

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FARMACIA
Departamento de Bioquímica y Biología Molecular II



**ANÁLISIS DE LA EXPRESIÓN GÉNICA GLOBAL
EN PACIENTES CON CÁNCER COLORRECTAL:
CLASIFICACIÓN, SIGNIFICADO BIOLÓGICO Y
RELEVANCIA CLÍNICA.**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR

Alejandro Romera López

Bajo la dirección de los doctores

Beatriz Pérez-Villamil Salgado
Eduardo Díaz-Rubio García

Madrid, 2010

ISBN: 978-84-693-4795-9

© Alejandro Romera López, 2009

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE FARMACIA
DPTO. DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR II



**ANÁLISIS DE LA EXPRESIÓN GÉNICA GLOBAL
EN PACIENTES CON CÁNCER COLORRECTAL:
CLASIFICACIÓN, SIGNIFICADO BIOLÓGICO Y
RELEVANCIA CLÍNICA**

TESIS DOCTORAL

ALEJANDRO ROMERA LÓPEZ

MADRID, 2009

HOSPITAL CLÍNICO SAN CARLOS
LABORATORIO DE ONCOLOGÍA MOLECULAR



Hospital Clínico San Carlos



Madrid

**ANÁLISIS DE LA EXPRESIÓN GÉNICA GLOBAL
EN PACIENTES CON CÁNCER COLORRECTAL:
CLASIFICACIÓN, SIGNIFICADO BIOLÓGICO Y
RELEVANCIA CLÍNICA**

DIRECTORES:

D^{RA.} BEATRIZ PÉREZ-VILLAMIL SALGADO

P^{ROF. DR.} EDUARDO DÍAZ-RUBIO GARCÍA

El presente estudio ha sido realizado en el Laboratorio de Oncología Molecular perteneciente al Servicio de Oncología Médica del Hospital Clínico San Carlos (Madrid) y en colaboración con el departamento de Bioinformática del Instituto de Salud Carlos III.

Trabajo presentado en el Departamento de Bioquímica y Biología Molecular II de la Facultad de Farmacia de la Universidad Complutense de Madrid por Alejandro Romera López para optar al grado de Doctor en Bioquímica y Biología Molecular. Madrid 2009.

AGRADECIMIENTOS

Durante la realización de estas tesis, son muchas las personas que me han brindado su apoyo y a las que sin duda debo un sincero agradecimiento:

- ❖ A mis directores de tesis, la Dra. Beatriz Pérez-Villamil y el Dr. Eduardo Díaz Rubio, por haber apostado por mí para la realización de este trabajo y por haberme apoyado en todo momento. Gracias Beatriz por enseñarme lo que es la investigación.
- ❖ A la Dra. Trinidad Caldés por permitirme formar parte de su grupo y por su ayuda durante estos años.
- .
- ❖ A la gente del labo, Ana, Juanma, Alicia, Pilar, Carmen, Atocha, Inma, Paula, Vanesa, Patricia, Olivia, Verónica y el resto de gente que ha pasado por aquí y con la que he podido contar para cualquier cosa. Gracias en especial a Miguel, por haberme enseñado gran parte de lo que se sabe sobre ciencia, por haberme culturizado un poco durante las sobremesas y por tu paciencia infinita (siempre con gesto amable a pesar de haberme repetido la misma cosa 100 veces).
- ❖ A la parte “clínica” y no “clínica” del departamento. Gracias Antonio y Javier por ayudarme a comprender un poco mejor esos jeroglíficos que son las historias de los pacientes.
- ❖ Al servicio de Anatomía Patológica y sobre todo a Susana, Jose, Julián. Gracias Susana por estar siempre dispuesta a ayudar con la mejor de las sonrisas y por el trato “preferencial” que me has dado.
- ❖ Al departamento de Bioinformática del Carlos III y en especial a Guillermo, Fernando y a Oscar. Gracias Guillermo por introducirme al “maravilloso” mundo del análisis de datos y por estar siempre dispuesto a contestar mis preguntas y dudas malévolas.

- ❖ Al Dr. Eschrich S y al Dr. Yeatman J por habernos cedido amablemente su matriz de expresión y sus datos de supervivencia, haciendo posible así, una parte importante de esta tesis.
- ❖ A los chicos del servicio de Inmunología y sobre todo a Elena, Jose Luís y Alfonso, siempre dispuestos a ayudar y a compartir los pequeños placeres de “Mediterránea de Catering”.
- ❖ Al departamento de Bioquímica y Biología Molecular II de la Facultad de Farmacia, con el Dr. Manuel Benito a la cabeza por permitirme presentar la tesis en su departamento y por su colaboración en todos los trámites.
- ❖ Gracias a mis padres, a Oscar, a Alberto, a Cris (¡¡qué larga ha sido mi tesis!!) y al resto de la familia por aguantarme, con mis humos, durante estos años.
- ❖ A Rafa (y sus rencores), David, Benito, Tere, Cris (y todos los respectivos. Lo siento Mayo, te toca ser novio de) y a todos lo demás amigos del colegio, por haberme soportado y animado durante estos años.
- ❖ A Ramón, Bea, Fran, Edu, Jana, Elena, Sandra, Marta, Garde y demás gente de la uni. Por esos momentos en Palacios de Goda o Navas de Oro, por esas cervezas, licores de madroño y esos San Patricios.
- ❖ A la gente de inglés (en especial a Laura y a Arancha) y de francés (merci ma sorcière pour me demander toujours si ça va), por preocuparos por mi en los momentos de bajón.
- ❖ A los olvidados. Porque seguro que hay personas que se merecen este agradecimiento y de las cuales ahora mismo no me acuerdo. Espero que no me lo tengáis en cuenta.

**ANÁLISIS DE LA EXPRESIÓN GÉNICA GLOBAL
EN PACIENTES CON CÁNCER COLORRECTAL:
CLASIFICACIÓN, SIGNIFICADO BIOLÓGICO Y
RELEVANCIA CLÍNICA**

Listado de Abreviaturas

- ❖ #: Número.
- ❖ #-NN: #-Nearest-Neighbor.
- ❖ ‘ : minutos.
- ❖ 5-FU: 5-Fluorouracilo.
- ❖ A: Adenina.
- ❖ AJCC: American Joint Committee on Cancer.
- ❖ ANOVA: Analysis of Variance.
- ❖ B&H: Benjamini and Hochberg method.
- ❖ BG: Background.
- ❖ C: Citosina.
- ❖ CCP: Compound Covariate Predictor.
- ❖ CCR: Cáncer colorrectal.
- ❖ cDNA: Complementary/copy DNA.
- ❖ CEA: Carcinoembryonic Antigen.
- ❖ CGH: Comparative Genome Hybridization.
- ❖ CIMP: CpG Island Methylator Phenotype.
- ❖ CIMP-0: CpG Island Methylator Phenotype-null.
- ❖ CIMP-H: CpG Island Methylator Phenotype-High.
- ❖ CIN: Chromosomal Instability.
- ❖ CN: Colon Normal.
- ❖ CNV: Copy Number Variation.
- ❖ cRNA: Complementary/copy RNA.
- ❖ CT: Colon Tumor.
- ❖ Cy-3/Cy-5: Cianina 3/Cianina 5.
- ❖ DAB: Diamino Bencidina.
- ❖ D-Index: Discrepancy Index.
- ❖ DLDA: Diagonal Linear Discriminant Analysis.
- ❖ DNA: Desoxiribonucleic Acid.
- ❖ DS: Desviación estándar.
- ❖ DWD: Distance Weighted Discrimination method.
- ❖ FE: Feature Extraction.
- ❖ FISH: Fluorescence In Situ Hybridization.

- ❖ FOLFIRI: Leucovorin + 5-Fluorouracilo + Irinotecan.
- ❖ G: Guanina.
- ❖ GL: Ganglios Linfáticos.
- ❖ GO: Gene Ontology.
- ❖ GSEA: Gene Set Enrichment Analysis.
- ❖ HC: Hierarchical Clustering..
- ❖ HCSC: Hospital Clínico San Carlos.
- ❖ HNPCC: Hereditary Non Polyposis Colorectal Cancer.
- ❖ HR: Hazard Ratio.
- ❖ HRP: Peroxidasa de rábano.
- ❖ IC: Intervalo de confianza.
- ❖ ICC: Intra-Class Correlation.
- ❖ IHQ: Inmunohitoquímica.
- ❖ KEGG: Kyoto Encyclopedia of Genes and Genomes.
- ❖ KW: Kruskal Wallis Test.
- ❖ LOH: Loss of Heterozigosity.
- ❖ LOOCV: Leave One Out Cross Validation.
- ❖ LOWESS: Locally Weighted Scatterplot Smoothing.
- ❖ LR: Likelihood Ratio Test.
- ❖ LV: Leucovorin.
- ❖ M: Metástasis.
- ❖ MAQC: Microarray Quality Control.
- ❖ MAT: Macrófagos Activados por Tumor.
- ❖ ME: Matriz de Ensayo.
- ❖ MEC: Matriz Extracelular.
- ❖ MHC: Major Histocompatibility Complex.
- ❖ MIAME: Minimum Information About Microarrays Experiment.
- ❖ miRNA: MicroRNA.
- ❖ MMR: MisMatch Repair.
- ❖ mRNA: Messenger RNA.
- ❖ MSI: Microsatellite Instability.
- ❖ MSI-H: Microsatellite Instability-High.
- ❖ MSI-L: Microsatellite Instability-Low.

- ❖ MSS: Microsatellite Stability.
- ❖ MV: Matriz de Validación.
- ❖ N: Grado de Infiltración del tumor en los ganglios linfáticos.
- ❖ NC: Nearest Centroid.
- ❖ OMS: Organización Mundial de la Salud.
- ❖ OXL: Oxaliplatino.
- ❖ PAF: Poliposis Adenomatosa Familiar.
- ❖ pb: Pares de Bases.
- ❖ PCA: Principal Component Analysis.
- ❖ PCP: Planar Cell Polarity.
- ❖ PCR: Polimerase Chain Reaction.
- ❖ PLE: Período Libre de Enfermedad.
- ❖ PMT: PhotoMultiplier Tube.
- ❖ QC: Quality Control.
- ❖ QT: Quimioterapia.
- ❖ R: Recurrencia.
- ❖ RIN: RNA Integrity Number.
- ❖ R-Index: Reproducibility Index.
- ❖ RNA: Ribonucleic Acid.
- ❖ RT: Retro-Transcripción.
- ❖ RTx: Radioterapia.
- ❖ S: Sano
- ❖ SC: Scan Control software.
- ❖ SG: Supervivencia Global.
- ❖ SNP: Single Nucleotide Polimorfism.
- ❖ SOM: Self Organizing Maps.
- ❖ SVM: Support Vector Machines.
- ❖ T: Grado de infiltración de la pared del intestino.
- ❖ TEM: Transición Epitelio-Mesénquima.
- ❖ Ti: Timina.
- ❖ TMA: Tissue-MicroArray.
- ❖ TNM: Tumor-Node-Metastasis.
- ❖ Treg: Linfocitos T reguladores.

- ❖ VPN: Valor Predictivo Negativo.
- ❖ VPP: Valor Predictivo Positivo.
- ❖ W: indica Adenina o Timina.
- ❖ Wt: Wild Type.
- ❖ B-ME: beta-Mercapto-Etanol.
- ❖ χ : test Chi-Cuadro.

ÍNDICE

Introducción.

1. Cáncer de colon	2
1.1. Incidencia	2
1.2. Cáncer esporádico vs. cáncer familiar	3
1.3. Factores de riesgo no genéticos	4
1.4. Clasificación del cáncer colorrectal	4
1.5. Marcadores pronósticos y predictivos	7
1.5.1. <i>Marcadores histopatológicos</i>	7
1.5.2. <i>Marcadores moleculares</i>	8
1.6. Modelos moleculares del cáncer colorrectal	11
1.6.1. <i>Modelo general del desarrollo del CCR de tipo esporádico.</i>	
<i>Fenotipo CIN</i>	12
1.6.2. <i>Fenotipo mutador (MSI) y fenotipo metilador (CIMP)</i>	14
1.6.3. <i>Vía Wnt</i>	17
2. Microentorno. Relevancia en la progresión tumoral	19
2.1. Células del sistema inmune	20
2.2. Fibroblastos	21
2.3. Inflamación	22
2.4. Matriz Extracelular (MEC) y Membrana Basal	24
3. Tratamientos en el cáncer colorrectal esporádico	25
3.1. 5-Fluorouracilo (5-FU)	26
3.2. Capecitabina	26
3.3. UFT	26
3.4. Leucovorin (LV)	27
3.5. Oxaliplatino (OXL)	27
3.6. Irinotecan (CPT-11)	28
3.7. Gefitinib y Erlotinib	28
3.8. Cetuximab	29
3.9. Bevacizumab	30
4. Microarrays	31
4.1. Microarrays de expresión. Introducción	32
4.1.1. <i>Fabricación de los microarrays</i>	32
4.1.2. <i>Marcaje e hibridación de las muestras</i>	34

4.1.3. <i>Análisis de datos</i>	35
4.2. <i>Fiabilidad y reproducibilidad de la técnica de los microarrays</i>	38
4.3. <i>Aplicaciones de los microarrays de expresión en cáncer</i>	39
4.4. <i>Papel de los perfiles de expresión en el cáncer colorrectal</i>	40
4.4.1. <i>Desarrollo de la enfermedad: “mucosa normal – adenoma – adenocarcinoma”</i>	40
4.4.2. <i>¿Diferente localización del tumor = Diferente pronóstico?</i>	41
4.4.3. <i>Perfiles de expresión e inestabilidad genética</i>	41
4.4.4. <i>Perfiles de expresión y clasificación de tumores</i>	43
4.4.5. <i>Perfiles de expresión y metástasis</i>	43
4.4.6. <i>Perfiles de expresión y pronóstico</i>	44
4.4.7. <i>Perfiles de expresión y tratamiento</i>	46

Antecedentes y objetivos.

1. Hipótesis y objetivos	50
---------------------------------------	----

Materiales y métodos.

1. Diseño experimental	53
2. Selección y estudio de los pacientes	54
3. Obtención y estudio histológico de los tumores	55
4. Extracción de los ácidos nucleicos	57
4.1. <i>Extracción del RNA: Método del Trizol</i>	57
4.2. <i>Extracción del DNA</i>	58
4.3. <i>Cuantificación de los ácidos nucleicos</i>	59
4.4. <i>Control de calidad del RNA</i>	59
4.5. <i>Control de calidad del DNA</i>	61
5. Obtención de la muestra de referencia (“pool” de colon normal)	61
6. Marcaje e hibridación de las muestras a los microarrays de expresión	61
6.1. <i>Marcaje y amplificación de las muestras</i>	62

6.2. Control del marcaje y la amplificación de las muestras	63
6.3. Hibridación y lavado del microarray	64
7. Escaneado de los microarrays	66
8. Cuantificación de la imagen	66
9. Análisis de datos	69
9.1. Control de calidad. “QC-Report”, “QC-Chart” y PCA	69
9.2. Filtrado de datos	71
9.3. Clasificación de los pacientes mediante métodos no supervisados	73
9.4. Comparación de clases y análisis funcional	75
10. Estudio de inestabilidad en secuencias microsatélites	76
11. Estudio de mutaciones en el gen KRAS	78
12. Estudio de asociación de las agrupaciones con los parámetros clínicos y patológicos	82
13. Análisis inmunohistoquímico	82
13.1. Construcción de los Tissue- Microarrays	82
13.2. Estudio inmunohistoquímico de las muestras	85
13.2.1. <i>Protocolo FISH para EGFR</i>	85
13.2.2. <i>Protocolo para tinción inmunohistoquímica (Ki-67, M-30 y β-catenina)</i>	86
14. Validación externa de la clasificación molecular	88
15. Análisis de los perfiles de supervivencia	92

Resultados.

1. Puesta a punto de la técnica y controles de calidad	94
❖ Análisis histológico de la composición de las muestras tumorales	94
❖ Control de calidad del RNA extraído	95
❖ Control de calidad de las hibridaciones (microarrays)	97
❖ Estudio de comportamientos anómalos de las muestras mediante Análisis de Componentes Principales o PCA	98
2. Características de los pacientes	99
3. Análisis de microarrays	102
3.1. Caracterización del patrón de expresión de muestras normales y tumorales	102

3.1.1. Comparación de clases entre tejido tumoral vs. tejido normal ...	105
3.2. Clasificación molecular de los tumores colorrectales	108
3.2.1. Clasificación de las muestras tumorales mediante métodos no supervisados	108
3.2.2. Comparación de subtipos y análisis funcional	113
3.2.3. Estudio de asociación entre la clasificación molecular y los parámetros clínicos y patológicos	120
3.2.4. Análisis inmunohistoquímico (Ki-67, M-30, β -catenina) y por FISH (EGFR) y correlación con la clasificación molecular	123
3.2.5. Estudios de asociación entre la clasificación histológica (Dukes), los parámetros analizados y los perfiles de expresión ..	125
3.2.6. Validación de la clasificación molecular mediante una matriz de datos externa	128
3.2.7. Análisis de supervivencia	136
3.2.7.1. Aplicación de predictores pronóstico	136
3.2.7.2. Análisis de supervivencia en pacientes HCSC	139
3.2.7.3. Análisis de supervivencia de la matriz de validación (Eschrich)	141

Discusión.

1. Introducción	147
2. Clasificación de tumores colorrectales	147
2.1. Validación de la clasificación molecular	152
2.2. Alteraciones de los perfiles de expresión en el proceso Tumoral (Tejido normal vs. Tejido tumoral)	153
2.3. Subtipos moleculares propuestos	155
2.3.1. Estroma y clasificación. Subtipos “de bajo estroma” y “de alto estroma”	155
2.3.2. Subtipo “relacionado con sistema inmune”	157
2.3.3. Subtipo “MSI-Mucinoso”	158
2.4. Supervivencia asociada a los subtipos moleculares	160

Conclusiones.

1. Conclusiones	165
------------------------------	-----

Anexo I. Análisis de microarrays.

1. Análisis de la imagen	168
1.1. Puntos problemáticos	168
1.2. Procesamiento de la fluorescencia	169
2. Análisis de datos	170
2.1. Análisis no supervisados	170
2.1.1 <i>Distancias</i>	170
2.1.2 <i>Métodos de agrupación de elementos</i>	173
2.1.3 <i>Centrado de genes</i>	175
2.1.4 <i>Métodos de análisis no supervisado</i>	176
2.1.4.1. <i>“Clustering” jerárquico</i>	176
2.1.4.2. <i>“k-means”</i>	179
2.1.4.3. <i>Análisis de componentes principales (PCA)</i>	181
2.2. <i>Métodos supervisados</i>	182
2.2.1. <i>Comparación de clases</i>	182
2.2.1.1. <i>Métodos de corrección para comparaciones múltiples</i>	183
2.2.1.2. <i>Análisis funcional</i>	184
2.2.2. <i>Predicción de clases</i>	185
2.2.2.1. <i>Técnicas de validación cruzada</i>	187

Anexo II. Predictores de supervivencia publicados.

1. Coincidencia entre los predictores de supervivencia y nuestra matriz de datos	189
---	-----

Anexo III. Lista genes meta-análisis de Chan SK et al.

1. Listado de genes publicados en el estudio de Chan SK et al	191
--	-----

Bibliografía.

1. Bibliografía 193

Anexo IV. Listado matriz de expresión HCSC.

1. Listado de 1722 genes que componen la matriz de datos HCSC 221

INTRODUCCIÓN

1. Cáncer de colon.

1.1. Incidencia.

El cáncer colorrectal (CCR) es uno de los tumores más frecuentes en los países desarrollados y, aunque su incidencia en los países en vías de desarrollo como India o los países africanos puede ser hasta 25 veces menor [1], se está empezando a observar un aumento del impacto de la enfermedad sobre su población. La Organización Mundial de la Salud (OMS) estima que al año se detectan aproximadamente 945.000 nuevos casos en todo el mundo y se producen en torno a 492.000 muertes debido a esta enfermedad [2].

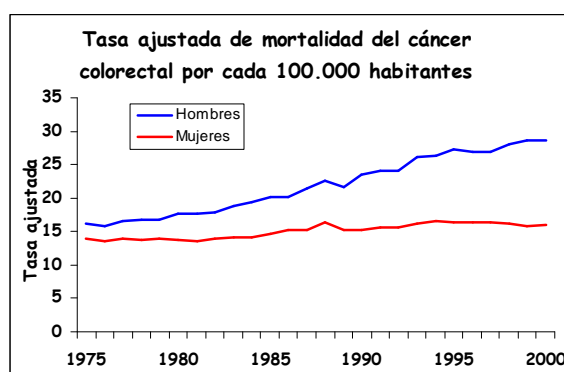


Fig.1. Representación gráfica del aumento de la mortalidad del cáncer colorrectal en España entre los años 1975 y 2000.

