(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 652 497**

(51) Int. CI.:

C12Q 1/68 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **17.12.2012 PCT/FR2012/052954**

(87) Fecha y número de publicación internacional: **27.06.2013 WO13093315**

(96) Fecha de presentación y número de la solicitud europea: **17.12.2012 E 12810390 (0)**

(97) Fecha y número de publicación de la concesión europea: **20.09.2017 EP 2794913**

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

(30) Prioridad:

20.12.2011 FR 1162049

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
02.02.2018

(73) Titular/es:

BIOMÉRIEUX (50.0%)
69280 Marcy l'Étoile, FR y
HOSPICES CIVILS DE LYON (50.0%)

(72) Inventor/es:

MALLET, FRANÇOIS;
MUGNIER, NATHALIE y
PEROT, PHILIPPE

(74) Agente/Representante:

CARPINTERO LÓPEZ, Mario

Observaciones:

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

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

DESCRIPCIÓN

Procedimiento para el diagnóstico *in vitro* del cáncer de colon

Los retrovirus endógenos constituyen la descendencia de retrovirus infecciosos que están integrados en su forma provírica en las células de la línea germinal y que han sido transmitidos de esta forma en el genoma de la descendencia del hospedador.

La secuenciación del genoma humano ha permitido revelar la gran abundancia de elementos transponibles o de sus derivados. De hecho, las secuencias repetidas representan cerca de la mitad del genoma humano, y los retrovirus endógenos y los retrotransposones suponen un 8 % con un número que se eleva, a día de hoy, hasta más de 400.000 elementos.

La abundancia de elementos retrovíricos endógenos (ERV) presentes actualmente en el genoma humano es el resultado de una centena de endogenizaciones logradas en el transcurso de la evolución de la línea humana. Las diferentes olas de endogenización se extienden durante un periodo que va desde los 2 hasta los 90 millones de años antes de nuestra era, y han estado seguidas por la expansión del número de copias mediante fenómenos de tipo « copiar/pegar » con la posibilidad de la aparición de errores, que conducen, a partir de un provirus ancestral, a la formación de una familia de los HERV, es decir, un conjunto de elementos que presentan homologías en las secuencias. Los elementos más antiguos, los de la familia HERV-L, estarían integrados antes de la aparición de los mamíferos. Hay dos familias HERV-F y HERV-H que aparecen en el periodo en el que hacen su aparición los primeros primates. Las familias HERV-FRD y HERV-K (HML-5) integradas hace entre 40 y 55 millones de años, son específicas de los primates superiores. Por el contrario, las familias HERV-W y HERV-E, por ejemplo, se integraron entre 5 y 10 millones de años más tarde, después de la separación con los simios del nuevo mundo, y son específicas de los Catarrinos (Homínidos y Cercopitecos).

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

Los ERV se han considerado durante mucho tiempo como parásitos o como simples desechos de ADN. No obstante, el impacto de los ERV en el organismo no está limitado únicamente a su participación pasada en el modelado del genoma o en recombinaciones perjudiciales que todavía podrían producirse.

La abundancia y la complejidad estructural de los ERV hacen que los análisis de su expresión sean muy complicados y a menudo difícilmente interpretables. La detección de la expresión de un HERV puede reflejar la activación de la transcripción de uno o de varios *loci* en el seno de una misma familia. El o los *loci* activados pueden variar además en función del tejido y/o del contexto.

Los presentes inventores ahora han descubierto y demostrado que las secuencias de ácidos nucleicos correspondientes a los *loci* identificados de forma precisa de elementos retrovíricos endógenos, están asociadas al cáncer de colon, y que estas secuencias son marcadores moleculares de la patología. Las secuencias identificadas son bien provirus, es decir, unas secuencias que contienen la totalidad o parte de los genes gag, pol y env flanqueados en 5' y en 3' por repeticiones terminales largas (LTR o « Long Terminal Repeat » según la terminología anglosajona), bien la totalidad o parte de las LTR o de los genes aislados. Las secuencias de ADN identificadas están referenciadas respectivamente en las ID. SEC. nº: 1 hasta 285 en la lista de secuencias, su localización cromosómica está identificada en la siguiente tabla (NCBI 36/hg18), así como su expresión, sobreexpresión o subexpresión representadas por la « proporción de expresión » entre la muestra cancerosa y la muestra normal. Cuando la expresión del ácido nucleico o el cambio en la expresión del ácido nucleico es específica del tejido de colon, esta información se indica mediante el símbolo « x » en la columna de tejido objetivo. Esto significa que si una expresión o un cambio en la expresión del ácido nucleico en cuestión es determinada en un compartimento biológico distinto al tejido de colon, esto representa, de lejos, una firma del cáncer de colon. Las secuencias de ADN identificadas como específicas del tejido de colon están referenciadas respectivamente en las ID. SEC. nº: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97 en la lista de secuencias. Las secuencias de ADN identificadas como no específicas del tejido de colon están referenciadas respectivamente en las ID. SEC. nº: 6, 14, 15, 16, 18, 20, 21, 23, 26, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 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, 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, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284 y 285.

Tabla

ID. SEC. nº:	Localización cromosómica	Tejido objetivo	Proporción de expresión cáncer/normal
1	(-) chr 5: 135904531-135912105	x	43,8
2	(-) chr X: 4468515-4474361	x	34,7
3	(-) chr 20: 19680691-19685420	x	19,4
4	(+) chr 4: 183972457-183977066	x	12,5
5	(+) chr 13: 108715439-108721465	x	10,9
6	(-) chr 1: 204451670-204454441		-8,6
7	(-) chr 7: 26030345-26035880	x	8,4
8	(+) chr 14: 73239795-73245370	x	8,3
9	(-) chr 5: 179775292-179781337	x	7,7
10	(-) chr 14: 30785319-30790095	x	7,6
11	(-) chr 3: 45773285-45774220	x	6,7
12	(+) chr 16: 84869202-84872386	x	6,4
13	(-) chr 11: 130128601-130134242	x	5,9
14	(-) chr 10: 6711545-6717509		5,7
15	(+) chr 8: 86591572-86592049		-5,6
16	(+) chr 12: 25207414-25213718		5,6
17	(+) chr 8: 98256433-98262143	x	4,9
18	(+) chr 1: 201726731-201727141		-4,8
19	(-) chr 8: 80460894-80463081	x	4,8
20	(+) chr 8: 105367316-105373215		4,8
21	(+) chr 17: 11815798-11820989		4,7
22	(-) chr 18: 24526989-24532843	x	4,5
23	(+) chr 22: 15472270-15476039		4,2
24	(-) chr 8: 91171628-91177272	x	4,2
25	(+) chr 1: 176275999-176281686	x	4,2
26	(-) chr 19: 48519503-48525376		4,1
27	(+) chr 9: 25658927-25666384	x	3,7
28	(+) chr 13: 34271392-34275485		3,7
29	(-) chr 11: 60237235-60238528		-3,6
30	(+) chr 8: 116896352-116900252		3,2
31	(+) chr 19: 60146916-60147844		3,2
32	(+) chr 2: 188084458-188084785		-3,1
33	(+) chr 10: 98482752-98486365		2,9
34	(-) chr 8: 99671725-99672015		2,9
35	(-) chr 13: 55586741-55590394		2,8
36	(-) chr X: 93840473-93846744		-2,8
37	(-) chr 4: 93411664-93417822		2,5
38	(+) chr 4: 121134267-121140271		2,5
39	(-) chr 5: 135912105-135915610		2,5
40	(-) chr 6: 82110364-82117596		-2,4
41	(-) chr 6: 123945178-123950851		2,4
42	(-) chr 2: 71238414-71246247		2,4
43	(-) chr 1: 171684177-171684627		-2,3
44	(+) chr 10: 65936583-65936972		2,3
45	(+) chr 14: 50429630-50433350		2,3
46	(+) chr 14: 41054824-41060836		2,3
47	(-) chr 5: 34514678-34514916		2,2
48	(+) chr 8: 115366935-115369140		2,2
49	(-) chr 11: 73926652-73927082		2,2
50	(-) chr 14: 22736836-22737207		2,2
51	(-) chr 1: 79726879-79732396		2,2
52	(+) chr 3: 130161351-130164778		2,1
53	(-) chr 12: 11656097-11659027		-2,1
54	(-) chr 17: 72113663-72119188		2,1
55	(-) chr 11: 123241237-123246982		2,1
56	(-) chr 8: 136848614-136854437		2,1
57	(+) chr 3: 135222763-135223084		-2,1
58	(+) chr 12: 94666771-94672799		2,1
59	(+) chr 6: 63559535-63565316		-2,1

