(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 429 299**

(51) Int. CI.:

***G01N 33/574*** (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **29.08.2008 E 08782839 (8)**

(97) Fecha y número de publicación de la concesión europea: **31.07.2013 EP 2191272**

(54) Título: **Conjunto de marcadores tumorales**

(30) Prioridad:

**30.08.2007 AT 13592007**

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
**14.11.2013**

(73) Titular/es:

**AIT AUSTRIAN INSTITUTE OF TECHNOLOGY GMBH (100.0%)**
**TECH GATE VIENNA WISSENSCH. U. TECH. PARK DONAU-CITY-STRASSE 1**
**1220 WIEN, AT**

(72) Inventor/es:

**VIERLINGER, KLEMENS;**
**LAUSS, MARTIN;**
**KRIEGNER, ALBERT y**
**NOEHAMMER, CHRISTA**

(74) Agente/Representante:

**CARPINTERO LÓPEZ, Mario**

## DESCRIPCIÓN

Conjunto de marcadores tumorales

La presente invención se refiere al campo del diagnóstico del cáncer y los medios diagnósticos para ello.

Los nódulos tiroideos son endémicos en zonas deficientes en yodo, del tipo de las regiones alpinas de Europa, en las que tienen una prevalencia del 10-20%. Se clasifican por su histología en 2 tipos benignos, bocio nodular (SN) Adenoma Folicular de Tiroides (FTA) y las entidades neoplásicas malignas Carcinoma Folicular de Tiroides (FTC), Carcinoma Papilar de Tiroides (PTC), Carcinoma Medular de Tiroides (MTC) y Carcinoma Anaplásico de Tiroides (ATC). Convencionalmente, la discriminación entre nódulos tiroides benignos y malignos se lleva a cabo mediante gammagrafía y aspiración por boquilla fina seguida por histología. A pesar de los muchos avances en el diagnóstico y el tratamiento de los nódulos tiroideos y el cáncer de tiroides, estos procedimientos tienen una carencia bien conocida es especificidad, particularmente para la discriminación entre FTA y FTC, que conduce a que un gran número de pacientes sean tratados de forma innecesaria para la enfermedad maligna

Dadas las limitaciones diagnósticas de los diversos procedimientos, en particular, la aspiración por boquilla seguida por citología, múltiples investigadores han llevado a cabo estudios de perfilación de la expresión con la esperanza de identificar nuevas herramientas diagnósticas. Dichos análisis intentan identificar los genes expresados de forma diferencial con un importante papel en el desarrollo o progresión de la enfermedad utilizando tecnologías de perfilación de la expresión a nivel de transcrito a gran escala tales como micromatrices de ADNc, matrices de oligonucleótidos y Análisis en Serie de la Expresión Génica (SAGE). Normalmente, se identifican docenas o cientos de genes, muchos de los cuales se espera que sean falsos positivos, y solo una pequeña útil como marcadores diagnósticos/de pronóstico o dianas terapéuticas (Griffith y col., J Clin Oncol 24(31): 5043-5051 (2006)).

En otros tipos de cáncer se ha mostrado que la perfilación de la expresión génica puede añadir un valor sustancial a la discriminación de las diferentes entidades tumorales clínicamente relevantes. El documento US 2006/183141 A describe, por ejemplo, la clasificación de marcadores tumorales a partir de una firma nuclear de respuesta en suero. Diferentes estudios han intentado clasificar las diferentes entidades del carcinoma de tiroides sobre la base de sus perfiles de expresión génica donde cada uno de los cuales discrimina entre 2 de 5 entidades. Sin embargo, los estudios bien no tienen o bien tienen muy pocos genes en común y aplicar un clasificador de un estudio a los datos de otro estudio da como resultado generalmente malos resultados de clasificación.

Es una meta de la presente invención proporcionar marcadores distintivos fiables para el diagnóstico del cáncer, en particular para distinguir nódulos tiroideos benignos procedentes de carcinoma folicular de tiroides maligno (FTC) y carcinoma papilar de tiroides (PTC).

Por tanto, la presente invención proporciona un procedimiento para la detección de marcadores del cáncer de tiroides en una muestra, que comprende detectar la presencia o la cantidad medida de la incidencia del ARNm de al menos 3 marcadores tumorales seleccionados entre los marcadores tumorales PI-1 a PI-33 en la muestra utilizando oligonucleótidos específicos de los ácidos nucleicos de los marcadores tumorales. Estos marcadores tumorales están relacionados con diferentes genes que se expresan de manera anómala en tumores y que se proporcionan en la tabla 1 y se pueden identificar mediante su firma de identificación génica, su nombre descriptivo del gen, pero de forma más ambigua por su UniGenID o su número de acceso que se refiere a secuencias específicas en las bases de datos de secuencias habituales tales como NCBI GenBank, la base de datos EMBL-EBI, EnsEMBL o el Banco de datos de ADN de Japón. Se proporcionan marcadores tumorales adicionales en las tablas 2 a 6. Estos marcadores se han identificado en forma de conjuntos preferidos (PI a PV, FI).

Tabla 1: Conjunto P1-1 a PI-33 marcador de PCT

| Número P I- | gen marcador | Descripción del gen | Nº de acceso. | UniGeneID |
|---|---|---|---|---|
| 1 | BBS9 | Síndrome 9 de Bardet-Biedl | NM_198428<br>NM_001033605<br>NM_001033604<br>NM_014451 | Hs.372360 |
| 2 | C13orf1 | Marco de lectura abierto 1 del cromosoma 13 | NM_020456 | Hs.44235 |
| 3 | CBFA2T3 | Factor de unión a núcleo, dominio enano, subunidad alfa 2 | NM_005187<br>NM_175931 | Hs.513811 |
| 4 | CDT1 | Licenciamiento de la cromatina y factor de replicación DANN 1 | NM_030928 | Hs.122908 |
| 5 | CRK | Homólogo del oncogén CT10 del virus del sarcoma V-crk (aves) | NM_016823<br>NM_005206 | Hs.638121 |
| 6 | CTPS | Sintasa de CTP | NM_001905 | Hs.473087 |

(continuación)

| Número P I- | gen marcador | Descripción del gen | Nº de acceso. | UniGeneID |
|---|---|---|---|---|
| 7 | DAPK2 | Proteína quinasa 2 asociada a muerte | NM_014326 | Hs.237886 |
| 8 | EIF5 | Factor 5 de inicio de la traducción eucariota | NM_001969 NM_183004 | Hs.433702 |
| 9 | EREG | Epirequlina | NM_001432 | Hs.115263 |
| 10 | GK | Glicerol quinasa | NM_203391 NM_000167 | Hs.1466 |
| 11 | GPATCH8 | Dominio del parche G que contiene 8 | NM_001002909 | Hs.463129 |
| 12 | HDGF | Factor de crecimiento derivado de hepatoma (proteína de tipo 1 del grupo de alta movilidad) | NM_004494 | Hs.506748 |
| 13 | IRF2BP1 | Proteína 1 de unión al factor 2 regulador del interferón | NM_015649 | Hs.515477 |
| 14 | KRT83 | Queratina 83 | NM_002282 | Hs.661428 |
| 15 | MYOD1 | Diferenciación miogénica 1 | NM_002478 | Hs.181768 |
| 16 | NME6 | Proteína expresada en células 6 no metastásicas, (nucleósido- difosfato quinasa) | NM_005793 | Hs.465558 |
| 17 | POLE3 | Polimerasa (dirigida a DNA), épsilon 3 (subunidad p17) | NM_017443 | Hs.108112 |
| 18 | PPP1R13B | Proteína fosfatasa 1, subunidad reguladora (inhibidora) 13B | NM_015316 | Hs.436113 |
| 19 | PRPH2 | Periferina 2 (degeneración retinal, lenta) | NM_000322 | Hs.654489 |
| 20 | RASSF7 | Asociación Ras (familia 7 del dominio (RaIGDS/AF-6) | NM_003475 | Hs.72925 |
| 21 | ROCK2 | Proteína quinasa 2 que contiene cola enrollada asociada a Rho | NM_004850 | Hs.591600 |
| 22 | RTN1 | Reticulón 1 | NM_021136 NM_206857 NM_206852 | Hs.368626 |
| 23 | S100B | Proteína B de unión a calcio S100 | NM_006272 | Hs.422181 |
| 24 | SLIT2 | Homólogo de hendidura 2 (Drosophila) | NM_004787 | Hs.29802 |
| 25 | SNRPB2 | Polipéptido B de la ribonucleoproteína nuclear pequeña | NM_003092 NM_198220 | Hs.280378 |
| 26 | SPAG7 | Antígeno 7 asociado a esperma | NM_004890 | Hs.90436 |
| 27 | STAU1 | Homólogo 1 de la proteína Staufende de unión a ARN (Drosophila) | NM_017453 NM_001037328NM_ 004602 NM_01745 NM_017454 | Hs.596704 |
| 28 | SUPT5H | Supresor del homólogo Ty 5 (S. cerevisiae) | NM_003169 | Hs.631604 |
| 29 | TBX10 | Secuencia T10 | NM_005995 | Hs.454480 |
| 30 | TLK1 | Quinasa 1 de tipo desordenado | NM_012290 | Hs.655640 |
| 31 | TM4SF4 | Miembro de la familia seis de transmembrana 4 L | NM_004617 | Hs.133527 |
| 32 | TXN | Tioredoxina | NM_003329 | Hs.435136 |
| 33 | UFD1L | Degradación de tipo 1 de fusión a ubiquitina (levadura) | NM_005659 NM_001035247 | Hs.474213 |

Tabla 2: Conjunto PII a PII-64 del marcador de PTC

| Número P II- | Gen marcador | Descripción del gen | Nº de acceso. | UniGeneID |
|---|---|---|---|---|
| 1 | ADH1B | Alcohol deshidrogenasa IB (clase I), polipéptido beta | NM_000668 | Hs.4 |
| 2 | AGR2 | Homólogo 2 del gradiente anterior (Xenopus laevis) | NM_006408 | Hs.530009 |
| 3 | AGTR1 | Receptor de tipo 1 de la Angiotensina II | NM_031850<br>NM_004835<br>NM_009585<br>NM 032049 | Hs.477887 |
| 4 | AGTR1 | Receptor de tipo 1 de la Angiotensina II | NM_000685 | Hs.654382 |
| 5 | ALDH1A 1 | Miembro A1 de la familia de la aldehído deshidrogenasa 1 | NM_000689 | Hs.76392 |
| 6 | ALDH1A 3 | Miembro A3 de la familia de la aldehído deshidrogenasa 1 | NM_000693 | Hs.459538 |
| 7 | AMIG02 | Molécula de adhesión al dominio 2 de tipo Ig | NM_181847 | Hs.121520 |
| 8 | ATP2C2 | Miembro 2 de tipo 2C del transportador de Ca++ de la ATPasa | NM_014861 | Hs.6168 |
| 9 | BID | Agonista de muerte del dominio de interacción BH3 | NM_197966<br>NM_001196<br>NM_197967 | Hs.591054 |
| 10 | C7orf24 | Marco de lectura abierto 24 del cromosoma 7 | NM_024051 | Hs.530024 |
| 11 | CA4 | Anhidrasa carbónica IV | NM_000717 | Hs.89485 |
| 12 | CCL21 | Ligando 21 de la quimioquina (motivo C-C) | NM_002989 | Hs.57907 |
| 13 | CD55 | Factor de aceleración de la desintegración de la molécula para el complemento (Grupo sanguíneo Cromer) | NM_000574 | Hs.527653 |
| 14 | CDH16 | Caderina 16, caderina KSP | NM_004062 | Hs.513660 |
| 15. | CDH3 | Caderina 3, P-caderina P de tipo 1 (placental) | NM_133458<br>NM_001793 | Hs.461074 |
| 16 | CFI | Factor I del complemento | NM_000204 | Hs.312485 |
| 17 | CH13LI 1 | Quitinasa 3-de tipo 1 (glicoproteína 39 de cartílago) | NM_001276 | Hs.382202 |
| 18 | CHST2 | Carbohidrato (N-acetilglucosamina-6-0) sulfotransferasa 2 | NM_004267 | Hs.8786 |
| 19 | CITED2 | Transactivador interactuante Cbp/p300 con dominio 2 del extremo carboxi rico en Glu/Asp | NM_006079 | Hs.82071 |
| 20 | CLCNKB | Canal Kb de cloruro | NM_000085 | Hs.352243 |
| 21 | C0MP | Proteína de matriz oligomérica de cartílago | NM_000095 | Hs.1584 |
| 22 | CTSH | Catepsina H | NM_004390<br>NM_148979 | Hs.148641 |
| 23 | DI02 | Yodotironina desyodinasa de tipo II | NM_013989<br>NM_000793<br>NM_001007023 | Hs.202354 |
| 24 | DIRAS3 | Familia DIRAS, RAS de tipo 3 de unión a GTP | NM_004675 | Hs.194695 |
| 25 | DUSP4 | Fosfatasa 4 de doble especificidad | NM_057158<br>NM_001394 | Hs.417962 |
| 26 | DUSP5 | Fosfatasa 5 de doble especificidad | NM_004419 | Hs.2128 |

(continuación)

| Número P II- | Gen marcador | Descripción del gen | Nº de acceso. | UniGeneID |
|---|---|---|---|---|
| 27 | EDN3 | Endotelina 3 | NM_207032<br>NM_207034<br>NM_207033<br>NM_000114 | Hs.1408 |
| 28 | ENTPD1 | Ectonucleósido trifosfato difosfohidrolasa 1 | NM_001776<br>NM_001098175 | Hs.576612 |
| 29 | FHL1 | Dominios LIM 1 cuatro y medio | NM_001449 | Hs.435369 |
| 30 | GDF15 | Factor 15 de diferenciación del crecimiento | NM_004864 | Hs.616962 |
| 31 | GPM6A | Glicoproteína M6A | NM_201591<br>NM_005277<br>NM_201592 | Hs.75819 |
| 32 | HBA1 | Hemoglobina, alfa 1 | NM_000558 | Hs.449630 |
| 33 | IRS1 | Sustrato 1 del receptor de la insulina | NM_005544 | Hs.471508 |
| 34 | KCNJ2 | Miembro 2 de la subfamilia J de internalización/rectificación del canal del potasio | NM_000891 | Hs.1547 |
| 35 | KCNN4 | Miembro 4 de la subfamilia N del canal del potasio de conductancia intermedia/pequeña activado por calcio | NM_002250 | Hs.10082 |
| 36 | KLK10 | Peptidasa 10 relacionada con Kallikreína | NM_002776<br>NM_001077500<br>NM_145888 | Hs.275464 |
| 37 | LAMB3 | Laminina, beta 3 | NM_001017402<br>NM_000228 | Hs.497636 |
| 38 | LCN2 | Lipocalina 2 (oncogén 24p3) | NM_005564 | Hs.204238 |
| 39 | LMOD1 | Leyomodina 1 (músculo liso) | NM_012134 | Hs.519075 |
| 40 | MATN2 | Matrilina 2 | NM_002380<br>NM_030583 | Hs.189445 |
| 41 | MPPED2 | Dominio de la metalofosfoesterasa que contiene 2 | NM_001584 | Hs.289795 |
| 42 | MVP | Proteína vault mayor | NM_017458<br>NM_005115 | Hs.632177 |
| 43 | NELL2 | NEL de tipo 2 (pollo) | NM_006159 | Hs.505326 |
| 44 | NFE2L3 | Factor nuclear (derivado 2 de eritroide) de tipo 3 | NM_004289 | Hs.404741 |
| 45 | NPC2 | Enfermedad de Niemann-Pick de tipo C2 | NM_006432 | Hs.433222 |
| 46 | NRCAM | Molécula de adhesión a células neuronales | NM_001037132<br>NM_005010<br>NM_001037133 | Hs.21422 |
| 47 | NRIP1 | Proteína 1 interactuante con el receptor nuclear | NM_003489 | Hs.155017 |
| 48 | PAPSS2 | 3'-fosfoadenosina 5'- fosfosulfato sintasa 2 | NM_001015880<br>NM_004670 | Hs.524491 |
| 49 | PDLIM4 | Dominio 4 de PDZ y LIM | NM_003687 | Hs.424312 |
| 50 | PDZK11P1 | Proteína 1 interactuante con PDZK1 | NM_005764 | Hs.431099 |
| 51 | PIP3-E | Proteína PIP3-Ede unión a fosfoinositido | NM_015553 | Hs.146100 |
| 52 | PLAU | Uroquinasa activadora del plasminógeno | NM_002658 | Hs.77274 |
| 53 | PRSS2 | Serina proteasa 2 (tripsina 2) | NM_002770 | Hs.622865 |
| 54 | PRSS23 | Serina proteasa 23 | NM_007173 | Hs.25338 |
| 55 | RAP1GAP | Proteína activadora de la RAP1 GTPasa | NM_002885 | Hs.148178 |
| 56 | S100A11 | Proteína A11 de unión a calcioS100 | NM_005620 | Hs.417004 |