En España, los últimos datos epidemiológicos publicados [3] referentes al año 2000, confirmaron que el cáncer colorrectal fue la primera causa de muerte por cáncer (11.971 personas) y mostraron un aumento del número de nuevos casos diagnosticados que ascendió hasta los 25.665. De acuerdo con estos datos, en el período de tiempo comprendido desde 1975 hasta el año 2000, se produjo un incremento de la tasa ajustada de mortalidad anual por cáncer colorrectal (Fig.1.) tanto en hombres como en mujeres (2.6% y 0.8% respectivamente).

A pesar de que en la actualidad la mejora en las técnicas de detección precoz del cáncer y el aumento del abanico terapéutico han podido invertir esta tendencia, dando lugar a un estancamiento o incluso una disminución en las cifras de mortalidad del CCR, este sigue teniendo una gran incidencia en la población española.

1.2. Cáncer esporádico vs. Cáncer familiar.

El cáncer colorrectal es una enfermedad en la que confluyen tanto causas genéticas como ambientales. La mayoría de los tumores colorrectales (70-85%) se encuadran dentro de los tumores de tipo esporádico y que por tanto no poseen una carga genética heredada (Fig.2). Sin embargo, se estima que hasta un 30% del total de tumores detectados podrían tener un componente hereditario con familiares en primer o segundo grado afectados (Fig.2). Asimismo, entre un 5-6% de la población afectada, desarrollará tumores asociados con síndromes hereditarios autosómicos dominantes (principalmente el síndrome de la Poliposis Adenomatosa Familiar (PAF) y el síndrome de Cáncer Colorrectal Hereditario No Polipósico (HNPPC)) [4].

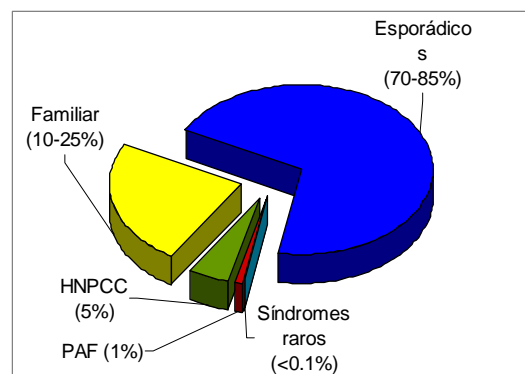


Fig.2. Distribución de los subtipos de cáncer colorrectal en la población.

Tiempo de progresión del cáncer colorrectal			
Colonocito normal	Inicio del tumor	Progresión tumoral	Carcinoma
CCR esporádico	→ Normal (30-50 años de edad)	→ Normal (10-20 años)	→ 6% riesgo de CCR (edad media ~ 68 años)
FAP	→ Temprana (5-20 años de edad)	→ Normal (10-20 años)	→ 100% riesgo de CCR (edad media ~ 40 años)
HNPPC	→ Normal (30-50 años de edad)	→ Rápida (1-3 años)	→ 80% riesgo de CCR (edad media ~ 40 años)

Tabla.1. Esquema de tiempo necesario para la aparición del cáncer colorrectal en función del tipo de tumor desarrollado. Adaptado de [5].

Este tipo de síndromes hereditarios han sido asociados a mutaciones germinales en genes como APC (síndrome PAF) o en los genes reparadores del DNA (principalmente MLH1, MSH2 y MSH6 en el caso del síndrome HNPPC) produciendo,

o bien una aparición temprana del inicio del tumor, o bien una disminución del tiempo necesario para la progresión del mismo (Tabla.1.), provocando la aparición de la enfermedad en torno a los 40 años en lugar de en torno a los 68 años como ocurre en el cáncer colorrectal esporádico [5].

1.3. Factores de riesgo no genéticos.

Los principales factores de riesgo para la aparición de tumores esporádicos son de tipo ambiental (Tabla.2.) y están particularmente asociados a comportamientos propios del estilo de vida occidental. Los últimos estudios de tipo prospectivo parecen confirmar la asociación de dietas ricas en carnes rojas, la elevada ingesta de alcohol o la obesidad con un mayor riesgo de desarrollar este tipo de tumores, mientras que las dietas ricas en fibra, calcio, vitamina D, ácido fólico o selenio podrían ser protectoras frente al desarrollo del mismo [1,6]. Otras características del estilo de vida occidental como la edad tardía del primer embarazo o la vida sedentaria han sido igualmente identificadas como factores de riesgo. Junto a estos, hay que añadir otros que ya se conocían previamente como el tabaquismo, la edad avanzada o diversas enfermedades asociadas al intestino grueso como la enfermedad de Crohn o la colitis ulcerosa [2].

Factores de Riesgo asociados a CCR en tumores esporádicos	
• Edad avanzada	• Sexo masculino
• Colectomía	• Anastomosis ureterocólica
• Colitis Ulcerosa	• Enfermedad de Crohn
• Obesidad	• Dieta rica en carnes y grasa
• Estilo de vida sedentario	• Dieta baja en fibra, folato y calcio.
• Diabetes melitus	• Tabaquismo
• Irradiación previa	• Elevado consumo de alcohol
• Trabajos peligrosos para la salud: exposición a asbestos	
• F. hormonales: multipara, edad tardía del primer embarazo, menopausia temprana.	

Tabla.2. Principales factores de riesgo asociados al desarrollo del cáncer colorrectal. Adaptada de [2].

1.4. Clasificación del cáncer colorrectal.

Desde su establecimiento por Dukes en la primera mitad del siglo XX [7], el sistema de clasificación de los tumores colorrectales ha tenido siempre como referencia

la extensión anatómica de los mismos, describiéndose en la actualidad mediante tres parámetros (Fig.3.):

- ❖ Extensión (T): indica el grado de infiltración tumoral en la pared del intestino grueso.
- ❖ Ganglios afectados (N): indica si el tumor se ha extendido a ganglios situados en las proximidades del tumor.
- ❖ Metástasis (M): indica si ha existido invasión de otros órganos por parte del tumor.

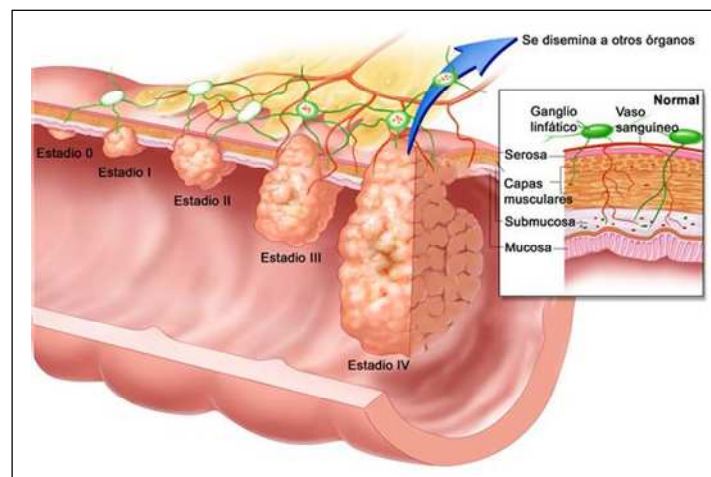


Fig.3. Representación gráfica de la evolución de un tumor colorrectal desde el estadio inicial hasta que produce la invasión de otros órganos.

Las modificaciones realizadas hasta nuestros días tanto en la propia clasificación [8] como en los métodos de evaluación (Ej. en la actualidad es necesario el estudio anatomopatológico de un mínimo de 12 ganglios linfáticos de la zona que rodea el tumor para realizar la correcta clasificación del paciente) [9], han contribuido a establecer una clasificación más precisa (Tabla.3 y 4). Fruto de esto, la clasificación que inicialmente fue diseñada para predecir el pronóstico de los pacientes, también es utilizada en la actualidad para la toma de decisiones sobre qué pacientes deben o no recibir tratamiento [10].

T-tumor primario	
Tx	No se puede valorar la infiltración del tumor
T0	No hay evidencias de tumor primario
Tis	Carcinoma in situ: intrepitelial o invasión de la lamina
T1	Tumor invade la capa submucosa
T2	Tumor invade la capa muscular propia
T3	Tumor invade la subserosa
T4	El tumor invade otros órganos o estructuras
N- Invasión de ganglios linfáticos	
Nx	No se puede valorar el nº de ganglios afectados
N0	No hay ganglios linfáticos regionales afectados
N1	Metástasis en 1-3 ganglios linfáticos regionales
N2	Metástasis en más de 4 ganglios linfáticos regionales
M- Metástasis distantes	
Mx	No se puede valorar la existencia de metástasis en órganos
M0	No hay metástasis distantes
M1	Metástasis Distante

Tabla.3. Descripción de los criterios histológicos y subcategorías en los que está basado el actual sistema de clasificación de los tumores colorrectales.

AJCC		Dukes	Astler-Coler	Supervivencia
TNM	Etapa	—	—	—
Tis, N0, M0	0	A	A	—
T1-T2, N0, M0	I	A	B1	80-95%
T3, N0, M0	IIA	B	B2	72-75%
T4, N0, M0	IIB	B	B3	65-66%
T1-T2, N1, M0	IIIA	C	C1	55-60%
T3-T4, N1, M0	IIIB	C	C2	35-42%
Cualq T, N2, M0	IIIC	C	C3	25-27%
Cualq T, Cualq N, M1	IV	D	D	0-7%

Tabla.4. Equivalencias de los tres sistemas de clasificación de tumores colorrectales más utilizados. Los datos de supervivencia proceden de [11].

Pese a la utilidad de esta clasificación en lo que se refiere al pronóstico, este sistema no es siempre capaz de predecir de manera clara cuál será la evolución de los pacientes, especialmente en los estadios intermedios (Dukes B y C), si se beneficiarán de un tratamiento con quimioterapia o radioterapia (Dukes B) y cuál de todas las posibilidades terapéuticas será la más adecuada (Dukes, B, C y D) [12].

1.5. Marcadores pronósticos y predictivos en CCR.

En respuesta a esta situación, se han realizando gran número de investigaciones cuyo objetivo es descubrir otros marcadores clínicos, patológicos o moleculares que aporten una mayor precisión a la clasificación. Dependiendo de su función se clasifican en [12]:

- ❖ Marcadores pronósticos: aportan información sobre la evolución del paciente en función de las características de su tumor.
- ❖ Marcadores predictivos: aportan información sobre la respuesta del tumor ante un determinado tratamiento.

Todo marcador, además de poder ser medido de manera sencilla y mediante técnicas poco invasivas, debe aportar información que suponga una mejora sobre las decisiones que se toman para el cuidado del paciente [13]. Entre los marcadores más estudiados en el cáncer colorrectal encontramos [12,14,15]:

1.5.1. Marcadores histopatológicos.

- ❖ Invasión venosa y linfática. La infiltración de células tumorales tanto en los vasos sanguíneos como en las vías linfáticas ha sido descrita como factor de mal pronóstico independiente para la formación de micrometástasis y para el crecimiento de tumores macroscópicos en una segunda localización [16].
- ❖ Grado tumoral: indica el grado de diferenciación de la célula tumoral y su participación en la formación de estructuras glandulares en comparación con el tejido colorrectal normal. En función de estos parámetros, los tumores se clasifican en [17]:
 - Indiferenciados (< 25% de estructuras glandulares).
 - Pobrementemente diferenciados (25-50% de estructuras glandulares).
 - Moderadamente diferenciados (50-75% de estructuras glandulares).

- Bien diferenciado (>75% de estructuras glandulares).

Un grado más bajo de diferenciación se asocia con un peor pronóstico del paciente.

- ❖ Invasión perineural: se produce cuando las células tumorales infiltran la cubierta de los nervios (perineuro). Su presencia ha sido asociada con mal pronóstico [18].
- ❖ Tipo histológico: diferencia los tumores en base al tipo de célula epitelial que lo compone. De entre todas las clases definidas por la OMS, tres han sido las relacionadas con mal pronóstico [19]: tumores con células en anillo de sello, tumores mucinosos y carcinomas de células pequeñas.
- ❖ Borde tumoral: hace referencia a la frontera de crecimiento del tumor. Se divide en infiltrante (asociado con mal pronóstico), mixto y expansivo (asociado con buen pronóstico).
- ❖ Infiltración linfocitaria del tumor: la presencia de linfocitos en la zona de desarrollo tumoral parece ser indicativa de una respuesta inmune del organismo frente al cáncer, siendo considerada un marcador pronóstico favorable [16].

1.5.2. Marcadores moleculares.

- ❖ Inestabilidad en microsatélites (MSI): la aparición de inestabilidad en secuencias microsatélites (ganancia o pérdida de estas repeticiones), debidas a fallos en el funcionamiento de los mecanismos reparadores del DNA (principalmente debido a la ausencia de la proteína MLH1 como consecuencia de la hipermetilación de su promotor en CCR esporádico), provoca el desarrollo del tumor mediante una vía diferente a la de los tumores de fenotipo estable (MSS). Se ha descrito por un lado, como factor pronóstico favorable [20] y por otro, como factor predictivo negativo para el tratamiento con 5-Fluorouracilo (5-FU) [21].
- ❖ Antígeno Carcino-Embrionario (CEA): Es una glicoproteína expresada, entre otras regiones, en el aparato digestivo del feto y que desaparece tras el nacimiento. En adultos se ha observado su expresión en el 90% de los tumores colorrectales, aunque también se han encontrado altos niveles de CEA en fumadores, en personas con

cirrosis, con pancreatitis, con colitis ulcerosa o con úlcera peptídica. Como factor pronóstico, se ha visto que niveles elevados preoperatorios de este marcador (más de 5 ng/ml en suero) confieren una menor probabilidad de supervivencia libre de enfermedad a los 5 años y que incluso niveles mayores a 15 ng/ml podrían ser indicativos de presencia de metástasis ocultas [22]. Al ser una proteína cuya expresión suele normalizarse tras la cirugía y que aumenta según avanza el estadio de la enfermedad, es un buen marcador de la evolución del paciente tras la operación y su determinación continuada durante la fase de seguimiento postoperatoria permite una detección temprana de recaídas y en concreto de la aparición de metástasis en hígado [23]. Niveles bajos de CEA también se han relacionado con un aumento de la supervivencia global de los pacientes a los 5 años [15]. También se ha descrito su posible utilidad como marcador predictivo para monitorizar la respuesta a tratamiento de pacientes con enfermedad metastásicas [15].

- ❖ LOH 18q: La pérdida de la región cromosómica 18q y, en concreto, del gen DCC es muy frecuente en el cáncer colorrectal y su presencia se ha asociado con mal pronóstico [24]. Sin embargo, la pérdida de esta región también podría afectar a otros genes importantes en el desarrollo de los tumores como los de la vía TGF- β (como son el TGF- β -R-2, SMAD2 o SMAD4) y que codifican en esa misma región.
- ❖ p53: está codificada en el brazo corto del cromosoma 17 y su actividad normal está relacionada con la parada del ciclo celular en respuesta a daño en el DNA posibilitando su reparación y, en caso de fallo de los procesos de reparación, con apoptosis. Su inactivación supone la supresión de su función y se ha correlacionado con peor pronóstico [25].
- ❖ KRAS: proteína de membrana con actividad GTPasa cuya activación provoca una cascada de señalización hacia el núcleo mediante la cual se regulan vías implicadas en la proliferación celular, diferenciación, migración, apoptosis, etc. Como factor pronóstico solo se ha podido demostrar la correlación de la mutación “Gly > Val” en el codon 12 con un mal pronóstico [26] y el beneficio de la quimioterapia adyuvante en pacientes en estadio C de Dukes con mutaciones que no implicasen cambio de Glicina por Aspártico [27]. Sin embargo, en la actualidad, se está ya utilizando la

detección de mutaciones en el gen KRAS como marcador predictivo de la respuesta al tratamiento con Cetuximab (fármaco dirigido contra el EGFR). Diversos estudios señalan a este respecto, que la presencia de mutaciones que activan constitutivamente el gen, correlaciona con falta de respuesta de los pacientes a ese fármaco [28].

- ❖ Timidilato Sintasa (TS): es la enzima encargada de transformar la deoxiuridina monofosfato (dUMP) en 2-deoxi-timidina-5-monofosfato (dTMP), precursor de uno de los cuatro nucleótidos requeridos para la síntesis de DNA. Su papel en los procesos de síntesis y replicación del DNA ha hecho que sea el gen diana de algunos de los agentes citotóxicos que se utilizan en el tratamiento del cáncer como el 5-FU. Niveles elevados de la TS, así como de otras enzimas relacionadas con la transformación o la degradación de fármacos que bloquean la actividad de la TS, como son Timidina Fosforilasa (TP) o Dihidropirimidina Deshidrogenasa (DPD) han sido correlacionadas con mal pronóstico en pacientes con cáncer de colon [29,30].
- ❖ EGFR: es una glicoproteína transmembrana codificada por el proto-oncogén c-ErbB-1 que está implicada en la activación de cascadas de señalización intracelulares relacionadas con adhesión celular, proliferación, diferenciación, apoptosis y metástasis. Estudios realizados sobre esta proteína han demostrado la existencia de asociación entre altos niveles de EGFR, una mayor probabilidad de recurrencia y una menor probabilidad de supervivencia [30].
- ❖ Células tumorales circulantes: la detección de células tumorales circulantes en sangre periférica, mediante técnicas de separación inmunomagnéticas, por encima de un determinado umbral (3 o más células circulantes/7.5 ml de sangre) se ha visto asociada con una menor supervivencia libre de progresión y una menor supervivencia global en pacientes con cáncer colorrectal metastásico [31].

La mayoría de los marcadores anteriormente descritos han producido resultados prometedores. Sin embargo, la aparición de otros estudios que contradecían los resultados previos, la falta de consenso en cuanto a la técnica más precisa para medir estos marcadores y la falta de experimentos de validación de carácter prospectivo en los que se utilice un tamaño muestral adecuado ha frenado su inclusión en la práctica

clínica rutinaria. La única excepción se ha producido en el uso del CEA como factor pronóstico previo a la cirugía, como factor pronóstico en la evolución del paciente tras cirugía y como factor predictivo en pacientes que reciben tratamiento paliativo con quimioterapia, además de la reciente incorporación de la detección de mutaciones en KRAS para predecir la respuesta del paciente frente a tratamiento con cetuximab.

Muchos otros marcadores han sido propuestos para su uso pronóstico y predictivo con el mismo resultado (Tabla 5). La falta de resultados positivos, en la mayoría de los casos, parece deberse a la “dificultad” de que de estos marcadores puedan prever de forma individual y precisa la evolución de un mecanismo tan complejo y heterogéneo como el cáncer. La aparición de trabajos en los que se establece mediante perfiles de expresión con microarrays un pronóstico para los pacientes de determinados tipos de cáncer, ha abierto un nuevo campo de investigación en el área de la oncología y parecen confirmar la necesidad de utilizar combinaciones de marcadores para establecer el pronóstico y la posible respuesta a tratamiento de los pacientes.

MARCADORES MOLECULARES CON POSIBLE POTENCIAL PRONOSTICO.

- Genes supresores de tumores y Oncogenes (c-myc, nm23).
- Apoptosis y genes relacionados con suicidio celular (bcl-2, BAX).
- Genes relacionados con la síntesis de DNA (*TOP1*).
- Factores de Crecimiento y Receptores de Factores de Crecimiento (TGF α , TGF β , hr-2/NEU, EGFR).
- Genes reparadores del DNA (MSH2, MLH1, ERCC1).
- Ciclinas y Cdks (p27, p21, p16).
- Moléculas de Adhesión y genes que codifican para glicoproteínas (cd44, E-cadherina, ICAM-1).
- Marcadores de invasión (MMPs, UTPA).
- Factores de proliferación (ki-67, Mib-1, antígeno de proliferación del núcleo celular).
- Antioxidantes (Superóxido dismutasa, GST-pi).
- Detección de células circulantes.
- Estudio de polimorfismos en genes implicados en cáncer (UGT1A1, TS)
- Longitud de los telómeros.

Tabla.5. Marcadores moleculares estudiados por su posible potencial pronóstico (modificada de [2]).

1.6. Modelos moleculares del cáncer colorrectal.

Como toda neoplasia, el cáncer colorrectal es un proceso secuencial en el que cada paso, representado por cambios genéticos o epigenéticos resulta normalmente en la activación de un oncogén o en la inactivación de un gen supresor de tumores [32,33]. Estos cambios alteran el comportamiento normal de la célula y le proporcionan ventajas selectivas que favorecen su expansión clonal y su evolución hacia un fenotipo

malignizado [33], caracterizado por una serie de manifestaciones fisiológicas entre las que destacan la capacidad de producir señales propias de crecimiento de manera independiente del entorno, la insensibilización ante las señales antiproliferativas, la resistencia a la apoptosis, la adquisición de un potencial replicativo infinito y la capacidad de inducir angiogénesis y metástasis [34].