(continuación)

ID. SEC. nº:	Localización cromosómica	Tejido objetivo	Proporción de expresión cáncer/normal
60	(-) chr 2: 54587807-54590183		-2,0
61	(+) chr 2: 173115073-173117251		2,0
62	(+) chr 5: 755586-759445		2,0
63	(+) chr 8: 129995887-130001750		2,0
64	(+) chr 4: 127503905-127504662		2,0
65	(-) chr 21: 41718929-41719369		2,0
66	(+) chr 20: 40489715-40490155		2,0
67	(+) chr 7: 100274888-100276065		2,0
68	(+) chr 7: 35702274-35703153		2,0
69	(+) chr 6: 131686129-131689771		-2,0
70	(-) chr 1: 226879232-226884835		2,0
71	(-) chr 10: 17712042-17714165		1,9
72	(-) chr 10: 53463521-53469327		1,9
73	(-) chr X: 112238600-112244308		1,9
74	(-) chr 1: 146722974-146723946		1,9
75	(+) chr 1: 53671142-53673436		1,9
76	(-) chr X: 91414738-91420449		-1,9
77	(-) chr 7: 16683220-16684043		1,9
78	(+) chr 10: 100097920-100098685		1,9
79	(+) chr 22: 31228741-31234490		1,8
80	(-) chr 13: 112501931-112502014		1,8
81	(+) chr 1: 146832410-146833382		1,8
82	(-) chr 10: 20449793-20453869		1,8
83	(+) chr 9: 21911892-21921934		1,8
84	(-) chr 3: 186574589-186580188		-1,8
85	(-) chr 10: 62463482-62469677		-1,8
86	(-) chr 15: 97959745-97960384		-1,7
87	(+) chr 17: 75988189-75996283		-1,7
88	(+) chr 2: 88760597-88762729		1,7
89	(+) chr 7: 29791133-29794522		1,7
90	(+) chr 18: 50528195-50528890		-1,7
91	(+) chr 1: 89162808-89170138		-1,7
92	(-) chr 3: 178866314-178871614		-1,7
93	(-) chr 19: 52245264-52250506		1,7
94	(+) chr 17: 30852752-30853047		-1,7
95	(-) chr 11: 69581214-69582655		1,7
96	(+) chr 5: 138886787-138888726		-1,7
97	(-) chr 3: 167701085-167706900	x	1,7
98	(+) chr 1: 101404629-101404696		1,7
99	(+) chr 8: 113277279-113282810		-1,6
100	(-) chr 20: 740491-741481		1,6
101	(-) chr 2: 58194071-58199769		1,6
102	(+) chr 13: 35788462-35794330		-1,6
103	(+) chr 9: 12938344-12944129		1,6
104	(-) chr 9: 116976750-116982362		-1,6
105	(-) chr 4: 11264289-11269976		-1,6
106	(-) chr 1: 223094264-223100407		-1,6
107	(+) chr 7: 92862465-92866875		1,6
108	(-) chr 17: 72296838-72297401		-1,6
109	(+) chr X: 86823336-86830473		1,6
110	(-) chr X: 89718079-89718557		-1,6
111	(-) chr 3: 58397972-58398349		-1,6
112	(-) chr 2: 215413351-215413804		-1,6
113	(+) chr 4: 87614948-87620699		1,6
114	(-) chr 1: 54874382-54875358		-1,6
115	(+) chr 1: 237429952-237430254		1,6
116	(-) chr 7: 79651365-79652053		-1,6
117	(-) chr 4: 180711319-180712287		-1,5
118	(-) chr 5: 146727162-146727562		-1,5

(continuación)

ID. SEC. nº:	Localización cromosómica	Tejido objetivo	Proporción de expresión cáncer/normal
119	(+) chr 3: 156178559-156179784		-1,5
120	(+) chr 1: 51465235-51468757		-1,5
121	(-) chr 3: 100178327-100178459		1,5
122	(-) chr 4: 67932305-67932737		1,5
123	(-) chr 18: 24532993-24533490		1,5
124	(-) chr 12: 20862483-20866818		1,5
125	(-) chr 12: 8334466-8337953		-1,5
126	(-) chr 13: 94759022-94759378		-1,5
127	(-) chr 1: 144779633-144780605		1,5
128	(+) chr 1: 143314771-143315743		1,5
129	(+) chr 2: 231645891-231645949		1,5
130	(-) chr 7: 65721380-65722348		-1,5
131	(+) chr 2: 188606470-188612115		-1,5
132	(-) chr 3: 176879333-176879730		-1,5
133	(-) chr 3: 186765708-186772209		1,5
134	(-) chr 11: 67377518-67381008		1,5
135	(-) chr 12: 123157590-123158004		1,5
136	(+) chr 7: 134507781-134510522		-1,5
137	(+) chr 12: 112980007-112980079		-1,5
138	(-) chr 10: 101571639-101577563		1,5
139	(+) chr 8: 81814506-81817982		-1,5
140	(-) chr 12: 29178820-29185001		-1,5
141	(-) chr 5: 119558433-119559245		-1,5
142	(-) chr 13: 23891789-23892035		1,5
143	(-) chr 1: 51845967-51846990		-1,5
144	(-) chr 19: 58630272-58630359		1,5
145	(-) chr 3: 193273756-193277341		-1,5
146	(+) chr 6: 149002601-149008339		-1,5
147	(+) chr 3: 186986788-186987217		1,5
148	(-) chr 3: 192866741-192872861		1,5
149	(+) chr 4: 69952306-69955060		-1,4
150	(-) chr 14: 38837505-38837576		1,4
151	(+) chr 2: 57102769-57103266		-1,4
152	(-) chr 3: 130032422-130036475		-1,4
153	(+) chr 4: 54112155-54112817		-1,4
154	(-) chr 3: 146311752-146313676		1,4
155	(-) chr 12: 14769094-14769447		1,4
156	(+) chr 13: 55512386-55518412		1,4
157	(+) chr 15: 72433538-72439180		-1,4
158	(+) chr 3: 40577391-40583175		-1,4
159	(-) chr 4: 120675771-120676048		-1,4
160	(-) chr 8: 12395268-12398823		-1,4
161	(-) chr X: 101106329-101106710		1,4
162	(-) chr 3: 190138663-190139459		-1,4
163	(-) chr 3: 98473995-98479084		-1,4
164	(+) chr 10: 5070454-5070498		-1,4
165	(+) chr 5: 114710170-114711150		1,4
166	(+) chr 3: 128381736-128382403		-1,4
167	(+) chr 9: 22813029-22813838		-1,4
168	(-) chr 2: 231673651-231673769		1,4
169	(+) chr 3: 116946112-116948448		1,4
170	(-) chr 11: 92482373-92482685		1,4
171	(+) chr 4: 167867005-167868159		1,4
172	(-) chr 1: 144779633-144780605		1,4
173	(+) chr 20: 58157372-58158054		1,4
174	(-) chr 4: 138339335-138340015		-1,4
175	(-) chr 12: 71688770-71689742		1,4
176	(+) chr 13: 63310225-63315979		-1,4
177	(-) chr 1: 110201971-110202287		-1,4
178	(+) chr 10: 92557026-92562997		-1,4

(continuación)

