa, citosólica, 2B, miembro 1	SULT2B1	19q13.3	metabolismo de esteroides

ES 2 396 967 T3

Cociente de Fisher	Número de GB	Descripción	Símbolo	Locus	Función
10,53	D76444	homólogo de la proteína con dedo de zinc 103 (ratón)	ZFP103	2p11.2	desarrollo del sistema nervioso central
10,50	X02761	fibronectina 1	FN1	2q34	matriz extracelular/ adhesión y motilidad celulares
10,20	AF001891	similar a proteína con dedo de zinc 1	ZFPL1	11q13	desconocida
9,74	AI400326	EST		2	Grupo UniGen Hs. 356456

Veintidós genes aumentados en L1 en comparación con L0

Cociente de Fisher	Número de GB	Descripción	Símbolo	Locus	Función
40,49	AI362017	cistatina C	CST3	20p11.21	Inhibidor de cisteín proteasa
21,66	L13977	prolilcarboxipeptidasa (angiotensinasa C)	PRCP	11q14	metabolismo/proteína relacionada con lisosomas
20,59	D32053	lisil-ARNt sintetasa	KARS	16q23-q24	biosíntesis de proteínas
13,70	AF038962	canal aniónico dependiente de voltaje 3	VDAC3	8p11.2	transporte de nucleótidos de adenina
11,90	AL008726	proteína protectora para beta-galactosidasa (catepsina A)	PPGB	20q13.1	proteína lisosomal/ activador enzimático
11,71	J03909	interferón, proteína inducible por gamma 30	IFI30	19p13.1	tiol reductasa lisosomal/ inducible por IFN
11,32	Z69043	receptor de secuencia señal, delta	SSR4	Xq28	translocación de polipéptidos recién sintetizados
11,17	AL080080	proteína transmembrana relacionada con tiorredoxina	TXNDC	14q21.3	reacción redox
11,15	M63138	catepsina D	CTSD	11p15.5	aspartil proteasa lisosomal/proteólisis
11,12	L09159	familia de genes homólogos a ras, miembro A	ARHA	3p21.3	oncogénesis/ reorganización del citoesqueleto de actina
10,99	AF017115	citocromo c oxidasa subunidad IV isoforma 1	COX4I1	16q22-qter	ruta de energía
10,76	M13560	antígeno CD74	CD74	5q32	respuesta inmune
10,22	M36035	receptor de benzodiazapina	BZRP	22q13.31	flujo de colesterol en la mitocondria
10,08	U47101	similar a grupo de fijación de nitrógeno	NIFU	12q24.1	desconocida
9,70	U81554	proteína quinasa dependiente de calcio/calmodulina II gamma	CAMK2G	10q22	transducción de señales
9,59	M21186	citocromo b-245, polipéptido alfa	CYBA	16q24	generación de energía
9,47	D32129	complejo mayor de histocompatibilidad, clase I, A	HLA-A	6p21.3	respuesta inmune
9,44	AL022723	complejo mayor de histocompatibilidad, clase I, F	HLA-F	6p21.3	respuesta inmune
9,41	M83664	complejo mayor de histocompatibilidad, clase II, DP beta 1	HLA-DPB1	6p21.3	respuesta inmune
9,16	U50523	complejo de proteínas relacionadas con actina 2/3, subunidad 2	ARPC2	13q12-q13	motilidad celular y citoesqueleto
9,02	M81757	proteína ribosómica S19	RPS19	19q13.2	oncogénesis/proteína de unión a ARN
8,89	AF102803	catenina (proteína asociada a cadherina), alfa 1	CTNNA1	5q31	adhesión celular

Tabla 4. 40 mejores genes distinguidores en L1 y G1

Veintiocho genes disminuidos en G1 en comparación con L1

Cociente de Fisher	Número de GB	Descripción	Símbolo	Locus	Función
26,84	M93221	receptor de manosa, C tipo 1	MRC1	10p13	fagocitosis y pinocitosis
26,08	AF079221	similar a proteína de interacción con BCL2/ adenovirus E1B de 19kD 3	BNIP3L	8p21	supresor tumoral/ inducción de apoptosis
21,46	V01512	homólogo del oncogén vírico de osteosarcoma murino FBJ v-fos	FOS	14q24.3	oncogénesis/transcripción
21,45	D88587	ficolina 3 (antígeno Hakata)	FCN3	1p35.3	espacio extracelular
20,15	U12022	calmodulina 1	CALM1	14q24-q31	transducción de señales/ proteína de unión a calcio
19,73	AF055376	homólogo del oncogén de fibrosarcoma musculoponeurótico v-maf	MAF	16q22-q23	oncogénesis/transcripción
19,19	R93527	metalotioneína 1H	MT1H	16q13	detoxificación
18,19	R92331	metalotioneína 1E	MT1E	16q13	detoxificación
17,65	U83460	familia de transportadores de soluto 31, miembro 1	SLC31A1	9q31-q32	transportador del ion cobre
17,30	AF052113	RAB14, miembro de la familia de oncogenes RAS	RAB14	9q32-q34.11	miembros de la superfamilia Ras de proteínas que unen GTP
15,26	H68340	proteína relacionada con la ARN helicasa	RNAHP	17q22	alteración de la estructura secundaria del ARN
14,96	M10943	metalotioneína 1F	MT1F	16q13	detoxificación
14,18	M13485	metalotioneína 1B	MT1B	16q13	detoxificación
13,34	U75744	similar a desoxirribonucleasa I 3	DNASE1L3	3p21.1-3p14.3	metabolismo de ADN
12,65	X02544	orosomucoide 1	ORM1	9q31-q32	respuesta inmune/ respuesta de fase aguda
11,95	M93311	metalotioneína 3	MT3	16q13	detoxificación
11,58	Z24725	inducible por mitógeno 2	MIG2	14q22.1	ciclo celular y proliferación celular
11,52	U22961	desconocida			desconocida
11,45	M62403	proteína de unión al factor de crecimiento insulínico 4	IGFBP4	17q12-q21.1	transducción de señales/ proliferación celular
11,01	M35878	proteína de unión al factor de crecimiento insulínico 3	IGFBP3	7p13-p12	transducción de señales/ proliferación celular
10,80	U84011	amilo-1,6-glucosidasa, 4-alfa-glucanotransferasa	AGL	1p21	degradación de glucógeno
10,74	AF055030	proteína con dedo de zinc PHD XAP135, isoforma b	XAP135	6q27	desconocida
10,29	L13977	prolilcarboxipeptidasa (angiotensinasa C)	PRCP	11q14	metabolismo/proteína relacionada con lisosomas
10,02	D13891	inhibidor de unión a ADN 2	ID2	2p25	regulador negativo de diferenciación celular
9,95	M63175	receptor del factor de motilidad autocrino	AMFR	16q21	transducción de señales/ motilidad celular
9,94	AB023157	proteína KIAA0940	KIAA0940	10q23.33	desconocida
9,76	U20982	proteína de unión al factor de crecimiento insulínico 4	IGFBP4	17q12-q21.1	transducción de señales/ proliferación celular
9,09	M14058	componente del complemento 1, subcomponente r	C1R	12p13	respuesta inmune