Los análisis de los genes y rutas moleculares implicados en el CCR han llevado al establecimiento de diferentes modelos (fenotipo Supresor o CIN, fenotipo mutador o MSI y fenotipo metilador o CIMP) de progresión a través de los cuales se desarrolla la enfermedad.

1.6.1. Modelo general del desarrollo del CCR de tipo esporádico. Fenotipo CIN.

La posibilidad de estudiar tanto las propias lesiones tumorales como las lesiones previas al tumor malignizado y las lesiones metastásicas, permitieron desarrollar un modelo molecular e histológico de la progresión del CCR [32,35] consistente en 7 etapas (Fig.4):

- ❖ Epitelio normal - Epitelio hiperproliferativo: se produce la mutación del gen APC, que se vuelve inactivo en un 80% de los pacientes [35,36]. Además se ha observado la aparición de procesos de pérdida de heterocigosidad en la región que codifica para este gen (5q21) en al menos un 30% de los tumores [37]. Esta inactivación provoca la desregulación de la vía Wnt [38] y parece afectar también a los procesos de migración y adhesión celular y a la segregación cromosómica, aunque todavía se desconoce en qué medida estos procesos afectan al desarrollo tumoral [39,40].
- ❖ Epitelio hiperproliferativo - Adenoma temprano: se produce la aparición de fenómenos epigenéticos entre los que destacan los fenómenos de hipometilación génica y de hipermetilación de la región promotora, que contribuyen a la sobreexpresión e inhibición de oncogenes y genes supresores de tumores respectivamente [32,41].
- ❖ Adenoma temprano - Adenoma intermedio: se caracteriza por la aparición, en un número elevado de casos (aproximadamente un 30-45%), de mutaciones activadoras en los codones 12, 13 y 61 del oncogén KRAS (12p12.1) [27,42] y

consecuentemente por la activación de su cascada de señalización que provoca la transcripción de genes diana implicados en proliferación celular e inhibición de la apoptosis.

- ❖ Adenoma intermedio - Adenoma tardío: en esta fase se produce la mutación y pérdida de heterocigosidad de genes codificados en la región del cromosoma 18q como DCC, Smad2 y Smad4.

Caracterizado inicialmente como un gen supresor de tumores (papel que, no obstante, fue puesto en duda [43]), el estudio del gen DCC ha recobrado interés debido al descubrimiento de su interacción con el gen Netrina-1 y su posible implicación en procesos de proliferación y diferenciación celular [44,45], de unión célula-célula [46], de apoptosis [47] y de metástasis [48].

La pérdida de los genes Smad2 y Smad4 altera la actividad de la vía TGF- β que parece jugar un papel doble durante la progresión tumoral. En las fases iniciales actúa como supresor de tumores, inhibiendo la proliferación e induciendo la apoptosis, mientras que en estadios más avanzados potencia el desarrollo tumoral, la invasión y la metástasis [49] a través de su implicación en los procesos de transformación epitelio-mesénquima [50], de angiogénesis [50] y de evasión de la respuesta inmune [51].

- ❖ Adenoma tardío - Adenocarcinoma: se produce la inactivación del gen p53 (gen pro-apoptótico situado en cromosoma 17p) en aproximadamente el 80% de los casos [35]. El gen p53 activo juega un papel clave en el control de la integridad y estabilidad del genoma y en la respuesta a situaciones de estrés, hipoxia, daño en el DNA o activación de oncogenes, todas ellas comunes durante el desarrollo tumoral. Además es capaz de provocar la parada del ciclo celular o activar mecanismos de senescencia y apoptosis que impiden la transformación celular [52,53].
- ❖ Adenocarcinoma - Metástasis: se produce la mutación tanto de oncogenes como de genes supresores de tumores en cromosomas adicionales, lo que proporciona al tumor la capacidad de progresar y de escapar de su localización original, pasando al torrente sanguíneo o al linfático, pudiendo así llegar a otros órganos donde se desarrollarán nuevos focos tumorales.

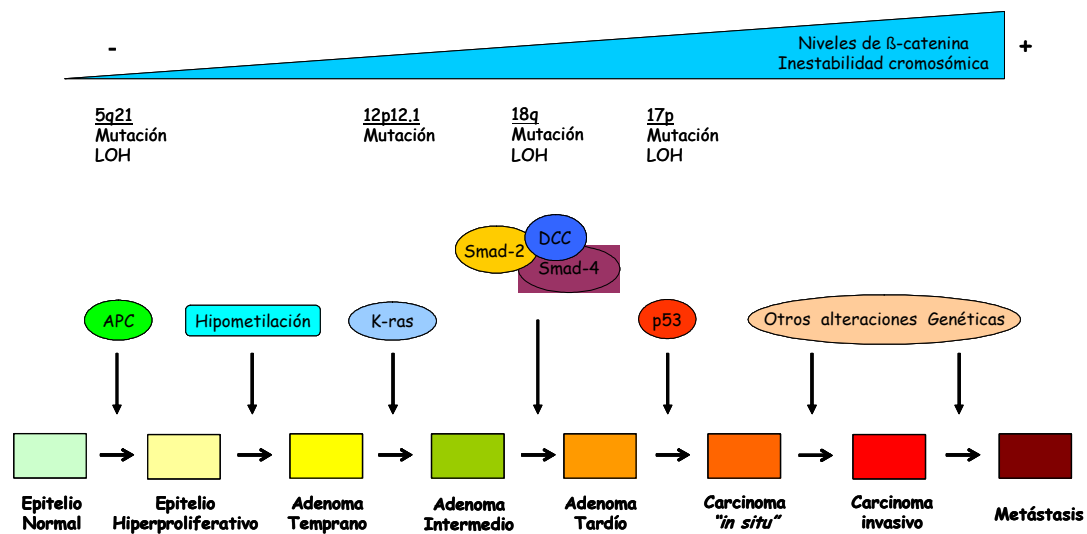


Fig.4. Modelo de progresión tumoral propuesto por Fearon y Vogelstein, en el que cada alteración histológica está asociada con la aparición de mutaciones que producen la ganancia de funciones necesarias para el desarrollo del cáncer colorrectal.

La adquisición del fenotipo tumoral no es un fenómeno frecuente, debido a la baja tasa de mutaciones y pérdidas alélicas que se producen en condiciones normales. El desarrollo de esta enfermedad implica por tanto la aparición de fenómenos de inestabilidad genómica que aumenten la tasa de alteraciones genéticas [54]. Según el modelo descrito por Fearon y Vogelstein, en el cáncer colorrectal esporádico se producen fenómenos de inestabilidad cromosómica que conllevan la ganancia o pérdida total o parcial de cromosomas debidos, en su mayoría, a la existencia de fallos en los procesos de recombinación mitótica y de segregación de los cromosomas durante la división celular [35].

1.6.2. Fenotipo Mutador (o MSI) y fenotipo Metilador (CIMP).

En paralelo a la aparición del modelo molecular de progresión tumoral propuesto por Vogelstein et al. [32], surgieron una serie de trabajos [55-57] que proponían un modelo diferente de desarrollo del cáncer colorrectal que podía aplicarse a, aproximadamente, un 12-15% de los tumores de tipo esporádico, además de explicar la aparición de tumores asociados a algún síndrome hereditario como el HNPCC. Este modelo, denominado “modelo del fenotipo mutador” o fenotipo MSI, se caracteriza por la aparición de inserciones o deleciones en secuencias microsatélites (secuencias

nucelotídicas cortas repetidas en tándem), que no pueden ser reparadas debido a fallos en los mecanismos de reparación del DNA. La alteración de estas secuencias en regiones codificantes produce un cambio de fase de lectura del gen, provocando la aparición de proteínas truncadas o con la función alterada. En el caso de tumores HNPCC, la falta de función de las enzimas reparadoras del DNA (MLH1, MSH2 y MSH6) se debe a la aparición de mutaciones, mientras que en los tumores esporádicos la inactivación del mecanismo reparador es causada principalmente (en más de un 80% de los casos) por la hipermetilación de la región promotora del gen MLH1 [58]. Los pacientes cuyos tumores pueden ser asociados a este modelo de progresión tumoral se caracterizan por una baja frecuencia de mutaciones en genes como p53, APC y KRAS, así como por una baja frecuencia de pérdida de heterocigosidad de las regiones cromosómicas 5q, 17p y 18q [59]. Sin embargo, no es la mutación de los genes MMR, si no la aparición de mutaciones en genes de interés, lo que lleva al desarrollo tumoral [60].

El número de genes afectados es elevado, conociéndose más de 30 en la actualidad y que comprenden numerosas funciones entre las que destacan el crecimiento celular (TGF β -R-II, IGF2-R), la apoptosis (BAX, Caspasa-5), la reparación del DNA (MSH3, MSH6), la señalización de la vía Wnt (Axin, β -catenina, Tcf4), etc. [60]. Los genes alterados en el desarrollo de tumores MSI, aunque distintos, parecen actuar afectando las mismas vías moleculares que las que se desregulan en los tumores de tipo CIN [61,62]. Desde un punto de vista clínico-patológico, los tumores desarrollados por la modificación de esta vía se caracterizan entre otras cosas por ser mucinosos, estar localizados en colon derecho o tener mejor pronóstico (Tabla.6).

<i>Características de los MSI</i>	
· Poco diferenciados	· Mucinosos
· Localizados en colon derecho	· Peor respuesta a tratamiento
· Diploides	· Mejor pronóstico
· Infiltración de células de respuesta inmune	· Menor afectación de gánглиos linfáticos

Tabla.6. Característica clínico-patológicas más frecuentes de los tumores colorrectales con inestabilidad en microsatélites [63,64].

A raíz del descubrimiento de los fenómenos de hipermetilación en las regiones promotoras de genes como MLH1, se llevaron a cabo trabajos que revelaron la

existencia de tumores que desarrollaban una hipermetilación masiva en las regiones promotoras de genes supresores de tumores, lo que ha llevado a proponer la existencia de un tercer fenotipo en el CCR denominado fenotipo CIMP (CpG Island Methylator Phenotype) o fenotipo metilador. Este fenotipo estaría asociado a diversos parámetros clínico y patológicos, como la alta tasa de adenocarcinomas serrados (no en vano se ha propuesto que los tumores CIMP tienen su origen en este tipo de tumores [65,66]), tumores proximales, pobremente diferenciados, alta presencia en mujeres, alta tasa de mutaciones en b-Raf o baja tasa de mutaciones en p53 y asociación inversa con β -catenina nuclear [67-70]. También se ha sugerido que salvo en el caso de presentarse junto con inestabilidad en microsatélites, estarían asociados con mal pronóstico [71]. Teniendo en cuenta estas observaciones y tras la publicación de los estudios de los grupos de Ogino y de Jass [72,73], en los que se analizaron largas series de pacientes con tumores colorrectales, se ha propuesto una clasificación de tumores colorrectales que los divide en grupos molecularmente homogéneos en función de su estatus MSI, CIMP y CIN. Aunque existe alguna diferencia entre las clasificaciones propuestas, parece probable la existencia de al menos de 4 grupos moleculares (Tabla.7): G1. (MSI-H, CIMP-H); G2. (MSI-H, CIMP-0); G3. (MSS, CIMP-H); G4. (MSS, CIMP-0). Cabe destacar la asociación de los tumores HNPCC con el grupo 2, la asociación de los pacientes con PAF en el grupo 4 o la metilación del gen MLH1 en tumores del grupo 1. Las características más relevantes de cada uno de estos grupos se resumen en la Tabla.7.

		Pacientes con tumores colorrectales			
		Grupo 1	Grupo 2	Grupo 3	Grupo 4
Fenotipo		CIMP-H y MSI-H	CIMP-0 y MSI-H	CIMP-H y MSS	CIMP-0 y MSS
% de pacientes		10	5	5-10	75-80
Características	CIN	Negativo	Negativo	Negativo	Positivo*
	APC	Nativo	Mutado	Nativo	*
	K-Ras	Nativo	Mutado	Nativo	Mutado*
	B-Raf	Mutado	Nativo	Mutado	Nativo
	TP53	Nativo	Nativo	Nativo	Mutado
	Localización	Proximal	Proximal	Proximal	Distal*
	Sexo	Mujeres	Hombres	Mujeres	Hombres*
	Precursor	Polipos serrados	Adenoma	Polipos serrados	Adenoma*
	Mucinoso	Si	Si*	Si	No
	Diferenciación	Pobremente dif.	Pobremente dif.*	Pobremente dif.	Diferenciados*
Infiltr. Linfocitario	Si	Si	Baja	Bajo*	

Tabla.7. Clasificación molecular en base a los fenotipos CIN, MSI y CIMP. Modificada de la clasificación realizada por Jass y Ogino [72,73]. El asterisco (*) indica la existencia de indicios que apuntan hacia esa característica aunque los datos aportados por los estudios no son concluyentes y en algún caso pueden ser contradictorios.

1.6.3. Vía Wnt.

En condiciones normales, la vía Wnt juega un papel esencial en la regulación de los procesos de desarrollo embrionario y de control de la homeostasis en el adulto. La interacción específica de ligandos Wnt con sus receptores [74] o, posiblemente, solo el tipo de receptor [75] van a ser los encargados de activar una de las tres rutas diferentes de esta vía: ruta Wnt/Ca²⁺, ruta “Planar Cell Polarity” (PCP) y ruta canónica.

Aunque las tres rutas han sido implicadas en el desarrollo de cáncer [76], parece que sólo la vía canónica juega un papel crucial en la aparición y progreso del cáncer colorrectal. Esta vía controla la capacidad de la proteína β -catenina de activar la transcripción de un número importante de genes diana (Fig.5). En ausencia de la señal activadora de la vía, la β -catenina que se encuentra libre en el citoplasma es fosforilada por el complejo proteico GSK3 β /APC/Axina/CKI, lo que la prepara para ser ubiquitinada y posteriormente degradada por la vía del proteosoma. Este proceso impide la acumulación de β -catenina en el citoplasma y su posterior translocación al núcleo en donde su unión al factor de transcripción TCF/LEF y al complejo activador de la transcripción iniciaría la transcripción de sus genes diana. Otras proteínas como las Proteínas Secretadas Relacionadas con Frizzled (SFRP) o la proteína Dickkopf van a bloquear la transmisión de la señal de la vía Wnt en otros puntos específicos de la vía [74,77].

La unión de proteínas tipo Wnt a los receptores Frz/LRP produce la activación de la ruta, desencadenando el secuestro de Axina a la membrana y, en consecuencia, la inhibición de la formación del complejo β -catenina-(GSK3 β /APC/Axina/CK1). β -catenina se acumulará en el citoplasma y translocará al núcleo en donde producirá la transcripción de sus genes diana.

Uno de los procesos clave durante el desarrollo del CCR es la activación constitutiva de la vía Wnt debido a mutaciones en las proteínas implicadas en ellas. Así, tiene lugar la mutación e inactivación de determinadas proteínas como APC o Axina (mutada al menos en el 25% de los tumores con inestabilidad [77]) o la mutación y activación constitutiva de β -catenina, que se produce en un 10% de los tumores

esporádicos [78] y que parece ser excluyente con la presencia de la mutación de APC [79].

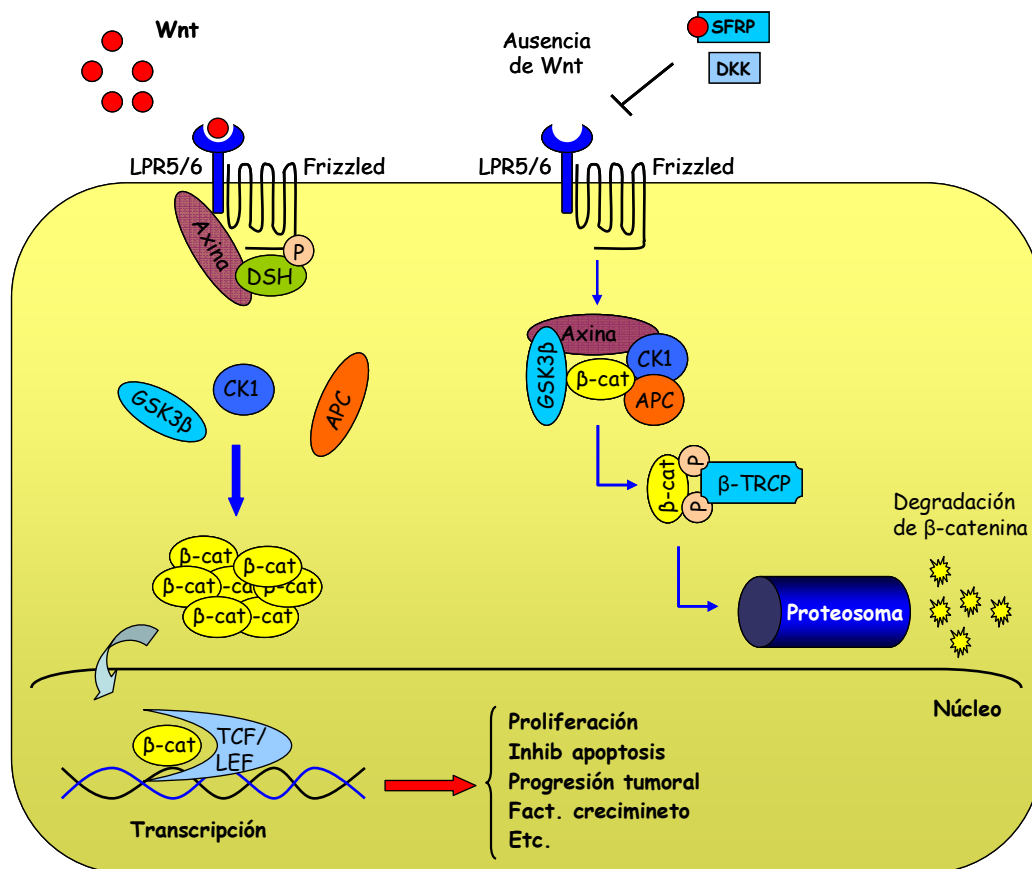


Fig.5. Representación esquemática de la vía Wnt.

La desregulación de la vía Wnt en el cáncer colorrectal conduce a un incremento en la expresión de genes implicados en proliferación celular como c-myc, ciclina D1 o Gastrina [80-82], en la progresión tumoral como MMPs, Nr-CAM, uPA [83-85], en la supervivencia celular como survivina o COX-2 [86,87], en la angiogénesis como VEGF [88], en la regulación de factores de crecimiento como c-met [89] o en procesos de Transición Epitelio-Mesénquima y pérdida de E-Cadherina como Snail1.

Por otro lado, la vía Wnt va a regular y ser regulada a través de su interacción con otras rutas moleculares como la de KRAS [64], Cox2 [90], TGFβ [91,92] o Hedgehog [93], importantes en el desarrollo del cáncer colorrectal.

2. Microentorno. Relevancia en la progresión tumoral.

La enfermedad tumoral no se puede entender como un proceso caracterizado exclusivamente por la acumulación de alteraciones genéticas en las células cancerígenas, ignorándose el papel clave que juega el microentorno y su interacción con el epitelio cancerígeno en el inicio y la progresión del mismo [34,94].

Se conoce como microentorno al conjunto de fibroblastos, células del sistema inmune e inflamatorias (linfocitos B, linfocitos T y Natural Killers, células dendríticas, macrófagos, mastocitos y neutrófilos), células de los vasos sanguíneos (células endoteliales, pericitos) y matriz fibrilar (matriz extracelular (MEC) y membrana basal) que rodean a las células cancerígenas.

En condiciones normales, las interacciones epitelio-estroma regulan el correcto funcionamiento y desarrollo del tejido. Sin embargo, el inicio del desarrollo tumoral, de manera similar a como ocurre en los procesos de curación de heridas [95], va a provocar la activación del estroma a través de la acción paracrina de factores de crecimiento (bFGF, VEGF, PDGF, TGF β , etc.), citoquinas, quimioquinas, etc., secretados por las células cancerígenas, que inducen la alteración de la homeostasis del tejido normal y favorecen, entre otros, los procesos de angiogénesis y respuesta inflamatoria, que facilitarán a su vez, la progresión tumoral [94]. Así, se ha observado que el carcinoma invasivo está asociado con la expansión del estroma tumoral y el incremento de la deposición de MEC, proceso conocido como desmoplasia [96]. En algunos tumores, el estroma conforma la mayor parte de la masa tumoral [97], siendo su composición y volumen definido por la respuesta frente a la activación de los fibroblastos [98] y leucocitos.

Los componentes y procesos del estroma que van a estar implicados en el desarrollo del tumor se describen brevemente a continuación.

