(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 652 212**

(51) Int. CI.:

C12Q 1/68 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **20.12.2012 PCT/FR2012/053012**

(87) Fecha y número de publicación internacional: **27.06.2013 WO13093353**

(96) Fecha de presentación y número de la solicitud europea: **20.12.2012 E 12816746 (7)**

(97) Fecha y número de publicación de la concesión europea: **25.10.2017 EP 2794921**

(54) Título: **Procedimiento para el diagnóstico, in vitro, del cáncer de pulmón**

(30) Prioridad:

20.12.2011 FR 1162089

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
01.02.2018

(73) Titular/es:

BIOMÉRIEUX (100.0%)
69280 Marcy l'Étoile, FR

(72) Inventor/es:

PEROT, PHILIPPE;
MALLET, FRANÇOIS;
MONTGIRAUD, CÉCILE y
MUGNIER, NATHALIE

(74) Agente/Representante:

CURELL AGUILÁ, Mireia

Observaciones :

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

DESCRIPCIÓN

Procedimiento para el diagnóstico, *in vitro*, del cáncer de pulmón.

Los retrovirus endógenos constituyen la descendencia de retrovirus infecciosos que se han integrado en su forma proviral en unas células de la línea germinal y que han sido transmitidos por esta vía en el genoma de la descendencia del hospedante.

La secuenciación del genoma humano ha permitido revelar la extrema abundancia de los elementos transposables o de sus derivados. De hecho, las secuencias repetidas representan cerca de la mitad del genoma humano y los retrovirus endógenos y los retrotransposones componen el 8% con un número que se eleva, en la actualidad, a más de 400000 elementos.

La abundancia de los elementos retrovirales endógenos (ERV) presentes actualmente en el genoma humano es el resultado de un centenar de endogenizaciones conseguidas durante la evolución de la línea humana. Las diferentes oleadas de endogenización se extienden durante un periodo que va de 2 a 90 millones de años antes de nuestra era y han ido seguidas de la expansión del número de copias por unos fenómenos de tipo “copiar/pegar” con posibilidad de aparición de errores, llevando a partir de un provirus ancestral a la formación de una familia de HERV, es decir un conjunto de elementos que presentan unas homologías de secuencias. Los elementos más antiguos, los de la familia HERV-L, se habrían integrado antes de la aparición de los mamíferos. Dos familias HERV-F y HERV-H aparecieron en el periodo en el que los primeros primates hacían su aparición. Las familias HERV-FRD y HERV-K (HML-5) integradas hace de 40 a 55 millones de años, son específicas de los primates superiores. Por el contrario, las familias HERV-W y HERV-E, por ejemplo, se han integrado de 5 a 10 millones de años más tarde, después de la separación con los monos del nuevo mundo, y son específicas de Catarhini (humanoides y cercopitecos).

Las secuencias ERV están representadas en el conjunto de los cromosomas, con una densidad que varía según las familias y no existe correlación entre la proximidad física de ERV y su proximidad filogenética.

Los ERV se han considerado durante mucho tiempo como unos parásitos o como simples restos del ADN. Sin embargo, el impacto de ERV sobre el organismo no se limita únicamente a su participación anterior en el moldeo del genoma o a recombinaciones perjudiciales que aún pueden aparecer.

La abundancia y la complejidad estructural de ERV hacen los análisis de su expresión muy complicados y de manera frecuente difícilmente interpretables. La detección de la expresión de HERV puede reflejar la activación transcripcional de uno o varios locus dentro de una misma familia. El o los locus activados pueden además variar en función del tejido y/o del contexto.

Los presentes inventores han descubierto y demostrado ahora que unas secuencias de ácidos nucleicos que corresponden a unos locus precisamente identificados de elementos retrovirales endógenos, están asociadas al cáncer de pulmón y que estas secuencias son unos marcadores moleculares de la patología. Las secuencias identificadas son unos provirus, es decir unas secuencias que contienen la totalidad o parte de los genes *gag*, *pol* y *env* flanqueados en 5’ y 3’ por largas repeticiones terminales (LTR o “Long Terminal Repeat” según la terminología anglosajona), o la totalidad o parte de LTR o de los genes aislados. Las secuencias ADN identificadas se referencian respectivamente como SEC ID: 1 a 242 en el listado de secuencias, su localización cromosómica se identifica en la tabla siguiente (NCBI 36/hg18), así como su expresión, sobreexpresión o subexpresión representadas por la “relación de expresión” entre muestra cancerosa y muestra normal. Cuando la expresión del ácido nucleico o el cambio de expresión del ácido nucleico son específicos del tejido del pulmón se indica esta información mediante el símbolo “x” en la columna de tejido diana. Esto significa que, si se determina una expresión o un cambio de expresión del ácido nucleico en cuestión en un compartimiento biológico diferente del tejido del pulmón, esto representa, de forma remota, una firma de un cáncer de pulmón. Las secuencias ADN identificadas como específicas del tejido del pulmón se referencian respectivamente como SEC ID nº: 4, 5, 7, 9, 10, 12, 20, 25 y 40 en el listado de secuencias. Las secuencias ADN identificadas como no específicas del tejido del pulmón se referencian respectivamente en las SEC ID nº: 1, 2, 3, 6, 8, 11, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241 y 242.

Tabla

SEC ID nº	Localización cromosómica	Tejido diana	Relación de expresión cáncer/normal
1	(+) chr 19: 57753757-57754726		-4,8
2	(-) chr 3: 32480101-32483411		-3,9
3	(+) chr 3: 135222763-135223084		-3,7
4	(+) chr 6: 131686129-131689771	x	-3,5
5	(-) chr X: 91414738-91420449	x	-3,2
6	(+) chr 7: 100274888-100276065		3,2
7	(-) chr 1: 154420719-154426128	x	-2,9
8	(+) chr 2: 231973667-231981798		-2,8
9	(+) chr 2: 188084458-188084785	x	-2,7
10	(+) chr 6: 131658190-131664031	x	-2,7
11	(+) chr 2: 201711970-201712935		-2,7
12	(-) chr 6: 82110364-82117596	x	-2,7
13	(+) chr 7: 93107183-93112802		2,6
14	(+) chr 6: 2833115-2833223		-2,5
15	(-) chr 2: 71238414-71246247		-2,5
16	(-) chr 17: 38571826-38572147		2,5
17	(+) chr 3: 133508751-133509387		-2,4
18	(+) chr 20: 24856581-24861663		-2,4
19	(+) chr 13: 108715439-108721465		-2,4
20	(+) chr 7: 17420212-17426910	x	-2,3
21	(-) chr 2: 54587807-54590183		-2,3
22	(-) chr 11: 49820917-49821387		-2,2
23	(-) chr 1: 79726879-79732396		-2,2
24	(-) chr 20: 15911118-15913833		2,2
25	(+) chr X: 93632019-93639453	x	-2,2
26	(-) chr 2: 58194071-58199769		2,1
27	(-) chr 1: 176376976-176378817		-2,1
28	(-) chr 13: 112501931-112502014		-2,0
29	(+) chr 3: 95139589-95145594		-2,0
30	(-) chr 6: 27901601-27902447		2,0
31	(+) chr 8: 86591572-86592049		-2,0
32	(+) chr 16: 55266229-55266680		2,0
33	(-) chr 2: 165222667-165224367		-2,0
34	(+) chr 6: 152853219-152859441		-2,0
35	(+) chr 6: 111558919-111565969		-1,9
36	(-) chr 11: 73926652-73927082		-1,9
37	(-) chr 19: 51297295-51305072		1,9
38	(-) chr 8: 138907724-138911942		-1,9
39	(-) chr 1: 144779633-144780605		1,9
40	(-) chr 6: 32732881-32733838	x	-1,9
41	(-) chr 1: 144779633-144780605		1,9
42	(-) chr 3: 130037918-130043560		-1,8
43	(-) chr Y: 13388258-13388740		1,8
44	(+) chr 9: 34993290-34999332		-1,8
45	(-) chr 12: 84386849-84387812		-1,8
46	(+) chr 1: 13551727-13561236		-1,7
47	(-) chr Y: 21547357-21548314		-1,7
48	(+) chr 15: 49439487-49440297		-1,7
49	(+) chr 2: 30592323-30596634		1,7
50	(+) chr 14: 73239795-73245370		-1,7
51	(-) chr 12: 11656097-11659027		-1,7
52	(-) chr 19: 46119961-46120936		-1,7
53	(+) chr 17: 75988189-75996283		-1,7
54	(+) chr 1: 204107496-204110550		-1,7
55	(-) chr 1: 223094264-223100407		1,7
56	(-) chr 1: 89419377-89419778		-1,7
57	(+) chr 7: 122244019-122250290		-1,7
58	(+) chr 2: 231645891-231645949		1,7
59	(+) chr 1: 201726731-201727141		-1,6
60	(-) chr 1: 12762845-12768665		-1,6