(continuación)

| Número P II- | Gen marcador | Descripción del gen | Nº de acceso. | UniGeneID |
|---|---|---|---|---|
| 57 | SFTPB | Proteína B asociada a tensioactivo pulmonar | NM_198843<br>NM_000542 | Hs.512690 |
| 58 | SLPI | Inhibidor de la leucocito peptidasa secretora | NM_003064 | Hs.517070 |
| 59 | SOD3 | Superóxido dismutasa 3, extracelular | NM_003102 | Hs.2420 |
| 60 | SPINT1 | Inhibidor Kunitz de tipo 1 de la serina peptidasa | NM_181642<br>NM_003710<br>NM001032367 | Hs.233950 |
| 61 | SYNE1 | Envoltura nuclear 1 que contiene la repetición de la espectrina | NM_182961<br>NM_033071<br>NM_015293<br>NM_133650 | Hs.12967 |
| 62 | TACSTD2 | Transductor 2 de la señal del calcio asociada a tumor | NM_002353 | Hs.23582 |
| 63 | UPP1 | Uridina fosforilasa 1 | NM_181597<br>NM_003364 | Hs.488240 |
| 64 | WASF3 | Miembro 3 de la familia de la proteína WAS | NM_006646 | Hs.635221 |

Tabla 3: Conjunto PIII a PIII-70 del marcador de PTC

| Número P III- | Gen marcador | Descripción del gen | Nº de acceso. | UniGeneID |
|---|---|---|---|---|
| 1 | APOE | Apolipoproteína E | NM_000041 | Hs.654439 |
| 2 | ATIC | 5-aminoimidazol-4-carboxamida ribonucleótido formiltransferasa / IMP ciclohidrolasa | NM_004044 | Hs.90280 |
| 3 | BASP1 | Proteína 1 de la señal unida a membrana, abundante en el cerebro | NM_006317 | Hs.201641 |
| 4 | C9orf61 | Marco de lectura abierto 61 del cromosoma 9 | NM_004816 | Hs.118003 |
| 5 | CCL13 | Ligando 13 de la quimioquina (motivo C-C) | NM_005408 | Hs.414629 |
| 6 | CD36 | Molécula CD36 (receptor de la tromboespondina) | NM_001001548<br>NM_001001547<br>NM 000072 | Hs.120949 |
| 7 | CDH6 | Caderina 6, caderina K de tipo 2 (riñón fetal) | NM_004932 | Hs.171054 |
| 8 | CFB | Factor B del complemento | NM_001710 | Hs.69771 |
| 9 | CFD | Factor D del complemento (adipsina) | NM_001928 | Hs.155597 |
| 10 | CLDN10 | Claudina 10 | NM_182848<br>NM_006984 | Hs.534377 |
| 11 | COL11A1 | Colágeno de tipo XI, alfa 1 | NM_080629<br>NM_001854<br>NM080630 | Hs.523446 |

(continuación)

| Número P III- | Gen marcador | Descripción del gen | Nº de acceso. | UniGeneID |
|---|---|---|---|---|
| 12 | COL13A1 | Colágeno de tipo XIII, alfa 1 | NM_005203<br>NM_080804<br>NM_080798<br>NM_080803<br>NM_080802<br>NM_080799<br>NM_080800<br>NM_080801<br>NM_080808<br>NM_080809<br>NM_080805<br>NM_080807<br>NM_080806<br>NM_080811<br>NM_080810<br>NM_080812<br>NM_080813<br>NM_080814<br>NM_080815 | Hs.211933 |
| 13 | CORO2B | Coronina, proteína 2B de unión a actina | NM_006091 | Hs.551213 |
| 14 | CRLF1 | Factor 1 similar al receptor de la citoquina | NM_004750 | Hs.114948 |
| 15 | CXorf6 | Marco de lectura abierto 6 del cromosoma X | NM_005491 | Hs.20136 |
| 16 | DDB2 | Proteína 2 de unión a AND específica de daño de 2,48kDa | NM_000107 | Hs.655280 |
| 17 | DPP6 | Dipeptidil-peptidasa 6 | NM_001039350<br>NM_130797<br>NM_001936 | Hs.490684 |
| 18 | ECM1 | Proteína 1 de matriz extracelular | NM_004425<br>NM_ 022664 | Hs.81071 |
| 19 | EFEMP1 | Proteína 1 de matriz extracelular de tipo fibulina que contiene EGF | NM_004105<br>NM_001039348<br>NM_001039349 | Hs.76224 |
| 20 | ESRRG | Receptor gamma relacionado con estrógeno | NM_206594<br>NM_001438<br>NM_206595 | Hs.444225 |
| 21 | ETHE1 | Encefalopatía etilmalónica 1 | NM 014297 | Hs.7486 |
| 22 | FAS | Fas (miembro 6 de la superfamilia del receptor TNF) | NM_000043<br>NM_152872<br>NM_152871<br>NM_152873<br>NM_152875<br>NM_152874<br>NM_152877<br>NM_152876 | Hs.244139 |
| 23 | FMOD | Fibromodulina | NM_002023 | Hs.519168 |
| 24 | GABBR2 | Receptor 2 del ácido Gamma-aminobutírico (GABA) B | NM_005458 | Hs.198612 |
| 25 | GALE | UDP-galactosa-4-epimerasa | NM_000403<br>NM_001008216 | Hs.632380 |
| 26 | GATM | Glicina amidinotransferasa (L- arginina:glicina amidinotransferasa) | NM_001482 | Hs.75335 |
| 27 | GDF10 | Factor 10 de diferenciación del crecimiento | NM_004962 | Hs.2171 |
| 28 | GHR | Receptor de la hormona de crecimiento | NM_000163 | Hs.125180 |
| 29 | GPC3 | Glipicano 3 | NM_004484 | Hs.644108 |

(continuación)

| Número P III- | Gen marcador | Descripción del gen | Nº de acceso. | UniGeneID |
|---|---|---|---|---|
| 30 | ICAM1 | Receptor de la molécula 1 de adhesión intercelular (CD54), del rinovirus humano | NM_000201 | Hs.643447 |
| 31 | ID3 | Proteína dominante negativa hélice-bucle-hélice del inhibidor de la unión a ADN 3 | NM_002167 | Hs.76884 |
| 32 | IER2 | Respuesta 2 inmediata temprana | NM_004907 | Hs.501629 |
| 33 | IGFBP6 | Proteína 6 de unión al factor de crecimiento de tipo insulina | NM_002178 | Hs.274313 |
| 34 | IQGAP2 | Proteína 2' activadora de la GTPasa que contiene el motivo IQ | NM_006633 | Hs.291030 |
| 35 | ITGA2 | Integrina, alfa 2 (CD49B, receptor VLA-2 de la subunidad alfa2) | NM_002203 | Hs.482077 |
| 36 | ITGA3 | Integrina, alfa 3 (antígeno CD49C, receptor VLA-3 de la subunidad alfa 3) | NM_002204<br>NM_005501 | Hs.265829 |
| 37 | ITM2A | Proteína 2A integral de membrana | NM_004867 | Hs.17109 |
| 38 | KIAA0746 | Proteína KIAA0746 | NM_015187 | Hs.479384 |
| 39 | LRIG1 | Dominios 1 de tipo inmunoglobulina y repeticiones ricas en leucina | NM_015541 | Hs.518055 |
| 40 | LRP2 | Proteína 2 relacionada con la lipoproteína de baja densidad | NM_004525 | Hs.470538 |
| 41 | LY6E | Complejo del antígeno 6 de linfocitos, locus E | NM_002346 | Hs.521903 |
| 42 | MAPK13 | Proteína quinasa 13 activada por mitógeno | NM_002754 | Hs.178695 |
| 43 | MDK | Midkina (factor 2 promotor del crecimiento de neuritas) | NM_001012334<br>NM_001012333<br>NM_002391 | Hs.82045 |
| 44 | MLLT11 | Leucemia de linaje mieloide/linfoide o mixto (homólogo de Drosophila trithorax) | NM_006818 | Hs.75823 |
| 45 | MMRN1 | Multimerina 1 | NM_007351 | Hs.268107 |
| 46 | MTMR11 | Proteína 11 relacionada con miotubularina | NM_181873 | Hs.425144 |
| 47 | MXRA8 | Remodelación 8 asociada a matriz | NM_032348 | Hs.558570 |
| 48 | NAB2 | Proteína 2 de unión a NGFI-A (proteína 2 de unión a EGR1) | NM_005967 | Hs.159223 |
| 49 | NMU | Neuromedina U | NM_006681 | Hs.418367 |
| 50 | OCA2 | Albinismo II oculocutáneo (homólogo de dilución de ojo rosado, ratón) | NM_000275 | Hs.654411 |
| 51 | PDE5A | Fosfodiesterasa 5A, específico de cGMP | NM_001083<br>NM_033430<br>NM_033437 | Hs.647971 |
| 52 | PLAG1 | Gen 1 de adenoma pleiomórfica | NM_002655 | Hs.14968 |
| 53 | PLP2 | Proteína 2 proteolipida (epitelio colónico enriquecido) | NM_002668 | Hs.77422 |
| 54 | PLXNC1 | Plexina C1 | NM_005761 | Hs.584845 |
| 55 | PRKCQ | Proteína quinasa C, theta | NM_006257 | Hs.498570 |
| 56 | PRUNE | Homólogo de la ciruela (Drosophila) | NM_021222 | Hs.78524 |
| 57 | RAB27A | Miembro RAB27A de la familia del oncogén RAS, | NM_004580<br>NM_183234<br>NM_183235<br>NM_183236 | Hs.654978 |
| 58 | RYR2 | Receptor 2 de la rianodina receptor 2 (cardiaco) | NM_001035 | Hs.109514 |
| 59 | SCEL | Escielina | NM_144777<br>NM_003843 | Hs.534699 |

(continuación)

| Número P III- | Gen marcador | Descripción del gen | Nº de acceso. | UniGeneID |
|---|---|---|---|---|
| 60 | SELENBP 1 | Proteína 1 de unión a selenio | NM_003944 | Hs.632460 |
| 61 | SORBS2 | Sorbina y dominio de SH3 que contiene 2 | NM_021069 NM_003603 | Hs.655143 |
| 62 | STMN2 | Statmina de tipo 2 | NM_007029 | Hs.521651 |
| 63 | TBC1D4 | Miembro 4 de la familia del dominio TBC1 | NM_014832 | Hs.210891 |
| 64 | TM4SF4 | Miembro 4 de la familia 4 L seis transmembrana | NM_004617 | Hs.133527 |
| 65 | TNC | Tenascina C (hexabrachion) | NM_002160 | Hs.143250 |
| 66 | TPD52L1 | Proteína tumoral D52 de tipo 1 | NM_001003395 NM_003287NM_0 01003396 NM_001003397 | Hs.591347 |
| 67 | TSC22D1 | Miembro 1 de la familia del dominio TSC22 | NM_183422 NM_006022 | Hs.507916 |
| 68 | TTC30A | Dominio 30A de repetición de tetratricopéptido | NM_152275 | Hs.128384 |
| 69 | VLDLR | Receptor de lipoproteína de muy baja densidad | NM_003383 NM_001018056 | Hs. 370422 |
| 70 | WFS1 | Síndrome 1 de Wolfram (wolframina) | NM_006005 | Hs.518602 |

Tabla 4: Conjunto FI-1 a FI-147 del marcador de FTC

| Número FI- | Gen marcador | Descripción de gen | Nº de acceso | UniGeneID |
|---|---|---|---|---|
| 1 | AATF | Factor de transcripción antagonizante de la apoptosis | NM_012138 | Hs.195740 |
| 2 | ACOX3 | Acil-Coenzima A oxidasa 3, pristanoílo | NM_003501 | Hs.479122 |
| 3 | AHDC1 | Motivo de unión de ADN al gancho de AT, que contiene 1 | NM_001029882 | Hs.469280 |
| 4 | ALAS2 | Aminolevulinato, delta-, sintasa 2 (anemia sideroblástica/hipocrómica ) | NM_000032 NM_001037968 NM_001037967 NM_001037969 | Hs.522666 |
| 5 | ALKBH1 | AlkB, homólogo 1 de reparación de la alquilación (E. coli) | NM_006020 | Hs.94542 |
| 6 | ANGPTL2 | Angiopoyetina de tipo 2 | NM_012098 | Hs.653262 |
| 7 | AP2A2 | Complejo de proteína 2 relacionado con el adaptador, subunidad alfa 2 | NM_012305 | Hs.19121 |
| 8 | APOBEC3 G | Enzima editora del ARNm de la apolipoproteína B , polipéptido catalítico de tipo 3G | NM_021822 | Hs.660143 |
| 9 | APRIN | Inhibidor de la proliferación inducido por andrógeno | NM_015032 | Hs.693663 |
| 10 | ARNT | Translocador nuclear del receptor del hidrocarburo arilo | NM_001668 NM_178427 NM_178426 | Hs.632446 |
| 11 | AZGP1 | Glicoproteína 1 alfa-2-glicoproteína1, de unión a cinc | NM_001185 | Hs.546239 |
| 12 | BAT2D1 | Dominio BAT2 que contiene 1 | NM_015172 | Hs.494614 |
| 13 | BATF | Factor de transcripción básico de cremallera de leucina, de tipo ATF | NM_006399 | Hs.509964 |