2.1. Células del sistema inmune.

Se cree que durante las fases iniciales del desarrollo tumoral, el sistema inmune actúa detectando y eliminando las células cancerígenas [99,100] (Fig.6). Estudios llevados a cabo en tumores colorrectales han demostrado la existencia de una elevada respuesta adaptativa en la región tumoral capaz de controlar e inhibir la progresión del mismo, lo que se correlaciona con un buen pronóstico de los pacientes [101]. Sin embargo, se cree que durante la fase de progresión tumoral, las células cancerígenas son capaces de desarrollar algunos mecanismos, detallados a continuación, que les permiten evadir la actuación del sistema inmune y continuar con la progresión tumoral:

- ❖ Alteraciones en los procesos de presentación de antígenos: implican la inhibición de la expresión de antígenos tumorales en la superficie celular [102,103], alteraciones en la maquinaria transportadora de antígenos o en los componentes del inmunoproteosoma [104] y la inhibición de la expresión de las proteínas que conforman el complejo MHC-I [105-107].
- ❖ Resistencia tumoral a la apoptosis: la sobreexpresión de moléculas anti-apoptóticas (FLIP, Survivina, Bax, PI9/SPI6, etc.) [108,109], la expresión en la superficie de las células tumorales de proteínas anti-complemento como CD55 o CD59 [110] y la secreción de receptores solubles que bloquean los ligandos de los receptores de muerte en la membrana [111] van a contribuir a que la célula sea capaz de eludir los mecanismos apoptóticos.
- ❖ Red de inmunosupresión: las células tumorales utilizan varios mecanismos para destruir las células del sistema inmune:
 - La secreción de quimioquinas (Ej.: CCL22) capaces de atraer a la zona tumoral a linfocitos de tipo Treg, implicados en la supresión de la respuesta de los linfocitos T [112-114].
 - La capacidad de segregar moléculas inmunosupresoras como IL-10, TGF β , Galectina-1 o FASL capaces de provocar la apoptosis de los linfocitos T [115-118] y de modificar la respuesta inmunológica de Th1 a Th2, impidiendo el rechazo tumoral [119].

2.2. Fibroblastos.

Son células del estroma implicadas en diversos procesos entre los que se encuentran la síntesis de proteínas que forman parte de la membrana basal como colágeno IV y laminina, la homeostasis de la matriz extracelular tanto a través de la síntesis de sus componentes (colágeno I, III y V, fibronectina) como a través de la secreción de proteasas implicadas en su degradación y la homeostasis de las células epiteliales a través de la secreción de factores de crecimiento y la interacción directa entre células del epitelio y el mesénquima. Durante el desarrollo tumoral, estas células son activadas por factores de crecimiento como PDGF y TGF β , transformándose en miofibroblastos, reconocibles por la expresión de proteínas como α -smooth-muscle-actin y Fibroblast Activation Protein (FAP) [120]. Esta activación además va a inducir a un cambio funcional que confiere a los miofibroblastos mayor capacidad secretora y de proliferación [121], que da lugar a su vez a un aumento de su número en la zona tumoral en comparación con el tejido normal [97].

La gran importancia de los fibroblastos en la progresión del cáncer radica en la gran cantidad de procesos en los que intervienen [120,122-125] (Fig.6.):

- ❖ Remodelación de la MEC: la secreción de proteasas secretadas (MMP3, MMP1, MMP2, MMP9) colágeno, fibronectina y tesnascina C van a alterar la estructura de la MEC, liberando numerosas moléculas implicadas en el desarrollo tumoral y remodelándola para hacer posible la incorporación de nuevas células (ya sean estromales o epiteliales).
- ❖ Regulación de los procesos de inflamación mediante la liberación de quimioquinas como MCP1 e interleuquinas como IL1.
- ❖ Regulación de la angiogénesis a través de la secreción del factor VEGF.
- ❖ Estimulación de la proliferación, supervivencia y del poder invasivo de las células cancerígenas mediante la secreción de numerosos factores como TGF β , IGF, SDF1, HGF, etc.

- ❖ Secreción de factores, durante los procesos metastásicos, que facilitan la adaptación de las células cancerígenas al nuevo microentorno.

En cuanto al origen de los fibroblastos del estroma, se cree que una gran parte pueden ser producto de los procesos de transición epitelio-mesénquima (TEM) que las células cancerígenas tienen que sufrir para que puedan llevar a cabo procesos de movilización, invasión y metástasis [126,127]. Factores como TGF- β , TNF- α , FGF o HGF van a regular este proceso a través de factores de transcripción como NF κ β , β -catenina/TCF, Snail, Slug o Twist [100,127].

Entre los cambios más importantes [100,127-129] que sufren las células cancerígenas en este proceso, se encuentran la pérdida de polaridad celular, la pérdida de expresión de E-Cadherina, implicada en las uniones célula epitelial-célula epitelial a favor de la expresión de N-Cadherina, implicada en la interacción epitelio-estroma lo que aumenta la movilidad celular o la supresión de la expresión de citoqueratinas en beneficio de la vimentina, componente esencial del citoesqueleto, la expresión de fibronectina que normalmente solo es expresada por células del estroma y la expresión de la integrina α v β 6 en la superficie celular que le permite unirse a componentes de la MEC como fibronectina y tenascina C.

2.3. Inflamación.

Es un proceso fisiológico que se produce en respuesta a situaciones de daño, infección o invasión tisular, en el que el organismo tratará de eliminar el agente causante del daño y reparar el tejido, mediante la atracción de células especializadas como macrófagos, mastocitos, células de respuesta inmune, etc. y la liberación de mediadores solubles. La ineficacia de la respuesta inflamatoria en determinadas situaciones, va a llevar a la cronificación de la propia respuesta, favoreciendo en determinados casos el desarrollo tumoral como ocurre en pacientes con enfermedad de Crohn cuyo riesgo de desarrollar cáncer colorrectal es mayor que el de la población general [130].

Durante el desarrollo tumoral (Fig.6.), las células cancerígenas van a secretar una serie de factores inflamatorios entre los que se encuentran CCL2, TNF- α , TGF- β ,

COX2, LTB4 o CSF [131-133], implicados en la activación y atracción de monocitos, mastocitos, leucocitos, células del sistema inmune (linfocitos T) y fibroblastos a la zona tumoral, activando así la respuesta inflamatoria. Diversos estudios han observado que la composición celular de la respuesta inflamatoria, esta relacionada con el pronóstico clínico de los pacientes. Así, mientras que un alto grado de infiltración linfocitaria podría ser beneficiosa para el paciente y estar asociados con buen pronóstico [134,135], un alto grado de infiltración de macrófagos y mastocitos se asocia con angiogénesis y mal pronóstico [133,136,137].

En los tumores, la respuesta inflamatoria está mediada en gran parte por la respuesta de los macrófagos (Macrófagos o Macrófagos activados por tumor (MAT)) y, en menor medida, por el resto de células inflamatorias (neutrófilos y mastocitos). Las principales funciones en las que participan todas estas células son:

- ❖ Favorecer el desarrollo del tumor al activar una respuesta tipo M2, asociada con remodelación tisular, angiogénesis, inhibición de la respuesta inmune y crecimiento tumoral en contraposición a la respuesta M1 asociada con inflamación y moléculas microbianas. La respuesta M2, induce además la respuesta inmune Th2 (inmunosupresora) en perjuicio de la Th1 (inhibidora de la progresión tumoral) [138-140].
- ❖ Secreción de factores de supervivencia como EGF, IL6, IL-8, etc., factores angiogénicos como IL8, VEGF, PDGF, TGF β , Angiotensina-1 [141] y factores de remodelación de la MEC como MMP9, MMP2, etc. [142].
- ❖ Amplificación de la respuesta inflamatoria a través de la secreción de citoquinas como IL-1, TNF α , CSF, IL-8 etc.) [143,144].
- ❖ Formación de un nicho pre-metastásico mediante la secreción de factores de crecimiento o quimioquinas como VEGF, TGF- β o TNF- α [145] o la movilización de células inflamatorias [146] desde el tumor primario que favorecen la migración de las células cancerígenas al sitio de la metástasis.

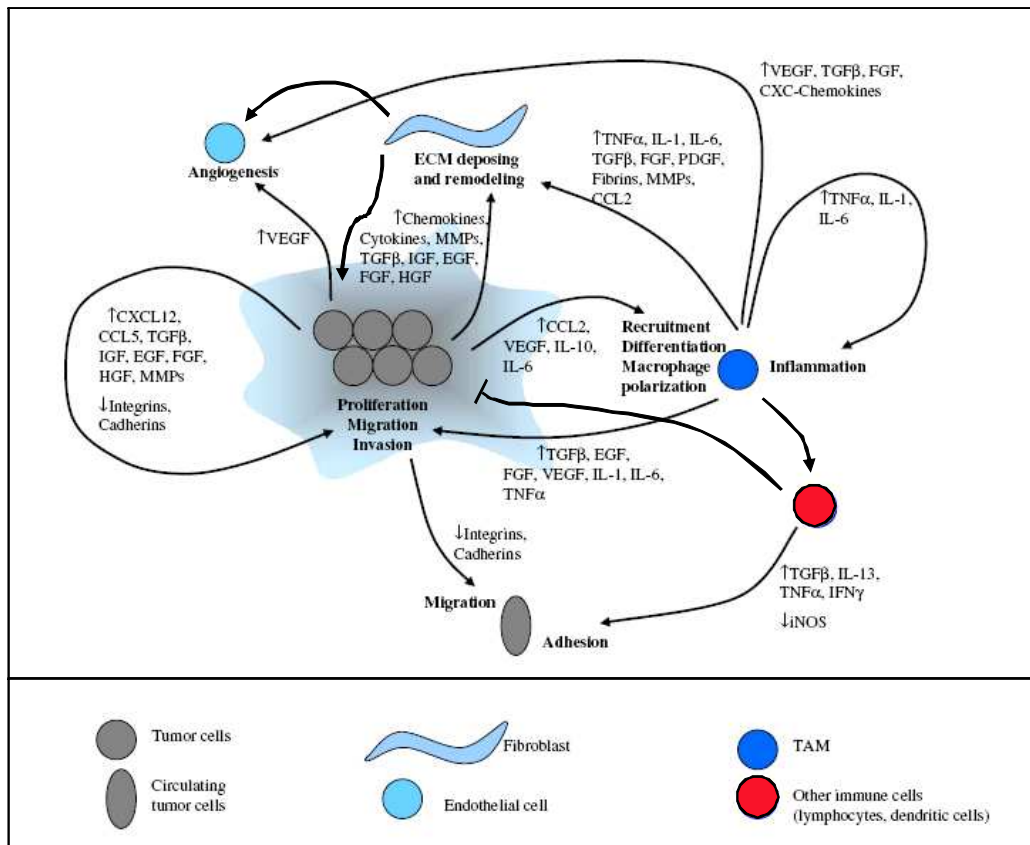


Fig.6. Esquema representativo de las principales interacciones entre las células epiteliales y estromales durante el desarrollo tumoral. Modificada de [140].

2.4. Matriz Extracelular (MEC) y Membrana Basal.

Es una matriz compuesta por proteínas, proteoglicanos y glicoproteínas de adhesión que da soporte estructural y mecánico a las células y tejidos [147]. La matriz extracelular da soporte a las células del estroma y se encuentra principalmente constituida por colágeno I, III y V y fibronectina. Separando el estroma de las células epiteliales (y dando soporte a estas últimas), se encuentra un tipo de matriz extracelular especializada denominada membrana basal que está compuesta principalmente por colágeno tipo IV, laminina, proteoglicanos de heparan-sulfato y otras moléculas en menor cantidad como SPARC u osteopontina [148].

La matriz extracelular, dependiendo de su integridad estructural y de su ensamblaje puede favorecer o inhibir el desarrollo tumoral. La degradación de la membrana basal va a suponer, ante todo, la eliminación de la barrera que limita la migración celular. Por otro lado, el proceso de degradación de la matriz extracelular,

llevado a cabo por metaloproteasas (MMPs) o proteinasas como Catepsina B y D, Enastasa o uPA, secretadas en su mayoría por células del estroma, liberará zonas de unión en la MEC con receptores de membrana que hasta ese momento habían permanecido bloqueados [149]. En este proceso, también se producirá la liberación de nuevos elementos procedentes de la degradación de compuestos de la MEC con actividad biológica implicada en el desarrollo tumoral [150,151] y se liberarán factores de crecimiento (TGF- β , b-FGF, PDGF, etc.) que se encontraban secuestrados por los componentes de la matriz que acaban de ser degradados [152,153]. Todo esto tendrá un gran impacto en los procesos de supervivencia, proliferación y migración celular [154].

La degradación de la membrana basal vascular va a ser primordial en los procesos de angiogénesis, ya que va a promover la proliferación y migración de las células endoteliales vasculares, además de liberar factores de crecimiento implicados en la generación de nuevos vasos (VEGF, PDGF, etc.). Los productos finales de la degradación (endostatina, arrestina, tumstatina, etc.) van a contribuir sin embargo a la inhibición de la angiogénesis [148].

3. Tratamientos en el cáncer colorrectal esporádico.

En la actualidad el tratamiento estándar en pacientes con cáncer colorrectal continúa siendo la resección del mismo mediante cirugía. La aplicación de tratamientos de quimioterapia o de radioterapia ha sido especialmente beneficiosa para pacientes con tumores colorrectales en estadio localmente avanzado (III) y avanzado (IV) y en algunos casos también para aquellos con estadios localizados (II). El descubrimiento de nuevos fármacos citotóxicos y de anticuerpos monoclonales dirigidos contra dianas moleculares específicas, junto con la posibilidad de realizar tratamientos multidiana en los que se combinan varios de estos fármacos a la vez (Ej.: Leucovorin + 5-Fluorouracilo + Irinotecan = FOLFIRI) ha conducido a una mejora en la supervivencia y en la calidad de vida de los enfermos. A continuación se presenta una descripción de los fármacos más utilizados en el tratamiento de dicha enfermedad:

3.1. 5-Fluorouracilo (5-FU) [155].

Es un antagonista de pirimidinas perteneciente al grupo de los antimetabolitos, que posee una estructura similar al uracilo, con la única diferencia de poseer un átomo de flúor en la posición 5 de la molécula. Su actividad citotóxica viene mediada por dos mecanismos específicos de la fase S del ciclo celular (Fig.7.)

- ❖ En condiciones normales, la Timidilato Sintetasa (TS) se encarga de transformar la deoxiuridina monofosfato (dUMP) en deoxitimidina monofosfato (dTMP) precursor de la deoxitimidina trifosfato (dTTP), uno de los nucleótidos esenciales para la síntesis del DNA. El 5-FU en su forma activa (F-dUMP) posee una estructura análoga al sustrato de la TS, lo que le permite unirse a la enzima, bloqueándola e inhibiendo por tanto su actividad. La inhibición prolongada de la enzima conducirá a la apoptosis de la célula.
- ❖ El 5-FU en su forma activa (F-dUMP/F-UMP) es capaz de fosforilarse hasta convertirse en un falso nucleótido trifosfato (F-dUTP o F-UTP) que se incorporará a la cadena creciente de DNA y RNA abortando los procesos de replicación del DNA y los procesos de transcripción del RNA, lo que llevará consigo la muerte celular.

3.2. Capecitabina [156].

Fluoropirimidina oral que es transformada mediante reacciones enzimáticas en 5-FU (Fig.7.). Posteriormente tras convertirse en su producto activo (F-dUMP/F-UMP) actuará como un fármaco citotóxico bloqueando los procesos de síntesis y replicación del DNA y de transcripción del RNA, disparando la apoptosis celular (Fig.7).

3.3. UFT [157]:

Fluoropirimidina oral compuesta por Tegafur y Uracilo en una proporción de 1:4. El Tegafur es una pro-droga que es convertida a 5-FU en el hígado. En dicho órgano, el uracilo se une e inhibe la acción de la enzima dihidropirimidina dehidrogenasa (DPD), encargada de catabolizar el 5-FU. En su forma activa el 5-FU actuará de manera citotóxica contra el tumor (Fig.7).

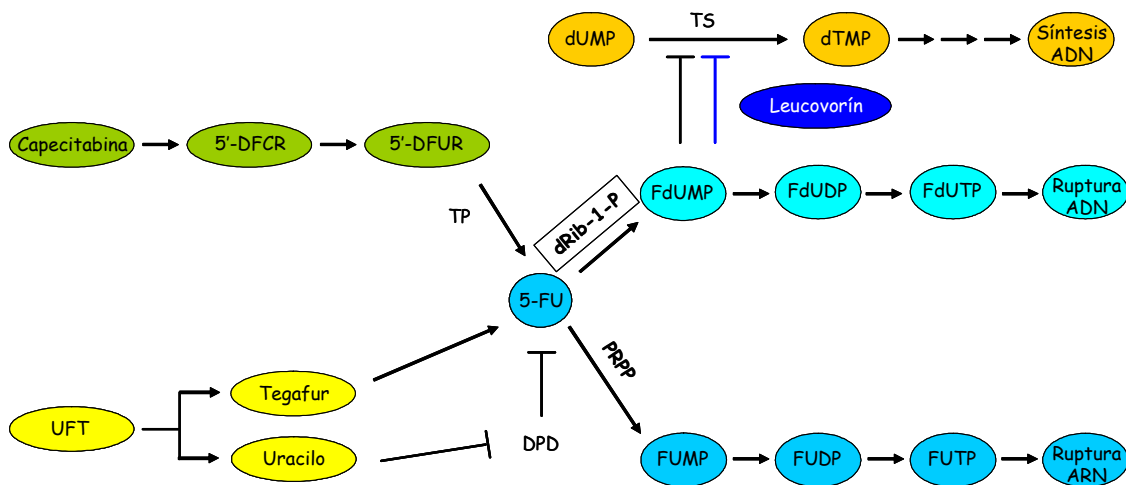


Fig.7. Mecanismos de acción del leucovorín y de las drogas citotóxicas pertenecientes al grupo del fluorouracilo (5-FU, Capecitabina y UFT). TS: Timidilato sintetasa; DPD: Dihidropirimidina dehidrogenasa; TP: timidina fosforilasa; PRPP: fosforribosil pirofosfato.

3.4. Leucovorin (LV) [158].

Metabolito proveniente del ácido fólico que normalmente se utiliza para paliar los efectos secundarios producidos por fármacos como el metotrexato en las células normales. En combinación con 5-FU aumenta el potencial inhibitorio de este último sobre la enzima TS. El leucovorín estabiliza la unión entre la TS y la forma activa del 5-FU, aumentando así el periodo de actividad de este último. Otros fármacos como el Levamisol realizan una función similar. (Fig.7.)

3.5. Oxaliplatino (OXL) [159].

Es un derivado del platino de tercera generación perteneciente al grupo de agentes alquilantes. Actúa uniéndose a la cadena de DNA en la que formará aductos Platino-DNA tanto inter- como intracatenarios, lo que impedirá la síntesis y replicación del DNA. Una característica importante de esta droga es su elevado sinergismo en cuanto a su acción citotóxica al ser administrado junto con 5-FU.

3.6. Irinotecan (CPT-11) [29].

Es un derivado de la camptotecina que ha demostrado ser eficaz en un amplio número de tumores. Su diana molecular es la Topoisomerasa I (TOP1), una enzima nuclear implicada en los procesos de enrollamiento y relajación de la torsión de las hebras del DNA, necesarios para que tengan lugar los procesos de replicación, recombinación, reparación y transcripción del DNA. El Irinotecan, tras ser metabolizado a su forma activa (SN-38) en el hígado, actúa estabilizando el complejo “ssDNA-TOP I” e inhibiendo el paso de ligación de las dos hebras del DNA. Esto no es letal para la célula, sin embargo el choque que se produce entre la maquinaria de replicación y el complejo “ssDNA-TOP1” produce numerosas roturas en el DNA que serán irreversibles y que llevarán a la muerte celular. Este mecanismo es, por tanto, específico de la fase S del ciclo celular (Fig.8).

3.7. Gefitinib [160] y Erlotinib [161]:

La unión del Factor de Crecimiento Endotelial (EGF) con su receptor (EGFR) produce la dimerización del mismo (con otro monómero miembro de la familia Erb-B), dando lugar a la activación de los dominios tirosina quinasa intracelulares y la consiguiente autofosforilación, que disparará una cascada de señalización intracelular implicada en procesos de crecimiento, proliferación, angiogénesis, inhibición de la apoptosis, metástasis, etc. Gefitinib es un medicamento oral que actúa como inhibidor selectivo de los dominios tirosina quinasa del receptor EGFR (también denominado ErbB-1) mediante el acoplamiento a la región de unión a ATP, evitando que esta molécula se pueda unir y por tanto producirse la autofosforilación del receptor y la activación de la cascada de señalización (Fig.9). El fármaco Erlotinib actúa mediante el mismo mecanismo que el Gefitinib (Fig.9). En la actualidad se están desarrollando otros fármacos cuya actividad sea efectiva en más de un tipo de receptor Erb-B, como es el caso de Lapatinib, capaz de afectar la activación de los dominios tirosina quinasa de EGFR y de Her2 a la vez [162].