SEC ID nº	Localización cromosómica	Tejido diana	Relación de expresión cáncer/normal
61	(-) chr 1: 196365449-196366430		-1,6
62	(-) chr 5: 170360805-170364300		-1,6
63	(+) chr 11: 76938868-76944155		-1,6
64	(-) chr 18: 64760486-64761450		-1,6
65	(+) chr 7: 76807190-76813131		-1,6
66	(-) chr 7: 128961001-128961758		-1,6
67	(-) chr 16: 69214387-69217522		-1,6
68	(-) chr 17: 59750387-59751101		-1,5
69	(+) chr 19: 45269688-45270401		1,5
70	(+) chr 14: 22268133-22268207		-1,5
71	(+) chr 10: 55596442-55602269		-1,5
72	(+) chr 13: 55512386-55518412		-1,5
73	(+) chr 19: 57816195-57826383		1,5
74	(-) chr 14: 24964620-24969975		-1,5
75	(+) chr 7: 125337255-125351402		-1,5
76	(-) chr 1: 171684177-171684627		-1,5
77	(+) chr 13: 35788462-35794330		-1,5
78	(+) chr 5: 143285967-143286088		1,5
79	(+) chr 18: 13646676-13647837		1,5
80	(-) chr 19: 60210007-60215603		-1,5
81	(-) chr 4: 4052713-4058389		-1,5
82	(-) chr 4: 62312529-62318338		-1,5
83	(+) chr 3: 147554294-147559942		-1,5
84	(-) chr 13: 60755266-60755331		-1,5
85	(-) chr 15: 65048616-65056011		1,5
86	(-) chr 4: 4030945-4038383		-1,5
87	(+) chr 13: 27913017-27913666		-1,5
88	(+) chr 14: 101776047-101781857		1,5
89	(-) chr 5: 95528434-95530534		1,4
90	(-) chr 15: 85890568-85891324		-1,4
91	(-) chr 6: 68639470-68641070		-1,4
92	(+) chr 6: 63353796-63359590		-1,4
93	(+) chr 6: 91950297-91950513		1,4
94	(-) chr 19: 20721466-20730278		-1,4
95	(+) chr 16: 10506850-10507201		-1,4
96	(+) chr 1: 146832410-146833382		1,4
97	(+) chr 8: 24149948-24150249		1,4
98	(+) chr 2: 34773168-34775448		-1,4
99	(-) chr 11: 23865384-23871043		-1,4
100	(+) chr 8: 98282174-98288062		1,4
101	(-) chr 19: 19756756-19757058		1,4
102	(-) chr 4: 135818540-135824157		-1,4
103	(-) chr 3: 184711418-184712409		-1,4
104	(+) chr 3: 75269085-75276706		1,4
105	(+) chr 2: 71468737-71476455		-1,4
106	(-) chr 9: 73768007-73768097		-1,4
107	(+) chr 5: 111807288-111815094		-1,4
108	(+) chr 17: 70760297-70765658		1,4
109	(-) chr 6: 16064468-16065182		1,4
110	(+) chr 1: 4781688-4782412		-1,4
111	(-) chr 16: 58106314-58114369		-1,4
112	(+) chr 8: 91057690-91058157		1,4
113	(+) chr 2: 215375861-215376821		1,4
114	(-) chr 12: 17764451-17768938		1,4
115	(-) chr X: 112238600-112244308		-1,4
116	(-) chr 5: 92818136-92819135		-1,4
117	(-) chr 1: 79877033-79882760		-1,4
118	(-) chr 4: 95433460-95438743		-1,4
119	(+) chr 16: 29617077-29617563		-1,4
120	(-) chr 5: 121980485-121987968		-1,4
121	(-) chr 13: 55050362-55056223		1,4
122	(+) chr 18: 31049990-31056476		-1,4

SEC ID nº	Localización cromosómica	Tejido diana	Relación de expresión cáncer/normal
123	(-) chr X: 73258716-73266192		-1,3
124	(+) chr 5: 55618832-55619003		-1,3
125	(+) chr 3: 156178559-156179784		1,3
126	(-) chr 3: 45334152-45335128		1,3
127	(-) chr 5: 100347893-100353827		1,3
128	(-) chr 6: 161967666-161968129		1,3
129	(-) chr 21: 20291919-20292024		-1,3
130	(-) chr 11: 3451335-3458971		-1,3
131	(+) chr 8: 105367316-105373215		1,3
132	(+) chr 18: 2829330-2829995		-1,3
133	(-) chr 13: 90298826-90304533		-1,3
134	(-) chr 3: 180722967-180726830		-1,3
135	(-) chr 8: 129693327-129699058		-1,3
136	(+) chr 5: 5852293-5852397		1,3
137	(-) chr 6: 122854248-122854377		1,3
138	(-) chr 7: 5033421-5033757		1,3
139	(+) chr 6: 130555149-130561062		-1,3
140	(+) chr 12: 94666771-94672799		-1,3
141	(-) chr 9: 29178033-29178600		1,3
142	(+) chr 11: 40847438-40848141		-1,3
143	(+) chr 11: 321893-322552		-1,3
144	(-) chr 4: 167866387-167867014		-1,3
145	(-) chr 1: 220215255-220218922		-1,3
146	(-) chr 22: 37122490-37122813		-1,3
147	(-) chr 2: 80956433-80960797		1,3
148	(+) chr 6: 63559535-63565316		-1,3
149	(-) chr 18: 3013633-3013763		-1,3
150	(-) chr 12: 20983776-20984452		1,3
151	(+) chr 6: 142192789-142193227		-1,3
152	(+) chr 6: 125253252-125260608		1,3
153	(+) chr 4: 175540733-175541728		-1,3
154	(-) chr 10: 65826993-65827758		-1,3
155	(-) chr 8: 12395268-12398823		-1,3
156	(+) chr 13: 61603944-61604298		1,3
157	(-) chr 3: 178866314-178871614		1,3
158	(-) chr 12: 20862483-20866818		-1,3
159	(-) chr 10: 59523861-59524645		1,3
160	(-) chr 7: 79651365-79652053		1,3
161	(-) chr 11: 69581214-69582655		-1,3
162	(+) chr 21: 44461476-44462142		1,3
163	(-) chr 2: 105967906-105968229		1,3
164	(+) chr 6: 123199111-123199590		1,3
165	(+) chr 9: 22813029-22813838		-1,3
166	(+) chr 14: 41740538-41741242		1,3
167	(-) chr 18: 1990815-1991782		1,3
168	(+) chr 3: 8686697-8686961		-1,3
169	(-) chr 2: 38161446-38167158		-1,3
170	(-) chr 10: 101570559-101571639		1,3
171	(+) chr 15: 31130887-31131553		-1,3
172	(+) chr 7: 65695119-65695525		-1,3
173	(-) chr 8: 828875-829287		1,3
174	(+) chr 6: 160569673-160575346		-1,3
175	(-) chr 10: 53463521-53469327		-1,3
176	(-) chr 4: 176619310-176625331		1,3
177	(+) chr 12: 122598918-122609265		1,3
178	(-) chr 3: 78475709-78476509		-1,3
179	(-) chr X: 37202604-37205422		-1,3
180	(-) chr 1: 23196587-23203483		-1,2
181	(+) chr 18: 38295804-38303710		-1,2
182	(+) chr 4: 54581003-54585874		-1,2
183	(-) chr 9: 35630300-35632824		1,2
184	(-) chr X: 8364892-8365592		1,2