(continuación)

| Número FI- | Gen marcador | Descripción de gen | Nº de acceso | UniGeneID |
|---|---|---|---|---|
| 14 | BPHL | Serina hidrolasa de tipo bifenil hidrolasa | NM_004332 | Hs.10136 |
| 15 | C13orf1 | Marco de lectura abierto 1 del cromosoma 13 | NM_020456 | Hs.44235 |
| 16 | C14orf1 | Marco de lectura abierto 1 del cromosoma 14 | NM_007176 | Hs.15106 |
| 17 | C2orf3 | Marco de lectura abierto 3 del cromosoma 2 | NM_003203 | Hs.303808 |
| 18 | CBFB | Subunidad beta del factor de unión a núcleo | NM_001755<br>NM 022845 | Hs.460988 |
| 19 | CBR3 | Carbonil reductasa 3 | NM_001236 | Hs.154510 |
| 20 | CBX5 | Homólogo 5 de Chromobox (homólogo alfa de HP1, Drosophila) | NM_012117 | Hs.632724 |
| 21 | CCNE2 | Ciclina E2 | NM_057749NM 057735 | Hs.567387 |
| 22 | CD46 | Molécula CD46 , proteína reguladora del complemento | NM_002389<br>NM_172354<br>NM_172351<br>NM_172355<br>NM_172352<br>NM_172359<br>NM_172357<br>NM_172360<br>NM_153826<br>NM_172358<br>NM_172356<br>NM_172353<br>NM_172361<br>NM_172350 | Hs.510402 |
| 23 | CHPF | Factor de polimerización de la condroitina | NM_024536 | Hs.516711 |
| 24 | CHST3 | Carbohidrato (condroitina 6) sulfotransferasa 3 | NM_004273 | Hs.158304 |
| 25 | CLCN2 | Canal del cloruro 2 | NM_004366 | Hs.436847 |
| 26 | CLCN4 | Canal del cloruro 4 | NM_001830 | Hs.495674 |
| 27 | CLIC5 | Canal del cloruro intracelular 5 | NM_016929 | Hs.485489 |
| 28 | CNOT2 | Complejo de transcripción CCR4-NOT, subunidad 2 | NM_014515 | Hs.133350 |
| 29 | COPS6 | Subunidad 6 del homólogo fotomorfogénico constitutiva de COP9 (Arabidopsis) | NM_006833 | Hs.15591 |
| 30 | CPZ | Carboxipeptidasa Z | NM_001014448<br>NM_001014447<br>NM_003652 | Hs.78068 |
| 31 | CSK | C-src tirosina quinasa | NM_004383 | Hs.77793 |
| 32 | CTDP1 | CTD (dominio del extremo carboxi, ARN polimerasa II, polipéptido A) fosfatasa, subunidad 1 | NM_004715<br>NM_048368 | Hs.465490 |
| 33 | DDEF2 | Factor 2 potenciador del desarrollo y diferenciación | NM_003887 | Hs.555902 |
| 34 | DKFZP586H2123 | Proteasa muscular asociada a regeneración | NM_015430<br>NM_001001991 | Hs.55044 |

(continuación)

| Número FI- | Gen marcador | Descripción de gen | Nº de acceso | UniGeneID |
|---|---|---|---|---|
| 35 | DLG2 | Homólogo 2 de discos grandes, capsina-110 (Drosophila) | NM_001364 | Hs.654862 |
| 36 | DPAGT1 | Dolicil-fosfato (UDP-N-acetilglucosamina) N-acetilglucosaminofosfotransferasa 1 (GlcNAc-1-P transferasa) | NM_001382<br>NM_203316 | Hs.524081 |
| 37 | DSCR1 | Gen 1 de la región crítica del síndrome de Down | NM_004414<br>NM_203418<br>NM_203417 | Hs.282326 |
| 38 | DUSP8 | Fosfatasa 8 de doble especificidad 8 | NM_004420 | Hs.41688 |
| 39 | E124 | Etopósido inducido por ARNm 2.4 | NM_004879<br>NM_001007277 | Hs.643514 |
| 40 | ENOSF1 | Miembro 1 de la superfamilia de la enolasa | NM_017512 | Hs.369762 |
| 41 | ERCC1 | Grupo 1 de complementación de la deficiencia en la reparación en roedores que complementa la reparación en roedores complementando la reparación cruzada por escisión (incluye la secuencia de sentido contrario del solapamiento) | NM_202001<br>NM_001983 | Hs.435981 |
| 42 | ERCC3 | Grupo 3 de complementación de la deficiencia en la reparación en roedores que complementa la reparación cruzada por escisión | NM_000122 | Hs.469872 |
| 43 | ERH | Potenciador del homólogo rudimentario (Drosophila) | NM_004450 | Hs.509791 |
| 44 | F13A1 | Polipéptido A1 del factor de coagulación XIII | NM_000129 | Hs.335513 |
| 45 | FAM20B | Miembro B de la familia 20 con similitud de la secuencia | NM_014864 | Hs.5737 |
| 46 | FBP1 | Fructosa-1,6-bisfosfatasa 1 | NM_000507 | Hs.494496 |
| 47 | FCGR2A | Receptor de baja afinidad IIa del fragmento Fc de la IgG, (CD32) | NM_021642 | Hs.352642 |
| 48 | FGF13 | Factor 13 de crecimiento de fibroblastos | NM_004114<br>NM_033642 | Hs.6540 |
| 49 | FGFR10P | Oncogén asociado a FGFR1 | NM_007045<br>NM_194429 | Hs.487175 |
| 50 | FLNC | Filamina C, gamma (proteína 280 de unión a actina) | NM_001458 | Hs.58414 |
| 51 | FMO5 | Monooxigenasa 5 que contiene flavina | NM_001461 | Hs.642706 |
| 52 | FRY | Homólogo de Furry (Drosophila) | NM_023037 | Hs.591225 |
| 53 | GADD45G | Paro en el crecimiento y daño inducible por ADN, gamma | NM_006705 | Hs.9701 |
| 54 | GCH1 | GTP ciclohidrolasa 1 (distonía sensible a dopa) | NM_000161<br>NM_001024024<br>NM_001024070<br>NM_001024071 | Hs.86724 |

(continuación)

| Número FI- | Gen marcador | Descripción de gen | Nº de acceso | UniGeneID |
|---|---|---|---|---|
| 55 | GFRA1 | Receptor alfa 1 de la familia GDNF | NM_005264<br>NM_145793 | Hs.591913 |
| 56 | GLB1 | Galactosidasa, beta 1 | NM_001039770<br>NM_000404<br>NM 001079811 | Hs.443031 |
| 57 | GOLGA8A | Autoantígeno de Golgi, golgina subfamilia a, 8A | NM_181077<br>NM_001023567 | Hs.182982 |
| 58 | HCLS1 | substrato1 de Lyn específico de células hematopoyéticas | NM_005335 | Hs.14601 |
| 59 | HDGF | Factor de crecimiento derivado de hepatoma (Grupo de proteína de tipo 1 de elevada movilidad) | NM_004494 | Hs.506748 |
| 60 | HRC | Proteína de unión a calcio rica en histidina | NM_002152 | Hs.436885 |
| 61 | ICMT | Isoprenilcisteína carboxil metil-transferasa | NM_012405 | Hs.562083 |
| 62 | IFNA5 | Interferón, alfa 5 | NM_002169 | Hs.37113 |
| 63 | IGF2BP3 | Proteína de unión 3 de unión al ARNm del factor 2 de crecimiento de tipo insulina | NM_006547 | Hs.648088 |
| 64 | IL12A | Interleuquina 12A (factor 1 estimulador de la célula asesina natural, factor 1 de maduración de los linfocitos citotóxicos, p35) | NM_000882 | Hs.673 |
| 65 | ITIH2 | Inhibidor H2 Inter-alfa (globulina) | NM_002216 | Hs.75285 |
| 66 | ITPKC | Inositol 1,4,5-trisfosfate 3-quinasa C | NM_025194 | Hs.515415 |
| 67 | JMJD2A | Dominio Jumonji que contiene 2A | NM_014663 | Hs.155983 |
| 68 | KCNJ15 | Miembro 15 de la subfamilia J del canal de internalización-rectificación del potasio | NM_170736<br>NM_002243<br>NM_170737 | Hs.411299 |
| 69 | KCTD12 | Dominio de tetramerización del canal del potasio que contiene 12 | NM_138444 | Hs.693617 |
| 70 | KIAA0652 | KIAA0652 | NM_014741 | Hs.410092 |
| 71 | KIAA0913 | KIAA0913 | NM_015037 | Hs.65135 |
| 72 | KLKB1 | Kallikreína B, plasma (factor de Fletcher) 1 | NM_000892 | Hs.646885 |
| 73 | KRT37 | Queratina 37 | NM_003770 | Hs.673852 |
| 74 | LAMB3 | Laminina, beta 3 | NM_001017402<br>NM_000228 | Hs.497636 |
| 75 | LPHN3 | Latrofilina 3 | NM_015236 | Hs.694758<br>Hs.649524 |
| 76 | LRIG1 | Dominios 1 de tipo inmunoglobulina y repeticiones ricas en leucina | NM_015541 | Hs.518055 |
| 77 | LSR | Lipolisis estimulada por el receptor de la lipoproteína | NM_205834<br>NM_015925<br>NM_205835 | Hs.466507 |
| 78 | MANBA | Manosidasa, beta A, lisosómica | NM_005908 | Hs.480415 |
| 79 | MAP7 | Proteína 7 asociada a microtúbulos | NM_003980 | Hs.486548 |

(continuación)

| Número FI- | Gen marcador | Descripción de gen | Nº de acceso | UniGeneID |
|---|---|---|---|---|
| 80 | MAPKAPK 5 | Proteína quinasa 5 activada por proteína quinasa activad por mitógeno | NM_139078 NM_003668 | Hs.413901 |
| 81 | MET | Protooncogén Met (receptor del factor de crecimiento de los hepatocitos) | NM_000245 | Hs.132966 |
| 82 | MMP14 | Metalopeptidasa de matriz 14 (insertada a membrana) | NM_004995 | Hs.2399 |
| 83 | MX1 | Mixovirus (virus de la gripe) resistencia 1, proteína p78 inducible por interferón (ratón) | NM_002462 | Hs.517307 |
| 84 | MYL9 | Miosina, cadena ligera 9, regulador | NM_006097 NM_181526 | Hs.504687 |
| 85 | MY09B | Miosina IXB | NM_004145 | Hs.123198 |
| 86 | NCOR1 | Correpresor 1 del receptor nuclear | NM_006311 | Hs.462323 |
| 87 | NDRG4 | Miembro 4 de la familia NDRG | NM_020465 NM_022910 | Hs.322430 |
| 88 | NDUFA5 | Subcomplejo alfa 1 de NADH deshidrogenasa (ubiquinona) 5, 13kDa | NM_005000 | Hs.651219 |
| 89 | NEUROD2 | Diferenciación neurogénica 2 | NM_006160 | Hs.322431 |
| 90 | NFKB2 | Factor nuclear del gen del polipéptido potenciador de la cadena kappa ligera en linfocitos B (p49/p100) | NM_001077494 NM_001077493 NM_002502 | Hs.73090 |
| 91 | NME6 | Proteína expresada en células no metastásicas 6, (nucleósido-difosfato quinasa) | NM_005793 | Hs.465558 |
| 92 | NPY1R | Receptor Y1 del neuropéptido Y | NM_000909 | Hs.519057 |
| 93 | NUP50 | Nucleoporina 50kDa | NM_007172 NM_153645 | Hs.475103 |
| 94 | PDGFRA | Polipéptido alfa del receptor del factor de crecimiento derivado de plaquetas | NM_006206 | Hs.74615 |
| 95 | PDHX | Componente X del complejo piruvato deshidrogenasa | NM_003477 | Hs.502315 |
| 96 | PDLIM1 | Dominio D1 de PDZ y LIM (elfin) | NM_020992 | Hs.368525 |
| 97 | PEX1 | Factor 1 de la biogénesis del peroxisoma | NM_000466 | Hs.164682 |
| 98 | PEX13 | Factor 13 de la biogénesis del peroxisoma | NM_002618 | Hs.567316 |
| 99 | PIB5PA | Fosfatidilinositol (4,5) bifosfato 5-fosfatasa, A | NM_014422 NM_001002837 | Hs.517549 |
| 100 | PICK1 | Proteína que interactúa con PRKCA 1 | NM_012407 NM_001039583 NM_001039584 | Hs.180871 |

(continuación)

| Número FI- | Gen marcador | Descripción de gen | Nº de acceso | UniGeneID |
|---|---|---|---|---|
| 101 | PLEC1 | Filamento intermedio de la proteína de unión a plectina 1, 500kDa | NM_201380<br>NM_201384<br>NM_000445 | Hs.434248 |
| | | | NM_201379<br>NM_201383<br>NM_201382<br>NM_201381<br>NM_201378 | |
| 102 | POLE2 | Polimerasa (dirigida a ADN), épsilon 2 (subunidad p59) | NM_002692 | Hs.162777 |
| 103 | POLE3 | Polimerasa (dirigida a ADN), épsilon 3 (subunidad p17) | NM_017443 | Hs.108112 |
| 104 | PPIF | Peptidilprolil isomerasa F (ciclofilina F) | NM_005729 | Hs.381072 |
| 105 | PPP2R5A | Subunidad reguladora B' de la proteína fosfatasa 2, isoforma alfa | NM_006243 | Hs.497684 |
| 106 | PSCD2 | Dominios 2 de la Sec y de cola enrollada de homología con la pleckstrina (citohesina-2) | NM_017457<br>NM_004228 | Hs.144011 |
| 107 | PSMA5 | Subunidad de tipo alfa 5 del proteosoma (proteosoma, macropain), 5 | NM_002790 | Hs.485246 |
| 108 | PTPN12 | No receptor de tipo 12 de la proteína tirosina fosfatasa | NM_002835 | Hs.61812 |
| 109 | PTPN3 | No receptor de tipo 3 de la proteína tirosina fosfatasa | NM_002829 | Hs.436429 |
| 110 | PTPRCAP | Proteína de tipo C de tipo receptor de la proteína tirosina fosfatasa | NM_005608 | Hs.155975 |
| 111 | QKI | Dominio KH de unión a ARN homólogo de Quaking (ratón) | NM_206855<br>NM_206854<br>NM_206853<br>NM_006775 | Hs.510324 |
| 112 | RASAL2 | Activador de la proteína RAS de tipo 2 | NM_170692<br>NM_004841 | Hs.656823 |
| 113 | RASSF7 | Familia 7 del dominio de asociación a Ras (RalGDS/AF-6) 7 | NM_003475 | Hs.72925 |
| 114 | RBM10 | Proteína 10 del motivo de unión a ARN | NM_005676<br>NM_152856 | Hs.401509 |
| 115 | RBM38 | Proteína 38 del motivo de unión a ARN | NM_017495<br>NM_183425 | Hs.236361 |
| 116 | RER1 | Homólogo de RER1 de retención en el retículo endoplasmático 1 (S. cerevisiae) | NM_007033 | Hs.525527 |
| 117 | RGL2 | Estimulador de tipo 2 de la disociación del nucleótido guanina | NM_004761 | Hs.509622 |
| 118 | RHOG | Miembro G de la familia del gen homólogo de Ras (rhoG) | NM_001665 | Hs.501728 |
| 119 | RNASE1 | Familia 1 de la ribonucleasa, ARNasa, 1 (pancreática) | NM_198235<br>NM_198234<br>NM_198232<br>NM_002933 | Hs.78224 |