5

Doce genes aumentados en G1 en comparación con L1

Cociente de Fisher	Número de GB	Descripción	Símbolo	Locus	Función
30,42	AL049650	ribonucleoproteína nuclear	SNRPB	20p13	procesamiento/

Cociente de Fisher	Número de GB	Descripción	Símbolo	Locus	Función
		pequeña polipéptidos B y B1			modificación de ARN/ ajuste de ARN
20,95	U61232	chaperona específica de tubulina e	TBCE	1q42.3	microtúbulos/ cochaperonina
11,95	AI991040	proteína asociada a DR1 1	DRAP1	11q13.3	transcripción
10,96	U64444	similar a degradación por fusión de ubiquitina 1	UFD1L	22q11.21	proteólisis
10,71	D63997	autoantígeno del golgi, subfamilia golgina a, 3	GOLGA3	12q24.33	estabilización de la estructura del Golgi
10,60	X55503	metalotioneína IV	MT4	16q13	detoxificación
10,23	AL080181	superfamilia inmunoglobulina, miembro 4	IGSF4	11q23.2	posee baja similitud al receptor viral
10,01	X76228	ATPasa, transporte de H ⁺ , lisosomal 31 kD, V1 subunidad E	ATP6V1E	22q11.1	transporte de protones
9,77	AB018330	proteína quinasa dependiente de calcio/calmodulina 2, beta	CAMKK2	12q24.2	transducción de señales/proteína de unión a calcio
9,41	D76444	homólogo de la proteína con dedo de zinc 103 (ratón)	ZFP103	2p11.2	desarrollo del sistema nervioso central
9,31	U70660	homólogo de la proteína antioxidante ATX1 1 (levadura)	ATOX1	5q32	homeostasis de cobre y transporte de iones
9,10	U10323	factor de unión potenciador de interleuquina 2, 45 kD	ILF2	1q21.1	transcripción

Tabla 5. 40 mejores genes distinguidores en G1 y G2

5 **Quince genes disminuidos en G2 en comparación con G1**

Cociente de Fisher	Número de GB	Descripción	Símbolo	Locus	Función
2,89	M87434	2'-5'-adenilato sintetasa 2	OAS2	12q24.2	proteína de respuesta antiviral/inducible por IFN
2,63	M12963	alcohol deshidrogenasa de clase I subunidad alfa	ADH1A	4q21-q23	detoxificación
2,51	AI625844	proteína hipotética FLJ20378			desconocida
2,43	M97936	transductor de señales y activador de transcripción 1	STAT1	2q32.2	transcripción/ruta de señalización de IFN
2,12	Z99129	factor de transcripción de choque término 2	HSF2	6q22.33	transcripción
2,08	L07633	activador del proteasoma subunidad 1	PSME1	14q11.2	proteólisis y peptidólisis/ inducible por IFN
2,06	D50312	canal de potasio rectificado hacia dentro subfamilia J, miembro 8	KCNJ8	12p11.23	transporte de potasio
2,02	U07364	activador del proteasoma subunidad 1	PSME1	14q11.2	proteólisis y peptidólisis/ inducible por IFN
2	AA883502	enzima conjugadora de ubiquitina E2L6	UBE2L6	11q12	proteólisis y peptidólisis
1,85	M97935	transductor de señales y activador de transcripción 1	STAT1	2q32.2	transcripción/ruta de señalización de IFN
1,83	AF061258	proteína LIM	LIM	4q22	transducción de señales
1,74	AB007447	producto del gen FLN 29	FLN29	12q	transducción de señales
1,72	M97935	transductor de señales y activador de transcripción 1	STAT1	2q32.2	transcripción/ruta de señalización de IFN
1,7	W28281	similar a proteína asociada al	GABARAPL	12p13.1	proteína asociada a

Cociente de Fisher	Número de GB	Descripción	Símbolo	Locus	Función
		receptor de GABA (A) 1	1		microtúbulos
1,66	M97935	transductor de señales y activador de transcripción 1	STAT1	2q32.2	transcripción/ruta de señalización de IFN