SEC ID nº	Localización cromosómica	Tejido diana	Relación de expresión cáncer/normal
185	(-) chr 6: 24784895-24791307		-1,2
186	(-) chr 14: 79937761-79938478		-1,2
187	(-) chr 2: 142963716-142969364		-1,2
188	(+) chr X: 30031651-30037293		1,2
189	(-) chr 2: 63928103-63928550		1,2
190	(+) chr X: 92571323-92580146		-1,2
191	(+) chr 7: 100353142-100354585		1,2
192	(+) chr 9: 113677404-113678370		-1,2
193	(+) chr 9: 129215427-129217168		-1,2
194	(+) chr 4: 171274156-171279886		-1,2
195	(-) chr 2: 52103626-52108242		-1,2
196	(-) chr 2: 16846720-16847190		1,2
197	(+) chr 3: 187581597-187582311		-1,2
198	(-) chr 12: 77758546-77759251		1,2
199	(-) chr 20: 740491-741481		1,2
200	(-) chr 3: 21205514-21210699		1,2
201	(-) chr 15: 49734758-49735530		-1,2
202	(-) chr 3: 103933971-103934748		-1,2
203	(-) chr 4: 177452380-177455182		1,2
204	(-) chr 9: 75103859-75104219		-1,2
205	(+) chr 1: 237334832-237336987		-1,2
206	(-) chr 13: 55584544-55586741		1,2
207	(-) chr 12: 29178820-29185001		-1,2
208	(-) chr 20: 36760391-36760718		1,2
209	(+) chr 20: 32476466-32477455		1,2
210	(+) chr 4: 8492255-8492953		-1,2
211	(+) chr 10: 65936583-65936972		1,2
212	(+) chr 11: 22761631-22761856		1,2
213	(+) chr 9: 133412422-133417146		-1,2
214	(-) chr 4: 41854992-41855740		1,2
215	(+) chr 12: 60267154-60272981		-1,2
216	(-) chr 3: 32477433-32480101		-1,2
217	(+) chr 18: 70111304-70117249		-1,2
218	(+) chr 19: 58554156-58559856		-1,2
219	(+) chr X: 53557364-53557692		1,2
220	(-) chr 5: 34514678-34514916		-1,2
221	(+) chr 6: 26107438-26108404		1,2
222	(-) chr 10: 6906147-6915609		1,2
223	(+) chr 1: 89162808-89170138		-1,2
224	(+) chr 5: 74636041-74638012		1,2
225	(-) chr 5: 115707988-115708051		-1,2
226	(-) chr X: 72938020-72941188		-1,1
227	(+) chr Y: 26648032-26648192		1,1
228	(+) chr 18: 6123610-6123972		1,1
229	(-) chr 3: 110377937-110378340		-1,1
230	(-) chr 9: 101942283-101948625		-1,1
231	(+) chr 10: 100097920-100098685		1,1
232	(-) chr 4: 135768442-135768594		1,1
233	(+) chr 11: 18656991-18657966		1,1
234	(-) chr 7: 68382377-68383085		1,1
235	(-) chr X: 85929452-85930118		-1,1
236	(+) chr 4: 1992875-1993855		-1,1
237	(+) chr 5: 64505763-64506756		1,1
238	(+) chr 18: 8944363-8944760		1,1
239	(-) chr 1: 225849586-225850738		1,1
240	(+) chr 2: 199752333-199758641		-1,1
241	(-) chr X: 137174181-137180160		-1,1
242	(+) chr 9: 12938344-12944129		-1,1

La presente invención tiene por lo tanto por objeto un procedimiento para el diagnóstico, *in vitro*, del cáncer de pulmón, en una muestra biológica extraída de un paciente, que comprende una etapa de detección de por lo menos un producto de expresión de por lo menos una secuencia de ácido nucleico, siendo dicha secuencia de

ácido nucleico seleccionada de entre las secuencias identificadas como específicas del tejido del pulmón, es decir seleccionadas de entre el grupo de las secuencias identificadas en SEC ID nº: 4, 5, 7, 9, 10, 12, 20, 25 y 40 o de entre las secuencias que presentan por lo menos el 99% de identidad, preferentemente por lo menos el 99,5% de identidad y ventajosamente por lo menos el 99,6% o por lo menos el 99,7% de identidad con una de las secuencias identificadas en las SEC ID nº 4, 5, 7, 9, 10, 12, 20, 25 y 40.

El diagnóstico permite establecer si un individuo está enfermo o no. El pronóstico permite establecer un grado de gravedad de la enfermedad (grados y/o fases) que tiene una incidencia sobre la supervivencia y/o la calidad de vida del individuo. En el marco de la presente invención, el diagnóstico puede ser muy precoz.

El porcentaje de identidad descrito anteriormente que define las variantes de las secuencias se ha determinado tomando en consideración la diversidad nucleotídica en el genoma. Se sabe que la diversidad nucleotídica es más elevada en las regiones del genoma ricas en secuencias repetidas que en las regiones que no contienen secuencias repetidas. A título de ejemplo, Nickerson D. A. *et al*. (1) han mostrado una diversidad de aproximadamente el 0,3% (0,32%) en unas regiones que contienen unas secuencias repetidas.

La capacidad de discriminación de un estado canceroso de cada una de las secuencias identificadas anteriormente se ha demostrado con la ayuda de un análisis estadístico que utiliza el procedimiento SAM (5) seguido de una correlación por el porcentaje de falso positivo (6) y de una sobreexpresión de los valores inferiores a 2^6. En consecuencia, cada una de las secuencias identificadas anteriormente presenta una diferencia de expresión significativa entre un estado tumoral y un estado normal. Como resultado de esto, una diferencia de expresión observada para una de las secuencias antes citadas constituye una firma de la patología. Por supuesto, es posible combinar las diferencias de expresión mostradas para varias de las secuencias referenciadas anteriormente, por ejemplo por una o más asociaciones de 2, 3, 4, 5, 6, 7, 8, 9, 10 y más, incluso hasta 242 de las secuencias enumeradas.