(continuación)

| Número FI- | Gen marcador | Descripción de gen | Nº de acceso | UniGeneID |
|---|---|---|---|---|
| 120 | RTN4 | Reticulón 4 | NM_020532<br>NM_207521<br>NM_207520<br>NM_153828<br>NM_007008 | Hs.645283 |
| 121 | RYR2 | Rianodina receptor 2 (cardiaco) | NM_001035 | Hs.109514 |
| 122 | SCC-112 | Proteína SCC-112 | NM_015200 | Hs.331431 |
| 123 | SDS | Serina deshidratasa | NM_006843 | Hs.654416 |
| 124 | SF3B2 | Subunidad 2 del factor 3b de corte y empalme, 145kDa | NM_006842 | Hs.406423 |
| 125 | SH3PXD2A | Dominios 2A SH3 y PX | NM_014631 | Hs.594708 |
| 126 | SIX6 | Homólogo 6 de homeobox de seno ocular (Drosophila) | NM_007374 | Hs.194756 |
| 127 | SLC10A1 | Familia 10 del vehículo del soluto (sodio/familia del cotransportador de ácidos biliares), miembro 1 | NM_003049 | Hs.952 |
| 128 | SLC6A8 | Familia 6 dl vehículo del soluto (transportador del neurotransmisor creatina), miembro 8 | NM_005629 | Hs.540696 |
| 129 | SMG6 | Factor de desintegración del ARNm mediado sin sentido del homólogo de Smg-6 (C. elegans) | NM_017575 | Hs.448342 |
| 130 | SNRPB2 | Polipéptido B" de la ribonucleoproteína nuclear pequeña | NM_003092<br>NM_198220 | Hs.280378 |
| 131 | SOX11 | SRY (región Y de determinación del sexo) secuencia 11 | NM_003108 | Hs.432638 |
| 132 | SPI1 | Oncogén spi1 de integración provírica del virus formador del foco del bazo (SFFV) | NM_001080547<br>NM_003120 | Hs.502511 |
| 133 | SRGAP3 | Proteína 3 activadora de la Rho GTPasa SLIT-ROBO | NM_014850<br>NM_001033117 | Hs.654743 |
| 134 | STX12 | Sintaxina 12 | NM_177424 | Hs.523855 |
| 135 | SYK | Tirosina quinasa del bazo | NM_003177 | Hs.371720 |
| 136 | TAF4 | TAF4 ARN polimerasa II, factor asociado a la proteína de unión a la secuencia TATA (TBP),135kDa | NM_003185 | Hs.18857 |
| 137 | TCN2 | Transcobalamina II | NM_000355 | Hs.417948 |
| 138 | TGOLN2 | Proteína 2 de red Trans-golgi | NM_006464 | Hs.593382 |
| 139 | TIA1 | Proteína de unión al ARN asociada a gránulo citotóxico TIA1 | NM_022173<br>NM_022037 | Hs.516075 |
| 140 | TOMM40 | Homólogo 40 de translocasa de membrana mitocondrial externa (levadura) | NM_006114 | Hs.655909 |
| 141 | TXN2 | Tioredoxina 2 | NM_012473 | Hs.211929 |
| 142 | UGCG | UDP-glucosa ceramida glucosiltransferasa | NM_003358 | Hs.304249 |
| 143 | USP11 | Peptidasa 11 específica de ubiquitina | NM_004651 | Hs.171501 |

(continuación)

| Número FI- | Gen marcador | Descripción de gen | Nº de acceso | UniGeneID |
|---|---|---|---|---|
| 144 | VDR | Receptor de la vitamina D (1,25-dihidroxivitamina D3 ) | NM_001017535 NM_000376 | Hs.524368 |
| 145 | VEGFC | Factor C de crecimiento endotelial vascular | NM_005429 | Hs.435215 |
| 146 | YWHAQ | Proteína de activación de la tirosina 3-monooxigenasa/triptófano 5, polipéptido teta | NM_006826 | Hs.74405 |
| 147 | ZNF140 | Proteína 140 de dedo de cinc | NM_003440 | Hs.181552 |

Tabla 5: Conjunto PIV-1 a PIV-9 del marcador de PTC

| Número PIV- | Gen marcador | Descripción del gen | Nº de acceso. | UniGeneID |
|---|---|---|---|---|
| 1 | WAS | Síndrome de Wiskott-Aldrich (eczema-trombocitopenia) | BC012738 | Hs.2157 |
| 2 | LRP4 | Proteína 4 relacionada con el receptor de lipoproteína de baja densidad | BM802977 | Hs.4930 |
| 3 | TFF3 | factor 3 de Trébol (intestinal) | BC017859 | Hs.82961 |
| 4 | ST3GAL6 | ST3 beta-galactósido alfa-2,3-sialiltransferasa 6 | BC023312 | Hs.148716 |
| 5 | STK39 | Homólogo de serina treonina quinasa 39 (STE20/SPS1 , levadura) | BM455533 | Hs.276271 |
| 6 | DPP4 | Dipeptidil-peptidasa 4 (CD26, proteína 2 complejante de la adenosina desaminasa) | BC065265 | Hs.368912 |
| 7 | CHI31 | Quitinasa 3 de tipo 1 (glicoproteína-39 de cartílago | BC038354 | Hs.382202 |
| 8 | FABP4 | Proteína 4 de unión a ácido, adipocito | BC003672 | Hs.391561 |
| 9 | LAMB3 | Laminina, beta 3 | BC075838 | Hs.497636 |

Tabla 6: Conjunto PV-1 a PV-11 del marcador de PTC

| Número PV- | Gen marcador | Descripción del gen | Nº de acceso. | UniGeneID |
|---|---|---|---|---|
| 1 | GPR4 | Receptor 4 acoplado a proteína G | BC067535 | Hs.17170 |
| 2 | STAM2 | Molécula 2 adaptadora transductora de la señal (dominio SH3 y motivo ITAM) | BC028740 | Hs.17200 |
| 3 | QPCT | Glutaminil-peptido ciclotransferasa (glutaminil ciclasa) | BC047756 | Hs.79033 |
| 4 | CDK7 | Quinasa 7 dependiente de ciclina (homólogo de MO15 , Xenopus laevis, quinasa activadora de cdk) | BC000834 | Hs.184298 |
| 5 | SFTPD | Proteína D asociada a tensioactivo pulmonar | BC022318 | Hs.253495 |
| 6 | CYB5R1 | Reductasa 1 del Citocromo b5 | BC018732 | Hs.334832 |
| 7 | VWF | Factor de Von Willebrand | BI490763 | Hs.440848 |
| 8 | VWF | Factor de Von Willebrand | BQ888783 | Hs.440848 |
| 9 | PDHX | Componente X del complejo de la Piruvato deshidrogenasa | BC010389 | Hs.502315 |
| 10 | HOXA4 | Homeobox A4 | BM996071 | Hs.654466 |
| 11 | HOXA4 | Homeobox A4 | BI521357 | Hs.654466 |

Se pueden utilizar los oligonucleótidos inventivos para detectar el cáncer de tiroides o células tumorales e incluso para distinguir nódulos tiroideos benignos del carcinoma de tiroides folicular maligno (FTC) y del carcinoma papilar de tiroides (PTC). En las realizaciones preferidas, los oligonucleótidos específicos de al menos 3 marcadores tumorales seleccionados entre los marcadores tumorales PI-1 a PI-33 son específicos del carcinoma papilar de tiroides (PTC) y el cáncer de tiroides diagnosticado se pueden caracterizar como PTC.

Se da a conocer un conjunto que comprende oligonucleótidos específicos de al menos 3 marcadores tumorales seleccionados entre los marcadores tumorales FI-1 a Fi-147. Estos marcadores son específicos del carcinoma folicular de tiroides (FTC) y el cáncer de tiroides diagnosticado se puede caracterizar como FTC.

También se da a conocer un oligonucleótido específico para el marcador tumoral SERPINA1 (inhibidor de la serina (o cisteína) proteasa, clado A (antiproteinasa alfa-1, antitripsina), miembro 1; NM_000295, NM_001002236, NM_001002235), que es un marcador muy potente para el PTC. Este marcador como miembro único puede distinguir el PTC de la forma benigna de la dolencia.

Se pueden usar preferentemente al menos 5 o al menos 10, preferentemente al menos 15, de forma más preferida al menos 20, de forma particularmente preferida al menos 25, lo más preferido al menos 30, oligonucleótidos específicos para los marcadores tumorales de las tablas 1 a 6. Se pueden seleccionar oligonucleótidos específicos para cualquiera al menos 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 33, 35, 40, 45, 50, 55, 60, 64, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 145, 147, 150, 160, 170, 180, 190 o 200 de los anteriores marcadores tumorales desde PI-1 a PI-33 opcionalmente de forma adicional con los marcadores tumorales PII-1 a PII-64, PIII-1 a PIII-70, FI-1 a FI-147, PIV-1 a PIV-9, preferentemente PIV-4 o PIV-5, y PV-1 a PV-11, preferentemente PV-1, PV-2 y PV-4 a PV-11, en particular desde uno cualquiera de PI-1, PI-2, PI-3, PI-4, PI-5, PI-6, PI-7, PI-8, PI-9, PI-10, PI-11, PI-12, PI-13, PI-14, PI-15, PI-16, PI-17, PI-18, PI-19, PI-20, PI-21, PI-22, PI-23, PI-24, PI-25, PI-26, PI-27, PI-28, PI-29, PI-30, PI-31, PI-32, PI-33, opcionalmente de forma adicional con PII-1, PII-2, PII-3, PII-4, PII-5, PII-6, PII-7, PII-8, PII-9, PII-10, PII-11, PII-12, PII-13, PII-14, PII-15, PII-16, PII-17, PII-18, PII-19, PII-20, PII-21, PII-22, PII-23, PII-24, PII-25, PII-26, PII-27, PII-28, PII-29, PII-30, PII-31, PII-32, PII-33, PII-34, PII-35, PII-36, PII-37, PII-38, PII-39, PII-40, PII-41, PII-42, PII-43, PII-44, PII-45, PII-46, PII-47, PII-48, PII-49, PII-50, PII-51, PII-52, PII-53, PII-54, PII-55, PII-56, PII-57, PII-58, PII-59, PII-60, PII-61, PII-62, PII-63, PII-64, PIII-1, PIII-2, PIII-3, PIII-4, PIII-5, PIII-6, PIII-7, PIII-8, PIII-9, PIII-10, PIII-11, PIII-12, PIII-13, PIII-14, PIII-15, PIII-16, PIII-17, PIII-18, PIII-19, PIII-20, PIII-21, PIII-22, PIII-23, PIII-24, PIII-25, PIII-26, PIII-27, PIII-28, PIII-29, PIII-30, PIII-31, PIII-32, PIII-33, PIII-34, PIII-35, PIII-36, PIII-37, PIII-38, PIII-39, PIII-40, PIII-41, PIII-42, PIII-43, PIII-44, PIII-45, PIII-46, PIII-47, PIII-48, PIII-49, PIII-50, PIII-51, PIII-52, PIII-53, PIII-54, PIII-55, PIII-56, PIII-57, PIII-58, PIII-59, PIII-60, PIII-61, PIII-62, PIII-63, PIII-64, PIII-65, PIII-66, PIII-67, PIII-68, PIII-69, PIII-70, FI-1, FI-2, FI-3, FI-4, FI-5, FI-6, FI-7, FI-8, FI-9, FI-10, FI-11, FI-12, FI-13, FI-14, FI-15, FI-16, FI-17, FI-18, FI-19, FI-20, FI-21, FI-22, FI-23, FI-24, FI-25, FI-26, FI-27, FI-28, FI-29, FI-30, FI-31, FI-32, FI-33, FI-34, FI-35, FI-36, FI-37, FI-38, FI-39, FI-40, FI-41, FI-42, FI-43, FI-44, FI-45, FI-46, FI-47, FI-48, FI-49, FI-50, FI-51, FI-52, FI-53, FI-54, FI-55, FI-56, FI-57, FI-58, FI-59, FI-60, FI-61, FI-62, FI-63, FI-64, FI-65, FI-66, FI-67, FI-68, FI-69, FI-70, FI-71, FI-72, FI-73, FI-74, FI-75, FI-76, FI-77, FI-78, FI-79, FI-80, FI-81, FI-82, FI-83, FI-84, FI-85, FI-86, FI-87, FI-88, FI-89, FI-90, FI-91, FI-92, FI-93, FI-94, FI-95, FI-96, FI-97, FI-98, FI-99, FI-100, FI-101, FI-102, FI-103, FI-104, FI-105, FI-106, FI-107, FI-108, FI-109, FI-110, FI-111, 112, FI-113, FI-114, FI-115, FI-116, FI-117, FI-118, FI-119, FI-120, FI-121, FI-122, FI-123, FI-124, FI-125, FI-126, FI-127, FI-128, FI-129, FI-130, FI-131, FI-132, FI-133, FI-134, FI-135, FI-136, FI-137, FI-138, FI-139, FI-140, FI-141, FI-142, FI-143, FI-144, FI-145, FI-146, FI-147, PIV-1, PIV-2, PIV-3, PIV-4, PIV-5, PIV-6, PIV-7, PIV-8, PIV-9, PV-1, PV-2, PV-3, PV-4, PV-5, PV-6, PV-7, PV-8, PV-9, PV-10, PV-11. Preferentemente, los oligonucleótidos son específicos del subconjunto completo seleccionado entre PI, opcionalmente de forma adicional entre PII, PIII, PIV, PV o FI. Sin embargo, es también posible repicar un pequeño número de estos subconjuntos o conjuntos combinados tal como se ha mencionado anteriormente y se ha definido en las reivindicaciones debido a que también se puede llevar a cabo una distinción entre estados benignos y malignos o el diagnóstico de cáncer también se puede conseguir con una certeza aceptable. Por ejemplo, en una realización preferida, el procedimiento inventivo comprende además utilizar al menos 5 (o cualquiera de los números anteriormente mencionados) de oligonucleótidos específicos para los marcadores tumorales seleccionados entre FI-1 a FI-147. Las Figs. 4 y 5 muestran dichas probabilidades diagnósticas de clasificación para PTC y FTC. Por ejemplo, un conjunto específico de cualquier número de marcadores de la tabla 2 (subconjunto PII) específico para 5 marcadores tiene solo un margen de error del 4%, es decir, el 96% de todos los casos se clasificaría de forma correcta. Se consiguió un valor de error de un 1% (99% de certeza) con al menos 20 miembros. En el caso de los marcadores específicos de FTC, se consiguió un valor estable de errores del 8% con al menos 11 marcadores diferentes seleccionados a partir del subconjunto FI.

Los oligonucleótidos de acuerdo con la invención son moléculas adecuadas para el reconocimiento específico de los marcadores inventivos. Dicho reconocimiento molecular puede ser en el nivel del nucleótido. Los oligonucleótidos o cebadores son específicos para los ácidos nucleicos marcadores tumorales. De acuerdo con la invención no es esencial qué porción de la secuencia de los ácidos es la reconocida por los oligonucleótidos siempre que esté facilitado el reconocimiento molecular. Son adecuados los oligonucleótidos ya conocidos en la técnica, especialmente los dados a conocer en las referencias citadas en el presente documento.