ID. SEC. nº:	Localización cromosómica	Tejido objetivo	Proporción de expresión cáncer/normal
179	(+) chr 1: 38146074-38152175		-1,4
180	(+) chr 6: 153045773-153046224		1,4
181	(-) chr 1: 79934615-79940335		-1,4
182	(+) chr 4: 155197148-155204405		-1,4
183	(+) chr 13: 40927858-40928571		1,4
184	(+) chr 4: 11974611-11980403		-1,3
185	(+) chr 18: 4676671-4682409		-1,3
186	(-) chr 21: 14575132-14576104		-1,3
187	(+) chr 5: 180186875-180187838		-1,3
188	(+) chr 4: 68579259-68585888		-1,3
189	(-) chr 11: 27607193-27613049		-1,3
190	(+) chr 18: 6123610-6123972		-1,3
191	(+) chr 14: 57658752-57662615		-1,3
192	(-) chr 1: 144213068-144214036		-1,3
193	(-) chr 4: 62312529-62318338		-1,3
194	(+) chr 1: 45751770-45752724		-1,3
195	(-) chr 17: 6577008-6581495		-1,3
196	(-) chr 1: 244819996-244820439		-1,3
197	(-) chr 9: 100067414-100073409		1,3
198	(+) chr 3: 196356095-196356926		-1,3
199	(+) chr 12: 31851416-31851846		-1,3
200	(-) chr 5: 118311697-118316179		-1,3
201	(-) chr 1: 94806343-94810970		-1,3
202	(+) chr 2: 185332604-185332980		-1,3
203	(+) chr 4: 107250443-107255274		-1,3
204	(-) chr 18: 31472178-31472680		-1,3
205	(+) chr 6: 38501905-38502660		-1,3
206	(-) chr 15: 49734758-49735530		1,3
207	(-) chr X: 84258326-84258441		-1,3
208	(+) chr 5: 55618832-55619003		-1,3
209	(-) chr 17: 16678919-16679896		1,3
210	(+) chr 2: 193715387-193716358		-1,3
211	(+) chr 5: 115014416-115019003		1,3
212	(+) chr 17: 22039285-22039750		-1,3
213	(+) chr 5: 108685641-108685853		-1,3
214	(+) chr 10: 65474521-65475382		-1,3
215	(-) chr 6: 89180631-89187940		1,3
216	(+) chr 1: 244312610-244313561		1,3
217	(+) chr 5: 76822592-76823033		-1,3
218	(-) chr Y: 4283432-4289271		-1,3
219	(+) chr 5: 71075676-71076376		-1,3
220	(+) chr 2: 24861412-24861769		-1,3
221	(-) chr 1: 110196377-110201971		-1,3
222	(+) chr 3: 177261846-177267467		-1,3
223	(+) chr 7: 6908153-6909119		-1,3
224	(-) chr 12: 33132053-33135058		-1,3
225	(-) chr 7: 146321019-146321687		1,3
226	(+) chr 4: 54581003-54585874		-1,3
227	(+) chr X: 138823266-138823672		-1,3
228	(+) chr 18: 2829330-2829995		-1,3
229	(+) chr 6: 66335631-66336029		-1,3
230	(-) chr 6: 121983265-121989145		-1,3
231	(-) chr 2: 226039745-226040459		-1,3
232	(+) chr 4: 109853300-109855458		1,3
233	(+) chr 19: 20126744-20133095		-1,3
234	(-) chr 11: 14839752-14840420		-1,2
235	(-) chr 3: 20380459-20380671		-1,2
236	(+) chr 4: 53395726-53396112		-1,2
237	(+) chr 19: 9553879-9554314		1,2
238	(-) chr 2: 63928103-63928550		-1,2

(continuación)

ID. SEC. nº:	Localización cromosómica	Tejido objetivo	Proporción de expresión cáncer/normal
239	(+) chr 3: 187581597-187582311		-1,2
240	(-) chr 5: 76211738-76218598		-1,2
241	(+) chr 17: 18280923-18281897		1,2
242	(-) chr 4: 178566070-178566839		-1,2
243	(+) chr 12: 60267154-60272981		-1,2
244	(+) chr 18: 27623185-27623837		-1,2
245	(-) chr 5: 135094585-135101074		-1,2
246	(+) chr 10: 4962301-4962771		-1,2
247	(-) chr 5: 153996685-154003250		-1,2
248	(-) chr 6: 5981329-5981780		-1,2
249	(-) chr 8: 828875-829287		-1,2
250	(+) chr 18: 38295804-38303710		-1,2
251	(-) chr 3: 54643549-54649271		-1,2
252	(-) chr 18: 68050597-68051144		-1,2
253	(-) chr 3: 46260865-46268560		-1,2
254	(-) chr 21: 40033061-40040223		-1,2
255	(+) chr 1: 148879269-148880889		1,2
256	(-) chr 2: 44343693-44344377		1,2
257	(-) chr 11: 80087903-80088662		-1,2
258	(+) chr 3: 46140179-46140255		-1,2
259	(+) chr 10: 45388242-45391541		-1,2
260	(-) chr 10: 52210114-52210776		1,2
261	(-) chr 3: 163923299-163927462		-1,2
262	(-) chr 18: 22761239-22762201		-1,2
263	(+) chr 5: 64505763-64506756		-1,2
264	(-) chr 16: 18909966-18916124		-1,2
265	(+) chr 5: 119114744-119115548		-1,2
266	(+) chr 1: 69101725-69102525		-1,2
267	(+) chr 1: 161827723-161832945		-1,2
268	(+) chr 1: 223140681-223147414		-1,2
269	(+) chr Y: 26648032-26648192		-1,2
270	(-) chr 2: 195907646-195913044		-1,2
271	(-) chr 7: 54710388-54711318		-1,2
272	(-) chr 8: 34837245-34837699		1,2
273	(+) chr 4: 27585609-27586295		-1,2
274	(+) chr 6: 160569673-160575346		-1,2
275	(-) chr 2: 167330763-167331232		-1,1
276	(-) chr 21: 16686151-16692725		-1,1
277	(-) chr 3: 180722967-180726830		-1,1
278	(+) chr 13: 61603944-61604298		-1,1
279	(-) chr Y: 15720288-15724345		-1,1
280	(-) chr 1: 115355475-115355826		1,1
281	(+) chr Y: 26648032-26648192		-1,1
282	(+) chr 11: 94682231-94690615		-1,1
283	(-) chr 18: 62088302-62089092		1,1
284	(+) chr 19: 49834194-49835177		-1,1
285	(-) chr 11: 69737702-69743910		-1,1

La presente invención tiene por lo tanto por objeto un procedimiento para el diagnóstico, *in vitro,* del cáncer de colon en una muestra biológica recogida de un paciente, que comprende una etapa de detección de al menos un producto de expresión de al menos una secuencia de ácido nucleico, eligiéndose dicha secuencia de ácido nucleico entre las secuencias identificadas en las ID. SEC. nº: 1, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97 o entre las secuencias que presentan una identidad de al menos el 99 %, preferiblemente una identidad de al menos el 99,5 % y ventajosamente una identidad de al menos el 99,6 % o de al menos el 99,7 % con una de las secuencias identificadas en las ID. SEC. nº: 1, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97.

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 estadios), 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 más arriba ha sido determinando teniendo en consideración la diversidad nucleotídica del 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. Como ejemplo, Nickerson D. A. y al. (1) han demostrado una diversidad de aproximadamente el 0,3 % (el 0,32 %) en las regiones que contienen secuencias repetidas.

La capacidad de discriminación de un estado canceroso de cada una de las secuencias identificadas más arriba ha sido evidenciada con la ayuda de un análisis estadístico que utiliza el procedimiento SAM (5) seguido de una corrección del índice de falsos positivos (6) y de una supresión de los valores inferiores a 2^6. Por consiguiente, cada una de las secuencias identificadas más arriba presenta una diferencia de expresión significativa entre un estado tumoral y un estado normal. Como resultado, una diferencia de expresión observada para una de las secuencias mencionadas anteriormente constituye una firma de la patología. Según la invención, es posible combinar las diferencias de expresión mencionadas para varias de las secuencias referenciadas más arriba, por ejemplo, para una o las asociaciones de 2, 3, 4, 5, 6, 7, 8, 9, 10 incluso más, y hasta 285 de las secuencias recogidas, preferiblemente para una o las asociaciones de 2, 3, 4, 5, 6, 7, 8, 9, 10 incluso más, y hasta 18 de las secuencias identificadas respectivamente en las ID. SEC. nº: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97, y para las que presentan una identidad de al menos el 99 % con las secuencias identificadas en las ID. SEC. nº: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97.

En particular las secuencias identificadas en las ID. SEC. nº: 2, 3 y 8, solas o en combinación (de 2 a 2 o 3) constituyen una o las firmas privilegiadas.