En un modo de realización del procedimiento según la invención, se detecta el producto de expresión de por lo menos dos secuencias de ácido nucleico, siendo dichas por lo menos dos secuencias de ácido nucleico seleccionadas de entre las secuencias identificadas como específicas del tejido del pulmón, es decir seleccionadas de entre el grupo de secuencias identificadas en las SEC ID nº 4, 5, 7, 9, 10, 12, 20, 25 y 40, en particular seleccionadas de entre las secuencias identificadas en las SEC ID nº 4, 5 y 7, o de entre las secuencias que presentan por lo menos el 99% de identidad, preferentemente por lo menos el 99,5% de identidad y ventajosamente por lo menos el 99,6% o por lo menos el 99,7% de identidad con una de las secuencias identificadas en las SEC ID nº 4, 5, 7, 9, 10, 12, 20, 25 y 40, y en particular las que presentan por lo menos el 99% de identidad, preferentemente por lo menos el 99,5% de identidad y ventajosamente por lo menos el 99,6% o por lo menos el 99,7% de identidad con una de las secuencias identificadas en las SEC ID nº 4, 5 y 7.

En otro modo de realización del procedimiento de la invención, se detecta además el producto de expresión de por lo menos una secuencia seleccionada de entre las secuencias identificadas como no específicas del tejido del pulmón, es decir seleccionadas de entre el grupo de secuencias identificadas en las SEC ID nº 1, 2, 3, 6, 8, 11, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241 y 242, en particular las secuencias seleccionadas de entre las secuencias identificadas en las SEC ID nº 1, 6 y 50, o seleccionadas de entre las secuencias que presentan por lo menos el 99% de identidad, preferentemente por lo menos el 99,5% de identidad y ventajosamente por lo menos el 99,6% o por lo menos el 99,7% de identidad con una de las secuencias identificadas en las SEC ID nº 1, 2, 3, 6, 8, 11, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241 y 242 y en particular de entre las secuencias que presentan por lo menos el 99% de identidad, preferentemente por lo menos el 99,5% de identidad y ventajosamente por lo menos el 99,6% o por lo menos el 99,7% de identidad con una de las secuencias identificadas en las SEC ID nº 1, 6 y 50.

Preferentemente, en el procedimiento de la invención, se detecta además el producto de expresión de por lo menos una secuencia de ácido nucleico, preferentemente de por lo menos dos secuencias de ácido nucleico o de tres secuencias de ácido nucleico, siendo dichas secuencias de ácido nucleico seleccionadas de entre el grupo de secuencias identificadas en las SEC ID nº 1, 6 y 50, o de entre las secuencias que presentan por lo menos el 99% de identidad, preferentemente por lo menos el 99,5% de identidad y ventajosamente por lo menos el 99,6% o por lo menos el 99,7% de identidad con las secuencias identificadas en las SEC ID nº 1, 6 y 50.

El producto de expresión detectado es por lo menos un transcrito ARN, en particular por lo menos un ARNm o por lo menos un polipéptido.

Cuando el producto de expresión es un transcrito ARNm, éste se detecta mediante cualquier método apropiado, tal como la hibridación, la secuenciación o la amplificación. El ARNm se puede detectar directamente poniéndolo en contacto con por lo menos una sonda y/o por lo menos un cebador que están diseñados para hibridarse en condiciones experimentales predeterminadas con los transcritos ARNm, demostrando la presencia o la ausencia de hibridación al ARNm y eventualmente cuantificando el ARNm. Entre los métodos preferidos, se puede citar la amplificación (por ejemplo la RT-PCR, la NASBA, etc.), la hibridación sobre chip o también la secuenciación. El ARNm también se puede detectar indirectamente a partir de ácidos nucleicos derivados de dichos transcritos, como las copias ADNc, etc.

Generalmente, el procedimiento de la invención comprende una etapa inicial de extracción de ARNm de la muestra a analizar.

Así, el procedimiento puede comprender:

(i) una etapa de extracción del ARNm de la muestra a analizar,

(ii) una etapa de detección y de cuantificación del ARNm de la muestra a analizar,

(iii) una etapa de extracción del ARNm en una muestra de referencia, que puede ser una muestra sana o que proviene del mismo individuo, o

(iv) una etapa de detección y de cuantificación del ARNm de la muestra sana,

(v) una etapa de comparación de la cantidad de ARNm expresado en la muestra a analizar y en la muestra de referencia; la determinación de una cantidad de ARNm expresado en la muestra a analizar diferente de la cantidad de ARNm expresado en la muestra de referencia sana se puede correlacionar con el diagnóstico de un cáncer de pulmón (siendo la diferencia de la cantidad de ARNm en el tejido de pulmón canceroso con respecto a la cantidad de ARNm expresado en el tejido de pulmón sano indiferentemente una expresión, una sobreexpresión o una sub-expresión);

y en particular:

(i) una extracción del ARNm a analizar de la muestra,

(ii) una determinación, en el ARN a analizar, de un nivel de expresión de por lo menos una secuencia ARN en la muestra, siendo la secuencia ARN el producto de transcripción de por lo menos una secuencia de ácido nucleico tal como se ha identificado anteriormente, y

(iii) una comparación del nivel de expresión de la o de las secuencias ARN definidas en (ii) con un nivel de expresión de referencia; pudiendo la determinación de un nivel de expresión del ARN a analizar que presenta una diferencia con respecto al nivel de expresión de referencia correlacionarse con el diagnóstico de un cáncer de pulmón (como se ha determinado anteriormente);

o

(i) una etapa de extracción del ARNm de la muestra a analizar,

(ii) una etapa de detección y de cuantificación del ARNm de la muestra a analizar,

(iii) una etapa de comparación de la cantidad de ARNm expresado en la muestra a analizar con respecto a una cantidad de ARNm de referencia, pudiendo la determinación de una cantidad de ARNm expresado en la muestra a analizar diferente de la cantidad de ARNm de referencia correlacionarse con el diagnóstico de un cáncer de pulmón (siendo la diferencia de la cantidad de ARNm en la muestra a analizar con respecto a la cantidad de ARNm de referencia indiferentemente una expresión, una sobreexpresión o una subexpresión).

En un modo de realización del procedimiento de la invención, se preparan unas copias ADN del ARNm, se ponen en contacto las copias ADN con por lo menos una sonda y/o por lo menos un cebador en unas condiciones predeterminadas que permiten la hibridación y se detecta la presencia o la ausencia de hibridación a dichas copias ADN.

El producto de expresión detectado puede también ser un polipéptido que es el producto de la traducción de por lo menos uno de los transcritos descritos anteriormente. En tal caso, se detecta el polipéptido expresado por la puesta en contacto con por lo menos una pareja de unión específica de dicho polipéptido, en particular un anticuerpo o un análogo de anticuerpo o un aptámero. La pareja de unión es preferentemente un anticuerpo, por ejemplo un anticuerpo monoclonal o un anticuerpo policlonal altamente purificado o un análogo de un anticuerpo, por ejemplo una proteína de afinidad a las propiedades competitivas (nanofitine™).