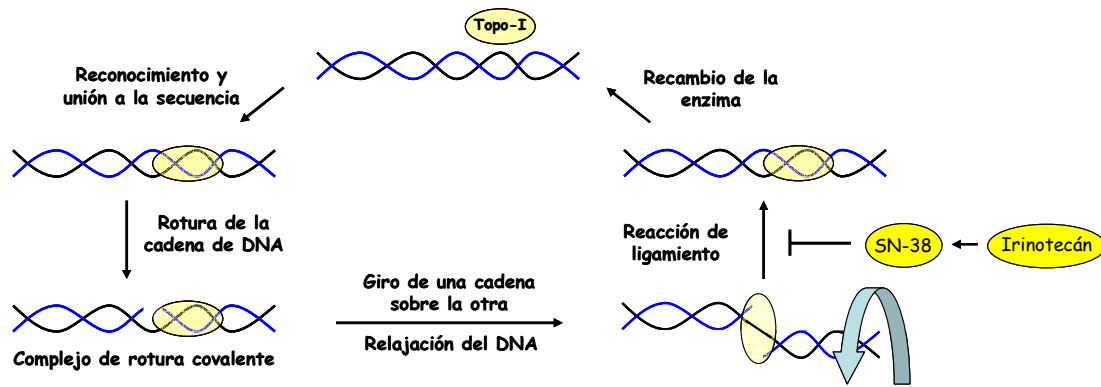


Fig.8. Mecanismos de acción de Irinotecan sobre la actividad de TOPO-I descrita.

3.8. Cetuximab [163]:

Es un anticuerpo monoclonal quimérico (ratón/humano) cuya diana es el Receptor del Factor de Crecimiento Epidérmico (EGFR) al que se une con una afinidad aproximadamente 5-10 veces superior a la de sus ligandos endógenos como el EGF. Esta unión bloquea la interacción ligando-receptor, impidiendo la activación del mismo y la consecuente activación de la cascada de señalización intracelular que llevará a la activación de vías de proliferación, angiogénesis, invasión y metástasis, etc. (Fig.9). En la actualidad se han desarrollado anticuerpos totalmente humanizados, como el Panitumumab [164], que mimetizan la acción del Cetuximab.

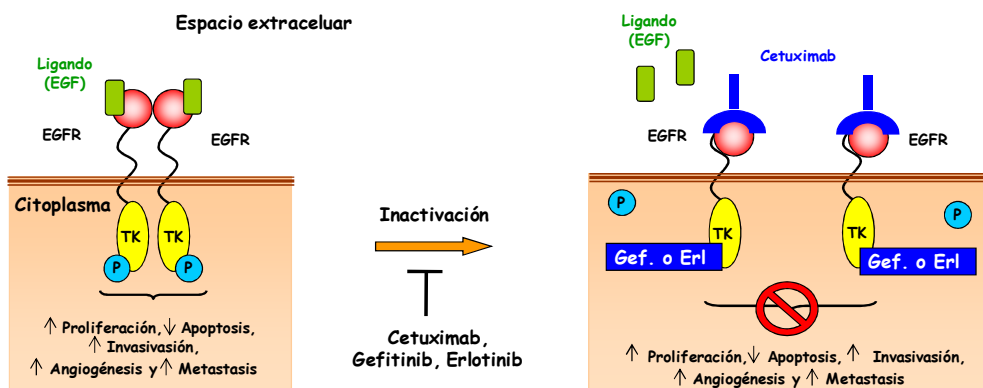


Fig.9. Representación de los mecanismo de acción de los fármacos anti-EGFR.

3.9. Bevacizumab [165]:

Es un anticuerpo monoclonal humanizado que se une al Factor de Crecimiento del Endotelio Vascular (VEGF), inhibiendo así la unión de éste a sus receptores Flt-1 (VEGFR-1) y KDR (VEGFR-2), situados en la superficie de las células endoteliales. Al neutralizar la actividad biológica del VEGF se reduce la vascularización de los tumores y otros procesos implicados en el crecimiento del tumor (Fig.10).

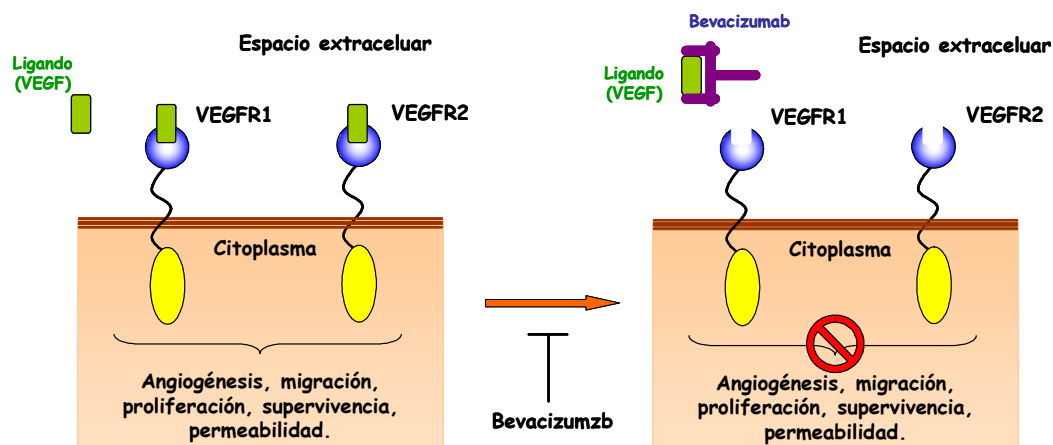


Fig.10. Representación gráfica del mecanismo de acción del fármaco Bevacizumab.

La disponibilidad de nuevos fármacos ha supuesto un abanico de posibilidades a la hora de tratar el CCR. Sin embargo, a la hora de administrar estos fármacos hay que tener en cuenta que:

- ❖ Un porcentaje elevado de pacientes (en Estadio II y III) se curará completamente mediante cirugía, siendo la aplicación de tratamientos de QT o RTx innecesarios y nocivos para el paciente debido a la elevada toxicidad que presentan algunos de estos fármacos.
- ❖ Un porcentaje importante de pacientes mostrará desde un principio resistencia a determinados fármacos.
- ❖ Un porcentaje importante de pacientes que en un comienzo eran sensibles a determinados fármacos, acabarán haciéndose resistentes a los mismos con el paso del tiempo.

Así, sería importante poder descubrir qué pacientes van a necesitar de un tratamiento complementario a la cirugía, que fármacos serán los más adecuados en cada caso y cuáles son los mecanismos que determinarán el estado “sensibilidad/resistencia” al tratamiento, para poder actuar más eficazmente en la lucha contra el cáncer.

4. Microarrays.

La técnica de los microarrays surge a mediados de los años 90 [166,167] como un método para estudiar la expresión de miles de genes de manera simultánea. La secuenciación del genoma humano [168,169] y la evolución de las técnicas de análisis molecular impulsaron enormemente el desarrollo de esta nueva metodología, haciendo que en la actualidad no sólo sea posible el estudio global del RNA que codifica para proteínas (variaciones en los niveles de mRNA mediante “Gene Expresión Profiling Analysis” [170]), sino también del genoma (análisis de mutaciones y polimorfismos mediante “Genome Wide Association Studies” [171] o de alteraciones en el número de copias de regiones de DNA mediante “Comparative Genome Hibridization” [172]) e incluso de parte del RNA no codificante (perfiles de miRNAs [173]) (Fig.11). Otras aplicaciones que se han ido sumando a este amplio catalogo de estudios que ofrece la técnica de los microarrays son el estudio de masivo de los procesos de metilación de las regiones promotoras de los genes [174] o el estudio de los factores transcripcionales mediante técnicas de inmunoprecipitación de cromatina [175].

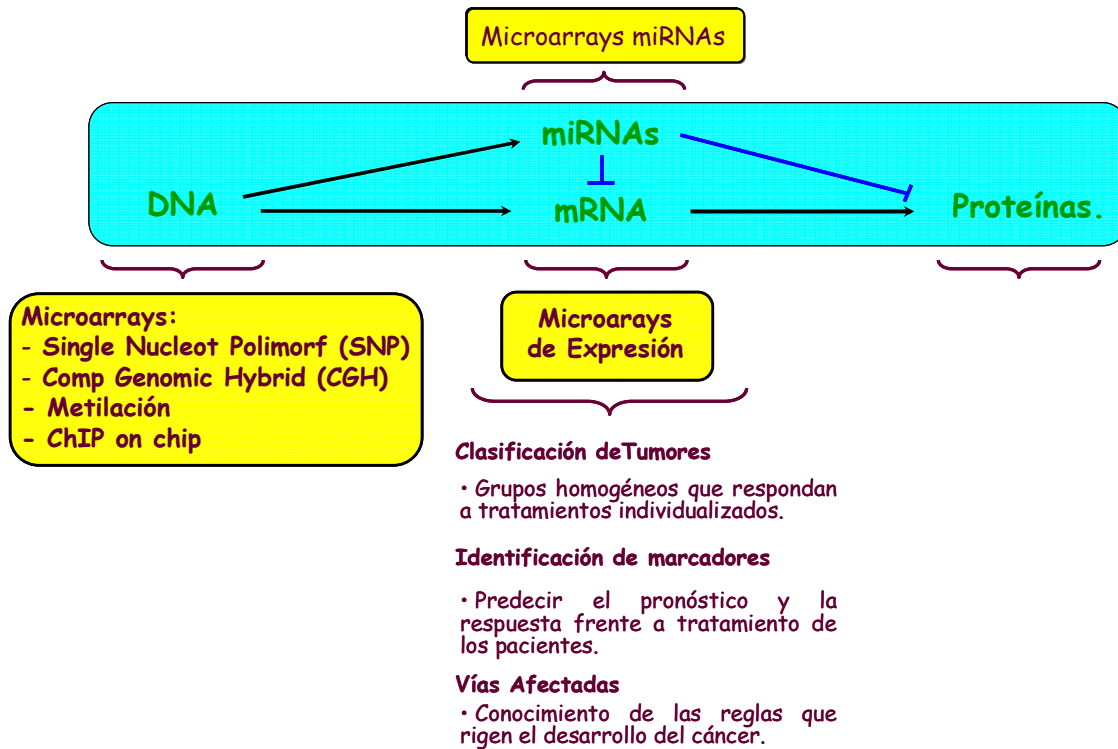


Fig.11. Aplicaciones de la técnica de los microarrays en las diferentes áreas de la biología molecular aplicadas a cáncer.

4.1. Microarrays de expresión. Introducción.

Los microarrays de expresión (Fig.12) se definen como un conjunto de miles de sondas (cDNAs u oligonucleótidos de DNA de 25 a 70 pb) dispuestas de manera ordenada sobre una superficie sólida, ya sea cristal o plástico, que permite analizar simultáneamente la expresión del genoma completo de un organismo. Cada una de las sondas representa específicamente un gen determinado al poseer una secuencia complementaria al mRNA transcrito por dicho gen posibilitando así, la medición de los niveles de expresión de todos los genes que conforman el genoma al mismo tiempo y en un único experimento. Los estudios de microarrays constan de 3 fases diferenciadas:

4.1.1. Fabricación de los microarrays.

Hasta hace poco tiempo, uno de los pasos más importante en este tipo de estudios era la síntesis del propio microarray, ya que la mayoría de ellos eran sintetizados por los propios laboratorios utilizando como sondas cDNAs de diferentes tamaños. Sin embargo desde hace unos años este tipo de chips está siendo sustituido por

microarrays de oligonucleótidos elaborados comercialmente por distintas empresas. Este nuevo tipo de microarrays, utiliza oligonucleótidos de un mismo tamaño, aumentando así la homogeneidad de la hibridación, disminuyendo los procesos de hibridación cruzada y maximizando la reproducibilidad obtenida en los resultados.

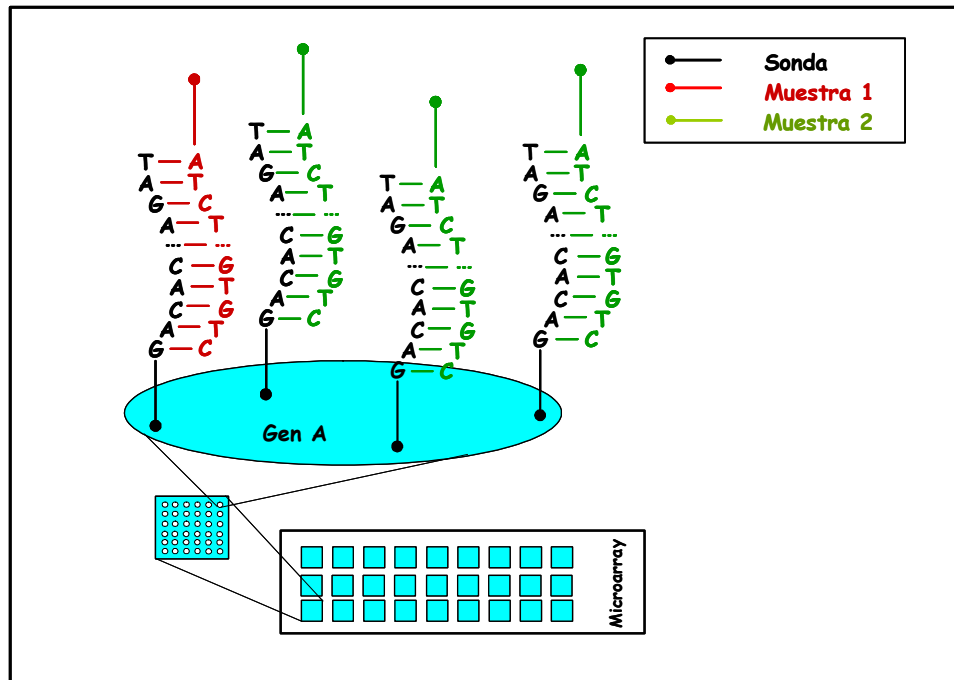


Fig.12. Esquema de un microarray en el que se observa la disposición ordenada de los *spots* o puntos sobre un soporte sólido. Cada *spots* está compuesto por numerosas copias de una sonda específica para un único gen (Ej. Agilent microarray). El mRNA del gen A para la muestra a estudio se unirá por complementariedad a las sondas dispuestas en el array que reconozcan de manera específica dicho gen. En este ejemplo en el que se muestra una hibridación doble, las dos muestras hibridadas competirán por la unión a las sondas. La cantidad de mRNA transcrito por cada gen es proporcional a la cantidad que se une a las sondas en el array.

Dentro de estos microarrays de oligonucleótidos existen dos *corrientes* en función de la longitud del oligonucleótido utilizado:

- ❖ Oligonucleótidos cortos: de unas 25 pb aproximadamente, como los utilizados por la plataforma de Affymetrix [176]. Este tipo de oligonucleótidos tienen el problema de que debido a su pequeño tamaño tienen una menor fuerza de unión al mRNA. Para solucionar el efecto que pueda tener este problema en los resultados obtenidos, la compañía Affymetrix utiliza una técnica de “mismatch-perfect match” y varios oligonucleótidos diferentes por cada gen.

- ❖ Oligonucleótidos largos: de entre 50-70 pb. Al ser más largos, tienen una mayor fuerza de unión y una mayor especificidad. Este tipo de sondas son las que más se están extendiendo y son utilizadas por casas comerciales como Agilent, Applied Biosystems, Illumina o recientemente Biotools.

Las estrategias a la hora de sintetizar los oligonucleótidos depende de la casa comercial, utilizando en general dos posibilidades: o bien la síntesis in situ del oligo sobre el propio microarray como es el caso de Agilent Technologies y de Affymetrix o bien la síntesis previa del oligo que luego será anclada a la matriz como es el caso de Applied Biosystems.

4.1.2. Marcaje e Hibridación de las muestras.

El primer paso es el proceso de marcaje del RNA que consiste en “etiquetar” todos los RNAs mensajeros del tumor, con una molécula fluorescente que luego seamos capaces de medir, sin alterar su proporción relativa y con una actividad específica suficiente para que produzcan una señal robusta [177,178]. Este proceso se puede realizar de manera directa, para lo que es necesario partir de concentraciones muy elevadas de RNA (entre 10 y 40 µg). De esta manera el mensajero es marcado durante su retrotranscripción a cDNA, molécula la cual será hibridada al microarray. El marcaje indirecto, permite sin embargo y gracias a la introducción de un pequeño paso de amplificación, partir de una menor cantidad de RNA inicial (0.1-1 µg). En este caso el mensajero suele ser transformado a cDNA y posteriormente a cRNA, paso en el cuál se produce el marcaje y la amplificación de la muestra. Este último método suele ser el más utilizado cuando se trabaja con muestras humana ante la usual escasez de material disponible.

Una vez realizado el marcaje se produce la hibridación del mensajero al microarray. En función de cómo se quiera estudiar la expresión diferencial entre las muestras, se utilizará o bien la hibridación simple o bien la hibridación doble (o competitiva):

- ❖ Hibridación simple: las muestras que se van a comparar se hibridan individualmente a un microarray, marcando todas las muestras con el mismo fluorocromo (Ej.:

Cianina 3). Esto hace que el error de incorporación o la velocidad de degradación del fluorocromo afecten por igual a todas ellas. Este diseño permite una mayor flexibilidad a la hora de realizar las comparaciones, aunque sin embargo dificulta los procesos de normalización.

- ❖ Hibridación doble: se marcan las dos muestras a comparar con compuestos diferentes (generalmente Cianina-3 y Cianina-5) y se hibridan juntas a un solo microarray. La ventaja de este tipo de hibridación es que al usar el mismo microarray, el ruido del fondo de cristal y de los “spots” de DNA son los mismos para ambas muestras y que en caso de usar una muestra de referencia, los procesos de normalización son más fáciles de llevar a cabo al contar con una referencia común para todas las muestras. Las desventajas principales son que exige corregir los datos de expresión debido a las diferencias de incorporación de cada fluorocromo y que las posibilidades experimentales también se ven más limitadas ante la necesidad de tener que contar con muestra de referencia suficiente, para cualquier tipo de comparación que se quiera llevar a cabo.

4.1.3. Análisis de datos.

La unión de cada RNA a su sonda (dispuesta en un “spot” o círculo en el microarray) produce una intensidad de fluorescencia que es directamente proporcional al nivel de expresión de ese gen. Una de las razones por las que el análisis de datos presenta una gran complejidad es debido al gran número de variables (genes) que convergen en cada experimento (Ej.: del orden de varios miles cuando se estudia el genoma humano completo) y al bajo tamaño muestral, en comparación, utilizado en los estudios (alrededor de 15-100 muestras). Una vez escaneado el microarray hibridado y procesada la imagen obtenida, se genera la matriz de expresión sobre la que se realizará el análisis de datos.

En este tipo de estudios los datos se organizan en matrices en donde cada línea representa un gen y cada columna una muestra, pudiendo ser definidos, tanto muestras como genes, en forma de vectores (Fig.13.). En el caso de los genes, cada uno será descrito como un “vector de expresión génica” con “m” dimensiones o valores, siendo “m” el número de muestras que conforman la matriz (ej.: un gen “g1” cuyo valor ha

sido calculado para tres muestras, se expresará como un vector con tres coordenadas o dimensiones de la forma (m_1, m_2, m_3) .

$$\begin{array}{c}
 \mathbf{g}_1 \\
 \mathbf{g}_2 \\
 \mathbf{g}_3 \\
 \dots
 \end{array}
 \begin{pmatrix}
 m_1 & m_2 & m_3 & m_4 & \dots \\
 12 & -4 & 6 & 0 & \dots \\
 5 & 3 & -1 & 10 & \dots \\
 11 & 3 & 0 & -2 & \dots \\
 \dots & \dots & \dots & \dots & \dots
 \end{pmatrix}$$

Fig.13. Representación de la matriz de datos de expresión en donde cada columna representa una muestra “m” y cada fila un gen “g”.

El análisis de los datos consta de dos fases:

❖ Pre-procesado de datos:

- *Filtrado de datos [177]:* “Limpia” la matriz de datos generada mediante la eliminación de sondas cuya expresión o bien es tan baja que se confunde con el ruido de fondo o bien refleja problemas puntuales acontecidos durante el proceso de hibridación.
- *Normalización [179,180]:* corrige errores o sesgos sistemáticos introducidos durante los procesos de marcaje e hibridación de las muestras, originados por la distinta incorporación de los fluorocromos, por diferencias en la cantidad de RNA inicial de las muestras estudiadas, por la desigual distribución de la fluorescencia a lo largo del chip, etc. Estos sesgos suelen ser de tipo espacial y/o dependientes de la intensidad. Para corregirlos se calculan factores normalización que asumen dos hechos: por un lado que la cantidad de muestra hibridada es la misma y por tanto el número total de transcritos también y por otro, que existe un número elevado de genes cuya expresión va a permanecer invariable. Dependiendo de si los sesgos encontrados varían o no en función de la intensidad de fluorescencia, se utilizarán métodos de normalización no lineales (Ej. LOWESS) o lineales (Ej. “mean normalization”, regresión lineal, etc.) respectivamente.