En un modo de realización del procedimiento según la invención se detecta el producto de expresión de al menos dos secuencias de ácido nucleico, eligiéndose dichas al menos dos secuencias de ácido nucleico entre las secuencias identificadas como específicas de tejido de colon, es decir, elegidas entre el grupo de secuencias identificadas en las ID. SEC. nº: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97 o entre las secuencias que presentan una identidad de al menos el 99 %, preferiblemente una identidad de al menos el 99,5 % y ventajosamente una identidad de al menos el 99,6 % o de al menos el 99,7 % con una de las secuencias identificadas en las ID. SEC. nº: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97.

En otro modo de realización del procedimiento de la invención, se detecta el producto de expresión de al menos una secuencia elegida entre las secuencias identificadas como específicas de tejido de colon, es decir, elegida entre el grupo de secuencias identificadas en las ID. SEC. nº: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97 o elegida entre las secuencias que presentan una identidad de al menos el 99 %, preferiblemente una identidad de al menos el 99,5 % y ventajosamente una identidad de al menos el 99,6 % o de al menos el 99,7 % con una de las secuencias identificadas en las ID. SEC. nº: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97, y el producto de expresión de al menos una secuencia elegida entre las secuencias identificadas como no específicas de tejido de colon, es decir, elegida entre el grupo de secuencias identificadas en las ID. SEC. nº: 6, 14, 15, 16, 18, 20, 21,23, 26, 28, 29, 30, 31,32, 33, 34, 35, 36, 37, 38, 39, 40, 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, 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, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271,272,273, 274, 275, 276, 277, 278, 279, 280, 281,282, 283, 284 y 285 o elegida entre las secuencias que presentan una identidad de al menos el 99 %, preferiblemente una identidad de al menos el 99,5 % y ventajosamente una identidad de al menos el 99,6 % o de al menos 99,7 % con una de las secuencias identificadas en las ID. SEC. nº: 6, 14, 15, 16, 18, 20, 21,23, 26, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 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, 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, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284 y 285.

En particular, en el procedimiento de la invención se detecta el producto de expresión de al menos una secuencia de ácido nucleico, preferiblemente de al menos dos secuencias de ácido nucleico o de tres secuencias de ácido nucleico, eligiéndose dichas secuencias de ácido nucleico entre el grupo de secuencias identificadas en las ID. SEC. nº: 2, 3 y 8, o entre las secuencias que presentan una identidad de al menos el 99 %, preferiblemente una identidad

de al menos el 99,5 % y ventajosamente una identidad de al menos el 99,6 % o de al menos el 99,7 % con las secuencias identificadas en las ID. SEC. nº: 2, 3 y 8.

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

Cuando el producto de expresión es un transcrito de ARNm, es detectado mediante cualquier procedimiento apropiado, tal como una hibridación, una secuenciación o una amplificación. El ARNm puede ser detectado directamente poniéndolo en contacto con al menos una sonda y/o al menos un cebador, que están concebidos para hibridar en unas condiciones experimentales predeterminadas con los transcritos de ARNm, evidenciando la presencia o la ausencia de hibridación con el ARNm y eventualmente la cuantificación del ARNm. Entre los procedimientos preferidos se pueden citar la amplificación (por ejemplo, la RT-PCR, la NASBA, etc.), la hibridación en chip o incluso la secuenciación. El ARNm puede ser detectado igualmente indirectamente a partir de ácidos nucleicos derivados de dichos transcritos, como las copias de ADNc, etc.

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

Así, el procedimiento puede comprender:

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

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

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

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

(iii) una etapa de comparación de la cantidad de ARNm expresado en la muestra que se va a analizar y en la muestra de referencia; la determinación de una cantidad de ARNm expresado en la muestra que se va a analizar diferente de la cantidad de ARNm expresado en la muestra de referencia sana puede ser correlacionada con el diagnóstico o el pronóstico de gravedad de un cáncer de colon (siendo la diferencia en la cantidad de ARNm en el tejido de colon canceroso con respecto a la cantidad de ARNm expresado en el tejido de colon sano indiferentemente una expresión, una sobreexpresión o una subexpresión);

y en particular:

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

(ii) una determinación, en el ARN que se va a analizar, de un nivel de expresión de al menos una secuencia de ARN en la muestra, preferiblemente de al menos dos secuencias de ARN en la muestra, siendo la secuencia de ARN y las secuencias de ARN respectivamente el producto de la transcripción de al menos una secuencia de ácido nucleico y de al menos dos secuencias de ácidos nucleicos elegidas entre las secuencias identificadas previamente, y

(iii) una comparación del nivel de expresión de la o de las secuencia(s) de ARN definida(s) en (ii) con un nivel de expresión de referencia; pudiendo ser correlacionada la determinación de un nivel de expresión del ARN que se va a analizar, que presenta una diferencia con respecto al nivel de expresión de referencia, con el diagnóstico o el pronóstico de un cáncer de colon (como se ha determinado más arriba); o

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

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

(iii) una etapa de comparación de la cantidad de ARNm expresado en la muestra que se va a analizar con respecto a una cantidad de ARNm de referencia, pudiendo ser correlacionada la determinación de una cantidad de ARNm expresado en la muestra que se va a analizar, diferente de la cantidad de ARNm de referencia, con el diagnóstico o el pronóstico de un cáncer de colon (siendo la diferencia en la cantidad de ARNm en la muestra que se va 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 copias de ADN del ARNm, se ponen en contacto las copias de ADN con al menos una sonda y/o al menos un cebador en unas condiciones predeterminadas que permiten la hibridación y porque se detecta la presencia o la ausencia de hibridación con dichas copias de ADN.

El producto de expresión que se detecta también puede ser un polipéptido, que es el producto de la traducción de al menos uno de los transcritos descritos más arriba. En ese caso, se detecta el polipéptido expresado poniéndolo en contacto con al menos un compañero de unión específico de dicho polipéptido, en particular un anticuerpo o un

análogo de anticuerpo o un aptámero. El compañero de unión es preferiblemente un anticuerpo, por ejemplo, un anticuerpo monoclonal o un anticuerpo policlonal muy purificado, o un análogo de un anticuerpo, por ejemplo, una proteína de afinidad con propiedades competitivas (nanofitine™).

Los anticuerpos policlonales pueden ser obtenidos mediante la inmunización de un animal con el inmunógeno apropiado, seguida de la recuperación de los anticuerpos buscados en forma purificada, mediante la recogida del suero de dicho animal, y la separación de dichos anticuerpos de otros constituyentes del suero, particularmente mediante una cromatografía de afinidad en una columna en la cual se ha fijado un antígeno que es reconocido específicamente por los anticuerpos.

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

En un primer momento, se inmuniza un animal, generalmente un ratón, con el inmunógeno apropiado, cuyos linfocitos B son entonces capaces de producir anticuerpos contra este antígeno. Estos linfocitos productores de anticuerpos se fusionan a continuación con células mielomatosas "inmortales" (murinas en el ejemplo) para dar lugar a hibridomas. A partir de la mezcla heterogénea de células así obtenida, se efectúa después una selección de las células capaces de producir un anticuerpo en particular y de multiplicarse indefinidamente. Cada hibridoma es multiplicado en forma de un clon, que conduce cada uno a la producción de un anticuerpo monoclonal cuyas propiedades de reconocimiento con respecto a la proteína podrían ensayarse, por ejemplo, en un ELISA, mediante una inmunotransferencia (Wéstern blot) en una o en dos dimensiones, una inmunofluorescencia o con la ayuda de un biocapturador. Los anticuerpos monoclonales así seleccionados son purificados a continuación, particularmente según la técnica de cromatografía de afinidad descrita más arriba.

Los anticuerpos monoclonales pueden ser igualmente anticuerpos recombinantes obtenidos mediante ingeniería genética, a través de técnicas bien conocidas por el experto en la materia.

Las nanofitines™ son pequeñas proteínas que, como los anticuerpos, son capaces de unirse a un objetivo biológico, permitiendo así su detección, su captura o simplemente su direccionamiento en el seno de un organismo. Se presentan, entre otros, como análogos de anticuerpos.