Los anticuerpos policlonales se pueden obtener por inmunización de un animal con el inmunógeno apropiado, seguida de la recuperación de los anticuerpos buscados en forma purificada, por extracción del suero de dicho animal, y separación de dichos anticuerpos de los otros constituyentes del suero, en particular por cromatografía de afinidad sobre una columna sobre la cual se fija un antígeno específicamente reconocido por los anticuerpos.

Los anticuerpos monoclonales se pueden obtener mediante la técnica de los hibridomas cuyo principio general se recuerda a continuación.

En un primer tiempo, se inmuniza un animal, generalmente un ratón con el inmunógeno apropiado, cuyos linfocitos B son entonces capaces de producir unos anticuerpos contra este antígeno. Se fusionan después estos linfocitos productores de anticuerpos con unas células mielomatosas “inmortales” (murinas en el ejemplo) para dar lugar a unos hibridomas. A partir de la mezcla heterogénea de las células así obtenida, se efectúa entonces una selección de las células capaces de producir un anticuerpo particular y de multiplicarse indefinidamente. Cada hibridoma se multiplica en forma de clon, conduciendo cada uno a la producción de un anticuerpo monoclonal cuyas propiedades de reconocimiento frente a la proteína se podrán ensayar, por ejemplo, en ELISA, por inmunotransferencia (transferencia Western) en una o dos dimensiones, en inmunofluorescencia, o con la ayuda de un biosensor. A continuación, se purifican los anticuerpos monoclonales así seleccionados, en particular según la técnica de cromatografía de afinidad descrita anteriormente.

Los anticuerpos monoclonales pueden ser también unos anticuerpos recombinantes obtenidos por ingeniería genética, mediante unas técnicas bien conocidas por el experto en la materia.

Las nanofitines™ son pequeñas moléculas que, como los anticuerpos, son capaces de unirse a una diana biológica que permiten así detectarla, capturarla o simplemente determinarla dentro de un organismo. Se les presenta, entre otros, como unos análogos de anticuerpos.

Los aptámeros son unos oligonucleótidos sintéticos capaces de fijar un ligando específico.

La invención se refiere también a la utilización de por lo menos una secuencia de ácido nucleico, aislada, como marcador molecular para el diagnóstico *in vitro* del cáncer de pulmón, caracterizada por que dicha secuencia de ácido nucleico consiste en:

(i) por lo menos una secuencia ADN seleccionada de entre las secuencias identificadas en las SEC ID nº :4, 5, 7, 9, 10, 12, 20, 25 y 40, o

(ii) por lo menos una secuencia ADN complementaria de una secuencia seleccionada de entre las secuencias SEC ID nº 4, 5, 7, 9, 10, 12, 20, 25 y 40, o

(iii) por lo menos una secuencia ADN que presenta respectivamente por lo menos el 99% de identidad, preferentemente por lo menos el 99,5% de identidad y ventajosamente por lo menos el 99,6% o por lo menos el 99,7% de identidad con una secuencia tal como como se define en (i) y (ii), o

(iv) por lo menos una secuencia ARN, que es el producto de transcripción de una secuencia seleccionada de entre las secuencias tales como se definen en (i), o

(v) por lo menos una secuencia ARN que es el producto de transcripción de una secuencia seleccionada de entre las secuencias que presentan por lo menos el 99% de identidad, preferentemente por lo menos el 99,5% de identidad y ventajosamente por lo menos el 99,6% o por lo menos el 99,7% de identidad con una secuencia tal como se define en (i).

En un modo de realización, se utilizan por lo menos dos secuencias de ácido nucleico que consisten en:

(i) por lo menos una secuencia ADN seleccionada de entre las secuencias SEC ID nº 4, 5, 7, 9, 10, 12, 20, 25 y 40, o sus variantes tales como se han definido anteriormente, y por lo menos una secuencia ADN

seleccionada entre las secuencias identificadas en las SEC ID nº 1, 2, 3, 6, 8, 11, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241 y 242 o sus variantes, en particular de entre las secuencias identificadas en las SEC ID nº 1, 6 y 50, o sus variantes tales como se han definido anteriormente, o por lo menos dos secuencias seleccionadas de entre las secuencias identificadas en las SEC ID nº 4, 5, 7, 9, 10, 12, 20, 25 y 40, en particular seleccionadas de entre las secuencias identificadas en las SEC ID nº 4, 5 y 7, o sus variantes tales como se han definido anteriormente, o

(ii) por lo menos dos secuencias ADN respectivamente complementarias de por lo menos dos secuencias tales como se definen en (i), o

(iii) por lo menos dos secuencias de ADN que presentan respectivamente por lo menos 99% de identidad, preferentemente por lo menos 99,5% de identidad y ventajosamente por lo menos 99,6% o por lo menos 99,7% de identidad con dos secuencias tales como se definen en (i) y (ii), o

(iv) por lo menos dos secuencias ARN que son respectivamente el producto de transcripción de dos secuencias seleccionadas de entre las secuencias tales como se definen en (i), o

(v) por lo menos dos secuencias ARN que son el producto de transcripción de dos secuencias seleccionadas de entre las secuencias que presentan por lo menos 99% de identidad, preferentemente por lo menos 99,5% de identidad y ventajosamente por lo menos 99,6% o por lo menos 99,7% de identidad con las secuencias tales como se definen en (i).

La invención se refiere también a un procedimiento para evaluar la eficacia de un tratamiento del cáncer de pulmón que comprende una etapa de obtención de una serie de muestras biológicas, una etapa de detección de por lo menos un producto de expresión de por lo menos una secuencia de ácido nucleico en dicha serie de muestras biológicas, seleccionándose dicha secuencia de ácido nucleico de entre las secuencias identificadas en las SEC ID nº 4, 5, 7, 9, 10, 12, 20, 25 y 40 o de entre las secuencias que presentan por lo menos el 99% de identidad, preferentemente por lo menos el 99,5% de identidad y ventajosamente por lo menos el 99,6% o por lo menos el 99,7% de identidad con las secuencias identificadas en las SEC ID nº 4, 5, 7, 9, 10, 12, 20, 25 y 40.

En otro modo de realización del procedimiento, se detecta el producto de expresión de por lo menos una secuencia de ácido nucleico, preferentemente de por lo menos dos secuencias de ácidos nucleicos o de tres secuencias de ácido nucleico, siendo dichas secuencias de ácido nucleico seleccionadas de entre el grupo de secuencias identificadas en las SEC ID nº 4, 5, 7, 9, 10, 12, 20, 25 y 40 y en particular de entre las secuencias identificadas en las SEC ID nº 4, 5 y 7, o de entre las secuencias que presentan por lo menos el 99% de identidad con las secuencias identificadas en las SEC ID nº 4, 5, 7, 9, 10, 12, 20, 25 y 40, y en particular de entre las secuencias que presentan por lo menos 99% de identidad con las secuencias identificadas en las SEC ID nº 4, 5 y 7.

En otro modo de realización más del procedimiento de la invención, se detecta además el producto de expresión de por lo menos una secuencia de ácido nucleico, preferentemente de por lo menos do secuencias de ácidos nucleicos o de tres secuencias de ácido nucleico, siendo dichas secuencias de ácidos nucleicos seleccionadas de entre el grupo de secuencias identificadas en las SEC ID nº 1, 6 y 50, o de entre las secuencias que presentan por lo menos 99% de identidad respectivamente, preferentemente por lo menos 99,5% de identidad y ventajosamente por lo menos 99,6% o por lo menos 99,7% de identidad con las secuencias identificadas en SEC ID nº 1, 6 y 50.