En una realización preferida, los oligonucleótidos se inmovilizan sobre un soporte sólido, preferentemente en la forma de una micromatriz o nanomatriz. El término “micromatriz”, igualmente “nanomatriz”, se utiliza para describir una matriz de una disposición microscópica (nanomatriz para una matriz a escala nanométrica) o se refiere a un

vehículo que comprende la mencionada matriz. Ambas definiciones no se contradicen entre sí y son aplicables en el sentido de la presente invención. Preferentemente, los oligonucleótidos se proporcionan en un chip mientras sobre el cual, los oligonucleótidos se pueden inmovilizar. Los chips pueden ser cualquier material adecuado para la inmovilización de biomoléculas tales como los oligonucleótidos, que incluyen vidrio, vidrio modificado (aldehído modificado) o chips metálicos.

De acuerdo con la presente invención, se proporciona un conjunto de los oligonucleótidos anteriormente mencionados para uso específico para el diagnóstico tumoral. Sin embargo, es también posible proporcionar conjuntos más grandes que incluyen oligonucleótidos adicionales para otros fines, en particular en una configuración de micromatriz, en la que es posible inmovilizar una multitud de oligonucleótidos. Sin embargo, se prefiere proporcionar un conjunto económico que incluye una cantidad limitada de oligonucleótidos para un único fin.

Por tanto, en una realización preferida, el conjunto comprende al menos 10%, al menos 15%, al menos 20%, al menos 25%, al menos 30%, al menos 35%, al menos 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, especialmente preferido al menos 100%, de los oligonucleótidos totales que se unen al analito del conjunto son oligonucleótidos, que son específicos de los marcadores tumorales seleccionados entre el grupo de PI-1 a PI-33, PII-l a PII-64, PIII-1 a PIII-70, FI-1 a FI-147, PIV-1 a PIV-9, y PV-1 a PV-11 (todos los marcadores dados a conocer en las tablas 1 a 6, anteriores) o entre al menos uno de los grupos de uno cualquiera de PI-1 a PI-33, PII-1 a PII-64, PIII-1 a PIII-70, FI-1 a FI-147, PIV-1 a PIV-9, PV-1 a PV-11 o cualquiera de sus combinaciones. Dichas combinaciones preferidas son, por ejemplo, todos los marcadores de los grupos PI-1 a PI-33, o adicionalmente con PII-1 a PII-64, PIII-1 a PIII-70, PIV-1 a PIV-9, y PV-1 a PV-11, siendo especialmente adecuados para el diagnóstico de PTC. Tal como se usa en el presente documento “restos de unión a analito” se refiere a todos los oligonucleótidos que se pueden utilizar para detectar específicamente un marcador, en particular un gen marcador o producto génico, que incluye ARNm. Los genes son preferentemente genes de un mamífero, en particular un ser humano. Los oligonucleótidos se incluyen en este término genérico de cualesquiera “restos de unión a analito” que pueden tener múltiples dianas diagnósticas. Por ejemplo, en la realización de una micromatriz, la matriz comprende al menos un 10% de oligonucleótidos específicos para los marcadores inventivos. Debido a que – de acuerdo con la tecnología actual – los medios de detección para los genes en un chip (moléculas de ácido nucleico) tales como ADN-EST o ADN-EST complementarios, respectivamente) permiten un diseño de matriz más fácil y más robusto, se prefieren los chips de genes que utilizan moléculas de ADN (para la detección del ARNm expresado en la muestra) de la realización de la presente invención. Dichos chips génicos permiten también la detección de un gran número de productos génicos, mientras que la detección de un gran número de proteínas que utilizan chips de proteínas (por ejemplo, chips de anticuerpos) es más difícil. Se da ha descrito la detección de proteínas, llevada a cabo usualmente utilizando técnicas ELISA (es decir, una placa, perla o chip de microvaloración basado en ELISA) como un chip de proteína. Un chip de proteína puede comprender medios adecuados para unirse específicamente a los productos génicos del gen de la lista de acuerdo con las tablas 1 a 6, por ejemplo, moléculas de afinidad tales como anticuerpos monoclonales o policlonales o lectinas.

En una realización adicional el conjunto comprende hasta 50000 oligonucleótidos de unión a analito, preferentemente hasta 40000, hasta 35000, hasta 25000, hasta 20000, hasta 15000, hasta 10000, hasta 7500, hasta 5000, hasta 3000, hasta 2000, hasta 1000, hasta 750, hasta 500, hasta 400, hasta 300, o incluso más preferido hasta 200 oligonucleótidos de unión a analito de cualquier tipo, tales como oligonucleótidos específicos para cualquier gen o producto génico.

En un aspecto adicional, la presente invención se refiere a un procedimiento para la detección de uno o más marcadores del cáncer de tiroides en una muestra que comprende utilizar el conjunto inventivo y detectar la presencia o medir la cantidad de incidencia del ARNm de al menos 3 marcadores tumorales seleccionados entre los marcadores tumorales PI-1 a PI-33 en la muestra tal como se define en las reivindicaciones. La incidencia del modelo de los marcadores seleccionados puede identificar de forma específica la presencia de estos marcadores que pueden ser relevantes para el diagnóstico del cáncer de tiroides o como una referencia de las muestras sanas, o sencillamente, una investigación genética de sujetos.

Preferentemente, la muestra comprende preferentemente células, células de mamíferos, siendo particularmente preferidas las células humanas, que pueden proporcionarse a partir de una biopsia o fluido corporal. En particular, la presencia o cantidad de marcadores tumorales se detecta o mide en estas células después de, por ejemplo, la desintegración celular.

El procedimiento puede comprender una detección o la medida mediante análisis de la expresión del ARN, preferentemente mediante la PCR en micromatriz o cuantitativa, preferentemente mediante detección de la micromatriz en tejido, detección del ARNm en micromatriz, ensayos multiplexados, o análisis del ADN, hibridación genómica comparativa, matrices (CGH) o análisis de polimorfismo de nucleótido único (SNP). Estos procedimientos son conocidos en la técnica y se pueden utilizar fácilmente para el procedimiento de la presente invención, como ejemplos del vasto campo de análisis de los marcadores genéticos.

En otro aspecto, la presente invención proporciona un procedimiento para el diagnóstico de cáncer seleccionado entre cáncer papilar de tiroides o cáncer folicular de tiroides en un paciente, que comprende proporcionar una

muestra de células del paciente, detectar los marcadores tumorales midiendo las señales de los marcadores tumorales con el procedimiento de acuerdo con la presente invención, comparar los valores de la señal medidos de los marcadores tumorales con valores de los marcadores tumorales en muestras sanas y diagnosticar el cáncer, si acaso, más del 50%, preferentemente más del 60%, de forma más preferida más del 70%, lo más preferido más del 80%, de los valores difieren en comparación con los valores de las muestras sanas en al menos la desviación estándar, preferentemente dos veces la desviación estándar, incluso de forma más preferida tres veces la desviación estándar, del procedimiento de medida. Las diferencias en la expresión genética entre las muestras de sujetos enfermos y sujetos sanos puede ser de cualquier tipo e incluye la regulación en exceso (por ejemplo, de oncogenes) o la regulación por defecto (por ejemplo, de genes supresores del tumor), es posible que en muestras sanas no se exprese un gen mientras que se produzca la expresión en muestras enfermas. Es también posible una vía aproximada diferente de que las muestras enfermas de un gen no se expresen mientras que se produce la expresión en muestras sanas.

Se puede diagnosticar también el cáncer si más del 50%, preferentemente más del 60%, de forma más preferida más del 70%, lo más preferido más del 80%, de los valores de la muestra difieren en comparación con los valores de las muestras sanas en al menos un factor 1,5, al menos un factor 2, al menos un factor 3 o al menos un factor 4. Normalmente, los productos de expresión del marcador tumoral están regulados por exceso o regulados por defecto en un factor de 2 a 6, pero también son posibles diferencias en un factor 60.

Se puede combinar el procedimiento de diagnóstico con la identificación de marcadores específicos de la enfermedad, como, por ejemplo, los proporcionados en las tabal 1 a 6, preferentemente los genes o los modelos de expresión génica, que comprende:

- proporcionar datos de la expresión génica sobre genes específicos de múltiples enfermedades potenciales de al menos dos conjuntos de datos de expresión diferentes,
- determinar los genes comunes de los conjuntos de datos,
- normalizar cada conjunto de datos de la expresión génica, preferentemente mediante normalización inferior o por cuantil
- combinar los conjuntos de bases de la expresión génica con un conjunto de datos combinados, y normalizar preferentemente el conjunto de datos combinados e integrar el conjunto de datos combinados.
- determinación de los genes del conjunto de datos combinados determinando su centroide comprimido, que incluye la determinación de un valor de error validado en cruzado de asignación de los genes de la enfermedad y minimizar el valor del error reduciendo el número de miembros del combinado, preferentemente el conjunto de datos normalizados.

en el que los genes del conjunto de datos reducidos son los marcadores específicos de la enfermedad. La validación cruzada puede ser, por ejemplo el procedimiento de omisión. Preferentemente, la etapa de determinación (la etapa de clasificación) comprende la determinación de un umbral maximizado de la diferencia del valor de expresión normalizado para cada gen al valor del centroide mediante la validación en cruzado. A continuación, los genes con los valores de expresión normalizados menores que el umbral se eliminan del conjunto reducido (o comprimido) y los genes con valores mayores que los del umbral al centroide son específicos de la enfermedad. La clasificación mediante los procedimientos de centroides comprimidos se da a conocer, por ejemplo en Tibshirani y col. (PNAS USA 99(10): 105-114 (2004)), Shen y col. (Bioinformatics 22(22) (2006): 2635-42) y en Wang y col. (Bioinformatics 23(8) (2007): 972-9), cuyas divulgaciones se incorporan en el presente documento por referencia.

La etapa de determinación se puede repetir múltiples veces omitiendo los marcadores resultantes de cada etapa anterior. El procedimiento del centroide comprimido más cercano dará como resultado un nuevo conjunto de resultados de marcadores adicionales que son específicos de la enfermedad. Preferentemente, la etapa de determinación se repitió 2, 3, 4, 5, 6, 7, 8, 9, 10 o más veces. Dependiendo del tamaño del conjunto de datos combinados se proporcionarán marcadores específicos adicionales. Preferentemente, se llevó a cabo una validación cruzada sobre cada resultado. La determinación se puede repetir hasta que la validación cruzada indica un valor de error de, por ejemplo, por debajo de 50%, 60%, 70% u 80%. A valores inferiores, se puede esperar que hayan de identificarse todos los marcadores.

Los datos de expresión génica iniciales son perfiles de expresión brutos, por ejemplo, cada uno obtenido de un análisis de micromatriz multigenética. Se espera que la mayoría de genes medidos no estén implicados en la enfermedad y el procedimiento inventivo es capaz de identificar genes marcadores característicos entre al menos dos, preferentemente al menos tres, al menos cuatro, al menos cinco, al menos seis, al menos siete, o al menos ocho conjuntos de datos de expresión. Por tanto, los datos de expresión de los conjuntos de datos iniciales comprenden preferentemente datos de al menos dos conjuntos de datos de micromatrices diferentes, en particular con los sesgos específicos del estudio o plataforma. Dichos sesgos se pueden producir utilizando solamente un conjunto específico durante la medida de los datos de expresión, por ejemplo, una micromatriz, que puede diferir significativamente de las configuraciones de otros conjuntos de datos. La presente invención tiene la ventaja de que durante la combinación de dichos conjuntos, se superan los problemas de dichos sesgos de la medida. Además, los datos de la expresión génica obtenidos (iniciales) son brutos, datos de la expresión génica no procesados, es decir, no se llevó a cabo el refino o la conversión de los datos antes del procedimiento inventivo.

La etapa de combinar e integrar el conjunto de datos combinados eliminó los sesgos específicos del estudio. En realizaciones preferidas, esta etapa se llevó a cabo mediante una combinación por etapas de dos conjuntos de datos de expresión génica por etapa y la integración del conjunto de datos combinados, preferentemente mediante DWD (Discriminación Ponderada de la Distancia). Por ejemplo, en el caso de 3 conjuntos de datos en primer lugar se combina el conjunto 1 con el conjunto 2 y el conjunto combinado 1+2 se combina con el conjunto 3. La integración puede, por ejemplo, incluir el cálculo del vector normal del conjunto de datos combinados y posteriormente un hiperplano que separa las agrupaciones (por ejemplo, de los conjuntos de datos iniciales) de los valores de datos del conjunto de datos y sustrae los promedios del conjunto de datos como en el procedimiento DWD. En principio, se puede utilizar para el procedimiento inventivo cualquier procedimiento de integración de datos que elimine los sesgos.

Preferentemente, el al menos uno, preferentemente dos, tres, cuatro, cinco, seis, siete u ocho conjuntos de datos de expresión obtenidos comprende los datos de al menos 10, preferentemente al menos 20, de forma más preferida al menos 30, incluso de forma más preferida, al menos 50, al menos 70, al menos 100, al menos 120, al menos 140, al menos 160 o incluso al menos 200 genes diferentes. El procedimiento inventivo es particularmente adecuado para filtrar a través de grandes conjuntos de datos e identificar los marcadores característicos en los anteriores. El conjunto obtenido de estos marcadores se denomina “clasificador”.

Este procedimiento de identificar marcadores específicos del cáncer y de esta manera, pueden también utilizarse en el anterior procedimiento de diagnóstico del cáncer oligonucleótidos específicos del cáncer, es decir, se identifican los marcadores que corresponden al conjunto de restos utilizados para el procedimiento diagnóstico (denominados también “clasificados”) de acuerdo con el anterior procedimiento que incluye el refino y el establecimiento de los valores del centroide de los valores medidos de los conjuntos de datos iniciales. Este modelo puede utilizarse a continuación para diagnosticar el cáncer si los valores de la muestra del paciente son más cercanos al valor del centroide agrupado de los marcadores tumorales. De acuerdo con esto, se proporciona un procedimiento para el diagnóstico del cáncer seleccionado entre cáncer papilar de tiroides o cáncer folicular de tiroides en un paciente, que comprende proporcionar una muestra de células, procedentes del paciente, detectar uno o más marcadores tumorales midiendo las señales de los marcadores tumorales con un conjunto de acuerdo con la presente invención, comparar los valores de la señal medidos de los marcadores tumorales con valores de los marcadores tumorales en muestras de cáncer mediante el procedimiento de identificación mencionado anteriormente y diagnosticar el cáncer si el centroide comprimido más cercano de los valores de la muestra del paciente para al menos un 50%, preferentemente al menos un 60%, de forma más preferida al menos un 70% o incluso al menos un 80%, lo más preferido un 90%, marcadores del conjunto está comprendido dentro de la desviación estándar, preferentemente dos veces la desviación estándar, incluso de forma más preferida tres veces la desviación estándar, del procedimiento de medida del centroide comprimido más cercano de los marcadores tumorales identificados con las muestras de cáncer.

La presente invención se ilustra adicionalmente mediante las siguientes figuras y los ejemplos sin restringirse de forma específica a los anteriores. Todas las referencias citadas en el presente documento se incorporan por referencia.