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

La invención concierne igualmente a la utilización de al menos una secuencia de ácido nucleico aislada como marcador molecular para el diagnóstico *in vitro* del cáncer de colon, caracterizada porque dicha secuencia de ácido nucleico consiste en:

(i) al menos una secuencia de ADN elegida entre las secuencias ID. SEC. nº: 1, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97, o

(ii) al menos una secuencia de ADN complementaria de una secuencia elegida entre las secuencias ID. SEC. nº: 1, , 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97, o

(iii) al menos una secuencia de ADN que presenta una identidad de al menos el 99 %, preferiblemente una identidad de al menos el 99,5 % y ventajosamente una identidad de al menos el 99,6 % o de al menos el 99,7 % con una secuencia tal como la definida en (i) y (ii), o

(iv) al menos una secuencia de ARN que es el producto de transcripción de una secuencia elegida entre las secuencias tales como las definidas en (i), o

(v) al menos una secuencia de ARN que es el producto de transcripción de una secuencia elegida entre las secuencias que presentan una identidad de al menos el 99 %, preferiblemente una identidad de al menos el 99,5 % y ventajosamente una identidad de al menos el 99,6 % o de al menos el 99,7 % con una secuencia tal como la definida en (i).

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

(i) al menos dos secuencias de ADN elegidas entre las secuencias identificadas en las ID. SEC. nº: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97, y en particular las secuencias ID. SEC nº: 2, 3 y 8, o

(ii) al menos dos secuencias de ADN respectivamente complementarias de al menos dos secuencias elegidas entre las secuencias identificadas en las ID. SEC. nº: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97, y en particular elegidas entre las secuencias ID. SEC nº: 2, 3 y 8, o

(iii) al menos dos secuencias de ADN que presentan respectivamente una identidad de al menos el 99 %, preferiblemente una identidad de al menos el 99,5 % y ventajosamente una identidad de al menos el 99,6 % o de al menos el 99,7 % con dos secuencias tales como las definidas en (i) y (ii), o

(iv) al menos dos secuencias de ARN que son respectivamente el producto de transcripción de dos secuencias

elegidas entre las secuencias tales como las definidas en (i), o

(v) al menos dos secuencias de ARN que son el producto de transcripción de dos secuencias elegidas entre las secuencias que presentan una identidad de al menos el 99 %, preferiblemente una identidad de al menos el 99,5 % y ventajosamente una identidad de al menos el 99,6 % o de al menos el 99,7 % con las secuencias tales como las definidas en (i).

La invención concierne también a un procedimiento para la evaluación de la eficacia de un tratamiento del cáncer de colon que comprende una etapa de obtención de una serie de muestras biológicas, una etapa de detección de al menos un producto de expresión de al menos una secuencia de ácido nucleico en dicha serie de muestras biológicas, eligiéndose dicha secuencia de ácido nucleico entre las secuencias identificadas en las ID. SEC. nº: 1, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97 o entre las secuencias que presentan una identidad de al menos el 99 %, preferiblemente una identidad de al menos el 99,5 % y ventajosamente una identidad de al menos el 99,6 % o de al menos el 99,7 % con las secuencias identificadas en las ID. SEC. nº 1, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97.

En un modo de realización, se detectan al menos dos productos de expresión de al menos dos secuencias de ácido nucleico, eligiéndose dichas dos secuencias de ácido nucleico entre las secuencias identificadas en las ID. SEC. nº: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97 o entre las secuencias que presentan una identidad de al menos el 99 % respectivamente, preferiblemente una identidad de al menos el 99,5 % y ventajosamente una identidad de al menos el 99,6 % o de al menos el 99,7 % con las secuencias identificadas en las ID. SEC. nº: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97.

En otro modo de realización del procedimiento, se detecta el producto de expresión de al menos una secuencia de ácido nucleico, preferiblemente de al menos dos secuencias de ácidos nucleicos o de tres secuencias de ácido nucleico, eligiéndose dichas secuencias de ácido nucleicos entre el grupo de secuencias identificadas en las ID. SEC. nº 2, 3 y 8 o entre las secuencias que presentan una identidad de al menos el 99 % respectivamente, preferiblemente una identidad de al menos el 99,5 % y ventajosamente una identidad de al menos el 99,6 % o de al menos el 99,7 % con las secuencias identificadas en las ID. SEC. nº: 2, 3 y 8.

Por muestra biológica se entiende un tejido, un fluido, los componentes de dicho tejido y fluido, tales como células o cuerpos apoptóticos, vesículas excretadas, que comprenden particularmente exosomas y microvesículas. Como ejemplo, la muestra biológica puede proceder de una biopsia de colon realizada previamente en un paciente sospechoso de padecer un cáncer de colon, o proceder de una biopsia practicada en un órgano distinto al colon en un paciente que presenta metástasis. En este segundo caso, cuando el cambio en la expresión del ácido nucleico (marcador molecular) es específico del órgano colon, es posible remontar hasta el cáncer primario, es decir, al cáncer de colon. La muestra biológica puede ser igualmente un fluido biológico, tal como sangre o una fracción sanguínea (suero, plasma), orina, saliva, líquido cefalorraquídeo, linfa, leche materna, esperma, al igual que los componentes de dichos fluidos, en particular vesículas excretadas tales como las definidas más arriba. Por ejemplo, la detección de un transcrito específico del tejido de colon en un exosoma o en una microvesícula, procedente de una célula epitelial, indica bien la presencia de un cáncer primario o bien de metástasis, sin que sea necesario realizar una toma de muestra en el órgano.

Figuras:

Las figuras 1 y 2 representan el diferencial de expresión observado en el cáncer de colon para un conjunto de secuencias del HERV. De forma más precisa, la figura 1 (clustering) reagrupa de forma exploratoria los elementos del HERV que tienen un tropismo de expresión asociado al cáncer de colon con respecto al conjunto de tejidos controlados, y la figura 2 muestra las diferencias estadísticas de expresión de los elementos del HERV entre el colon normal y el colon tumoral.

Las figuras 3 y 4 muestran la detección de las secuencias del HERV en dos fluidos biológicos: orina y suero.

La figura 5 presenta los resultados de la expresión obtenidos en una RT-PCR cuantitativa de las secuencias SEQ ID nº 2, SEQ ID nº 3 y SEQ ID nº 8, para 28 tumores de colon y 18 tejidos sanos de colon, y ponen especialmente en evidencia el interés de su combinación para aumentar el número de detecciones en las muestras de tumores colorrectales.

Ejemplos:

Ejemplo 1: identificación de las secuencias del HERV que presentan un diferencial de expresión en el cáncer de colon

Procedimiento:

La identificación de las secuencias del HERV que presentan un diferencial de expresión en el cáncer de colon se basa en la concepción y la utilización de un chip de ADN de alta densidad con el formato GeneChip, denominado HERV-V2, concebido por los inventores y cuya fabricación ha sido subcontratada a la compañía Affymetrix. Este

chip contiene sondas que se corresponden con las secuencias del HERV características en el seno del genoma humano. Estas secuencias se han identificado a partir de un conjunto de referencias prototípicas cortadas en regiones funcionales (LTR, *gag, pol* y *env)* y después mediante una investigación de la similitud a la escala de la totalidad del genoma humano (NCBI 36/hg18), se han identificado 10035 *loci* del HERV característicos, se han anotado y finalmente se han reagrupado en un banco de datos denominado HERVgDB3.

Las sondas que forman parte de la composición del chip se han definido a partir del HERVgDB3 y se han seleccionado aplicando un criterio de especificidad de hibridación, cuyo objetivo es excluir del procedimiento de creación aquellas sondas que presentan un riesgo de hibridación elevado con un objetivo no buscado. Para ello, en primer lugar se han segmentado las secuencias del HERVgDB3 en conjuntos de 25 nucleótidos (25-mers) solapantes, dando lugar a un conjunto de sondas candidatas. El riesgo de hibridación no específica se ha evaluado a continuación para cada sonda candidata realizando alineaciones sobre el conjunto del genoma humano con la ayuda del algoritmo KASH (2). Una puntuación experimental sanciona el resultado de la hibridación, la adición del impacto del número, del tipo y de la position de los errores en la alineación. El valor de esta puntuación se correlaciona con el potencial de hibridación objetivo/sonda. El conocimiento de todos los potenciales de hibridación de una sonda candidata sobre el conjunto del genoma humano permite la simulación de su especificidad de captura. Las sondas candidatas que presentan una buena afinidad de captura son conservadas y después reagrupadas en « probesets » y finalmente sintetizadas en el chip HERV-V2.