Por muestra biológica, se entiende un tejido, un fluido, unos componentes de dichos tejido y fluido, tales como unas células o unos cuerpos apoptóticos, unas vesículas excretadas, que comprenden en particular unos exosomas y unas microvesículas. A título de ejemplo, la muestra biológica puede proceder de una biopsia de pulmón realizada previamente en un paciente sospechoso de padecer un cáncer de pulmón o proceder de una biopsia practicada sobre un órgano diferente del pulmón en un paciente que presenta metástasis. En este segundo caso, cuando el cambio de expresión del ácido nucleico (marcador molecular) es específico del órgano pulmón, es posible remontarse al cáncer primario, es decir al cáncer de pulmón. La muestra biológica puede también ser un fluido biológico, tal como sangre o una fracción sanguínea (suero, plasma), orina, saliva, líquido cefalorraquídeo, linfa, leche materna, esperma, así como unos componentes de dichos fluidos, en particular vesículas excretadas tales como se han definido anteriormente. Por ejemplo, la detección de un transcrito

específico del tejido pulmón en un exosoma o una microvesícula, originaria de una célula epitelial, es signo o bien de la presencia de un cáncer primario, o bien de unas metástasis, sin que sea necesario realizar una extracción a nivel del órgano.

Figuras

Las figuras 1 y 2 representan el diferencial de expresión observado en el cáncer de pulmón para un conjunto de secuencias HERV. Más precisamente, la figura 1 (clustering) reagrupa de manera exploratoria los elementos HERV que tienen un tropismo de expresión asociado al pulmón normal con respecto al conjunto de los tejidos controles y cancerosos, y la figura 2 muestra las diferencias estadísticas de expresión de elementos HERV entre el pulmón normal y el pulmón tumoral.

Las figuras 3 y 4 muestran la detección de secuencias HERV en dos fluidos biológicos: las orinas y los sueros.

Ejemplos

Ejemplo 1: Identificación de secuencias HERV que presentan un diferencial de expresión en el cáncer de pulmón

Método

La identificación de secuencias HERV que presentan un diferencial de expresión en el cáncer de pulmón se basa en la concepción y la utilización de un chip de ADN de alta densidad en formato GeneChip, denominado HERV-V2, diseñado por los inventores y cuya fabricación se ha subcontratado a la compañía Affymetrix. Este chip contiene unas sondas que corresponden a unas secuencias HERV distintas dentro del genoma humano. Estas secuencias se han identificado a partir de un conjunto de referencias prototípicas recortadas en regiones funcionales (LTR, *gag*, *pol* y *env*) y después, mediante una búsqueda de similitud a escala del genoma humano entero (NCBI 36/hg18), se han identificado, anotado y finalmente reagrupado 10035 locus distintos en un banco de datos denominado HERVgDB3.

Las sondas que entran en la composición del chip se han definido a partir de HERVgDB3 y seleccionadas aplicando un criterio de especificidad de hibridación, cuyo objetivo es excluir del procedimiento de creación las sondas que presentan un riesgo de hibridación elevado con un objetivo no buscado. Para ello, las secuencias de HERVgDB3 se segmentaron en primer lugar mediante un conjunto de 25 nucleótidos (25-meros) solapantes, que conduce a un conjunto de sondas candidatas. El riesgo de hibridación inespecífica se evaluó después para cada sonda candidata realizando unas alineaciones sobre el conjunto del genoma humano con la ayuda del algoritmo KASH (2). Una puntuación experimental sanciona el resultado de la hibridación, adición del impacto del número, del tipo y de la posición de los errores en la alineación. El valor de esta puntuación se correlaciona con el potencial de hibridación diana/sonda. El conocimiento de todos los potenciales de hibridación de una sonda candidata sobre el conjunto del genoma humano permite evaluar su especificidad de captura. Las sondas candidatas que presentan una buena afinidad de captura se conservan y después se agrupan en “probesets” y finalmente se sintetizan sobre el chip HERV-V2.

Las muestras analizadas con la ayuda del chip de alta densidad HERV-V2 corresponden a unos ARN extraídos de tumores y a los ARN extraídos de los tejidos sanos adyacentes de estos tumores. Los tejidos analizados son el pulmón, con controles en el pecho, el ovario, el útero, la próstata, el colon, el testículo y la placenta. En el caso de la placenta, se utilizaron sólo unos tejidos sanos. Para cada muestra, 50 ng de ARN han servido para la síntesis de ADNc utilizando el protocolo de amplificación conocido bajo el nombre de WTO. El principio de la amplificación WTO es el siguiente: se añaden unos cebadores aleatorios, así como unos cebadores que tienen como diana el extremo 3’ del transcrito ARN, antes de una etapa de transcripción inversa seguida de una amplificación lineal y monocatenaria designada como SPIA. Los ADNc se determinan después, se caracterizan y se purifican, y después se fragmentan 2 µg, y se marcan con biotina en el extremo 3’ por la acción de la enzima *terminal transferasa.* El producto diana así preparado se mezcla con oligonucleótidos de control, después se realiza la hibridación según el protocolo recomendado por la compañía Affymetrix. Se muestran entonces los chips y se leen para adquirir la imagen de su fluorescencia. Se realiza un control calidad que se basa en los controles estándar, y un conjunto de indicadores (MAD, MAD-Med plots, RLE) sirven para excluir los chips no conformes a un análisis estadístico.

El análisis de los chips consiste en primer lugar en un pretratamiento de los datos por la aplicación de una corrección del ruido de fondo basada sobre la intensidad de las señales de las sondas triptófanos, seguida de una normalización RMA (3) basada en el método de los cuantiles. A continuación, se realiza una doble corrección de los efectos relacionados con los lotes de experimentos aplicando el método COMBAT (4) con el fin de garantizar que las diferencias de expresión observadas son de origen biológico, y no técnico. En esta etapa, se lleva a cabo un análisis exploratorio de los datos con la ayuda de herramientas de reagrupamiento de datos por partición euclidiana (*clustering*), y finalmente se aplica un análisis estadístico, que utiliza el procedimiento

SAM (5) seguido de una corrección por el porcentaje de falsos positivos (6) y de una supresión de los valores inferiores a 2^6 para la búsqueda de secuencias que presentan un diferencial de expresión entre el estado normal y el estado tumoral de un tejido.

Resultados

El tratamiento de los datos generados por el análisis de los chips a ADN HERV-V2 con la ayuda de este método ha permitido identificar un conjunto de "probesets" que presentan una diferencia de expresión estadísticamente significativa entre el pulmón normal y el pulmón tumoral. Los resultados del *clustering* así como la búsqueda de expresión diferencial dentro de las muestras controles ha demostrado, por otro lado, unos elementos HERV cuya expresión diferencial se asocia específicamente con el pulmón tumoral.

Las secuencias nucleotídicas de los elementos HERV que presentan un diferencial de expresión en el pulmón tumoral se identifican mediante las SEC ID nº 1 a 242, la localización cromosómica de cada secuencia se da en el referencial NCBI 36/hg18, y la mención "tejido diana" (una cruz) apunta a los elementos cuya expresión diferencial se ha observado sólo en la comparación entre pulmón normal y pulmón tumoral (con respecto a las comparaciones dentro de los tejidos controles). También se comunica un valor indicativo de la relación de expresión entre estado normal y estado tumoral, y sirve para ordenar las secuencias únicamente en interés de la presentación.