❖ Análisis de datos:

Una vez los datos filtrados, normalizados y por tanto listos para el estudio, su análisis dependerá de cual sea el objetivo del trabajo, pudiendo usarse métodos no supervisados o métodos supervisados [177,178].

- *Métodos no supervisados (Ver Anexo I. Pág.170):* son técnicas de agrupación que se basan exclusivamente en los datos de expresión para generar los resultados, sin hacer uso de ningún conocimiento previo relativo a la patología. Estos métodos se utilizan en los estudios de descubrimiento de clases en donde se intenta crear una nueva clasificación o descubrir nuevos subtipos moleculares con relevancia en la patología, basándose solamente en las similitudes entre los perfiles de expresión. Entre los métodos más frecuentes destacan la técnica de agrupación jerárquica (“clustering” jerárquico), el “k-means”, los “Self Organizing Maps” (SOM) o el análisis de componentes principales (“Principal Component Analysis” o PCA).

- *Métodos supervisados (ver Anexo I. Pág.182):* se utilizan parámetros externos a la matriz de expresión (Ej. estadiaje del tumor, estatus del receptor de estrógeno, etc.) para generar los supuestos que se van a analizar mediante las matrices de expresión. Cuando el objetivo final es caracterizar las diferencias entre grupos de muestras se utilizan los procesos de comparación de clases (t-test, ANOVA, etc.). Sin embargo si el objetivo último es establecer un clasificador que permita catalogar nuevas muestras en los grupos previamente definidos, se utilizarán los procedimientos de predicción de clases (“Support Vector Machines” (SVM), “k-nearest neighbor” (KNN), etc.). Estas técnicas crean algoritmos de clasificación en base a los genes que muestran una mayor expresión diferencial entre los diferentes grupos, de manera que al aplicarlo sobre los valores de expresión de una nueva muestra, se obtiene el grupo al que es más probable que la nueva muestra pertenezca.

Los tres supuestos anteriores (descubrimiento, comparación y predicción de clases) exigen, en cualquier caso, para confirmar la universalidad de sus resultados la

verificación de los hallazgos realizados en conjuntos de datos independientes de aquel utilizado para obtener los resultados [181].

4.2. Fiabilidad y reproducibilidad de la técnica de los microarrays.

La euforia creada ante la posibilidad de ser capaces, mediante microarrays, de caracterizar todas las enfermedades humanas en un corto período de tiempo [182], llevo a la aparición de artículos poco rigurosos que utilizaban esta metodología. Consecuencia de ello, en 2005 se publicó una comparativa de diversos estudios de microarrays en cáncer en las que se ponía en duda su reproducibilidad y se insinuaba que la mayoría de ellos no lograban obtener una clasificación más precisa que la que se podría establecer al azar [183]. Durante este tiempo y en consonancia con este estudio, otros trabajos habían señalado las numerosas carencias de este tipo de estudios [183-190].

A raíz de esto, a partir del año 2005 y en paralelo con la evolución de la técnica y la progresiva estandarización de los protocolos, fueron surgiendo una serie de trabajos llevados a cabo por grupos cooperativos [191-196] dentro del proyecto Microarray Quality Control (MAQC), que demostraron la reproducibilidad de esta técnica tanto entre laboratorios como entre plataformas siempre y cuando se estandarizasen los protocolos y se establecieran controles de calidad. Además, durante este tiempo, otros grupos consiguieron demostrar la validez funcional de los predictores moleculares, al probar como en el caso de cáncer de mama, que distintos predictores pronóstico reconocían los mismos grupos de pacientes [197] y que existía una similitud biológica entre los genes que conformaban los distintos predictores [198].

Como parte de los procesos de estandarización también se ha hecho hincapié sobre la información que debía ser hecha pública en este tipo de trabajos. A partir de 2001 diversos grupos colaborativos [199] comenzaron a desarrollar las denominadas reglas MIAME o Minimum Information About Microarray Experiments en donde se especifican la información mínima que se ha de proporcionar de cada análisis de microarrays (dato bruto, dato procesado, identificación y características de las muestras, anotación de los genes, etc.) para facilitar el intercambio de información entre grupos de investigación y la interpretación por otros grupos de los trabajos publicados.

4.3. Aplicaciones de los microarrays de expresión en cáncer.

La aparición de esta tecnología ha permitido un mayor conocimiento de las vías moleculares que regulan el desarrollo de gran cantidad de enfermedades, facilitando en algunos casos la toma de decisiones clínicas y el manejo de los pacientes. La oncología es una de las áreas en las que los microarrays han despertado mayor interés, quizás debido a la posibilidad de que esta técnica sea capaz de poner de manifiesto gran parte de la variabilidad inherente a esta enfermedad. Hasta ahora, los trabajos realizados con esta técnica en el campo de la oncología han permitido [177,200,201]:

- ❖ Establecer la individualidad del tumor (tumor-metástasis) frente a otros tumores [202], observándose que el tumor primario es más parecido a su metástasis que a otro tumor primario aunque este pertenezca al mismo estadio.
- ❖ Descubrir nuevas vías alteradas en el desarrollo de tumores, ya sean independientes del órgano en el que se desarrollen (se describió la existencia de un “cluster” de genes implicado en proliferación que se encontraba sobreexpresado en tumores de diferentes órganos) [203], o específicas del mismo [204].
- ❖ Descubrir nuevos subtipos tumorales basados en la composición celular del tumor [202,205-207] y en base a ello crear nuevos sistemas de clasificación con valor pronóstico [205,207,208].
- ❖ Desarrollar predictores pronóstico para diferentes tumores como el de cerebro [209], riñón [210], linfomas [211], pulmón [212] o mama [170,213-216]. Algunos de estos predictores han sido aprobados para su uso en la práctica clínica, como el MamaPrint (Agendia) [213], el OncotypeDX (Genomic Health) [214] y el H/I test (Aviara Dx) [216] en el caso de cáncer de mama. Dentro del desarrollo de perfiles pronósticos cabe destacar la importancia que están adquiriendo aquellos basados en la expresión de los componentes del estroma [217,218].
- ❖ Creación de perfiles de expresión predictivos, ya sea mediante trabajos en líneas celulares y posterior validación en tumores humanos [219,220], o bien directamente

mediante estudios de comparación entre pacientes respondedores y no respondedores frente a tratamiento [221-226].

- ❖ Determinar vías capaces de modular la sensibilidad o resistencia de las células ante un determinado tratamiento [227,228].
- ❖ Caracterizar nuevos marcadores diagnóstico [229] o posibles nuevas dianas terapéuticas para el tratamiento del cáncer [230].

4.4. Papel de los perfiles de expresión en el cáncer de colorrectal.

A pesar de que los resultados obtenidos mediante este tipo de trabajos en cáncer colorrectal no alcanzan la importancia de los descritos en otros tumores como el de mama para el que algunas de las firmas genéticas encontradas ya se están utilizando en la clínica, existe una gran cantidad de trabajos que vienen a sugerir el posible uso clínico de los perfiles de expresión en cáncer colorrectal en un futuro próximo. Hasta el momento, los estudios realizados han conseguido establecer con más o menos fortuna perfiles de expresión diferentes en función del pronóstico del paciente, de su posible respuesta a tratamiento o de parámetros tales como la localización tumoral, la presencia de inestabilidad en microsatélites, etc. [231]. Los resultados más destacados se describen brevemente a continuación.

4.4.1. Desarrollo de la enfermedad: “mucosa normal – adenoma – adenocarcinoma”.

Los primeros estudios demostraron que el uso de técnicas globales en expresión no sólo permitía separar muestras tumorales de muestras de mucosa de colon normal [232-237], si no que mediante el análisis de los genes que se expresaban de manera diferencial, se podía también conocer la composición celular de los tejidos (estroma, epitelio, músculo liso, etc.) [232,238] o ver qué genes se encontraban funcionalmente correlacionados entre sí [232]. Estudios posteriores, consiguieron añadir un grado más de definición al ser capaces de diferenciar a nivel molecular la evolución tumoral: “mucosa normal - adenoma - adenocarcinoma” [237,239,240].

Durante el desarrollo de estos estudios se fueron describiendo una serie de genes que caracterizaban y diferenciaban estos tres “estadios”. Así, se observó la desregulación de vías biológicas que proporcionan ventajas selectivas a las células tumorales (adenomas y adenocarcinomas) como la síntesis del RNA y proteínas, la expresión de oncogenes, crecimiento y división celular, la transducción de señales, la desregulación del ciclo celular, los procesos de muerte celular y de inmunidad, etc. También se encontraron genes implicados en crecimiento celular y proliferación cuya expresión iba aumentando a medida que aumentaba el proceso de malignización. Finalmente genes diferencialmente expresados entre adenomas y carcinomas se vio que estaban relacionados con situaciones de hipoxia y privación de alimentos, lo que podría indicar que las células tumorales están cambiando sus perfiles de expresión y se están adaptando a un entorno con baja concentración de oxígeno y de nutrientes.

4.4.2. ¿Diferente localización del tumor = Diferente pronóstico?

Algunos estudios insinuaron la posibilidad de que la distinta localización del tumor (colon derecho-colon izquierdo) pudiese estar relacionada con diferencias en el desarrollo del tumor y en el pronóstico del paciente [241,242], lo que a su vez podría venir reflejado por la aparición de perfiles moleculares diferentes para tumores de ambas localizaciones. Sin embargo y aunque se confirmó la existencia de genes capaces de diferenciar los tumores según su localización proximal o distal en el intestino grueso, y se encontró asociación entre dicha localización con parámetros como la invasión linfática o la invasión vascular, no se ha podido confirmar que el pronóstico de dichos pacientes sea significativamente diferente, ni que el índice de recurrencia sea mayor o menor o que vivan más o menos tiempo dependiendo de donde se localice la enfermedad.

4.4.3. Perfiles de expresión e Inestabilidad genética.

Aproximadamente un 10-15% de los tumores esporádicos de colon y los tumores de tipo HNPCC presentan inestabilidad en secuencias microsatélite (MSI). A lo largo de los años diversos trabajos han relacionado el estatus “MSI” con diferentes características clínico-patológicas [63] (Tabla.6) y con un patrón de desarrollo diferente al de los tumores estables [243].

Varios estudios han demostrado mediante técnicas no supervisadas [243,244] que existen diferencias moleculares entre los tumores MSS y los MSI. Los genes que caracterizan los perfiles de expresión de ambos grupos parecen apoyar la hipótesis de que estos dos grupos desarrollan el tumor por vías diferentes y además ambas vías estarían asociadas a distinto pronóstico [30,63,243-246]. Las diferencias más relevantes entre ambos tipos tumorales a nivel de la expresión génica son las siguientes:

- ❖ La sobreexpresión de un grupo de metalotioneinas en los MSI podrían estar detrás de la mayor resistencia de estos tumores a la quimioterapia [245].
- ❖ La sobreexpresión de genes como MUC5AC y MUC1 podrían ser los responsables de que los tumores MSI presenten un fenotipo mucinoso en comparación con los tumores MSS [243,246].
- ❖ La sobreexpresión de genes que codifican proteínas de respuesta inmune como la Granulinsina, el 4-1BBL, la Granzima A, etc. [243,244,246] se correlaciona con la característica de los tumores MSI de poseer infiltración de células del sistema inmune, lo que podría explicar, al menos en parte, el mejor pronóstico de este grupo frente a los MSS. Otro factor que podría contribuir a esta diferencia de pronóstico es el hecho de que genes que teóricamente estarían implicados en la progresión del tumor como VEGF β , DD5, NNOX1, PTP4A3, MMP7, etc. estarían inhibidos en los tumores MSI frente a los MSS, mientras que otros que teóricamente frenarían su progresión (Caspasa2, BCL10, SMAD4, HME, etc.) estarían sobreexpresados [244-246].

A este respecto cabe destacar el estudio de expresión realizados en adenocarcinomas serrados [247]. En este trabajo, Laiho et al. describen perfiles de expresión capaces de diferenciar entre tumores serrados y convencionales y apuntan hacia una posible conexión entre adenocarcinomas serrados y adenocarcinomas con inestabilidad en microsatélites y de fenotipo metilador.

4.4.4. Perfiles de expresión y clasificación de tumores.

El grupo de Friedericksen [248] intentó probar que era posible clasificar los tumores desde un punto de vista molecular en función de los estadios de Dukes. Sin embargo, sólo lo consiguieron con los tumores en estadios B y C, mientras que los estadios A y D quedaban sin agrupar. En sintonía con esto, el grupo de Kwong [249] no fue capaz de establecer diferencias entre los estadios B, C y D. En otros trabajos como los de Groene [250], aunque son capaces de diferenciar estadios II vs. III, se no observan vías moleculares diferencialmente alteradas entre los estadios de Dukes analizados.

Bertucci et al. [251] fueron capaces, mediante perfiles de expresión, de distinguir tumores con potencial metastásico (diagnosticados como tumores con metástasis o que durante el seguimiento la desarrollaron) de tumores sin dicho potencial (con ganglios afectados pero que no desarrollaron metástasis durante el seguimiento). Esta clasificación sin embargo distinguía peor si se dividía a los pacientes en estadios C o D de Dukes, reflejando de nuevo que la clasificación histológica no parece corresponderse con la clasificación molecular.

Estudios posteriores como los del grupo de Eschrich [252] y del grupo de Wang [253] apuntaban también hacía la existencia de heterogeneidad a nivel molecular dentro de los estadios clínicos, ya que los perfiles moleculares eran capaces de predecir buen y mal pronósticos cuando se analizaban solo tumores en estadios B o tumores en estadios B y C conjuntamente, cosa que no ocurría si se utilizaba parámetros como la propia clasificación de Dukes o el número de capas infiltradas por el tumor.

4.4.5. Perfiles de expresión y metástasis.

La aparición de metástasis es una de las causas principales de muerte por cáncer. Cuando se habla de investigar los patrones que rigen su desarrollo es necesario considerar dos aspectos fundamentales:

- ❖ La capacidad del tumor para producir metástasis: diversos estudios han puesto de manifiesto que esta capacidad viene definida por el propio tumor primario, siendo

los perfiles de expresión de los tumores y los de sus metástasis correspondientes prácticamente iguales, mientras que el perfil de los tumores primarios que metastatizan difieren de aquellos otros que no lo hacen [251,254-256]. Esta teoría parece confirmarse mediante los resultados obtenidos en los estudios que relacionan perfiles de expresión en tumores primarios con pronóstico y/o recurrencia [252,253,257-260]. A pesar de ello, otros estudios como el de Ki [261] y el de Yanagawa [262] han sido capaces de encontrar perfiles diferenciales (aunque con un bajo número de genes entre los que se encontraba MMP1, MMP2, TIMP1) entre tumores colorrectales y sus metástasis.

- ❖ *El tropismo hacia un determinado órgano:* en cáncer colorrectal, nada se sabe a este respecto, pero se cree que esta capacidad podría también venir marcada en los tumores primarios, como ocurre en otros tumores como el de mama [263].

4.4.6. Perfiles de expresión y pronóstico.

Diversos estudios ya habían demostrado en otros tipos de cáncer la eficacia del uso de microarrays en la determinación del pronóstico de los pacientes [170]. El primer trabajo que confirmó la utilidad de las técnicas de análisis masivo para predecir recurrencia en cáncer colorrectal fue realizado por el grupo de Wang en 2004 [253]. A partir de entonces diversos grupos han centrado sus investigaciones en éste área (Tabla 8), Las conclusiones más importantes aportadas por estos trabajos son [252,253,257-260]:

- ❖ Es posible establecer predictores moleculares que diferencien pacientes con un riesgo más elevado de recurrencia. Curvas de Kaplan-Meier realizadas en base a estos perfiles moleculares corroboraron que existe una diferencia estadísticamente significativa en el tiempo de supervivencia entre ambos grupos (es decir, entre aquellos clasificados como “recurrentes o de mal pronóstico” y “no recurrentes o de buen pronóstico” utilizando perfiles genéticos).
- ❖ Predictores individuales que se habían tratado de utilizar con anterioridad, o bien no eran capaces de predecir la recurrencia, como ocurría con la expresión de los genes TP53, RAS, el tamaño del tumor, el grado de diferenciación, la pérdidas

alélica en el 18q o la división de las muestras según su estadio de Dukes (B y C) o bien su poder predictivo era menor del que mostraban la mayoría de las firmas genéticas (Ej.: “invasión de ganglios linfáticos + edad”, TNM, etc.).

- ❖ A través del estudio del perfil molecular del tumor se podría establecer el comportamiento y evolución del paciente, lo que no es posible hacer de manera precisa a través de la morfología del tumor.
- ❖ La falta de concordancia entre los genes que conforman los diferentes predictores moleculares no invalida su poder pronóstico. Esta afirmación, viene además avalada por el hecho de que varias de estas firmas genéticas ya han sido validadas por los propios grupos en matrices de muestras provenientes de otros laboratorios o incluso de manera completamente independiente por otros grupos como sucedió con la firma genética establecida por el grupo de Wang y que fue confirmada por el laboratorio de Barrier. La eficacia de un predictor y, por tanto, la posibilidad de aplicarlo a la clínica no depende tanto de su composición sino del hecho de que con él se pueda predecir recurrencia de manera precisa en cualquier grupo de muestras.
- ❖ La interacción Epitelio tumoral - Estroma podría jugar un papel muy importante en el desarrollo de la metástasis como se deduce del hecho de que grupos como el de Barrier hayan sido capaces de crear un predictor de recurrencia basado en perfiles de mucosa de colon normal adyacente al tumor.

En la actualidad se está a la espera de la aparición en el mercado de la primera prueba que mediante perfiles de expresión génicos será capaz de predecir recurrencia en pacientes con cáncer de colon en estadio II (Agendia) [264]. Este test comercial sería capaz de diagnosticar el riesgo de recurrencia con mayor precisión que los actuales parámetros clínico-patológicos, facilitando la identificación de los pacientes que se beneficiarían de la quimioterapia adyuvante.

4.4.7. Perfiles de expresión y tratamiento.

Al igual que en otros tumores, los estudios realizados en cáncer colorrectal en relación con el tratamiento han estado enfocados a descubrir los patrones que rigen la sensibilidad o resistencia (ya sea innata o adquirida) de los pacientes frente a determinados fármacos.

Tras la obtención de resultados positivos para la determinación de resistencia o sensibilidad a 5-FU y Camptotecina mediante perfiles de expresión en un conjunto de líneas celulares de carcinoma de colon [265], se llevaron a cabo los primeros estudios en muestras humanas. Varios grupos se plantearon la posibilidad de estudiar, a través de los perfiles de expresión de biopsias de tumor primario [266-268] o del propio tumor primario una vez resecado [269,270], si era posible diferenciar aquellos pacientes que iban a responder a un determinado tratamiento de aquellos otros que serían resistentes al mismo (Tabla 9). La mayor probabilidad de respuesta al tratamiento, podría permitir llevar a cabo intervenciones quirúrgicas en pacientes con tumores inicialmente irreseables, incrementando la probabilidad de curación o mejorando la supervivencia. En los cinco estudios descritos, se obtuvieron firmas genéticas que eran capaces de diferenciar respondedores de no respondedores. Además, estos resultados, junto con lo encontrado en otros trabajos [271], han permitido obtener una visión más global de los mecanismos de acción de algunos de esos fármacos, que afectaban vías relacionadas con adhesión, metabolismo de DNA, matriz extracelular y citoesqueleto, proliferación y división celular, apoptosis, etc.).

Esto confirma la posibilidad de utilizar este tipo de aproximaciones en el estudio del cáncer colorrectal para valorar la respuesta futura de los pacientes frente a los fármacos, aunque siempre siendo cautos y teniendo en cuenta que deberían llevarse a cabo nuevos ensayos con series de pacientes más homogéneas y numerosas.

Comparación de los artículos publicados relacionados con pronóstico en cáncer colorrectal y firmas genéticas.