Las muestras analizadas con ayuda del chip de alta densidad HERV-V2 se corresponden con los ARN extraídos a partir de tumores y con los ARN extraídos a partir de tejidos sanos adyacentes a estos tumores. Los tejidos analizados son el colon, con controles de mama, de ovario, de útero, de próstata, de pulmón, de testículo y de placenta. En el caso de la placenta únicamente se utilizan tejidos sanos. Para cada muestra, 50 ng de ARN han servido para la síntesis del ADNc utilizando el protocolo de amplificación conocido con el nombre de WTO. El principio de la amplificación WTO es el siguiente: se añaden cebadores aleatorios, así como cebadores que se dirigen al extremo 3' del transcrito de ARN, antes de una etapa de transcripción inversa, seguida de una amplificación lineal y una hebra simple denominada SPIA. Los ADNc son medidos, caracterizados y purificados, y después se fragmentan 2 µg, se marcan con biotina en el extremo 3' por la acción de la enzima *terminal transferase.* El producto objetivo así preparado se mezcla con oligonucleótidos de control, después se lleva a cabo la hibridación según el protocolo recomendado por la compañía Affymetrix. Después se revelan los chips y se leen para adquirir la imagen de su fluorescencia. Se realiza un control de calidad que se basa en los controles de calidad estándares, y un conjunto de indicadores (MAD, MAD-Med plots, RLE) sirven para excluir los chips que no son conformes con un análisis estadístico.

El análisis de los chips consiste en primer lugar en un pretratamiento de los datos mediante la aplicación de una corrección del ruido de fondo basada en la intensidad de las señales de las sondas de triptófano, seguida por una normalización RMA (3) basada en el método de los cuantiles. Después se lleva a cabo 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 en la expresión observadas tienen un origen biológico y no técnico. En este estadio, se lleva a cabo un análisis exploratorio de los datos con ayuda de herramientas de reagrupación de los datos mediante un particionamiento euclidiano *(clustering),* y finalmente un análisis estadístico utilizando el procedimiento SAM (5) seguido de una corrección según el índice de falsos positivos (6) y se aplica 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 mediante el análisis de los chips de ADN HERV-V2 con la ayuda de este procedimiento ha permitido la identificación de un conjunto de « probesets » que presenta una diferencia de expresión estadísticamente significativa entre el colon normal y el colon tumoral. Los resultados del *clustering,* así como la investigación de la expresión diferencial en el seno de las muestras de control, han evidenciado además elementos del HERV cuya expresión diferencial está asociada específicamente con el colon tumoral.

Las secuencias nucleotídicas de los elementos del HERV que presentan un diferencial de expresión en el colon tumoral son identificadas mediante las SED ID nº 1 a 285, la localización cromosómica de cada secuencia se proporciona en el referencial NCBI 36/hg18, y la mención « tejido objetivo » (una cruz) indica los elementos cuya expresión diferencial no se ha observado más que en la comparación entre el colon normal y el colon tumoral (con respecto a las comparaciones en el seno de los tejidos de control). Igualmente se comunica un valor indicativo de la proporción de expresión entre el estado normal y el estado tumoral, y sirve para clasificar las secuencias únicamente con fines de presentación.

Ejemplo 2: detección de las secuencias del HERV en los fluidos biológicos

Principio:

Los inventores han demostrado que las secuencias del HERV son detectadas en los fluidos biológicos, lo que permite, entre otras, la caracterización de un cáncer de colon recurriendo a una detección a distancia del órgano

primario. Se ha llevado a cabo un estudio con 20 muestras de orina y 38 muestras de suero procedentes de individuos diferentes.

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

sueros: a 500 g durante 10 minutos a 4 ºC. El sobrenadante se ha recuperado y se ha centrifugado de nuevo a 16.500 g durante 20 minutos a 4 ºC. El sobrenadante de esta segunda centrifugación, desprovisto de células, pero que todavía comprende exosomas, microvesículas, ácidos nucleicos, proteínas, se ha analizado en los chips. El chip es el chip HERV-V2 utilizado según las modalidades descritas previamente.

orinas: después de su recogida, una centrifugación a 800 g durante 4 minutos a 4 ºC. El sedimento se ha recuperado con RNA protect cell reagent™. Después, una centrifugación a 5.000 g durante 5 minutos antes de la adición del tampón de lisis sobre el sedimento. El chip es el chip HERV-V2 utilizado según las modalidades descritas previamente.

Resultados:

Se ha detectado un número importante de señales positivas, incluyendo las señales de expresión correspondientes a las secuencias recogidas en la siguiente tabla, a la vez en los sobrenadantes de los sueros y en los sedimentos celulares procedentes de las orinas, según 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 del HERV. Está comúnmente aceptado que el umbral de positividad es del orden de 2^6, es decir, de 64.

Ejemplo 3: puesta en evidencia del interés de una combinación de dos, y de tres, secuencias del HERV para la detección del cáncer colorrectal.

Procedimiento:

Se han extraído los ARN procedentes de 28 tumores de colon (CT1, CT2, CT3, CT65, CT94, CC1, CC2, CC3, CC4, CC5, CC6, CC7, CC8, CC9, CC10, 549T, 551T, 553T, 556T, 558T, 559T, 560T, 561T, 601T, 602T, 604T, 607T y 615T) y de 18 tejidos sanos de colon (CN1, CN2, CN3, CN64, CN93, 549N, 551N, 553N, 556N, 558N, 559N, 560N, 561N, 601N, 602N, 604N, 607N y 615N) y se han utilizado para ayudar a demostrar que la combinación de los resultados de la expresión de dos, o de tres, secuencias del HERV, permite aumentar el número de detecciones en el seno de las muestras de tumores colorrectales.

Se ha llevado a cabo una amplificación mediante una RT-PCR cuantitativa correspondiente a las regiones 1723-1824 de la SEQ ID nº 2, 1540-1665 de la SEQ ID nº 3 y 1684-1923 de la SEQ ID nº 8, para el conjunto de las 46 muestras del estudio. Los resultados de la expresión de las secuencias del HERV se han tratado mediante el método de delta Ct y se han normalizado para los resultados de la expresión del gen constitutivo de la GAPDH. Los resultados se presentan en la figura 5.

Resultados:

La amplificación mediante una RT-PCR cuantitativa de la región 1723-1824 de la SEQ ID nº 2 es positiva para 11 muestras de tumores de colon (CT3, CT94, CC1, CC2, 558T, 559T, 560T, 561T, 602T, 604T y 607T). No se ha observado ninguna detección de esta región para las muestras de tejidos sanos de colon.

La amplificación mediante una RT-PCR cuantitativa de la región 1540-1665 de la SEQ ID nº 3 es positiva para 6 muestras de tumores de colon (CT1, CT3, CT94, CC5, CC8 y CC10). No se ha observado ninguna detección de esta región para las muestras de tejidos sanos de colon.

La amplificación mediante una RT-PCR cuantitativa de la región 1684-1923 de la SEQ ID nº 8 es positiva para 4 muestras de tumores de colon (CT3, CC2, CC6 y CC10). No se ha observado ninguna detección de esta región para las muestras de tejidos sanos de colon.

El tener en cuenta de forma aditiva las detecciones de la región 1723-1824 de la SEQ ID nº 2 y de la región 1540-1665 de la SEQ ID nº 3 permite una cobertura de 15 muestras de tumores de colon (CT1, CT3, CT94, CC1, CC2, CC5, CC8, CC10, 558T, 559T, 560T, 561T, 602T, 604T y 607T), mostrando así una ganancia en la sensibilidad de esta combinación de dos secuencias del HERV con respecto a la utilización de estas mismas secuencias del HERV consideradas por separado.

El tener en cuenta de forma aditiva las detecciones de la región 1723-1824 de la SEQ ID nº 2 y de la región 1684-1923 de la SEQ ID nº 8 permite una cobertura de 13 muestras de tumores de colon (CT3, CT94, CC1, CC2, CC6, CC10, 558T, 559T, 560T, 561T, 602T, 604T y 607T), mostrando así una ganancia en la sensibilidad de esta combinación de dos secuencias del HERV con respecto a la utilización de estas mismas secuencias del HERV consideradas por separado.

El tener en cuenta de forma aditiva las detecciones de la región 1540-1665 de la SEQ ID nº 3 y de la región 1684-1923 de la SEQ ID nº 8 permite una cobertura de 8 muestras de tumores de colon (CT1, CT3, CT93, CC2, CC5,

CC6, CC8 y CC10), mostrando así una ganancia en la sensibilidad de esta combinación de dos secuencias del HERV con respecto a la utilización de estas mismas secuencias del HERV consideradas por separado.