Ejemplo 2: Detección de secuencias HERV en los fluidos biológicos

Principio

Los inventores han demostrado que se detectan unas secuencias HERV en los fluidos biológicos, lo cual permite, entre otros, caracterizar un cáncer de pulmón recurriendo a una detección a distancia del órgano primario. Se ha llevado a cabo un estudio sobre 20 muestras de orina y 38 muestras de suero que provienen de individuos diferentes.

Los sueros y las orinas se han centrifugado en las condiciones siguientes:

Sueros: 500 g durante 10 minutos a 4ºC. El sobrenadante se ha recuperado y centrifugado de nuevo a 16500 g durante 20 minutos a 4ºC. El sobrenadante de esta segunda centrifugación, desprovisto de células, pero que comprende también unos exosomas, unas microvesículas, unos ácidos nucleicos, unas proteínas, se analizó sobre unos chips. El chip es el chip HERV-V2 utilizado según las modalidades descritas anteriormente.

Orinas: después de recogerla, se centrifuga a 800 g durante 4 minutos a 4ºC. el pelete se recuperó con RNA protect cell reagent™. Después, se centrifuga a 5000 g durante 5 minutos antes de la adición del tampón de lisis sobre el residuo. El chip es el chip HERV-V2 utilizado según las modalidades descritas anteriormente.

Resultados

Un número importante de señales positivas, que incluyen las señales de expresión que corresponden a las secuencias listadas en la tabla anterior, se ha detectado al mismo tiempo en los sobrenadantes de sueros y en los residuos celulares que provienen de orinas, como se ilustra en las figuras 3 y 4. Esto confirma que los fluidos biológicos, en particular el suero y la orina, son una fuente utilizable de material biológico para la detección de secuencias HERV. Está comúnmente aceptado que el límite de positividad es del orden de 2^6, es decir 64.

Referencias bibliográficas

1. Nickerson, D.A., Taylor, S.L., Weiss, K.M., Clark, A.G., Hutchinson, R.G., Stengard, J., Salomaa, V., Vartiainen, E., Boerwinkle, E. y Sing, C.F. (1998) DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nat. Genet., 19, 233-240.

2. Navarro, G. y Raffinot, M. (2002) Flexible Pattern Matching in Strings: Practical On-Line Search Algorithms for Texts and Biological Sequences. Cambridge University Press.

3. Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U. y Speed, T.P. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics (Oxford, England), 4, 249-264.

4. Johnson, W.E., Li, C. y Rabinovic, A. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics (Oxford, England), 8, 118-127.

5. Tusher, V.G., Tibshirani, R. y Chu, G. (2001) Significance analysis of microarrays applied to the ionizing

radiation response. Proceedings of the National Academy of Sciences of the United States of America, 98, 5116-5121.

6. Storey, J.D. y Tibshirani, R. (2003) Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences of the United States of America, 100, 9440-9445.

REIVINDICACIONES

1. Procedimiento para el diagnóstico, *in vitro*, específico del cáncer de pulmón en una muestra biológica extraída de un paciente, que comprende una etapa de detección de por lo menos un producto de expresión de por lo menos una secuencia de ácido nucleico, siendo dicha secuencia nucleica seleccionada de entre las secuencias identificadas en las SEC ID nº 4, 5, 7, 9, 10, 12, 20, 25 y 40 o de entre las secuencias que presentan por lo menos el 99% de identidad con una de las secuencias identificadas en las SEC ID nº 4, 5, 7, 9, 10, 12, 20, 25 y 40.

2. Procedimiento según la reivindicación 1, en el que se detecta además el producto de expresión de por lo menos una secuencia de ácido nucleico seleccionada de entre el grupo de secuencias identificadas en las SEC ID nº 1, 2, 3, 6, 8, 9, 11, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241 y 242 o de entre el grupo de secuencias que presentan por lo menos el 99% de identidad con las secuencias identificadas en las SEC ID nº 1, 2, 3, 6, 8, 9, 11, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74 , 75, 76, 77,78, 79, 80, 81, 82, 83, 84, 85, 86 , 87, 88, 89, 90, 91, 92, 93, 94,95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109,110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137,138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151,152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165,166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179,180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207,208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221,222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241 y 242

3. Procedimiento según la reivindicación 2, en el que se detecta el producto de expresión de por lo menos una, incluso dos, preferentemente tres secuencias de ácido nucleico, siendo dichas secuencias de ácido nucleico seleccionadas de entre el grupo de secuencias identificadas en las SEC ID 1, 6 y 50, o de entre las secuencias que presentan por lo menos el 99% de identidad con las secuencias identificadas en las SEC ID nº 1, 6 y 50.

4. Procedimiento según la reivindicación 1, en el que se detecta el producto de expresión de por lo menos dos secuencias de ácido nucleico, preferentemente de tres secuencias de ácido nucleico, siendo dichas secuencias de ácido nucleico seleccionadas de entre el grupo de secuencias identificadas en las SEC ID nº 4, 5 y 7, o de entre las secuencias que presentan por lo menos 99% de identidad con las secuencias identificadas en las SEC ID nº 4, 5 y 7.

5. Procedimiento según una de las reivindicaciones 1 a 4, en el que el producto de expresión detectado es por lo menos un transcrito ARN o por lo menos un polipéptido.

6. Procedimiento según la reivindicación 5, caracterizado por que el transcrito ARN es por lo menos un ARNm.

7. Procedimiento según la reivindicación 5 o 6, en el que el transcrito ARN, en particular el ARNm es detectado por hibridación, por amplificación o por secuenciación.

8. Procedimiento según la reivindicación 6 o 7, en el que el ARNm se pone en contacto con por lo menos una sonda y/o por lo menos un cebador en unas condiciones predeterminadas que permiten la hibridación, y por que se detecta la presencia o la ausencia de hibridación al ARNm.

9. Procedimiento según la reivindicación 6 o 7, caracterizado por que se preparan unas copias ADN del ARNm, se ponen en contacto las copias ADN con por lo menos una sonda y/o por lo menos un cebador en unas condiciones predeterminadas que permiten la hibridación, y por que se detecta la presencia o la ausencia de hibridación a dichas copias ADN.

10. Procedimiento según la reivindicación 5, en el que se detecta el polipéptido expresado por puesta en contacto con por lo menos una pareja de unión específica de dicho polipéptido, en particular un anticuerpo o un análogo de anticuerpo, una proteína de afinidad o un aptámero.

11. Utilización de por lo menos una secuencia de ácido nucleico, aislada, como marcador molecular para el

diagnóstico *in vitro* específico del cáncer de pulmón, caracterizada por que la secuencia de ácido nucleico consiste en:

(i) una secuencia de ADN seleccionada de entre las secuencias identificadas en las SEC ID nº 4, 5, 7, 9, 10, 12, 20, 25 y 40, o

(ii) por lo menos una secuencia de ADN complementaria de por lo menos una secuencia seleccionada de entre las definidas en (i), o

(iii) por lo menos una secuencia de ADN que presenta por lo menos el 99% de identidad, preferentemente por lo menos 99,5% de identidad y ventajosamente por lo menos 99,6% o por lo menos 99,7% de identidad con una secuencia tal como se define en (i) y (ii), o

(iv) por lo menos una secuencia de ARN que es el producto de transcripción de una secuencia seleccionada de entre las secuencias tales como se definen en (i), o

(v) por lo menos una secuencia de ARN que es el producto de transcripción de una secuencia seleccionada de entre las secuencias que presentan por lo menos el 99% de identidad, preferentemente por lo menos 99,5% de identidad y ventajosamente por lo menos 99,6% o por lo menos 99,7% de identidad con las secuencias tales como se definen en (i).