Veinticinco genes aumentados en G2 en comparación con G1

Cociente de Fisher	Número de GB	Descripción	Símbolo	Locus	Función
4,41	Y00281	riboforina I	RPNI	3q21.3-q25.2	modificación de proteínas/ unión a ARN
3,25	D28118	proteína con dedo de zinc 161	ZNF161	17q23.3	transcripción
2,83	AF104913	factor de iniciación de síntesis de proteínas eucariota 4 gamma	EIF4G1	3q27-qter	traducción
2,27	AA675900	proteína de unión a formina 3	FNBP3	2q23.3	proteólisis y peptidólisis
2,27	L27706	chaperonina que contiene TCP1, subunidad 6A (zeta 1)	CCT6A	7p14.1	chaperona/ plegamiento de proteínas
2,15	D32050	alanil-ARNt sintetasa	AARS	16q22	procesamiento de ARNt/ síntesis de proteínas
2,1	M63573	peptidilproil isomerasa B	PIIB	15q21-q22	chaperona/respuesta inmune
2,09	AF014398	inositol(mio)-1(o 4)-monofosfatasa 2	IMPA2	18p11.2	transducción de señales
2,08	X70944	factor de ajuste rico en prolina/glutamina	SFPQ	1p34.2	ayuste de ARN/ procesamiento de ARN
2,03	U70671	proteína relacionada con ataxina 2	A2LP	7	desconocida
1,89	AA447263	proteína de apilamiento de reensamblaje del golgi 2, 55 kD	GORASP2	2p24.3-q21.3	apilamiento del golgi
1,87	AB014569	producto del gen KIAA0669	KIAA0669	3	desconocida
1,85	M23115	ATPasa, transportador de Ca ⁺⁺ , músculo cardiaco, movimiento lento 2	ATP2A2	12q23-q24.1	transporte de moléculas pequeñas
1,83	D38521	activador del proteasoma 200 kD	PA200	2p16.2	proteólisis y peptidólisis
1,82	X00351	actina, beta	ACTB	7p15-p12	citoesqueleto
1,75	L11672	proteína con dedo de zinc 91	ZNF91	19p13.1-p12	transcripción
1,75	X82834	autoantígeno del golgi, subfamilia golgina a, 4	GOLGA4	3p22-p21.3	transporte de vesículas
1,74	AB007963	producto del gen KIAA0494	KIAA0494	1pter-p22.1	desconocida
1,74	U76247	homólogo 1 de seven in absentia (Drosophila)	SIAH1	16q12	proteólisis y peptidólisis/ apoptosis
1,73	X68560	factor de transcripción Sp3	SP3	2q31	transcripción
1,73	AB015344	ubiquilina 2	UBQLN2	Xp11.23-p11.1	ubiquitinación
1,73	AB018327	neuroprotector dependiente de actividad	ADNP	20q13.13-q13.2	desconocida
1,7	AF004430	similar a proteína tumoral D52 2	TPD52L2	20q13.2-q13.3	proliferación celular
1,67	D14697	farnesil difosfato sintasa	FDPS	1q21.2	biosíntesis de colesterol
1,67	AB028449	Dicer1, homólogo de Dcr-1 (Drosophila)	DICER1	14q32.2	ARN helicasa

Tabla 6. 40 mejores genes distinguidores en G2 y G3

5

Treinta genes disminuidos en G3 en comparación con G2

Cociente de Fisher	Número de GB	Descripción	Símbolo	Locus	Función
2,36	AA976838	apolipoproteína C-I	APOC1	19q13.2	metabolismo lipídico

ES 2 396 967 T3

Cociente de Fisher	Número de GB	Descripción	Símbolo	Locus	Función
2,20	Z11793	selenoproteína P, plasma, 1	SEPP1	5q31	actividad antioxidante
1,86	AB002311	factor de intercambio de nucleótidos de guanina que contiene el dominio PDZ 1	PDZ-GEF1	4q32.1	transducción se señales asociada a Ras/Rap 1A
1,80	Y18004	similar a sex comb on midleg 2 (Drosophila)	SCML2	Xp22	transcripción/embriogénesis y morfogénesis
1,76	AL031230	familia aldehído deshidrogenasa 5, miembro A1	ALDH5A1	6p22	transportador de electrones/ catabolismo de aminobutirato
1,71	AF002697	proteína que interacciona con BCL2/adenovirus E1B 19 kD 3	BNIP3	14q11.2-q12	apoptosis
1,65	AB014596	proteína con dominios F-box y WD-40 1B	FBXW1B	5q35.1	ubiquitinación
1,64	U49897	fenilalanina hidroxilasa	PAH	12q22-q24.2	biosíntesis de aminoácidos
1,62	AF070570	secuencia de ARN 24473 clon de homo sapiens		4	desconocida
1,59	M80482	sistema 4 de corte de aminoácidos básicos pareados	PACE4	15q26	señalización célula-célula/proteólisis
1,59	AI263099	FLJ31305 fis o clon LIVER1000104		16	similar al ARNm de proteína específica de riñón de Rattus norvegicus
1,57	U22961	desconocida			desconocida
1,57	Z24725	inducible por mitógeno 2	MIG2	14q22.1	control del ciclo celular
1,53	U77594	respondedor del receptor de ácido retinoico (inducido por tazaroteno) 2	RARRES2	7q35	receptor de ácido retinoico/inducible por ácido retinoico
1,49	L34081	ácido biliar coenzima A: aminoácido N-aciltransferasa	BAAT	9q22.3	enzima hepática para metabolismo de glicina y ácido biliar
1,49	M88458	receptor de la proteína de retención del retículo endoplásmico KDEL 2	KDEL2	7p22.2	tráfico intracelular de proteínas
1,48	U68723	supresor del punto de regulación 1	CHES1	14q24.3-q31	transcripción/ciclo celular
1,48	X92098	proteína de membrana de vesículas cubiertas	RNP24	12q24.31	tráfico intracelular de proteínas
1,44	D10040	ácido graso-Coenzima A ligasa, cadena larga 2	FACL2	4q34-q35	metabolismo de ácidos grasos
1,43	AB023194	proteína KIAA0977	KIAA0977	2q24.3	desconocida
1,42	AF001903	L-3-hidroxiacil-Coenzima A deshidrogenasa, cadena corta	HADHSC	4q22-q26	enzima mitocondrial/ generación de energía
1,40	X96752	L-3-hidroxiacil-Coenzima A deshidrogenasa, cadena corta	HADHSC	4q22-q26	enzima mitocondrial/ generación de energía
1,40	AB006202	complejo succinato deshidrogenasa, subunidad D	SDHD	11q23	proteína mitocondrial/ transporte de electrones
1,37	M75106	carboxipeptidasa B2	CPB2	13q14.11	proteólisis y peptidólisis
1,37	Y12711	componente de membrana 1 del receptor de progesterona	PGRMC1	Xq22-q24	proteína rica en hígado que se une a progesterona
1,36	D14662	proteína antioxidante 2	AOP2	1q23.3	actividad antioxidante/ no selenio glutatión peroxidasa
1,36	S87759	proteína fosfatasa 1A	PPM1A	14q23.1	respuestas al estrés celular
1,36	Z48199	sindecano 1	SDC1	2p24.1	adhesión celular y metástasis
1,35	AF088219	ligando de quimioquina (motivo C-C) 14	CCL14	17q11.2	proliferación celular