Finalmente, el tener en cuenta de forma aditiva las detecciones de la región 1723-1824 de la SEQ ID nº 2 y de la región 1540-1665 de la SEQ ID nº 3 y de la región 1684-1923 de la SEQ ID nº 8 permite una cobertura de 16 muestras de tumores de colon (CT1, CT3, CT94, CC1, CC2, CC5, CC6, CC8, CC10, 558T, 559T, 560T, 561T, 602T, 604T y 607T), mostrando así una ganancia en la sensibilidad de esta combinación de tres secuencias del HERV con respecto a la utilización de estas mismas secuencias del HERV consideradas por separado o en combinación de dos en dos.

Referencias bibliográficas

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

2. Navarro, G. and 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, KJ., Scherf, U. and 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. and Rabinovic, A. (2007) Adjusting batch effects in microarray expression data using empirical Bayes métodos. Biostatistics (Oxford, England), 8, 118-127.

5. Tusher, V. G., Tibshirani, R. and 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. and 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 colon en una muestra biológica recogida de un paciente que comprende una etapa de detección de al menos un producto de expresión de al menos una secuencia de ácido nucleico elegida entre las secuencias identificadas en las ID. SEC. nº: 1, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97 o entre las secuencias que presentan una identidad de al menos el 99 % con una de las secuencias identificadas en las ID. SEC. nº: 1, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97.

2. Procedimiento según la reivindicación 1, en el que se detecta el producto de expresión de al menos dos secuencias de ácido nucleico, eligiéndose dichas al menos dos secuencias de ácido nucleico entre el grupo de secuencias identificadas en las ID. SEC. nº: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97 o entre las secuencias que presentan una identidad de al menos el 99 % con las secuencias identificadas en las ID. SEC. nº: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97.

3. Procedimiento según la reivindicación 2, en el que una de las dos secuencias es la ID. SEC. nº: 2 o una secuencia que presenta una identidad de al menos el 99 % con la ID. SEC. nº: 2.

4. Procedimiento según la reivindicación 2 o 3, en el que se detecta el producto de expresión de al menos dos secuencias de ácido nucleico, preferiblemente de tres secuencias de ácido nucleico, eligiéndose dichas secuencias de ácido nucleico entre el grupo de secuencias identificadas en las ID. SEC. nº 2, 3 y 8, o entre las secuencias que presentan una identidad de al menos el 99 % con las secuencias identificadas en las ID. SEC. nº: 2, 3 y 8.

5. Procedimiento según una cualquiera de las reivindicaciones 1 a 3, **caracterizado porque** se detecta, además, el producto de expresión de al menos una secuencia de ácido nucleico elegida entre las secuencias identificadas en las ID. SEC. nº: 6, 14 a 16, 18, 20, 21, 23, 26, 28 a 96, 98 a 285 y las secuencias que presentan una identidad de al menos el 99 % con las secuencias identificadas en las ID. SEC. nº: 6, 14 a 16, 18, 20, 21, 23, 26, 28 a 96, 98 a 285.

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

7. Procedimiento según la reivindicación 6, **caracterizado porque** el transcrito de ARN es al menos un ARNm.

8. Procedimiento según la reivindicación 6 o 7, en el que el transcrito de ARN, en particular el ARNm, es detectado mediante una hibridación, una amplificación o una secuenciación.

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

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

11. Procedimiento según la reivindicación 6, en el que se detecta el polipéptido expresado poniéndolo en contacto con al menos un compañero de unión específico de dicho polipéptido, en particular un anticuerpo o un análogo de un anticuerpo, una proteína de afinidad o un aptámero.

12. Uso de al menos una secuencia de ácido nucleico, aislado, como marcador molecular para el diagnóstico *in vitro* específico del cáncer de colon, en una muestra biológica recogida de un paciente, **caracterizado porque** la secuencia de ácido nucleico consiste en

(i) al menos una secuencia de ADN elegida entre las secuencias identificadas en las ID. SEC. nº: 1, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97, o
(ii) al menos una secuencia de ADN complementario de al menos una secuencia elegida entre las secuencias identificadas en las ID. SEC. nº: 1, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97, o
(iii) al menos una secuencia de ADN que presenta una identidad de al menos el 99 % con una secuencia tal como la definida en (i) y (ii), o
(iv) al menos una secuencia de ARN que es el producto de transcripción de una secuencia elegida entre las secuencias tales como las definidas en (i), o
(v) al menos una secuencia de ARN que es el producto de transcripción de al menos una secuencia elegida entre las secuencias que presentan una identidad de al menos el 99 % con las secuencias tales como las definidas en (i).

13. Uso según la reivindicación 12 de al menos dos secuencias de ácido nucleico que consisten en:

(i) al menos dos secuencias de ADN elegidas entre las secuencias identificadas en las ID. SEC. nº: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97 y en particular entre las secuencias identificadas en las ID. SEC. nº: 2, 3 y 8, o

(ii) al menos dos secuencias de ADN complementarias respectivamente de al menos dos secuencias elegidas entre las secuencias identificadas en las ID. SEC. nº: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97 y en particular entre las secuencias identificadas en las ID. SEC. nº: 2, 3 y 8, o
(iii) al menos dos secuencias de ADN que presentan una identidad de al menos el 99 % con dos secuencias tales como las definidas en (i) y (ii), o
(iv) al menos dos secuencias de ARN que son respectivamente el producto de transcripción de dos secuencias elegidas entre las secuencias tales como las definidas en (i), o
(v) al menos dos secuencias de ARN que son el producto de transcripción de al menos dos secuencias elegidas entre las secuencias que presentan una identidad de al menos el 99 %: con las secuencias tales como las definidas en (i).

14. Procedimiento para evaluar la eficacia de un tratamiento del cáncer de colon que comprende una etapa de detección, en una muestra biológica recogida de un paciente, de al menos un producto de expresión de al menos una secuencia de ácido nucleico, eligiéndose dicha secuencia de ácido nucleico entre las secuencias identificadas en las ID. SEC. nº: 1, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97 o entre las secuencias que presentan una identidad de al menos el 99 % respectivamente con las secuencias identificadas en las ID. SEC. nº: 1, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97.

15. Procedimiento según la reivindicación 14, en el que se detecta el producto de expresión de al menos dos secuencias de ácido nucleico, eligiéndose dichas al menos dos secuencias de ácido nucleico entre el grupo de secuencias identificadas en las ID. SEC. nº: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97 o entre las secuencias que presentan una identidad de al menos el 99 % con las secuencias identificadas en las ID. SEC. nº: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 17, 19, 22, 24, 25, 27 y 97.

16. Procedimiento según la reivindicación 15, en el que una de las dos secuencias es la ID. SEC. nº: 2 o una secuencia que presenta una identidad de al menos el 99 % con la ID. SEC. nº: 2.

17. Procedimiento según la reivindicación 15 o 16, en el que se detecta el producto de expresión de al menos dos secuencias de ácido nucleico, preferiblemente de tres secuencias de ácido nucleico, eligiéndose dichas secuencias de ácido nucleicos entre el grupo de secuencias identificadas en las ID. SEC. nº 2, 3 y 8, o entre las secuencias que presentan una identidad de al menos el 99 % con las secuencias identificadas en las ID. SEC. nº: 2, 3 y 8.