**Figuras**

Figura 1. Los dos primeros componentes principales antes y después de la integración mediante DWD. Los conjuntos de datos están codificados por color y las entidades tumorales están codificadas por letras de acuerdo con la leyenda.

Figura 2 Dendograma de los datos integrados mediante DWD en todos los genes. Los colores de las ramas del dendograma indican el conjunto de datos de la muestra correspondiente, el color de la etiqueta indica la entidad tumoral.

Figura 3. Discriminación entre carcinoma papilar y nódulos benignos mediante cuatro conjuntos de datos diferentes por un único gen (SERPINA 1)

La Figura 4 muestra una gráfica de la probabilidad promedio de error durante la clasificación de PCT de los conjuntos atraídos (clasificador) de los marcadores de la tabla 2

La Figura 5 muestra una gráfica de la probabilidad promedio de error durante la clasificación de FCT de los conjuntos atraídos (clasificador) de los marcadores de la tabla 4.

**Ejemplos**

**Ejemplo 1 – Conjuntos de datos**

Los conjuntos de datos se descargaron tanto de los sitios web como de los depósitos públicos (GEO, ArrayExpress). La Tabla 7 muestra un resumen de los conjuntos de datos utilizados en este estudio (He y col, PNAS USA 102(52): 19075-80 (2005); Huang y col. PNAS USA 98(26): 15044-49 (2001); Jarzab Cancer Res 65(4): 1587-97 (2005); Lacroix Am J Pathol 167(1): 223-231 (2005); J Clin Endocrinol Metab 90(5): 2512-21 (2005)). Aquí, se utilizaron tres categorías diferentes de tejidos no cancerosos, contralateral (c.lat) para tejido sano que rodea emparejado con una muestra tumoral, otra enfermedad (o.d) para tejido tiroideo operado para otra enfermedad y SN (bocio nodular) para nódulos tiroideos benignos. Para todos los análisis posteriores se combinaron estos como sanos.

Tabla 7. Datos de la micromatriz utilizados para el metaanálisis

| Publicados | | FTA | FTC | PTC | SN | o.d. | c.lat | Plataforma |
|---|---|---|---|---|---|---|---|---|
| He | PNAS 2005 | 0 | 0 | 9 | 0 | 0 | 9 | Affy U133plus |
| Huang | PNAS 2001 | 0 | 0 | 8 | 8 | 0 | 0 | Affy U133A |
| Jarzab | Cancer Res 2005 | 0 | 0 | 23 | 0 | 11 | 17 | Affy U133A |
| Lacroix | Am J Path 2005 | 4 | 8 | 0 | 11 0 | | 0 | Agilent Custom |
| Reyes | ¿Sin publicar? | 0 | 0 | 7 | 0 | 0 | 7 | Affy U133A |
| Weber | J Clin Endocr Metabol 2005 | 12 | 12 | 0 | 0 | 0 | 0 | Affy U95A |

**Ejemplo 2: Hallazgo del solapamiento génico**

La primera etapa en cualquier metaanálisis de datos de micromatriz es encontrar el conjunto de genes que están compartidos por todas las plataformas de la micromatriz utilizadas en el análisis. Tradicionalmente, se evalúa el solapamiento encontrando identificadores UniGene comunes. Esto, sin embargo no tiene en cuenta todas las posibles variaciones de corte y empalme en los genes bajo investigación. Por ejemplo, si un gen tenía 2 variantes de corte y empalme, una de las cuales se expresó diferencialmente en el experimento y la otra no y si una plataforma pudiera contener un oligo específico solo de la variante expresada diferencialmente y la otra plataforma solo un oligo de la otra variante, entonces, un emparejamiento basado en UniGene podría combinar sondas que miden diferentes cosas.

Para superar este problema, la solución adoptada aquí combina solo sondas que se anotan al mismo conjunto de identificadores RefSeq. Con este fin todos los RefSeq correspondientes se descargaron para cada sonda (conjunto), tanto mediante los paquetes de anotaciones del bioconductor (hgu133a, hgu95a y hgul33plus2; disponibles en la www.-bioconductor.org) o mediante una búsqueda con BLAST de las secuencias de la base de datos NCBI. A continuación, para cada sonda de RefSeq se clasificaron y se concatenaros. Esto es la representación más precisa de la entidad medida en la matriz. Se utilizó el valor de la mediana, si un conjunto de RefSeq estaba representado por múltiples sondas sobre la matriz, Estuvieron presentes en todas las matrices 5707 conjuntos diferentes de RefSeq.

**Ejemplo 3 Preprocesamiento e integración de los datos**

En primer lugar, cada conjunto de datos se corrigió a fondo y se normalizó por separado, tal como se recomendó para cada plataforma (normalización inferior para el color doble y por cuantil para los experimentos de color individuales) (Bolstad y col. Bioinformatics 19(2): 185-193 (2003); Smyth y col. Methods 31(4): 265-273 (2003)), a continuación se combinaron y se normalizaron colectivamente por el cuantil. A pesar de todo el Preprocesamiento, se ha mostrado que los datos generados en diferentes plataformas de micromatriz o en diferentes generaciones de la misma plataforma pueden no ser comparables debido a los sesgos específicos de las plataformas (Eszlinger y col. Clin Endocrinol Metab 91(5): 1934-1942 (2006)). Esto es también evidente a partir de los principales análisis de componentes de los datos combinados tal como se muestra en la fig. 1. Con el fin de corregir estos sesgos, se han desarrollado procedimientos para la integración de los datos de la micromatriz. Uno de estos procedimientos es la Discriminación Ponderada por la Distancia (DWD) que se describe en detalle en otra parte (Benito y col. Bioinformatics 20(1): 105-114 (2004)). De forma breve, los datos de los proyectos de DWD apuntan sobre el vector normal de una clase (conjunto de datos)- separando el hiperplano como se calcula mediante una máquina de vector de soporte modificada (SVM) y sustrae los promedios de la clase (conjuntos de datos). Por tanto, para un problema multiclase (más de dos conjuntos de datos a combinar), los conjuntos de datos necesitan combinarse secuencialmente. Para conjuntos de datos, esto conduce a 720 posibilidades diferentes para combinar, que no incluyen las soluciones estructuradas, por ejemplo en vez de (((1 + 2) + 3), considerar ((1 + 2) (3 + 4)). Los órdenes de combinación aplicados aquí se escogen con la idea general de que deben combinarse conjuntos de datos similares y más grandes en primer lugar y separarse más tarde de los mismos. Es también un trabajo reseñable que ayudar a una muestra de un conjunto de datos combinado con DWD cambiará el conjunto de datos completo añadiendo así mismo un nuevo número a un vector de números que cambiará su promedio.

Se ilustra en la figura 1 la integración de los datos mediante DWD, que muestra el efecto del procedimiento integración de los datos sobre los primeros dos componentes principales. En este análisis, DWD fue capaz de eliminar la separación entre los conjuntos de datos tal como se ha indicado mediante las representaciones gráficas de PC y mediante la mezcla de las ramas en el dendograma (véase la fig. 2). Sin embargo, incluso en el conjunto de datos integrados con DWD, los datos de Lacroix están separados todavía parcialmente de los otros datos, Con más probabilidad, esto es debido a la plataforma, los datos de Lacroix se una plataforma no de Affymetrix, La Figura 2 muestra los dendogramas de los respectivos conjuntos de bases integrados. También, no se observa que la integración de DWD obstaculice la discriminación entre las entidades tumorales (véase la tabla 8 a continuación)

**Ejemplo 4: Clasificación**

Para la selección de la sonda se seleccionó el procedimiento del centroide comprimido más cercano y la clasificación y la validación cruzada (Tibshirani y col. PNAS USA 99(10): 105-114 (2004)) (implementado en el paquete pamr del bioconductor). Se seleccionó este por diversos motivos, permite la clasificación multiclase y lleva a cabo la selección de las características, la clasificación y la validación cruzada de una sola vez. De forma breve, calcula varios clasificadores posibles diferentes utilizando diferentes umbrales de contracción (es decir, diferentes números de genes) y encuentra el mejor umbral para la validación cruzada. El clasificador puede repicarse con el número más pequeños de genes (umbral más grande). Si acaso más de un umbral da como resultado los mismos resultados de validación cruzada.

**Ejemplo 5: Carcinoma papilar de tiroides**

En primer lugar, como una medida de la cantidad de cada estudio, se tomó por separado cada conjunto de datos (ante integración mediante DWD) y una clasificación de pamr y se llevó a cabo la validación cruzada por omisión (loocv). Los resultados de la validación cruzada son casi perfectos clasificándose con fuerzas las muestras individuales. Sin embargo, con la excepción del clasificador procedente del conjunto de datos He, ninguno de estos clasificadores puede aplicarse a ninguno de los otros conjuntos de datos. Los resultados de la clasificación son raramente muy superiores a los esperados debidos al azar. Si acaso, sin embargo, se utilizan datos integrados mediante DWD (a continuación), los clasificadores ya se ajustan mucho mejor (véase la tabla 8).

Tabla 8: resultados de la clasificación cuando se aplican los clasificadores procedentes de un estudio en otro estudio. Antes de la integración de los datos (izquierda) y después de la integración de DWD (derecha)

| serie de ensayos | he | huang | jarzab | reyes | serie de ensayos | he | huang | jarzab | reyes |
|---|---|---|---|---|---|---|---|---|---|
| he | 1,00 | 1,00 | 0,98 | 1,0 | he | 1,00 | 1,00 | 0,96 | 1,00 |
| huang | 0,50 | 1,00 | 0,55 | 0,50 | huang | 0,50 | 1,00 | 0,90 | 0,71 |
| jarzab | 0,50 | 0,81 | 1,00 | 0,57 | jarzab | 0,89 | 1,00 | 1,00 | 1,00 |
| reyes | 0,78 | 0,50 | 0,92 | 1,00 | reyes | 0,89 | 0,88 | 0,90 | 1,00 |

A continuación se construyó un clasificador pamr – para el conjunto de datos completos integrado por DWD y se validó en una validación cruzada por omisión. Esta identificó un clasificador del gen uno (|), que clasifica el 99% de las muestras correctamente en loocv. Este gen discriminador es SERPINA 1. La Fig.3 muestra la discriminación de PTC frente a SN antes y después de DWD. Se podrían añadir hasta 422 genes al clasificador y dar como resultado un 99% de precisión (a partir de loovc), Si se elimina la sonda SERPINA 1 procedente del análisis, se puede construir de nuevo un clasificador (denominado por consiguiente clasificador) con un 99% de precisión en loocv, utilizando durante este tiempo una firma de 9 genes (véase la Tabla 3). Eliminar estos 9 genes da como resultado otro clasificador de 9 genes con un rendimiento similar (99% de precisión), y además un clasificador de 11 genes con un 99% de precisión. Dichos clasificadores adicionales se proporcionan, por ejemplo en las tablas 1,3, 5 y 6 (anteriores) para PTC

Sin embargo, se obtuvieron resultados similares haciendo el mismo análisis sobre los datos no integrados. Teniendo en cuenta los resultados de PCA (fig. 1), en los que fue obvio que la varianza explicada por los diferentes conjuntos de bases es mucho más grande que la varianza explicada por la entidad tumoral, se podría imaginar que el sesgo introducido por los conjuntos de datos puede ayudar (o impedir) la clasificación. Por tanto, se llevó a cabo un estudio de validación cruzada, en el que se capturó secuencialmente un estudio a partir del conjunto de datos, se construyó un clasificador a partir de las muestras restantes y se ensayó sobre el conjunto de datos eliminados. Sobre los datos integrados mediante DWD, la precisión de la predicción fue del 100, 100, 98 y 100% dejando He, Huang, Jarzab y Reyes procedentes del clasificador, respectivamente. Para los datos no integrados, los resultados fueron similares (100, 100, 94 y 100%)

Tabla 9: Genes en el clasificador 2 (tras excluir SERPINA 1)

| Símbolo | Título | Grupo | Acceso |
|---|---|---|---|
| WAS | Síndrome de Wiskott-Aldrich (eczema-trombocitopenia | Hs. 2157 | BC012738 |
| LRP4 | Proteína 4 relacionada con el receptor de lipoproteína de baja densidad | Hs. 4930 | BM802977 |
| TFF3 | Factor 3 Trébol (intestinal) | Hs. 82961 | BC017859 |
| ST3GAL6 | ST3 beta-galactósido alfa-2,3-sialiltransferasa 6 | Hs.148716 | BC023312 |
| STK39 | Serina treonina quinasa 39 (homólogo STE20/SPS1 , levadura) | Hs.276271 | BM455533 |
| DPP4 | Dipeptidil-peptidasa 4 (CD26, proteína 2 complejante de la adenosina desaminasa) | Hs.368912 | BC065265 |
| CHI3L1 | Quitinasa 3 de tipo I (glicoproteína 39 de cartílago) | Hs.382202 | BC038354 |
| FABP4 | Proteína 4 de unión a ácido graso, adipocito | Hs.391561 | BC003672 |
| LAMB3 | Laminina, beta 3 | Hs.497636 | BC075838 |

**Ejemplo 6: Carcinoma folicular**

Se llevó a cabo también un análisis similar para los datos de FTC, pero se impidió la validación cruzada, debido a la muy limitada disponibilidad de datos. De nuevo, se construyó un clasificador para cada conjunto de datos (Lacroix y Weber). Consiguieron una precisión de loocv del 96% (Weber) y del 100% (Lacroix) sobre 25 y 3997 genes, El número de genes en los datos de Lacroix sugiere ya un ajuste por exceso, que se confirmó mediante la clasificación cruzada con los otros conjuntos de datos (25 y 35% de precisión, respectivamente). También, el solapamiento del gen entre los dos clasificadores es bajo (entre 0 y 10% dependiendo del umbral). Si acaso, sin embargo, los 2 conjuntos de datos se combinaron utilizando DWD, se podría construir un clasificador de 147 genes (tabla 4 anterior) que fue capaz de identificar correctamente las muestras (con un 92% de precisión).

**Ejemplo 7. Discusión**

La presente invención representa la cohorte más grande de los datos de la micromatriz del carcinoma de tiroides analizados hasta la fecha. Hace uso del novedoso procedimiento combinatorio que utiliza los algoritmos últimos para la integración y la clasificación de los datos de la micromatriz. Sin embargo, el metaanálisis de los datos de la micromatriz supone además un estímulo, debido principalmente a que las investigaciones individuales de la micromatriz son el objetivo al menos en cuestiones parcialmente diferentes y utilizan por tanto diferentes diseños experimentales. Además, el número de datos de la micromatriz del tumor de tiroides disponibles es todavía comparablemente bajo (en comparación, por ejemplo, con el cáncer de mama). Por tanto, cuando se llevan a cabo metaanálisis, el investigador se ve forzado a usar todos los datos disponibles, incluso si las cohortes de pacientes representan una población más bien heterogénea y potencialmente sesgada. De forma más específica, es difícil obtener una colección homogénea de material de control (procedente de pacientes sanos). Estos se toman normalmente de pacientes que han sido operados de otras enfermedades de tiroides, lo que a la vez produce muy probablemente un cambio en la expresión génica tal como se midió en las micromatrices. La generación de cohortes de pacientes homogéneas está impedida adicionalmente por la disponibilidad limitada de datos del paciente, edad, género, antecedentes genéticos, etc.