Cociente de Fisher	Número de GB	Descripción	Símbolo	Locus	Función
1,35	AA453183	EST			desconocida

Diez genes aumentados en G3 en comparación con G2

Cociente de Fisher	Número de GB	Descripción	Símbolo	Locus	Función
2,80	D31767	proteína asociada a DAZ 2	DAZAP2	2q33-q34	proteína de unión a ARN
2,57	AB000095	inhibidor de serín proteasa, tipo Kunitz 1	SPINT1	15q13.3	inhibidor específico para el activador de HGF
2,40	AB006782	galectina 9	LGALS9	17q11.1	adhesión celular y metástasis
2,18	M21186	citocromo b-245, polipéptido alfa	CYBA	16q24	generación de energía
1,96	AB002312	bromodominio adyacente al dominio de dedo de zinc 2A	BAZ2A	12q24.3-qter	proteína de unión a ADN
1,84	U44772	palmitoil-proteína tioesterasa 1	PPT1	1p32	maduración neuronal
1,77	AI541308	proteína de unión a calcio S100 A13	S100A13	1q21	estímulos extracelulares y respuestas celulares
1,53	Z49107	galectina 9	LGALS9	17q11.1	adhesión celular y metástasis
1,36	U77735	oncogén pim-2	PIM2	Xp11.23	proliferación celular
1,34	M38449	factor de crecimiento transformante, beta 1	TGFB1	19q13.2	crecimiento y adhesión celular

REIVINDICACIONES

- 5 1. Un método in vitro de definir el grado de diferenciación de un tumor con genes y/o proteínas seleccionados mediante análisis estadísticos basados en el nivel o patrón de expresión de los genes y/o las proteínas de tejidos tumorales humanos obtenibles de pacientes de cáncer, en donde los genes y/o proteínas se seleccionan en orden descendiente del cociente de Fisher y en donde el cociente de Fisher se determina sin el uso de una probabilidad previa, en donde el cociente de Fisher para un gen j está dado por

$$F(j) = \frac{(\hat{\mu}_j(A) - \hat{\mu}_j(B))^2}{\hat{\sigma}_j^2(A) + \hat{\sigma}_j^2(B)}$$

10 donde $\hat{\mu}_j(i)$ es la media de la muestra del nivel de expresión del gen j para las muestras en grado i , y $\hat{\sigma}_j^2(i)$ es la varianza de la muestra del nivel de expresión del gen j para las muestra en grado i .

- 15 2. El método in vitro según la reivindicación 1, en donde los tejidos humanos son tejidos hepáticos humanos.
3. El método in vitro según la reivindicación 2, en donde el grado de diferenciación del tumor se selecciona del grupo que consiste en hígado no canceroso, hígado precanceroso, carcinoma hepatocelular (CHC) bien diferenciado, CHC moderadamente diferenciado y CHC mal diferenciado.
- 20 4. El método in vitro según la reivindicación 3, en donde los genes y/o las proteínas se expresan diferencialmente entre hígado no canceroso e hígado precanceroso, hígado precanceroso y carcinoma hepatocelular (CHC) bien diferenciado, CHC bien diferenciado y CHC moderadamente diferenciado, o CHC moderadamente diferenciado y CHC mal diferenciado.
- 25 5. El método in vitro según cualquiera de las reivindicaciones 1 a 4, en donde el nivel o patrón de expresión de los genes y/o proteínas se examina por medio de micromatriz de ADN, transcripción inversa reacción en cadena de la polimerasa o matriz de proteínas.
6. El método in vitro según la reivindicación 5, en donde el número de los genes y/o proteínas está entre 40 y 100.
- 30 7. El método in vitro según la reivindicación 5, en donde el número de los genes y/o proteínas está entre 35 y 45.
8. El método in vitro según la reivindicación 7, en donde el número de los genes y/o proteínas es 40.
- 35 9. Un método in vitro de definir el grado de diferenciación de un tumor, el método comprende los pasos de:

- 40 (a) seleccionar genes y/o proteínas que tienen los mayores cocientes de Fisher en la comparación entre hígado no canceroso e hígado precanceroso, hígado precanceroso y carcinoma hepatocelular (CHC) bien diferenciado, CHC bien diferenciado y CHC moderadamente diferenciado, o CHC moderadamente diferenciado y CHC mal diferenciado; y
- (b) definir el grado de diferenciación de un tumor usando los genes y/o proteínas, en donde el cociente de Fisher se determina sin el uso de una probabilidad previa, en donde el cociente de Fisher para un gen j está dado por

$$F(j) = \frac{(\hat{\mu}_j(A) - \hat{\mu}_j(B))^2}{\hat{\sigma}_j^2(A) + \hat{\sigma}_j^2(B)}$$

45 donde $\hat{\mu}_j(i)$ es la media de la muestra del nivel de expresión del gen j para las muestras en grado i , y $\hat{\sigma}_j^2(i)$ es la varianza de la muestra del nivel de expresión del gen j para las muestra en grado i .