Estudio, Año y Revista	Estadio de las muestras	Nº de muestras	Período de recurrencia	Predicador	Matriz de datos	Precisión	Sensibilidad	Especificidad	Vías implicadas
<i>Wang Y et al. JCO 2004</i>	Estadio II	74 muestras (43 S y 31 R)	3 años	23 genes (ej: Cadheria 17)	36 ME y 36 MV	78%	72%	83%	Ciclo celular, proliferación, progresión tumoral y respuesta inmune
<i>Arango D et al. Gastroenterology 2005</i>	Estadio III	25 muestras (15 S y 10 R)	5 años	17 genes (ej: RHOA)	25 muestra (ME = MV) (LOOCV)	88%	80%	93%	Transporte proteico (mitocondrial, nuclear, ret. Endoplasmático), procesamiento de proteínas (plegamiento y aminoacilación de tRNAs)
<i>Eschrich S et al. JCO 2005</i>	Estadio II, III y IV	78 muestras (30 S y 48 R)	3 años	43 genes (ej: SPP1 y Neuregulina)	78 muestras (LOOCV) MV = muestras danesas	90%	93%	84%	La osteopontina (SPP1) parece estar implicada en adhesión celular, invasión, angiogénesis y apoptosis. Neuregulina es un ligando para el receptor ERBB que puede estar implicado en crecimiento tumoral
<i>Barrier A et al. JCO 2006</i>	Estadio II	50 muestras (25 S y 25 R)	5 años	30 genes (ej: prot ribosomales)	ME = 10-40 muestras MV = 40-10 muestras	76.3%	85.1%	67.5%	Prot. Ribosomales: traducción y podrían estar implicadas en progresión de ciclo celular, apoptosis, respuesta a daño en el DNA e invasión tumoral
<i>Lin YH et al. Clin Canc Res 2007</i>	Nueva Zelanda: Estadio I, II, III y IV	149 muestras (102 S y 45 R)	5 años	T, N y M	149 muestra (LOOCV) MV = muestras Alemania	77%	34%	97%	Progresión tumoral, transcripción, traducción y mitosis
				TNM + Expresión (ZBTB6, CASP1, CXCL14)		73%	49%	86%	
				22 genes (ej: PKB)		77%	53%	88%	
	Alemania: Estadio I y II	55 muestras (26 S y 29 R)	5 años	Inv de Gángl Linf y Edad	55 muestras (LOOCV) MV = muestras Nueva Zelanda	65%	84%	45%	Implicados en respuesta inmune
				Inv GL + Edad+ Expres (CA2, CXCL9, PKB)		80%	74%	86%	
				19 genes (ej: PKB)		84%	85%	83%	
<i>Barrier A et al. Oncogene 2007</i>	Estadio II	24 muestras colon normal (14 S y 10 R)	5 años	70 genes (ej: Anexina y CD24)	18 ME y 8 MV	82%	73%	87.1%	Receptores de membrana plasmática, crecimiento celular y/o mantenimiento y respuesta inmune

Tabla .8. Resumen de los trabajos realizados con microarrays en relación a pronóstico de pacientes con cáncer colorrectal ^(18, 19, 23-26).

Nota: todos los trabajos han sido realizados con microarrays de la casa Affymetrix (Affymetrix, Santa Clara, CA) a excepción del experimento realizado por el grupo de Nueva Zelanda (MWG Human 30k array) y el grupo de Eschrich (TIGR array).

Abreviaturas: S = sano; R = recurrencia; T = capas infiltrada por el tumor; N = ganglios linfáticos afectados; M = metástasis; GL = ganglios linfáticos; ME = matriz ensayo; MV = matriz de validación; LOOCV = "leave-one-out-cross validation".

Precisión = proporción de pronósticos correctamente realizados; Sensibilidad = proporción de recurrencias correctamente diagnosticadas; Especificidad = proporción de "no recurrencia" correctamente asignadas.

Comparación de los artículos publicados relacionados con tratamiento en cáncer colorrectal y firmas genéticas.

Resumen de los estudios relacionados con tratamiento					
Trabajo	<i>Del Rio M et al. JCO 2007</i>	<i>Matsuyama R et al. Int J Cancer 2006</i>	<i>Ghadimi BM et al. JCO 2005</i>	<i>Watanabe T et al. Canc Res 2006</i>	<i>Kim IJ. Dis Col Rect 2007</i>
Origen del Tumor	Colon (21 pacientes)	Colon (22 pacientes)	Recto (30 pacientes)	Recto (52 pacientes)	Recto (46 pacientes)
Lugar de la metástasis	Hígado	15 Hígado 4 "Hígado+pulmón" 3 "Hígado+peritoneo"	---	---	---
Tratamiento	FOLFIRI (post-cirugía)	Terapia basada en 5-FU (post-cirugía)	QT (5-FU) + RT (previo cirugía)	RT (previo cirugía)	QT + RT (previo cirugía)
Finalidad del estudio	Disminución de la metástasis con la finalidad de reseccarlas	Disminución de la metástasis con la finalidad de reseccarlas	Disminución del tumor primario con la intención de reseccarlo	Disminución del tumor primario con la intención de reseccarlo	Predicción de respuesta completa frente a tratamiento preoperatorio
Metodología	Microarrays Affymetrix	qRT-PCR & microarrays	Home-made arrays	Microarrays Affymetrix	Microarrays Affymetrix
Muestra de trabajo	Tumor primario	Tumor primario	Biopsia del tumor primario	Biopsia del tumor primario	Biopsia del tumor primario
Predictor	14 genes	3 genes	54 genes	33 genes	95 genes
Precisión	74.4%	86.3%	83.0%	82.4%	84.0%

Tabla.9. Cuadro resumen de los trabajos que analizan mediante perfiles la respuesta de los pacientes a determinados tratamientos [266-270].

Hipótesis y objetivos

El cáncer colorrectal es una de las neoplasias con mayor incidencia y mayor tasa de mortalidad en los países desarrollados. En la actualidad los sistemas de clasificación, de pronóstico y de tratamiento de estos pacientes están única y exclusivamente basados en criterios clínicos y anatomopatológicos que no son capaces de reflejar toda la heterogeneidad de una enfermedad tan compleja. Este hecho provoca que sujetos clasificados en un mismo estadio, evolucionen en la enfermedad o respondan a un mismo tratamiento de manera diferente.

La secuenciación del genoma humano y la aparición de técnicas de análisis masivo como los microarrays han permitido abordar de manera global y efectiva la complejidad molecular de esta enfermedad, aportando, en algunos casos, como en el cáncer de mama, herramientas efectivas para el manejo clínico de los pacientes.

Teniendo en cuenta estos antecedentes, nos planteamos la realización del presente estudio cuyos objetivos se detallan a continuación:

Objetivo principal:

- ❖ Estudio del patrón de expresión de los tumores colorrectales con el fin de realizar una clasificación de los pacientes en grupos molecularmente homogéneos y de establecer la correlación existente con la actual clasificación clínica y otros parámetros de posible interés para el tratamiento de los pacientes.
 - Aplicación de técnicas no supervisadas para la agrupación de muestras (descubrimiento de clases) en función de los perfiles de expresión y establecimiento de una clasificación molecular.
 - Caracterización de los patrones de expresión diferencial para cada uno de los grupos tumorales propuestos con la finalidad de identificar nuevos marcadores moleculares capaces de predecir el comportamiento del paciente.
 - Correlación de los subtipos moleculares definidos con los parámetros clínicos y patológicos estudiados, en especial con la clasificación histológica (Dukes).
 - Validación de la clasificación mediante el uso de grupos de muestras independientes.

- Realización de análisis de supervivencia para el estudio de una posible asociación entre el pronóstico de los pacientes y los subtipos de la clasificación molecular propuesta.

Objetivos secundarios:

- ❖ Caracterización del perfil de expresión diferencial que identifica el fenotipo malignizado (tejido normal vs. tejido tumoral), con la intención de determinar marcadores y funciones biológicas alteradas que permitan profundizar en el conocimiento de las reglas que regulan el cáncer y que pueden ser de interés para su futura aplicación en la práctica clínica .
- ❖ Implantación de la técnica de los microarrays en el Laboratorio de Oncología Molecular (Dpto. de Oncología Médica) del Hospital Clínico San Carlos, con el fin de posibilitar la realización de nuevos proyectos en el área de genómica:
 - Extracción de ácidos nucleicos de muestras biológicas (tejido congelado) y establecimiento de controles de calidad.
 - Marcaje e hibridación de la muestras. Establecimiento de controles de calidad.
 - Implantación y manejo de “softwares” de análisis de datos genómicos.

Materiales y Métodos

1. Diseño experimental .

El diseño experimental en este tipo de estudio es esencial para la obtención de resultados fiables y consistentes. En este trabajo se establecieron cuatro pilares básicos sobre los que desarrollar el proyecto:

- ❖ Selección de muestras: puesto que el objetivo del proyecto era realizar una clasificación molecular del cáncer colorrectal se intentó que el conjunto de muestras analizadas fuese representativo de los diferentes estadios tumorales, para lo que se seleccionaron muestras en los cuatro estadios de Dukes (A, B, C y D) en proporciones similares.

- ❖ Análisis del tejido tumoral. La importancia de la interacción epitelio-estroma en el desarrollo del tumor está cobrando cada vez mayor protagonismo en el desarrollo del cáncer, por lo que se decidió no utilizar técnicas como la microdissección con captura por láser para seleccionar solamente células epiteliales, si no analizar la expresión conjunta de las células del estroma y del epitelio, teniendo cuidado de que la proporción entre ambas fuese un fiel reflejo de la composición del tumor.

- ❖ Selección de la plataforma a utilizar: se utilizó la plataforma de microarrays de Agilent Technologies. Sus matrices de expresión están formadas por oligonucleótidos largos de secuencia específica (60pb) para cada gen sintetizados “in situ”, lo que aumenta la fuerza y especificidad de unión con los transcritos hibridados. Este sistema posee además un rango dinámico muy amplio y es capaz de detectar de manera fiable y con un alta reproducibilidad, diferencias mínimas de expresión de hasta 1,3 veces y niveles bajos de mRNA (1 transcrito/ célula / millón de células).

- ❖ Hibridación:
 - *Modelo de hibridación utilizado*. Se utilizó una técnicas de hibridación doble frente a una muestra de referencia común, facilitándose y homogenizándose los procesos de normalización de las muestras.

- *Selección de la muestra de referencia:* se usó un “pool” de tejidos normales, obtenidos de los propios pacientes incluidos en el estudio, como muestra de referencia. Aunque esta muestra de referencia se usó casi exclusivamente para los procesos de normalización, permitió desde un principio establecer el significado biológico de los datos de expresión que son un reflejo de las variaciones que se producen en el proceso de transformación neoplásica.

2. Selección y estudio de los pacientes.

Se seleccionaron como sujetos de estudio, pacientes operados en el Hospital Clínico San Carlos de adenocarcinomas colónicos y rectales primarios localizados (estadios I, II o III) y avanzados (estadio IV) entre los años 2001-2006. Pacientes tratados con quimioterapia neoadyuvante fueron rechazados para este estudio al considerarse que dicho tratamiento podría alterar el comportamiento natural de tumor.

Datos clínicos como la edad, el sexo, la localización y el estadiaje del tumor (T, N, M y Dukes), el tipo histológico y los niveles de Antígeno Carcinoembrionario (CEA), así como el seguimiento y el tratamiento del paciente fueron recogidos del historial clínico (Tabla.1). En relación con el tipo histológico, hay que señalar que no se tuvo en consideración el subtipo de los adenocarcinomas serrados, debido a la dificultad que entrañó su identificación.

Del análisis histológico del tumor también se recabó información en relación a una serie de parámetros histopatológicos de posible relevancia en el pronóstico de la enfermedad como el grado de diferenciación celular, la invasión perineural, la invasión vascular, etc. (Tabla.1).

Las muestras incluidas en el estudio fueron obtenidas a través del Banco de Tumores del Hospital Clínico San Carlos, el cual se rige por las normas establecidas desde el propio hospital y aprobadas por el comité de ética, entre las que se incluye la obtención del correspondiente consentimiento informado firmado por los pacientes.

Parámetro	Escala	Parámetro	Escala
· <u>Sexo</u>	Varon / Mujer	· <u>Margen Tumoral</u>	Infiltrante
· <u>Edad</u>	Años		Expansivo
· <u>Extensión</u>	T1-T4		Mixto
· <u>Nódulos</u>	N0-N1	· <u>Diferenciación</u>	Indiferenciado
· <u>Metástasis</u>	M0-M1		Poco
· <u>Estadio de Dukes</u>	Estadio A		Moderadamente
	Estadio B	· <u>Tipo microscópico</u>	Bien
	Estadio C		Convencional
	Estadio D		Mucinoso
· <u>Localización</u>	C. Ascendente	· <u>Mutaciones en k-ras</u>	Mutado (codon 12 o 13)
	C. Transverso		Silvestre
	C. Descendente	· <u>Inestabilidad</u>	MSS
	Sigma		MSI
	Recto	· <u>Tratamiento</u>	Regímenes de QR y RTx
· <u>CEA</u>	Unidades	· <u>Seguimiento</u>	SG y PLE
· <u>Inv. vascular</u>	Si / No	· <u>Proliferación (Ki67)</u>	Porcentaje
· <u>Inv. Perineural</u>	Si / No	· <u>Apoptosis (M30)</u>	Porcentaje
· <u>Ganglios analiz</u>	Menos de 12	· <u>β-catenina nuclear</u>	Ausente (< 1)
	12 o más		Medio (1= x < 2)
· <u>Infiltrado linfocitario</u>	Ausente (0)		Alto (= 2)
	Bajo (1)	· <u>EGFR (Polisomía)</u>	Si / No
	Medio (2)		
	Alto (3)		

Tabla.1. Parámetros clínicos, patológicos y moleculares analizados durante este estudio.

3. Obtención y estudio histológico de los tumores.

De cada paciente se obtuvo una muestra de tejido tumoral y otra de mucosa de colon normal (extraída a más de 10cm. del tumor). Todas las muestras utilizadas fueron incluidas en un medio criosolidificable (OCT compound Tissue-Tek. Sakura. Tokio, Japón) necesario para su procesamiento posterior y congeladas con isopentano (-160°C)

en un plazo inferior a 30 minutos desde su extracción en el quirófano. Hasta su uso se almacenaron en un congelador a una temperatura de -80°C.

Antes de llevar a cabo la extracción de ácidos nucleicos (RNA y DNA) se realizó un análisis histológico para valorar el porcentaje de componente tumoral existente en los tejidos. De cada pieza tumoral congelada se obtuvieron dos cortes de 4µm de grosor (Criostato Leica CM1850. Wetzlar, Germany) que se tiñeron con hematoxilina-eosina. El primer corte corresponde al tejido justamente anterior al material seleccionado para extraer los ácidos nucleicos y el segundo corte corresponde al tejido inmediatamente posterior a dicho material seleccionado para la extracción.

Con la finalidad de evitar una contaminación elevada con tejido normal, solo se seleccionaron muestras que en el análisis histológico contuvieron una media de al menos un 80% de componente tumoral (células epiteliales tumorales y estroma tumoral) en el análisis histológico (Fig.1). Durante dicho análisis se anotó también la proporción aproximada de estroma y de necrosis de cada muestra tumoral.

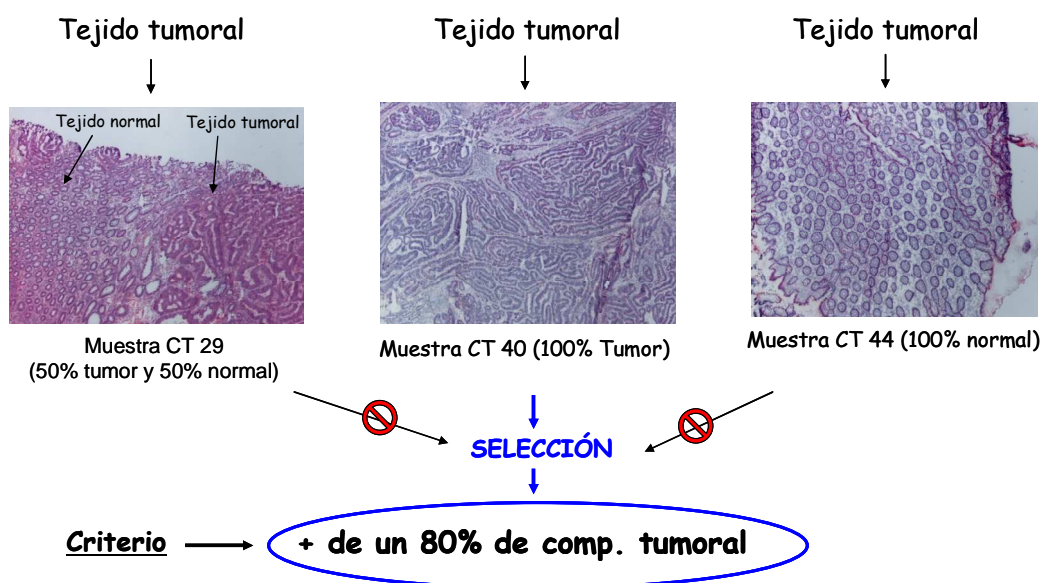


Fig.1. Estudio histológico para la selección de pacientes a estudio. El criterio estableció la necesidad de tener más de un 80% de componente tumoral en media (incluyendo epitelio tumoral y estroma tumoral) para que la pieza fuese incluida en el estudio.

En paralelo también se realizó un estudio histológico complementario de cada tumor a través de piezas independientes de tejido para confirmar que la composición

tisular del tejido congelado analizado era representativa del tumor en general. Como muestra independiente de tejido en cada uno de los tumores se utilizó tejido que había sido parafinado y almacenado para su uso en clínica.

4. Extracción de los ácidos nucleicos.

4.1. Extracción del RNA: Método del Trizol.

El Trizol® (Invitrogen, California, USA) es una solución monofásica de fenol e isotiocianato de guanidina que es capaz de mantener la integridad del RNA mientras lisa las células y disuelve los compuestos celulares permitiendo así la extracción de dicho ácido nucleico.

El protocolo de extracción fue el indicado por el fabricante con pequeñas modificaciones. Consta de los siguientes pasos:

- ❖ Homogenización de los cortes del tejido congelado (aprox. 20 cortes de 20µm cada uno) en frío usando Trizol® (1ml) y un homogenizador mecánico (“Ultra Turrax T8 (S8N-5G)”. Rose Scientific Ltd, Canadá).
- ❖ Separación del homogeneizado mediante cloroformo (200µl x ml de Trizol®) y centrifugación (13000rpm durante 10’) obteniéndose tres fases: acuosa (RNA), orgánica (restos celulares y DNA) y fenólica (restos de Trizol® y cloroformo).
- ❖ Precipitación del RNA disuelto en la fase acuosa mediante isopropanol (600µl x ml de Trizol®) durante toda la noche y centrifugación (13000rpm durante 30’), seguido de lavado con etanol al 75% (para eliminar impurezas y restos de sales).
- ❖ Resuspensión del RNA en H₂O libre de RNAsas (Quiagen, Hilden, Alemania).

Una parte del volumen de RNA obtenido de cada muestra fue sometido posteriormente a tratamiento con DNAsa y purificación en columnas (“RNeasy Micro Kit”. Qiagen, Hilden, Alemania) para eliminar cualquier rastro de contaminación con

DNA genómico. En este proceso se siguieron fielmente los protocolos indicados por el fabricante que aparecen resumidos a continuación:

- ❖ Unión del RNA (45µg) a la matriz de la columna. Incorporación de 45µg de RNA + 350µl Buffer RLT-(β-ME) + 250µl EtOH abs. al interior de la columna y posterior centrifugación (10.000rpm durante 15’’).
- ❖ Incubación de la membrana de la columna con DNAsa (30 unidades) + Buffer RDD (70µl) durante 15’.
- ❖ Elución del RNA con H₂O libre de DNAsas, tras lavados con dos Bufferes (RW1 y RLT) y con EtOH al 80%.

4.2. Extracción del DNA.

El material de partida para la extracción del DNA fueron los restos del tejido procedente de la purificación con Trizol. El método de extracción utilizado fue el siguiente:

- ❖ Eliminación de los restos de la fase acuosa mediante la adición de EtOH Absoluto (600µl por ml de Trizol) y centrifugación (8000rpm durante 10’) y aspiración del sobrenadante.
- ❖ Lavado del precipitado primero con citrato de sodio con etanol al 10% (1ml) para eliminar los restos fenólicos y luego con una solución de NaCl 0.075 Molar y EDTA 0.024 Molar para eliminar los restos de citrato de sodio con etanol al 10%.
- ❖ Digestión de los restos celulares (membranas, proteínas, etc.) con SDS al 1.2%, y Proteínas K (0.6 mg/ml) en solución B e incubación a 48°C durante 48 horas.
- ❖ Precipitación (con NaCl 2.1 Molar _{final}) de los restos celulares digeridos y obtención del DNA en disolución mediante centrifugación (14.000rpm durante 15’).

- ❖ Precipitación del DNA con isopropanol (vol. /vol.) durante 30' y centrifugación (14.000rpm durante 30'). Lavado con EtOH al 75%. y resuspensión del pellet en H₂O destilada.

4.3. Cuantificación de los ácidos nucleicos:

La cantidad de DNA y RNA obtenido así como su pureza (ratio 260/280 para descartar la contaminación con proteínas y ratio 260/230 para descartar la existencia de restos de sales), se evaluaron mediante uso del espectrofotómetro Nanodrop ND-1000 (Thermo Fisher Scientific Inc. Wilmington, USA) en 1µl de muestra. La concentración de los ácidos nucleicos se obtuvo mediante la medición de la absorbancia y su aplicación a la fórmula:

$$C = (A \times E) / D$$

Donde “C” es la concentración (ng/µl); “A” es la absorbancia (λ260nm); “E” es el coeficiente de extinción molar (50 en el caso de DNA y 40 en el caso de RNA); “D” es la dilución de la muestra.