12. Procedimiento para evaluar la eficacia de un tratamiento del cáncer de pulmón, que comprende una etapa de detección de por lo menos un producto de expresión de por lo menos una secuencia de ácido nucleico, siendo dicha secuencia de ácido nucleico seleccionada de entre las secuencias identificadas en las SEC ID nº 4, 5, 7, 9, 10, 12, 20, 25 y 40 o de entre las secuencias que presentan por lo menos el 99% de identidad respectivamente con las secuencias identificadas en las SEC ID nº 4, 5, 7, 9, 10, 12, 20, 25 y 40.

13. Procedimiento según la reivindicación 12, en el que se detecta además el producto de expresión de por lo menos una, incluso dos secuencias de ácido nucleico, preferentemente tres secuencias de ácido nucleico, siendo dichas secuencias de ácido nucleico seleccionadas de entre el grupo de las secuencias identificadas en las SEC ID nº 1, 6 y 50, o de entre las secuencias que presentan por lo menos el 99% de identidad con las secuencias identificadas en las SEC ID nº 1, 6 y 50.

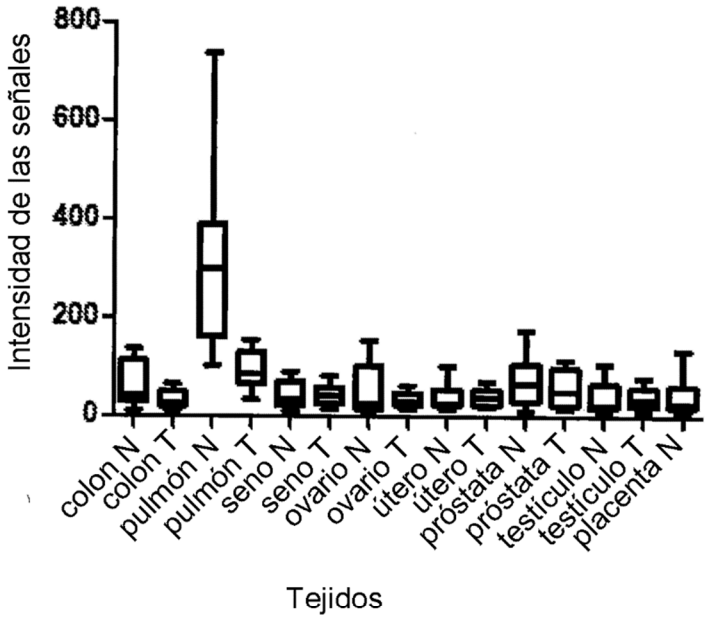

Figura 1

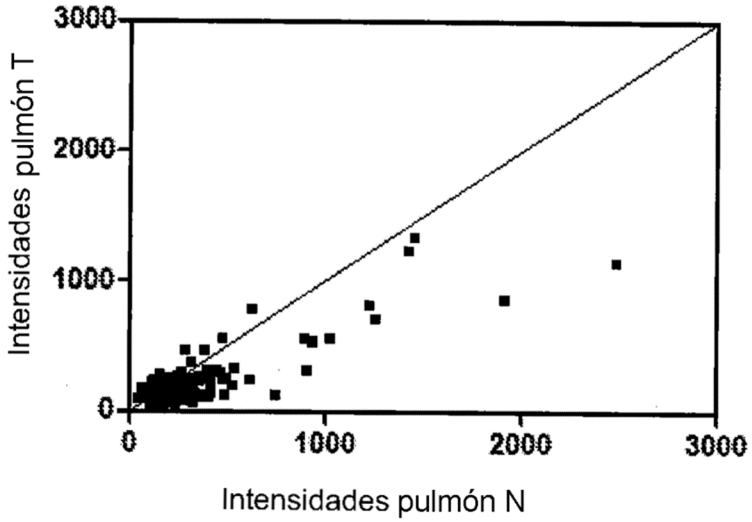

Figura 2

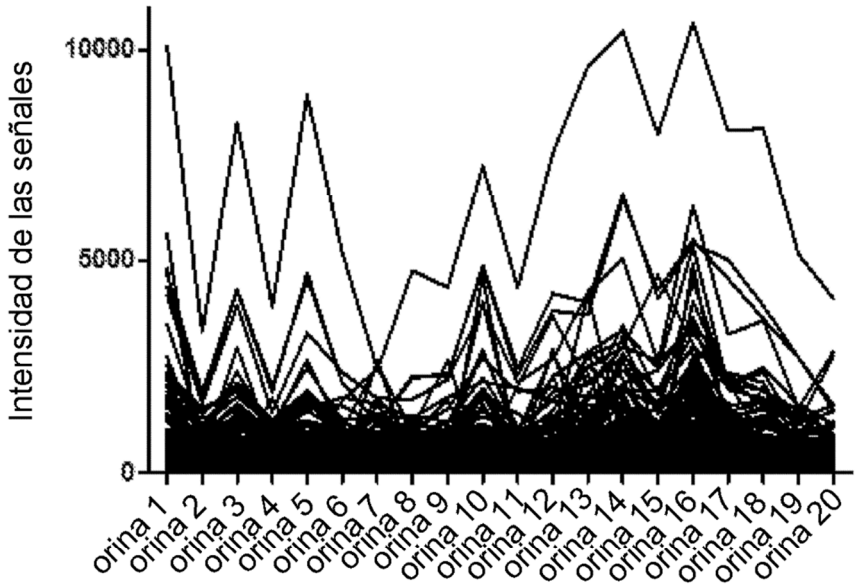

Figura 3

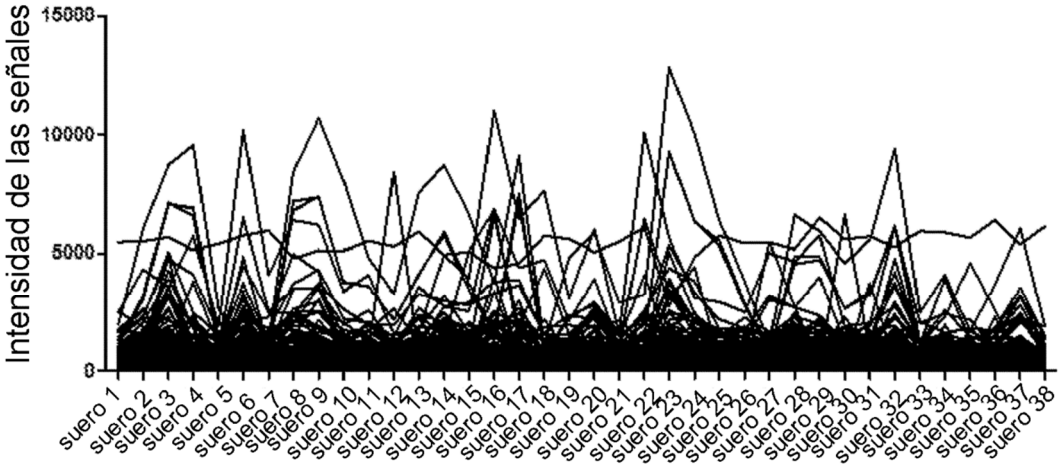

Figura 4