10. Un método in vitro de definir el grado de diferenciación de un tumor, el método comprende los pasos de:
- 50 (a) determinar el número de genes y/o proteínas para definir el grado de diferenciación de un tumor;
- (b) seleccionar genes y/o proteínas que tienen los mayores cocientes de Fisher en la comparación entre hígado no canceroso e hígado precanceroso, hígado precanceroso y carcinoma hepatocelular (CHC) bien diferenciado, CHC bien diferenciado y CHC moderadamente diferenciado, o CHC moderadamente diferenciado y CHC mal diferenciado;
- 55 (c) aplicar los datos de genes y/o proteínas seleccionados en el paso (b) a todas las muestras; y
- (d) definir el grado de diferenciación de un tumor,

en donde el cociente de Fisher se determina sin el uso de una probabilidad previa, en donde el cociente de Fisher para un gen j está dado por

$$F(j) = \frac{(\hat{\mu}_j(A) - \hat{\mu}_j(B))^2}{\hat{\sigma}_j^2(A) + \hat{\sigma}_j^2(B)}$$

donde $\hat{\mu}_j(i)$ es la media de la muestra del nivel de expresión del gen j para las muestras en grado i , y $\hat{\sigma}_j^2(i)$ es la varianza de la muestra del nivel de expresión del gen j para las muestra en grado i .

11. Un método in vitro de definir el grado de diferenciación de un tumor, el método comprende los pasos de:

- (a) determinar el número de genes y/o proteínas para definir el grado de diferenciación de un tumor;
- (b) seleccionar genes y/o proteínas que tienen los mayores cocientes de Fisher en la comparación entre hígado no canceroso e hígado precanceroso, hígado precanceroso y carcinoma hepatocelular (CHC) bien diferenciado, CHC bien diferenciado y CHC moderadamente diferenciado, o CHC moderadamente diferenciado y CHC mal diferenciado;
- (c) aplicar los datos de genes y/o proteínas seleccionados en el paso (b) a todas las muestras;
- (d) diseñar un clasificador de distancia mínima con los datos de los genes y/o proteínas seleccionados en el paso (b);
- (e) aplicar el clasificador de distancia mínima diseñado en el paso (d) a todas las muestras;
- (f) generar un mapa autoorganizado con los datos de todos los genes y/o proteínas seleccionados en el paso (b);
- (g) aplicar el mapa autoorganizado generado en el paso (f) a todas las muestras; y
- (h) definir el grado de diferenciación de un tumor,

en donde el cociente de Fisher se determina sin el uso de una probabilidad previa, en donde el cociente de Fisher para un gen j está dado por

$$F(j) = \frac{(\hat{\mu}_j(A) - \hat{\mu}_j(B))^2}{\hat{\sigma}_j^2(A) + \hat{\sigma}_j^2(B)}$$

donde $\hat{\mu}_j(i)$ es la media de la muestra del nivel de expresión del gen j para las muestras en grado i , y $\hat{\sigma}_j^2(i)$ es la varianza de la muestra del nivel de expresión del gen j para las muestra en grado i .