Cuando se lleva a cabo el metaanálisis de datos de la micromatriz, muchos investigadores han basado sus enfoques en comparar listas de genes procedentes de estudios no publicados (Griffith y col., citados anteriormente). Esto es muy útil ya que el investigador puede incluir todos los estudios en el análisis y no está limitado a los estudios en los que están disponibles datos brutos. Sin embargo, los estudios siguen generalmente estrategias de análisis muy diferentes, algunas más rigurosas que otras. No está bajo el control del metaanálisis como los autores han llegado a estas listas de genes. Por tanto, estos análisis pueden estar sesgados.

Con respecto a la integración de datos, de acuerdo con el papel del DWD original, el DWD se lleva a cabo mejor cuando están presentes al menos 25-30 muestras. Se llevó a cabo todavía el DWD comparablemente bien para eliminar los sesgos de la plataforma (véase la Tabla 8)

DWD mejoró mucho los resultados de PCA (figura 1), el agrupamiento jerárquico (figura 2) y la precisión de la clasificación cuando se aplica a un clasificador de un estudio a otro estudio (Tabla 9). A esta luz fue sorprendente observar que los datos no integrados se llevaron a cabo igualmente bien en el estudio de validación cruzada en comparación con los datos integrados mediante DWD. Una explicación para esto es que cualquier sesgo específico del estudio será menos importante que la mayoría de estudios que se están evaluando. Dado que los sesgos del estudio afectan a algunos genes más que a otros, los genes más afectados será menos probable que sobrevivan al umbral de pamr debido a la varianza introducida por el sesgo del estudio. Sin embargo, tal como se muestra anteriormente, existe una gran abundancia de genes que discriminan PTC y nódulos benignos. Siempre que uno (o unos pocos) de aquellos genes no se vean afectados por el sesgo del estudio, sobrevivirán al umbral y la discriminación entre entidades tumorales será todavía posible.

Existe una aparente discrepancia cuando el investigador observa la fig. 3: Antes de DWD, las muestras de PCT tienen una mayor expresión de SERPINA1 mientras que después de DWD es por otra parte redondo. Sin embargo, tal como se ha señalado en la sección de Materiales y Procedimientos, DWD sustrae los promedios de clase de cada muestra. Esto significa sencillamente que antes de DWD el sesgo del estudio par SERPINA 1 es mayor que la diferencia en la expresión entre las clases de tumor. Esto explica también, por qué en los datos no integrados SERPINA1 no es un clasificador que trabaja bien.

Un Metaanálisis y una Metarrevisión recientes de Griffith y col., (citados anteriormente) han resumido los genes con un diagnóstico potencial en el contexto de la enfermedad del tiroides. Publicaron listas de genes que aparecían en más de un estudio de alto rendimiento (Microarray, SAGE) que analizaba la enfermedad del tiroides y aplicaba un sistema de clasificación. En su análisis SERPINA 1 puntuó el tercero más alto, y TFF3, que es parte del clasificador 2 (cuando se omite SERPINA 1), puntuó segundo. Cuatro de 9 genes procedentes del clasificador 2 aparecieron en la lista de Griffith y col. (LRP4, TFF3, DPP4 y FABP4).

La mayoría de estas listas se generaron a partir del análisis de la micromatriz. Sin embargo, incluso cuando se comparaban los genes en los clasificadores con las listas de genes generadas con tecnologías independientes, existe un sustancial solapamiento en la generación de bibliotecas de ADNc similares. SERPINA1 aparece en sus listas así como cuatro de nueve genes procedentes del clasificador 2 (TFF3, DPP4, CHI3L1 y LAMB3).

Para el caso de la enfermedad folicular del tiroides, la construcción de un clasificador robusto es mucho más difícil. Esto es principalmente debido a la limitada disponibilidad de datos. También, los dos conjuntos de datos fueron muy diferentes en términos de las plataformas usadas; mientras que los otros conjuntos de datos se generaron en las matrices GeneChips de Affymetrix de diferentes generaciones, los datos de Lacroix se generaron en una plataforma Agilent. Sin embargo, el clasificador (conjunto) de la tabla 4 fue capaz de identificar la mayoría de muestras correctamente en loocv.

El poder de la solución del metaanálisis adoptada aquí se demostró por un 99% de precisión de loocv (97,9% de precisión promedio ponderada en el estudio de validación cruzada) para l distinción entre el carcinoma papilar de tiroides y los nódulos benignos. Esto se ha conseguido ampliamente en los conjuntos de datos más grandes y más diversos (99 muestras procedentes de 4 estudios diferentes).

Una muestra se clasificó erróneamente, y aunque no es posible cartografiar correctamente las muestras procedentes de este análisis a los análisis originales, la muestra clasificada incorrectamente es procedente del mismo grupo (PTC, grupo de validación), como la muestra que se clasificó erróneamente en el análisis original. De acuerdo con Jarzab y col., la muestra estuvo muy alejada debido a que contenía solo ≈ 20% de células tumorales.

**REIVINDICACIONES**

1. Procedimiento de detección de los marcadores del cáncer de tiroides en una muestra que comprende detectar la presencia o medir la cantidad de la ocurrencia del ARNm de al menos 3 marcadores tumorales seleccionados entre los marcadores tumorales PI-1 a PI-33 en la muestra utilizando oligonucleótidos específicos para ácidos nucleicos marcadores tumorales.

2. El procedimiento de la reivindicación 1, **caracterizado porque** la muestra comprende células, preferentemente células de mamífero, de forma particularmente preferida, células humanas.

3. El procedimiento de acuerdo con la reivindicación 1 o 2, **caracterizado porque** la detección o la medida se lleva a cabo mediante el análisis de la expresión del ARN, preferentemente mediante micromatriz o PCR cuantitativa, preferentemente por detección de tejido mediante micromatriz, detección de ARNm mediante micromatriz, ensayos multiplexados, o análisis del ADN, matrices de hibridación genómica comparativa (CGH) o análisis de polimorfismo de nucleótido único (SNP).

4. El procedimiento de acuerdo con una cualquiera de las reivindicaciones 1 a 3, **caracterizado porque** los oligonucleótidos son inmovilizados sobre un soporte sólido.

5. El procedimiento de la reivindicación 4, **caracterizado porque** los oligonucleótidos son inmovilizados en la forma de una micromatriz

6. Un procedimiento in vitro de diagnóstico del cáncer seleccionado entre cáncer papilar de tiroides o cáncer folicular de tiroides en un paciente, que comprende proporcionar una muestra de células procedente de un paciente, detectar los marcadores tumorales de acuerdo con un procedimiento de una cualquiera de las reivindicaciones 1 a 5, comparar los valores de la señal medidos de los marcadores tumorales con los valores de los marcadores tumorales en muestras sanas y diagnosticar el cáncer si (a) más de un 50%, preferentemente más de un 60%, de forma más preferida más de un 70%, lo más preferido más de un 80%, de los valores difieren en comparación con los valores de las muestras sanas en al menos la desviación estándar, preferentemente dos veces la desviación estándar, incluso más preferentemente tres veces la desviación estándar del procedimiento de medida y/o (b) más de un 50%, preferentemente más de un 60%, de forma más preferida más de un 70%, lo más preferido más de un 80%, de los valores de la muestra difieren en comparación con los valores de las muestras sanas en al menos un factor de 1,5.

7. Procedimiento in vitro para el diagnóstico del cáncer seleccionado entre cáncer papilar de tiroides o cáncer folicular de tiroides en un paciente, que comprende proporcionar una muestra de células, procedentes del paciente, detectar los marcadores tumorales de acuerdo con un procedimiento de una cualquiera de las reivindicaciones 1 a 5, comparar los valores de la señal medidos de los marcadores tumorales con los valores de los marcadores tumorales en muestras de cáncer mediante un procedimiento de identificación que comprende

- proporcionar datos de la expresión génica, preferentemente brutos, datos de expresión génica no procesados, sobre múltiples genes potenciales específicos de la enfermedad de al menos dos conjuntos de datos de la expresión diferentes,
- determinar los genes comunes de los conjuntos de datos,
- normalizar cada conjunto de datos de la expresión génica, preferentemente mediante normalización inferior o por cuantil,
- combinar los conjuntos de datos de la expresión génica en un conjunto de datos combinados, y normalizar preferentemente el conjunto de datos combinados, e integrar el conjunto de datos combinados,
- determinación de los genes del conjunto de datos combinados en el centroide comprimido más cercano, que incluye la determinación del valor del error validado en cruzado de asignar los genes a la enfermedad y minimizar el valor del error reduciendo el número de miembros del conjunto de datos combinados, preferentemente normalizados.

en el que los genes de los conjuntos de datos reducidos son los marcadores específicos para la enfermedad, siendo de manera preferible un trastorno genético, preferentemente un trastorno con la expresión génica alterada, se prefiere de forma particular cáncer,
y diagnosticar el cáncer en el centroide comprimido más cercano de los valores de la muestra del paciente para al menos un 50% de marcadores del conjunto está comprendido dentro de la desviación estándar, preferentemente dos veces la desviación estándar, incluso de forma más preferida tres veces la desviación estándar, del procedimiento de medida del centroide comprimido más cercano de los marcadores tumorales identificados con las muestras del cáncer.

8. El procedimiento de la reivindicación 7, **caracterizado porque** los datos de la expresión comprenden los datos de al menos dos conjuntos de datos diferentes de la micromatriz, en particular con los sesgos específicos.

9. El procedimiento de la reivindicación 7 u 8, **caracterizado porque** la etapa de combinación se lleva a cabo mediante una combinación por etapas de dos conjuntos de bases de la expresión génica e integración de los datos combinados, preferentemente mediante DWD.

10. El procedimiento de una cualquiera de las reivindicaciones 7 a 9, **caracterizado porque** cada conjunto de datos de la expresión comprende los datos de al menos 10, preferentemente al menos 20 de forma más preferida al menos 30, incluso de forma más preferida al menos 40, lo más preferido al menos 50, genes diferentes.

11. El procedimiento de una cualquiera de las reivindicaciones 7 a 10, **caracterizado porque** la etapa de determinación se repite en un conjunto combinado sin los genes determinados en una etapa de determinación previa y/o la etapa de determinación comprende la determinación de un umbral maximizado de la diferencia entre el valor de la expresión normalizada para cada gen al centroide mediante la validación cruzada, y en el que los genes con valores de expresión normalizados por debajo del umbral son eliminados del conjunto reducido.

Fig. 1

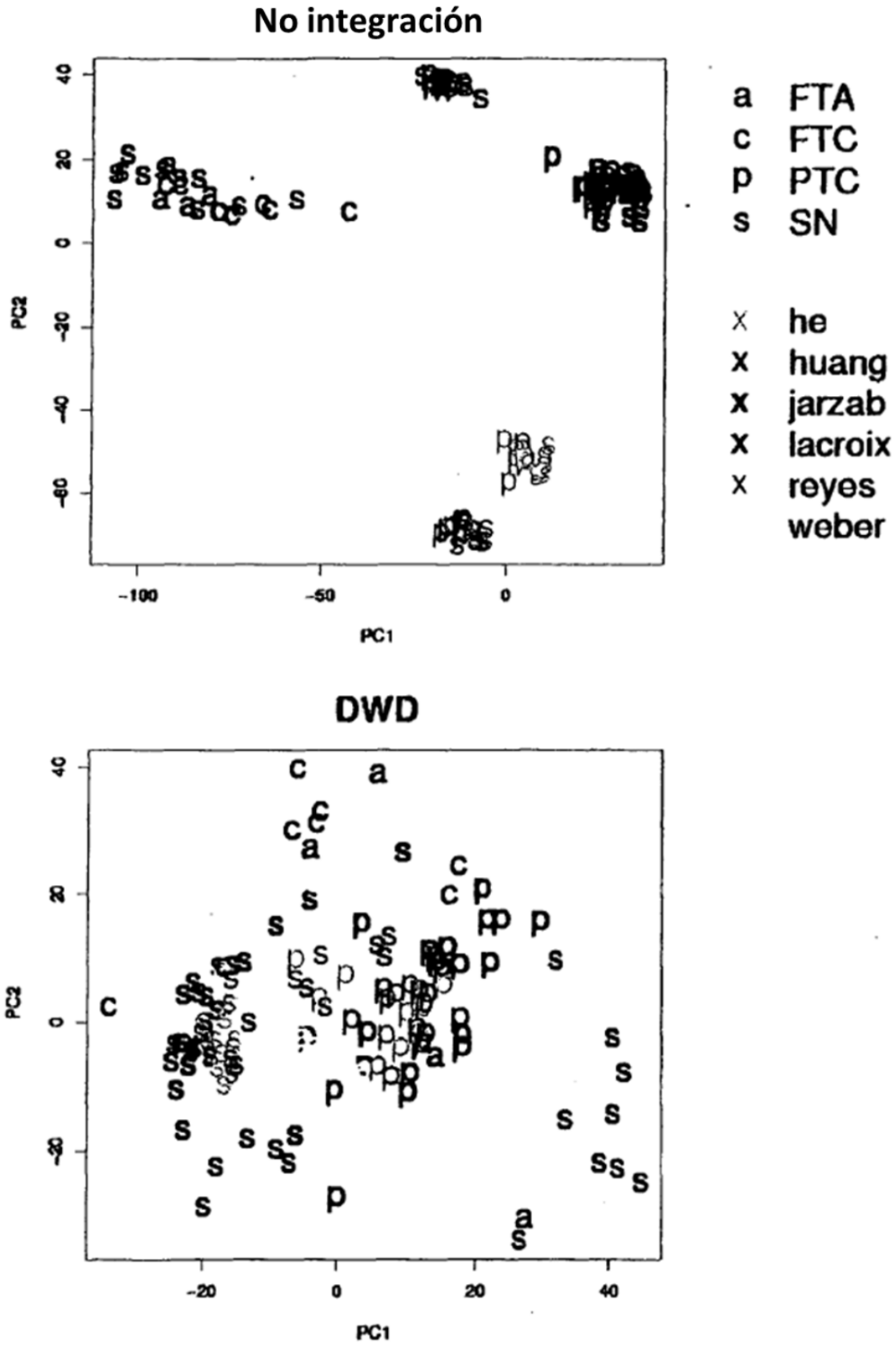

No integración
PC2
PC1
-100
-50
0
40
20
0
-20
-40
-60
a FTA
c FTC
p PTC
s SN
x he
x huang
x jarzab
x lacroix
x reyes
weber
DWD
PC2
PC1
-20
0
20
40