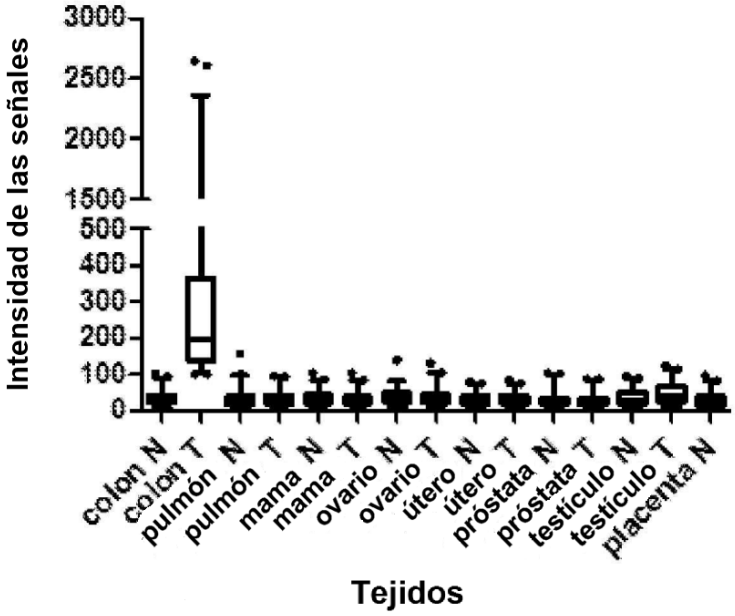

Figura 1

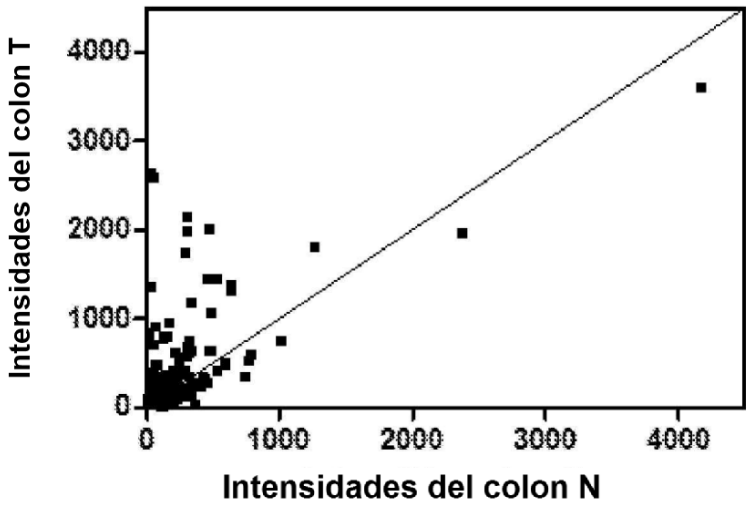

Figura 2

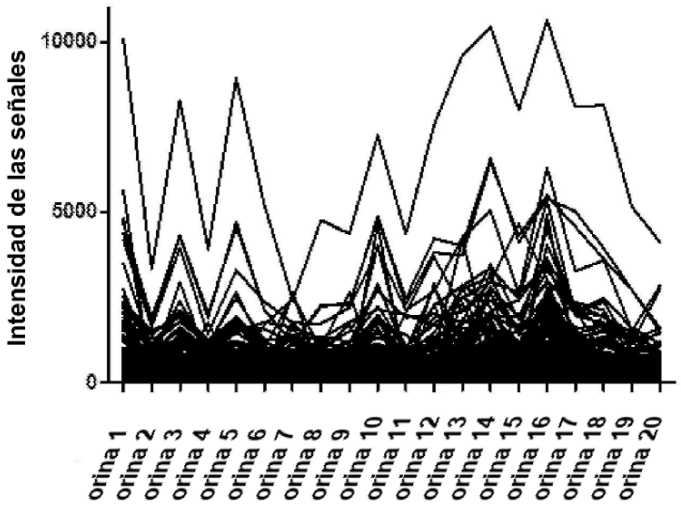

Figura 3

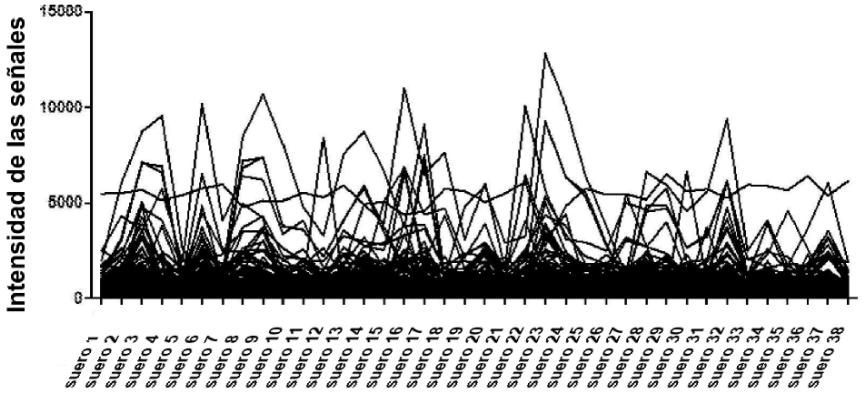

Figura 4

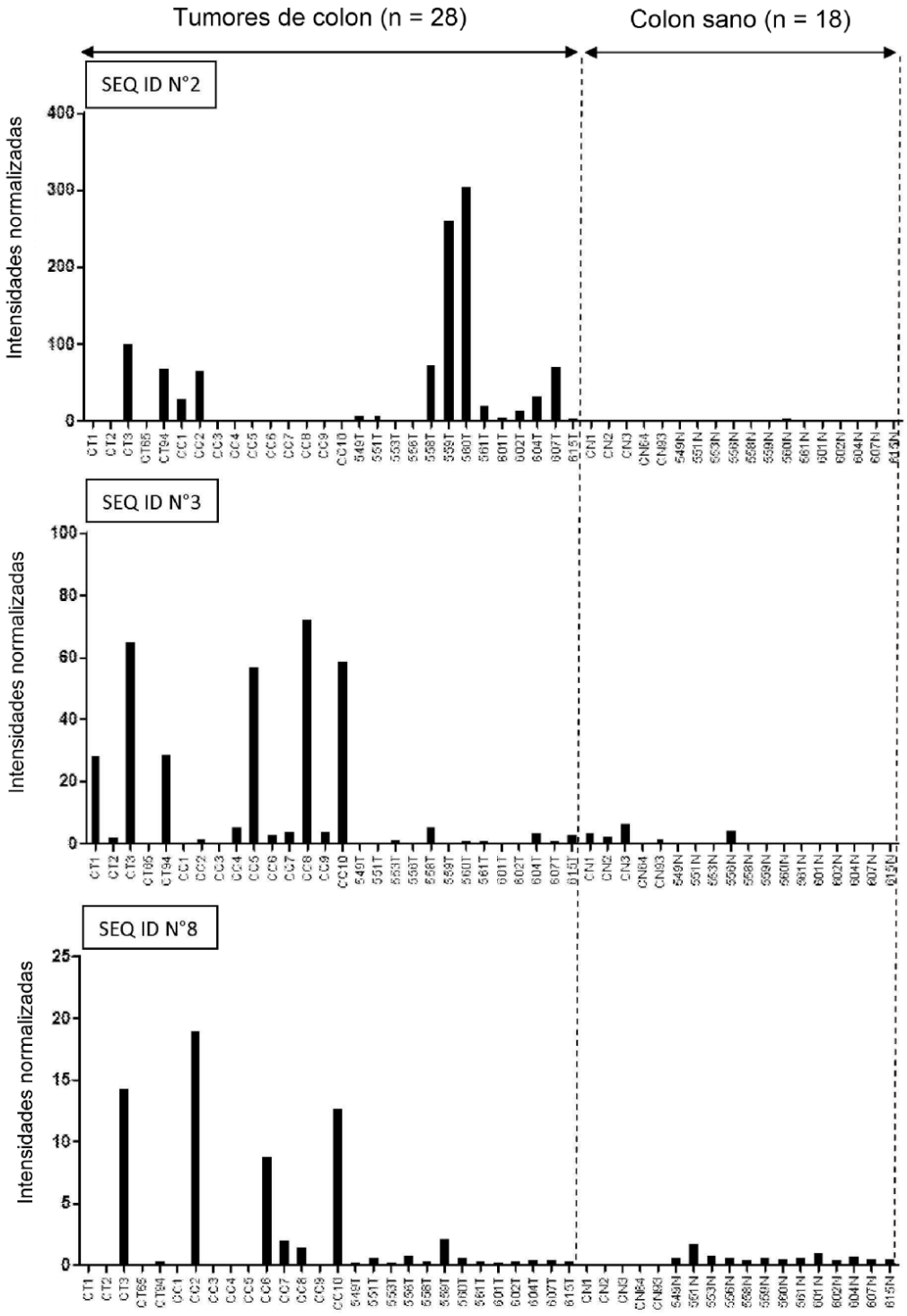
Tumores de colon (n = 28)
Colon sano (n = 18)
SEQ ID N°2
SEQ ID N°3
SEQ ID N°8
Intensidades normalizadas

Figura 5