Fig. 1a-d

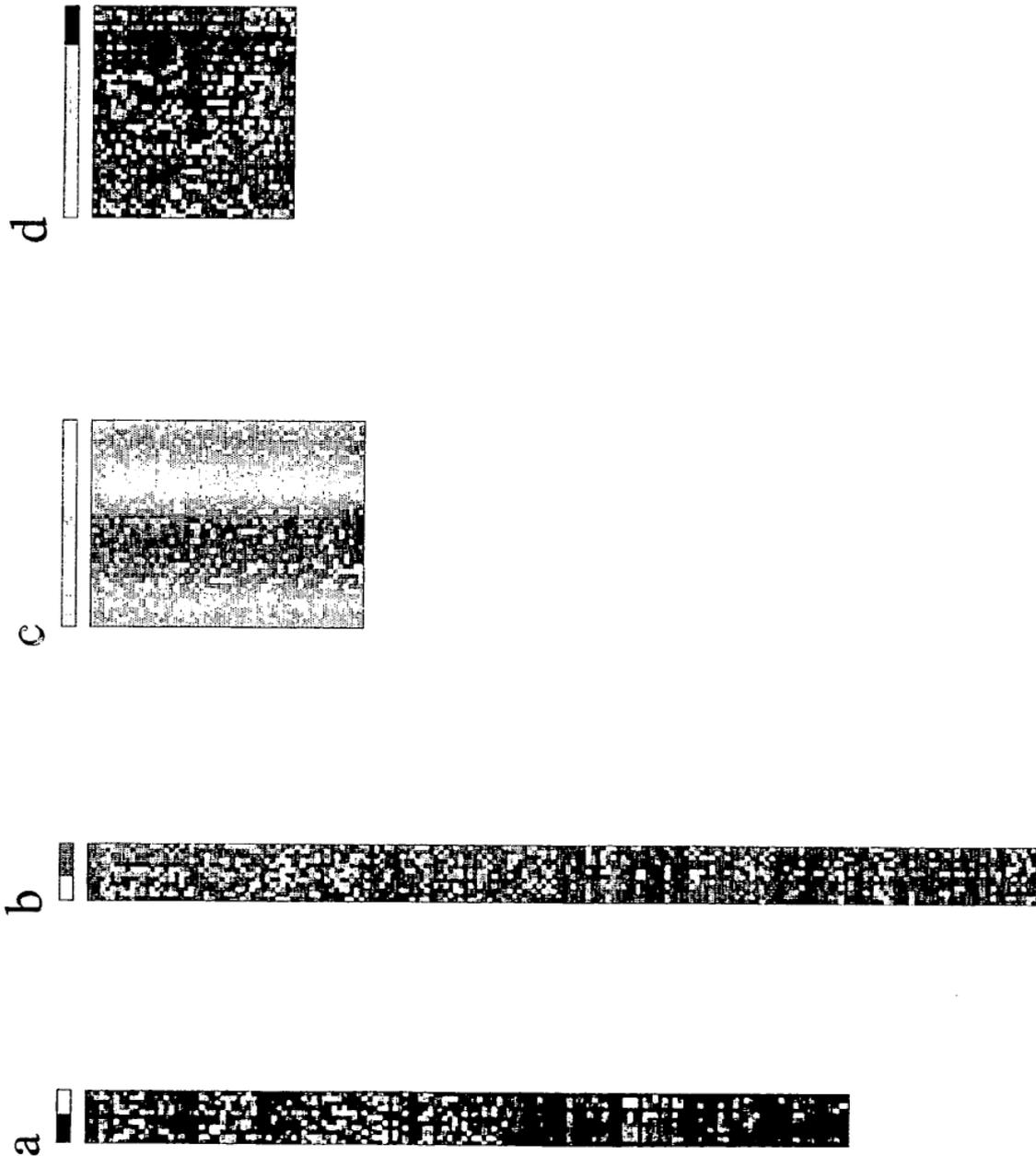
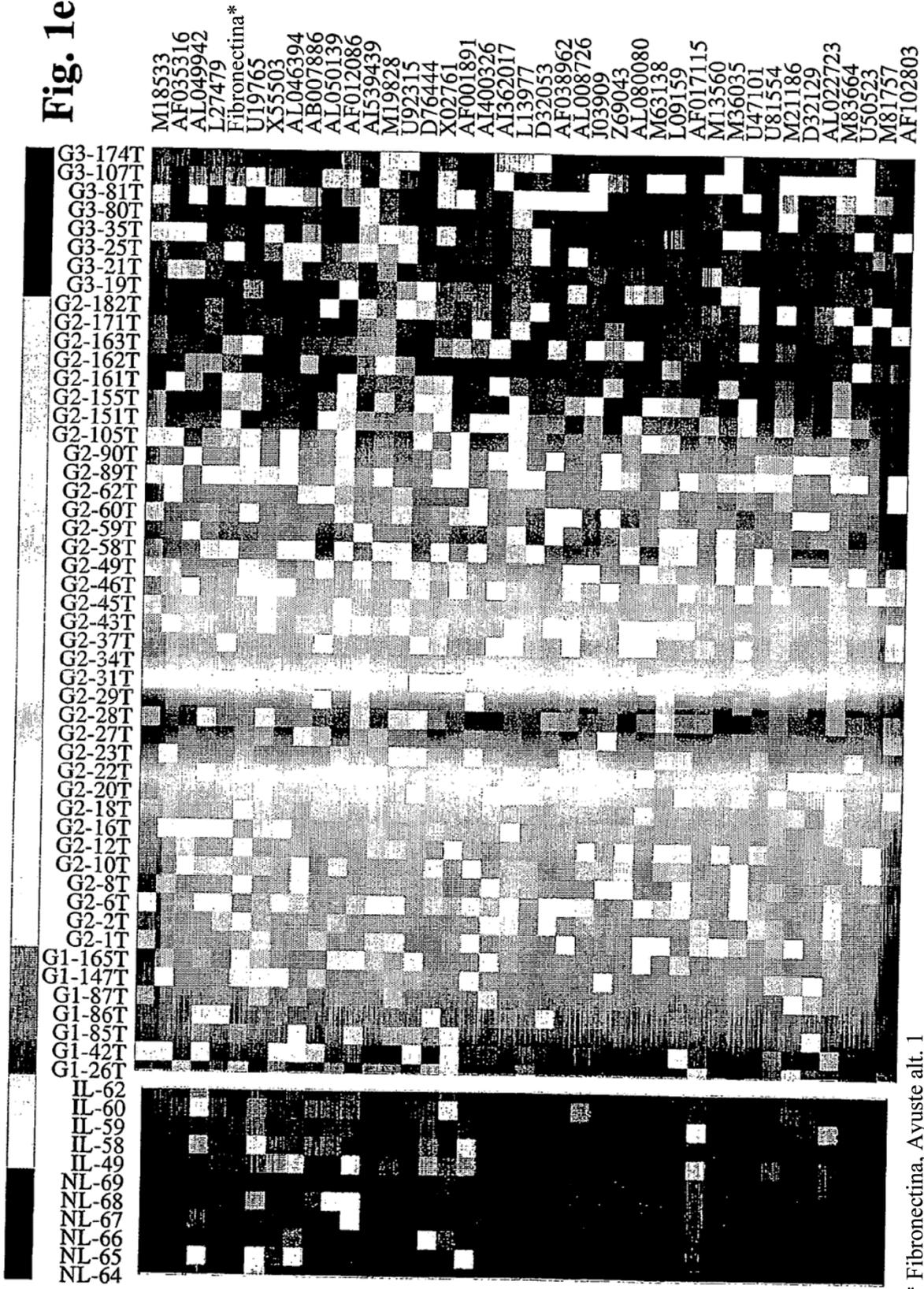