Fig.2

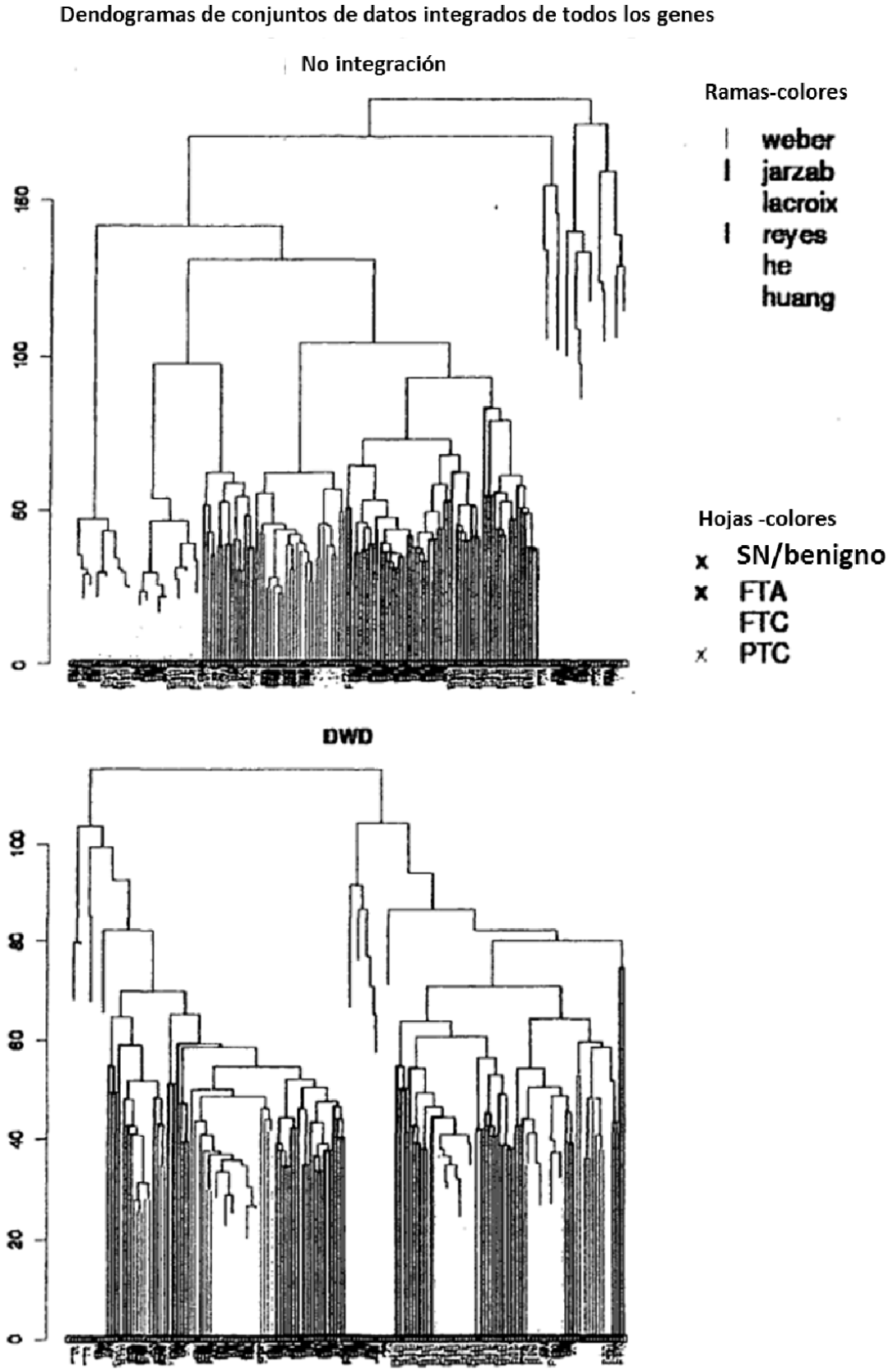

Dendogramas de conjuntos de datos integrados de todos los genes
No integración
Ramas-colores
weber
jarzab
lacroix
reyes
he
huang
Hojas -colores
SN/benigno
FTA
FTC
PTC
DWD
0
20
40
60
80
100
0
50
100
150

Fig-3

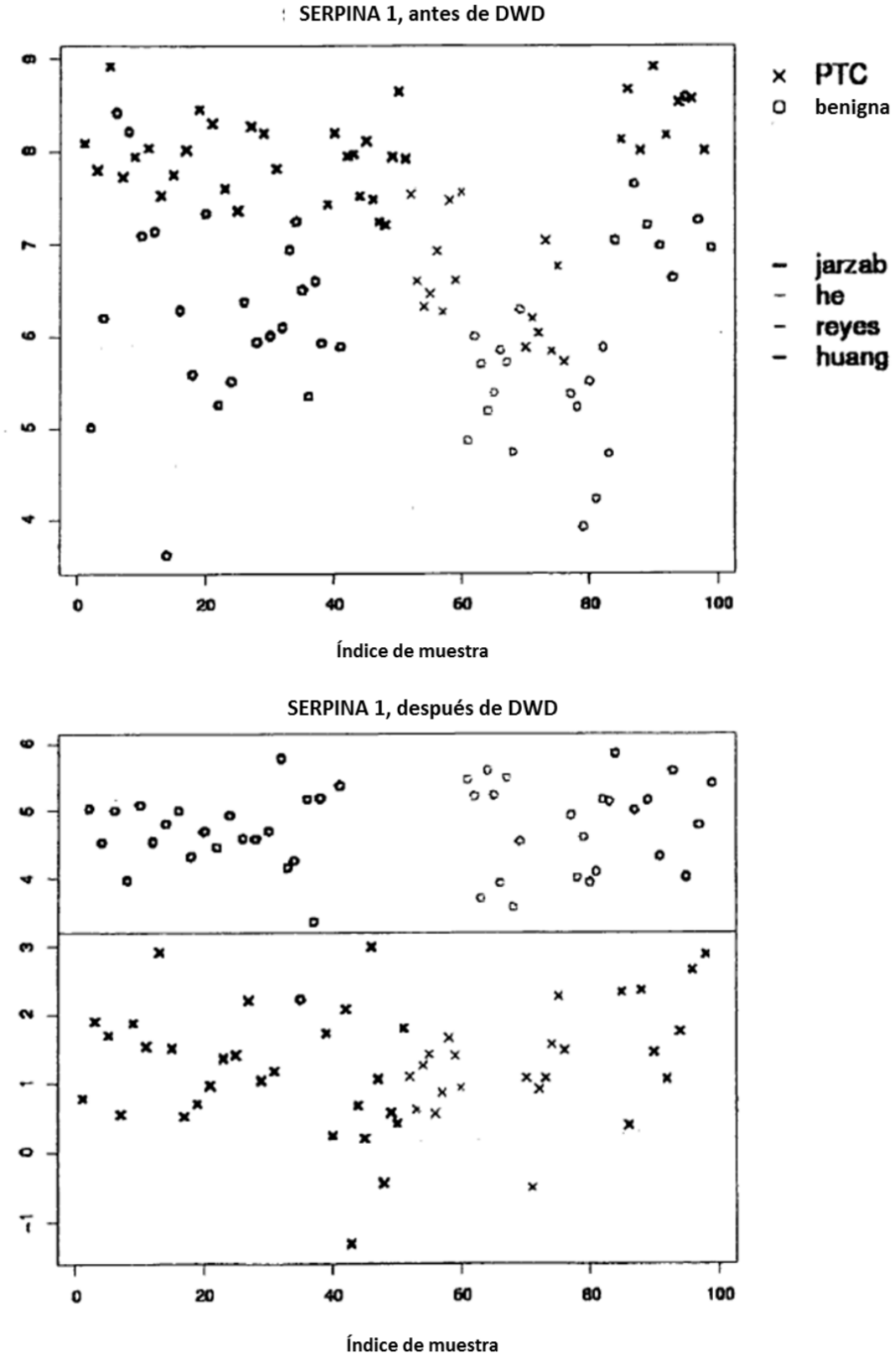

SERPINA 1, antes de DWD
PTC
benigna
jarzab
he
reyes
huang
Índice de muestra
SERPINA 1, después de DWD
Índice de muestra

Fig. 4

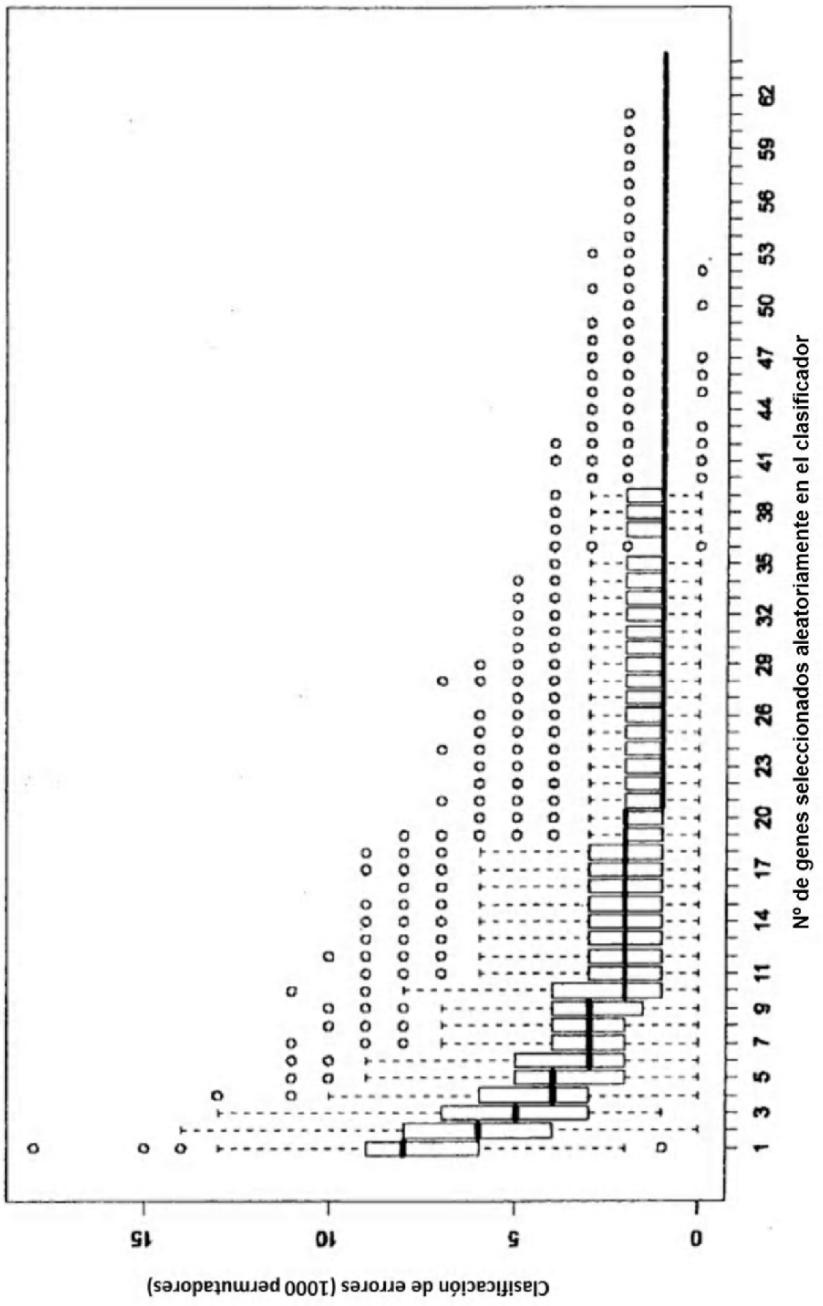
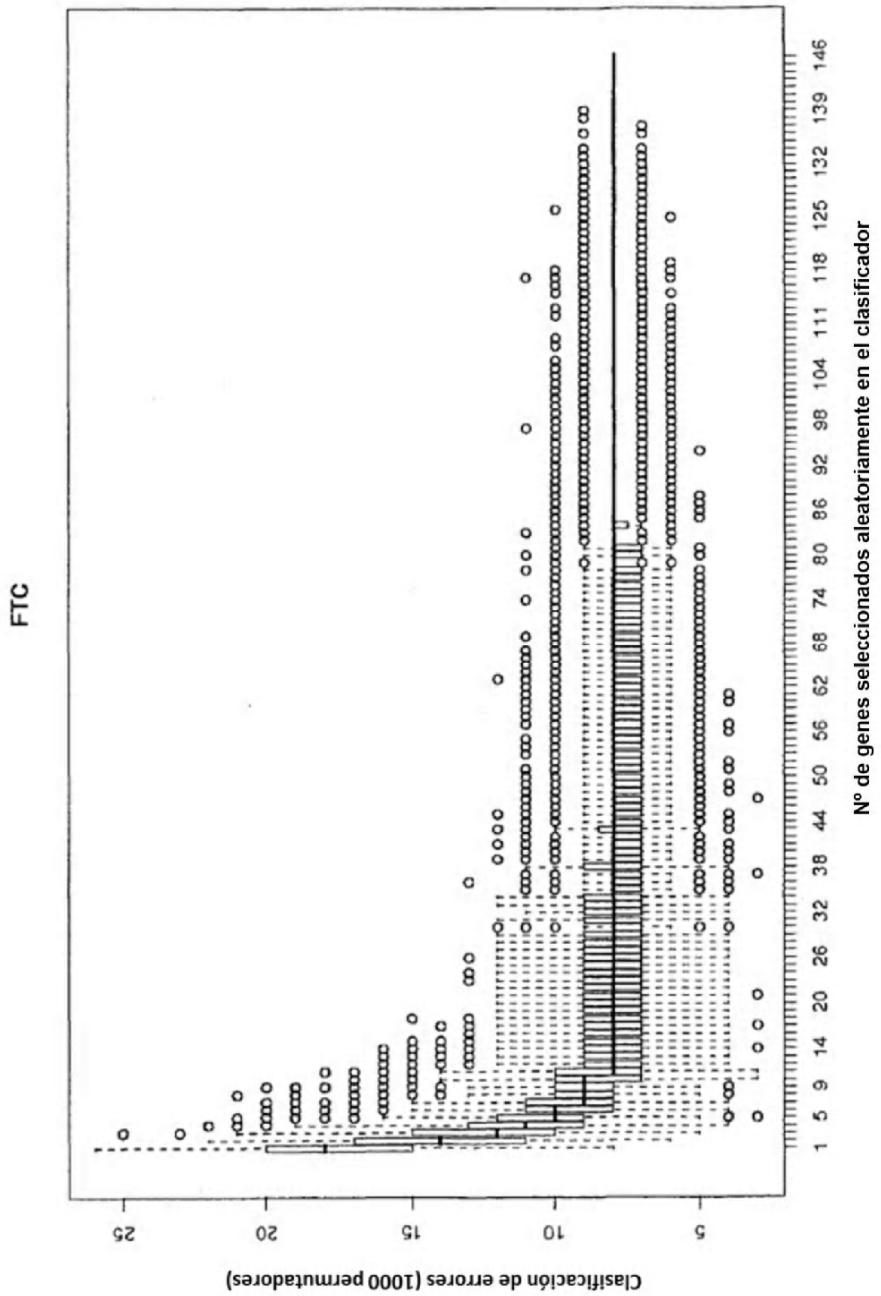

Clasificación de errores (1000 permutadores)
0
5
10
15
Nº de genes seleccionados aleatoriamente en el clasificador
1 3 5 7 9 11 14 17 20 23 26 29 32 35 38 41 44 47 50 53 56 59 62

Fig. 5


FTC
Clasificación de errores (1000 permutadores)
25
20
15
10
5
1 5 9 14 20 26 32 38 44 50 56 62 68 74 80 86 92 98 104 111 118 125 132 139 146
Nº de genes seleccionados aleatoriamente en el clasificador