Fig. 1c



* Fibronectina, Ayuste alt. 1

Fig. 1f

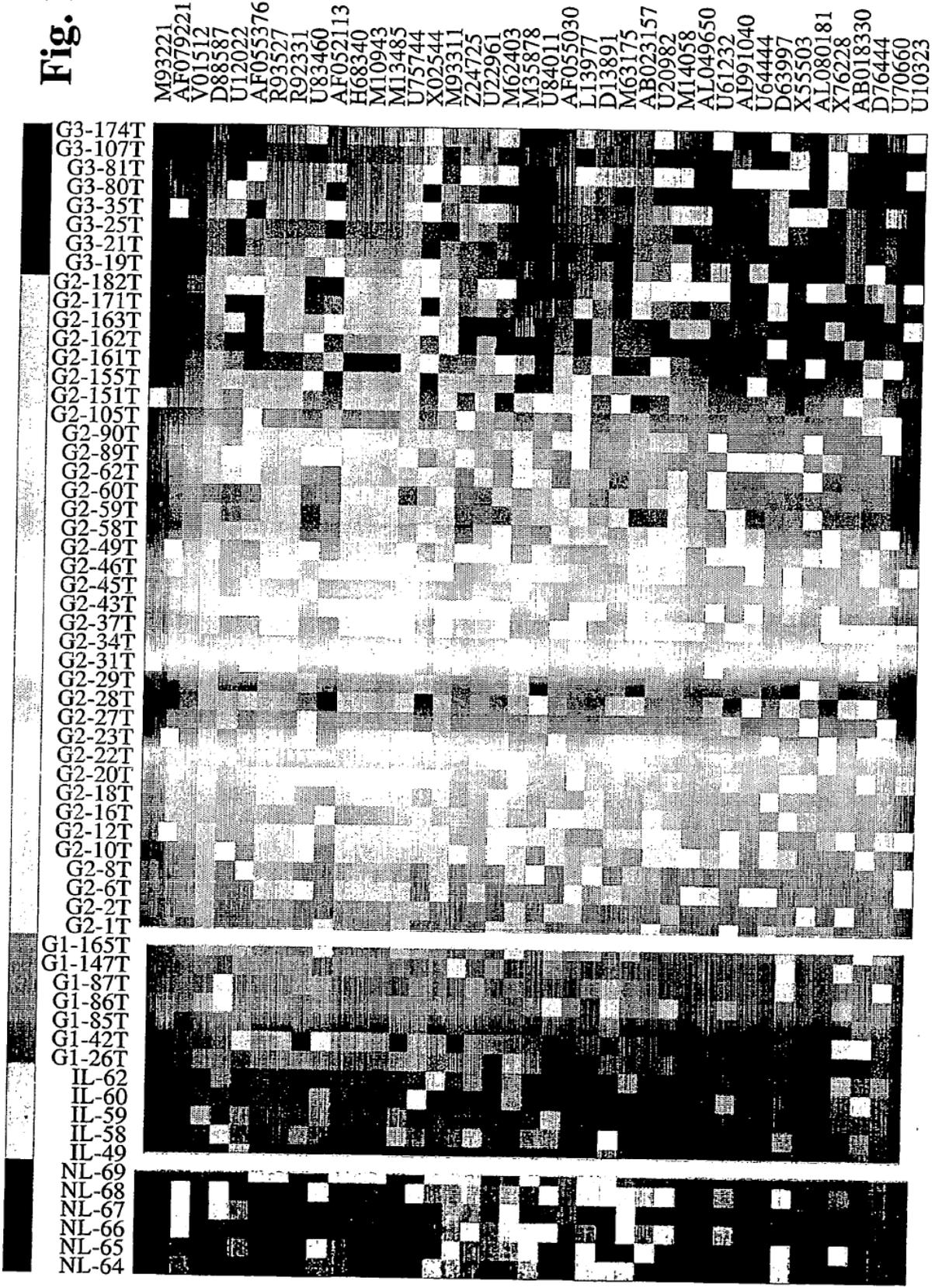


Fig. 1g

M87434
 M12963
 AI625844
 M97936
 Z99129
 L07633
 D50312
 U07364
 AA883502
 M97935
 AF061258
 AB007447
 M97935
 W28281
 M97935
 Y00281
 D28118
 AF104913
 AA675900
 L27706
 D32050
 M63573
 AF014398
 X70944
 U70671
 AA447263
 AB014569
 M23115
 D38521
 X00351
 L11672
 X82834
 AB007963
 U76247
 X68560
 AB015344
 AB018327
 AF004430
 D14697
 AB028449

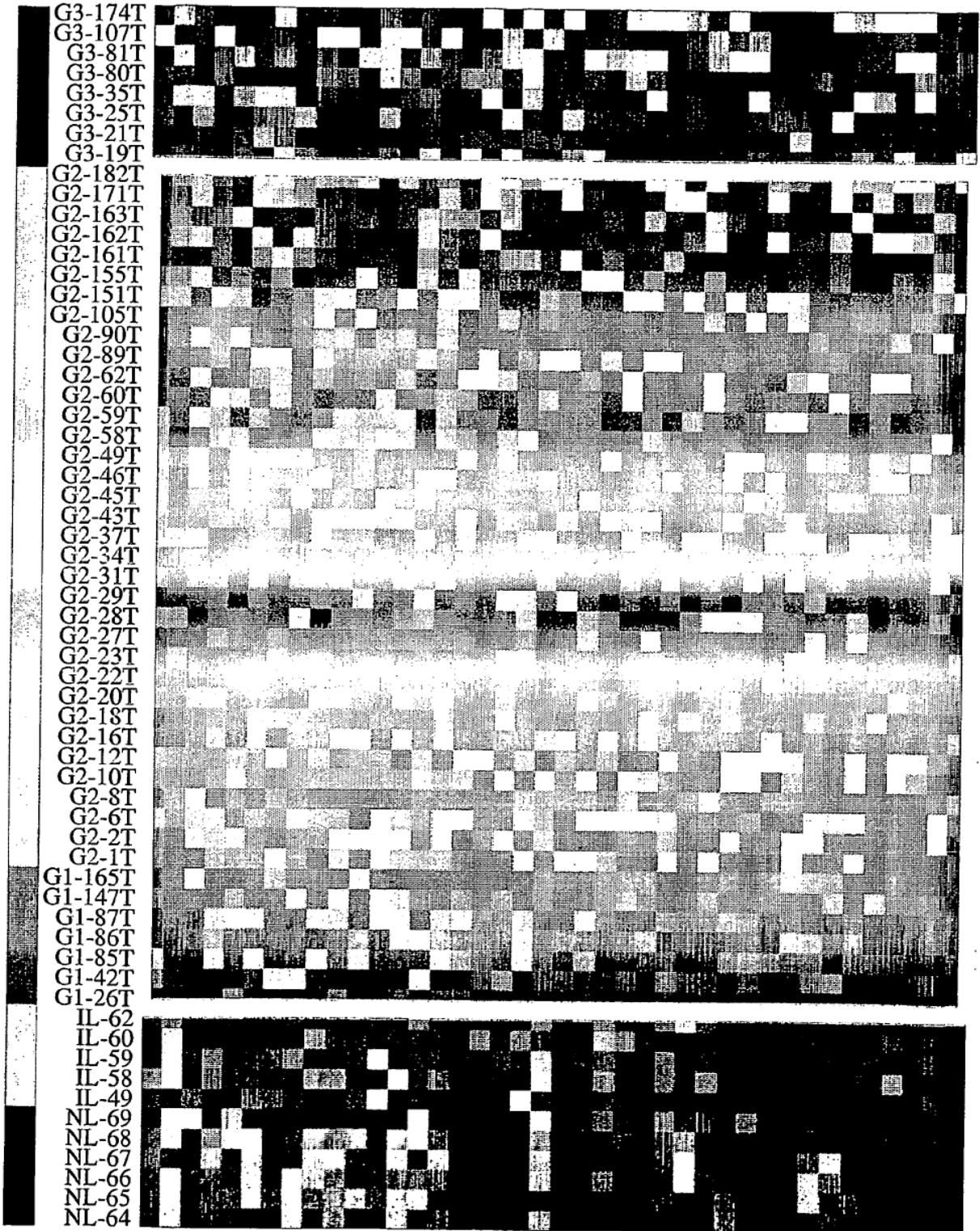


Fig. 1h

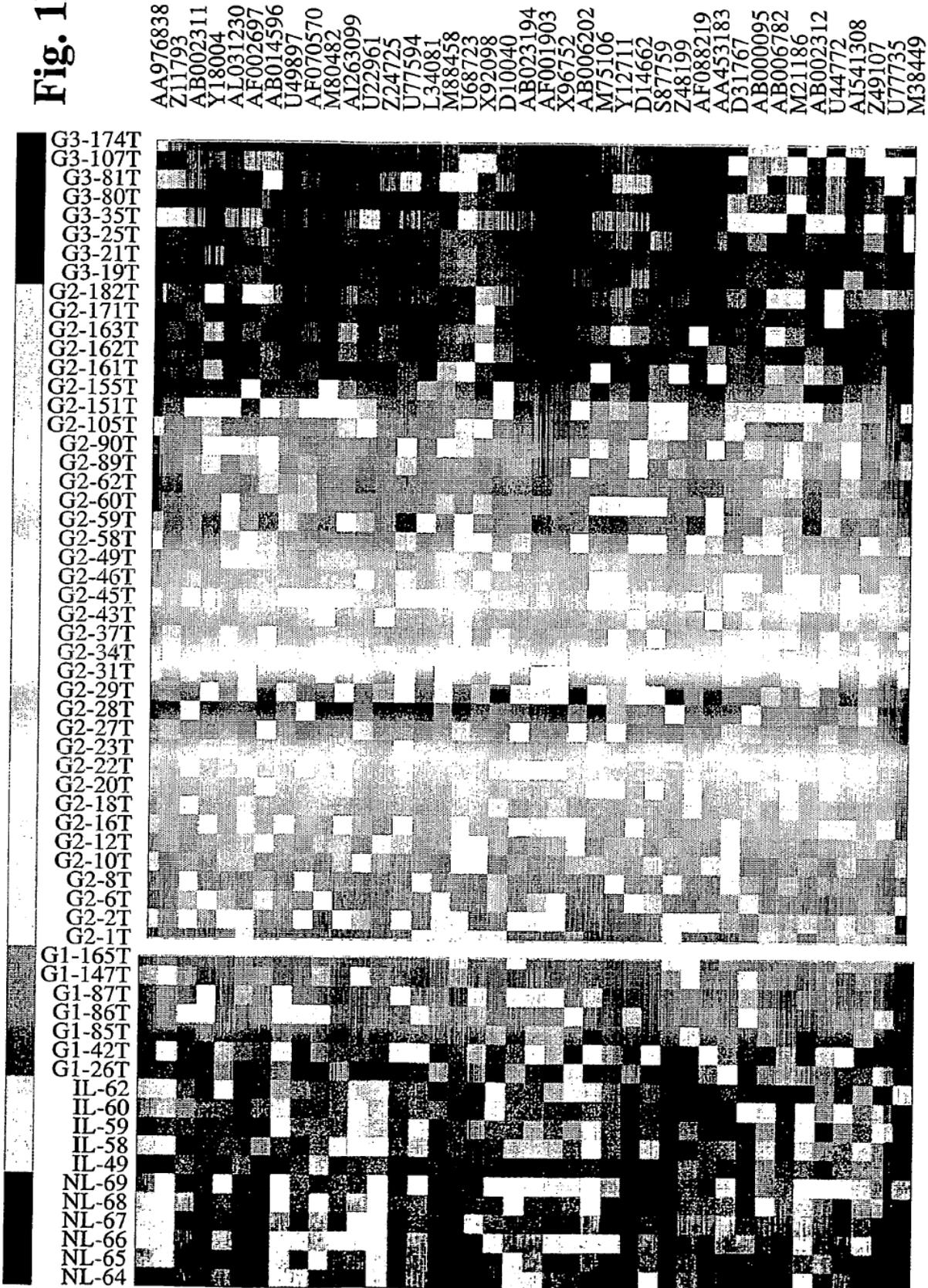


Fig. 3