4.4. Control de calidad RNA:

La degradación de la molécula de RNA afecta a los niveles de expresión del RNA mensajero. La valoración de la calidad (o nivel de integridad) del RNA se llevó a cabo utilizando métodos microfluídicos (“Bioanalyzer 2100”. Agilent Technologies, Palo Alto, USA) (Fig.2) en combinación con el software “RNA Integrity Number (RIN)”. Este software integra, mediante algoritmos matemáticos, diversos parámetros del perfil del RNA (Fig.2) y proporciona un valor (en una escala del 0-10, en donde el 0 indica degradación completa) que refleja la calidad de la muestra [272,273].

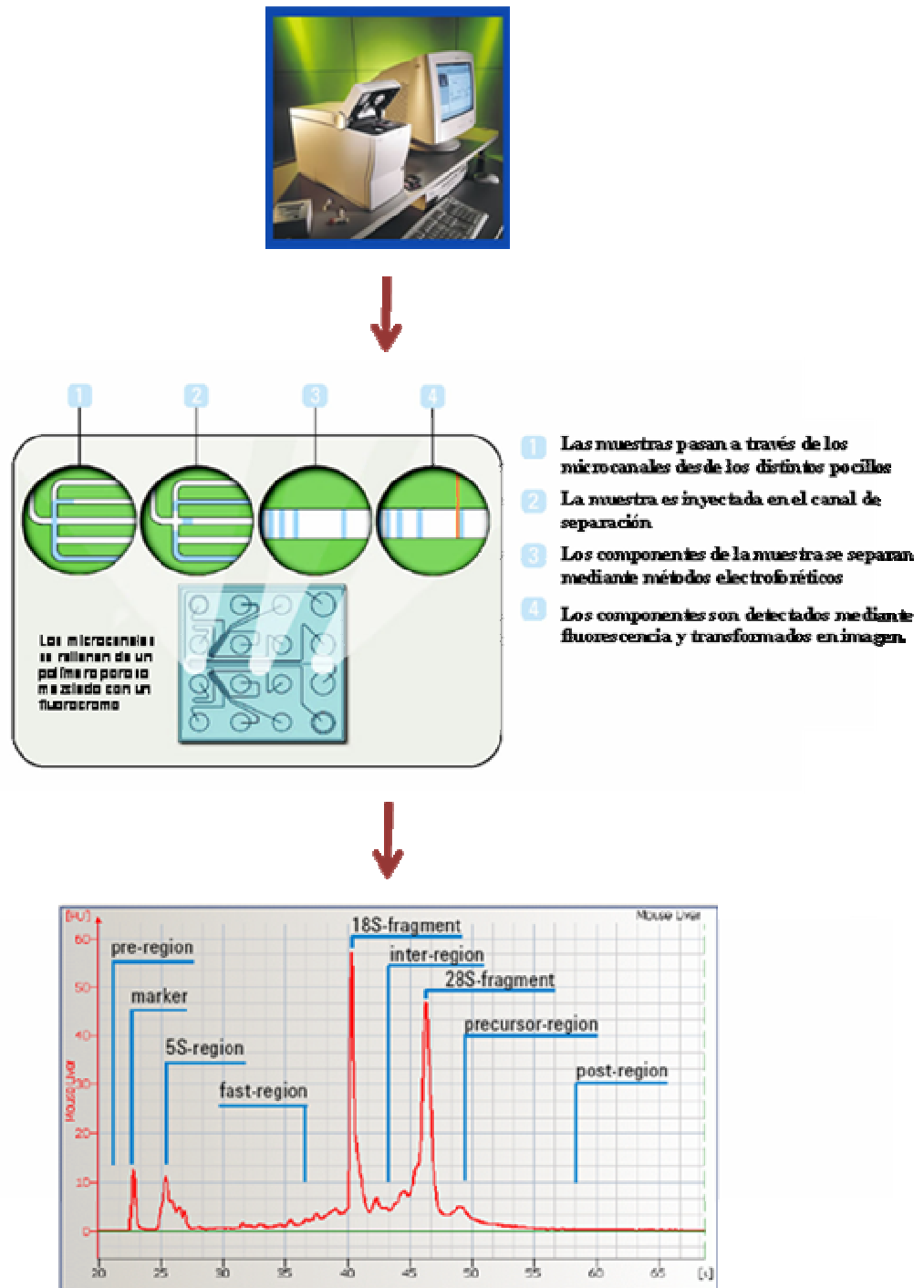


Fig.2. Esquema del proceso de microelectroforesis que realiza el “Bioanalyzer 2100” junto con el electroferograma resultante para una de las muestras. En dicho electroferograma se indican las regiones que son utilizadas para calcular el valor del RIN.

4.5. Control de calidad del DNA:

La integridad o calidad del DNA se analizó mediante técnicas de electroforesis convencional en geles de Agarosa al 0,8% -TAE (0.5X) con Bromuro de Etidio (0.2 mg/ml).

5. Obtención de la muestra de referencia (“pool” de colon normal).

Al tratarse de una hibridación doble, todas las muestras a estudio (Tumores y tejidos normales individuales), se van a hibridar frente a una única muestra común denominada “Referencia”. En este caso, como muestra de referencia (o “Pool” CN) se utilizó un “pool” de RNA obtenido a partir de 68 muestras de colon normal situadas a más de 10cm del tumor, todas ellas pertenecientes a pacientes incluidos en el estudio. En su fabricación, se descartaron aquellos tejidos que poseían alguna anomalía que pudiese afectar la expresión de los genes en la muestra normal (Ej.: colitis ulcerosa), aquellos con un RIN < 7.5 o aquellos con una cantidad de muestra baja que no permitiese destinar una parte a la creación del “pool” y otra a la realización de estudios posteriores individualizados.

6. Marcaje e hibridación de las muestras a los microarrays de expresión.

Se utilizaron microarrays “Human Whole Genome Oligo Microarray” (Agilent Technologies, Palo Alto, USA) con dos formatos diferentes (1x44K o microarray tipo G4112A y 4x44K o microarray tipo G4112F), cuyas características se presentan en la tabla.2.

	Whole Human Oligo Microarrays	
Formato del array	4x44k	1x44k
Dimensiones del array	25mm x 75mm	25mm x 75mm
Total de puntos/array	45,015	43,931
Tamaño del punto	65 micrones Ø	115 micrones Ø
Longitud de la sonda	60 pb	60 pb
Total de puntos a estudio	43,376	41,675
Total de sondas a estudio	41,000	41,000
Controles Positivos	1486	1942
Controles Negativos	153	314

Tabla.2. Características de los dos formatos de microarrays usados en el estudio.

- ❖ **Puntos a estudio:** representan el conjunto de RNA mensajero que se transcribe en las células.
- ❖ **Controles negativos:** son oligonucleótidos que al contacto con el líquido de hibridación van a formar un “hairpin” consigo mismos evitando que se pueda unir ningún transcrito marcado. Su fluorescencia es considerada como inespecífica de la técnica.
- ❖ **Controles positivos:** son oligonucleótidos de producción (utilizado como control de la síntesis del microarray por la compañía encargada de su fabricación) y de marcaje de los extremos del microarray (utilizado para identificar la posición exacta del array en el cristal).

El proceso de marcaje e hibridación, que se describe brevemente a continuación, fue igual para los dos modelos de micromatrices.

6.1. Marcaje y amplificación de la muestra.

Ambos procesos se llevaron a cabo siguiendo el protocolo del kit “Agilent low RNA Input Fluorescent Linear Amplification Kit” (Agilent Technologies, Palo Alto, USA) basado en la técnica de Eberwine [274]. De forma resumida, el proceso es el siguiente (Fig.3):

- ❖ 500ng de RNA total fueron retro-transcritos usando como cebador un oligo poli dT con la secuencia del promotor de la T7 RNA Polimerasa en su extremo 5’ y como enzima retro-transcriptasa una MMLV-RT.
- ❖ A partir del cDNA de doble cadena se obtuvo cRNA utilizando la enzima T7-RNA Polimerasa. Durante este proceso se produjo la amplificación y la incorporación del marcaje a la muestra. Las muestras a estudio (tumores y tejidos normales) se marcaron utilizando CTP-Cy5 (color rojo), mientras que la muestra de referencia (“Pool” CN) se marcó utilizando CTP-Cy3 (color verde).

- ❖ El cRNA marcado obtenido fue purificado mediante columnas (“RNeasy Mini Kit”, Quiagen. Hilden, Alemania) siguiendo el protocolo establecido por el fabricante para eliminar restos de reactivos y de cianina no incorporada a las muestras.

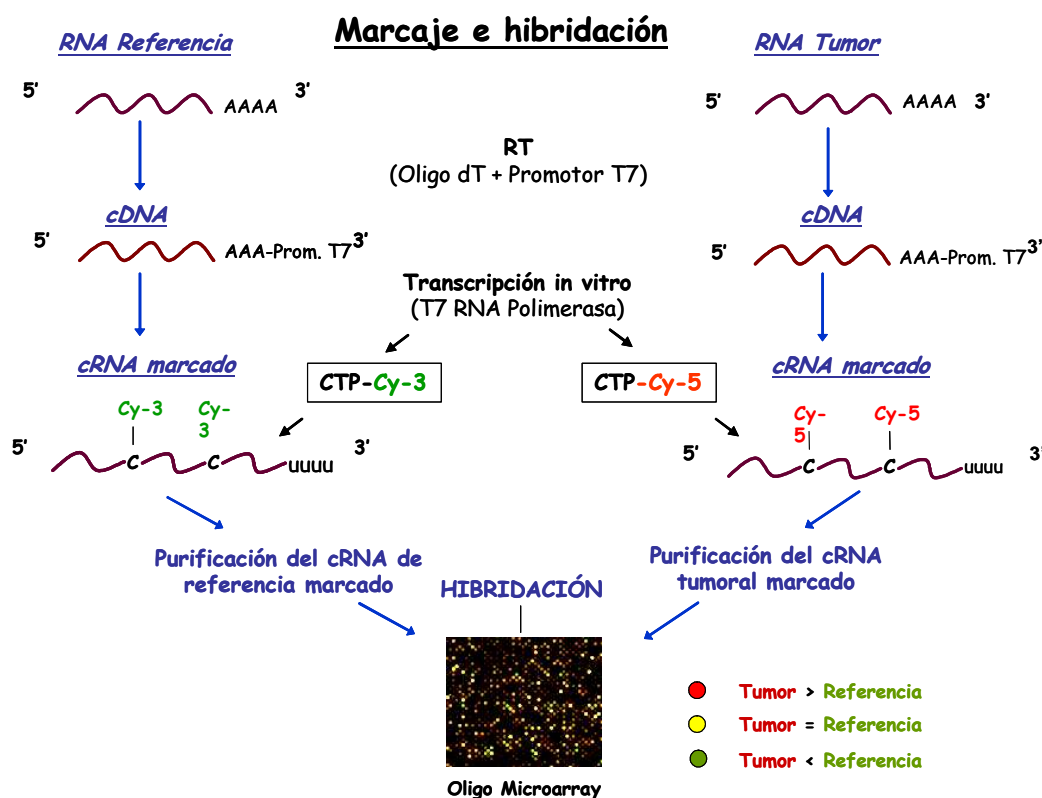


Fig.3. Resumen esquemático del proceso de amplificación, marcaje e hibridación de un tumor (proceso de hibridación competitiva frente a una muestra de referencia) a un microarray.

6.2. Control del marcaje y la amplificación de las muestras.

Para comprobar que todo el proceso se realizó correctamente se miden por duplicado tanto las concentraciones de cRNA marcado obtenido, así como la tasa de incorporación de cianina a la muestra (Fig.4). Todos estos controles se llevaron a cabo utilizando el espectrofotómetro “Nanodrop-ND1000” (Nanodrop Technologies. Delaware, EE.UU). Posteriormente, se comprobó que los tamaños de los fragmentos obtenidos (cRNA marcado) eran los esperados mediante su visualización con el “Bioanalyzer 2100” (Agilent Thecnologies. Palo Alto EE.UU) (Fig.4). Muestras con una cantidad de cRNA total menor de 0.95 μg o con una incorporación de cianina menor de 8 pmol / μg fueron descartadas.

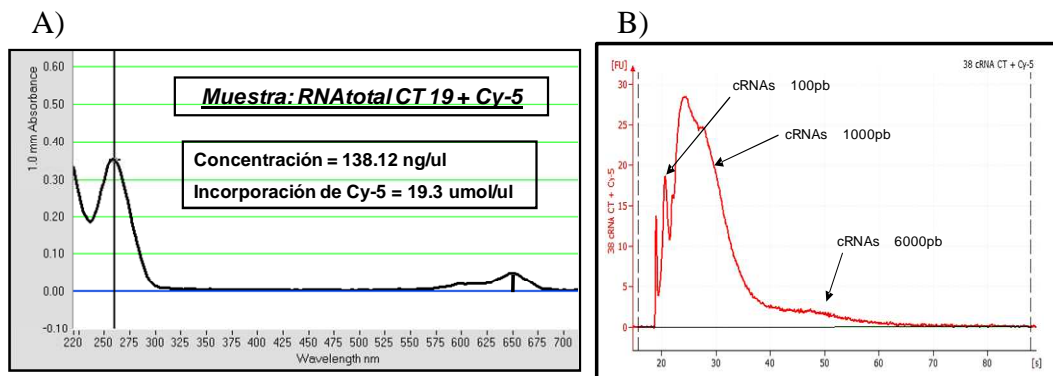


Fig.4. Resumen del control de marcaje y la amplificación de las muestras. A) medición de la concentración y de la incorporación de cianina a la muestra tras el proceso de amplificación y marcaje. B) Visualización mediante el “Bioanalyzer 2100” de la correcta síntesis de de los fragmentos de cRNA.

6.3. Hibridación y lavado del microarray.

Estos dos procesos constan brevemente de los siguientes pasos:

- ❖ El cRNA marcado es fragmentado en secuencias más pequeñas y es hibridado al microarray. El proceso de hibridación tiene lugar en un horno de hibridación (Agilent G2545A Hybridization Oven. Palo Alto, USA) durante un período de 17h (Tabla.3).
- ❖ Posteriormente el array es lavado para eliminar todo aquello que no se ha unido a los oligonucleótidos (Tabla 4).

Estos protocolos de hibridación y lavado de las muestras fueron variando según las recomendaciones de la casa comercial a lo largo del tiempo y dependiendo del formato del microarrays.

Diferencias en el Protocolo de Hibridación		
Parámetros	Formato 1x44k	Formato 4x44k
Cantd. inicial de cRNA-Cy5	750ng	825ng
Cantd. inicial de cRNA-Cy3	750ng	825ng
10 x Control Target (1x44k)	50µl	—
10 x Blocking Agent (4x44k)	—	11µl
25 x Fragmentation Buffer	10µl	2.2µl
H₂O destilada	hasta 250µl	hasta 55µl
Buffer de Hibridación	2 x Hibridization Buffer (55 µl)	GE x Hibridization Buffer HI-RPM (250µl)
Volumen final de Hibridación	490µl	100µl
Temperatura de hibridación	60°C	65°C
Tiempo de hibridación	17h	17h
Velocidad del rotor	4rpm	10rpm

Tabla.3. Diferencias principales entre los dos protocolos empleados en el proceso de hibridación de las muestras a los microarrays dependientes del formato de microarray usado.

Variaciones en el Protocolo de lavado			
Parámetros	Formato 1x44k (arrays 1-30)	Formato 1x44k (arrays 31-110)	Formato 1x 44k y 4x44k (arrays 111-116)
Buffer para el desensamblaje de la cámara de hibridación	6 x SSPE 0.005% N-Lauroylsarcosine	6 x SSC 0.005% Triton X-102	GE Expression Wash Buffer 1
Buffer de lavado 1	6 x SSPE 0.005% N-Lauroylsarcosine	6 x SSC 0.005% Triton X-102	GE Expression Wash Buffer 1
Tiempo de lavado 1	1 minuto	10 minutos	1 minuto
Temperatura de lavado 1	Temperatura Ambiente	Temperatura Ambiente	Temperatura Ambiente
Buffer de lavado 2	0.06 x SSPE 0.005% N-Lauroylsarcosine	0.1 x SSC 0.005% Triton X-102	GE Expression Wash Buffer 2
Tiempo de lavado 2	1 minuto	5 minutos	1 minuto
Temperatura de lavado 2	Temperatura Ambiente	4 °C	37 °C
Buffer de lavado 3	Stabilization and dry Solution	—	—
Tiempo de lavado 3	30 segundos	—	—
Temperatura de lavado 3	Temperatura Ambiente	—	—
Secado de la muestra	—	Nitrógeno comprimido	—

Tabla.4. Diferencias principales entre los protocolos de lavado de los microarrays, realizados a lo largo del estudio.

7. Escaneado de los microarrays.

El escaneado de los microarray se realizó utilizando el escáner confocal “Agilent Microarray Scanner” (modelo G2565BA) con autofocus y el programa “Agilent Scan Control” (Agilent Technologies, Palo Alto, USA). Este equipo es capaz de captar al mismo tiempo la intensidad de señal emitida por los dos fluorocromo (Cy-3 y Cy-5) incorporados a la muestra cuando son excitados mediante láser (SHG-TAG (532nm) y Helio-Neon (633nm) respectivamente), creando una imagen que refleja los niveles de expresión de cada punto para esa muestra (imagen de alta definición en formato TIFF). Los parámetros de escaneado variaron dependiendo del array y de la versión del software utilizado. Los cambios más significativos se describen en la Tabla.5.

Variaciones en el Protocolo de Escaneado		
Parámetros	SC v.6 (arrays 1-110)	SC v.7 (arrays 111-116)
Área de escaneado	61mm x 21mm	61mm x 21mm
"Extended Dinamic Range"	No	Si
PMT Canal Rojo	100%	100%-10%
PMT Canal Verde	100%	100%-10%
Resolución del escaner	10 µm	5 µm
Modo de escaneado	1 vez	2 veces

Tabla.5. Comparación de las dos versiones del programa “Scan Control” (SC) utilizadas en el estudio. El “Extended Dynamic Range” (XDR) permite definir dos medidas diferentes para la sensibilidad del “PhotoMultiplier Tube” (PMT). La combinación de mediciones a alta y baja sensibilidad (PMT 100% y 10%) evita la aparición de puntos saturados. La versión SC.6 no tenía habilitada esta opción.

8. Cuantificación de la imagen.

Una vez obtenida la imagen escaneada de los arrays (formato TIFF) se usó el programa “Feature Extraction” versiones 8.5 y 9.5 (Agilent Technologies, Palo Alto, EEUU) para la extracción de los datos contenidos en estas. Brevemente, el proceso de extracción fue el siguiente:

- ❖ Identificación y localización exacta, utilizando la plantilla de localización de genes (“grid”) diseñada por Agilent Technologies (Fig.5A).

- ❖ Extracción, mediante el algoritmo “Cookie Cutter”, de los valores de intensidad de fluorescencia y de ruido de fondo local para cada uno de los puntos del array (ver Fig.5B).
- ❖ Identificación de puntos “problemáticos” (Ej. aquellos que no poseen una medida de la intensidad uniforme, que están saturados o cuya relación señal/ruido era muy baja, etc.). Los parámetros utilizados para definir estas alteraciones técnicas fueron los siguientes: “IsSaturated”, “FeatureNonUnifOL”, “IsFeatPopOL”, “IsPosAndSig” y “IsWellAbove” (Fig. 5C. Para más información ver Anexo I. Pág. 168).

Al mismo tiempo, se dispuso de un estudio basado en los mismos criterios de selección para la intensidad de señal producida por el ruido de fondo de cada una de las sondas. Sin embargo este estudio no se tuvo en cuenta de manera directa ya que por un lado se consideró que eran menos indicativos del resultado de la hibridación y por otro lado por que se juzgó que ya habían sido tenidos en cuenta a la hora de calcular los parámetros “IsPosAndSig”, “IsWellAbove” y “P-valueLogRatio”.

- ❖ Además de los parámetros recogidos en el punto anterior, también se tuvo en cuenta otro parámetro:
 - “P-valueLogRatio” (p-valor de logRatio): es una medida de la confianza (vista como probabilidad) de que ese punto no esté expresado de manera diferentemente significativa en nuestro tejido a estudio con respecto a la muestra de referencia. Este cálculo esta influenciado por los posibles errores de medición que se hayan detectado para los dos canales en ese punto.
- ❖ Cálculo de la fluorescencia procesada del punto (Fig.5D). El cálculo de esta fluorescencia consta de tres pasos sucesivos en los que se elimina el efecto del ruido de fondo, se lleva a cabo una normalización de los valores de expresión y por último y sólo en el caso de sondas con señales muy bajas, se establece un valor de fluorescencia mínimo. Para el resto de casos el valor final de fluorescencia será el obtenido tras el proceso de normalización (más información en el Anexo I. Pág.169).

- ❖ Estos datos serán utilizados posteriormente para el análisis de los resultados (Fig.5D).

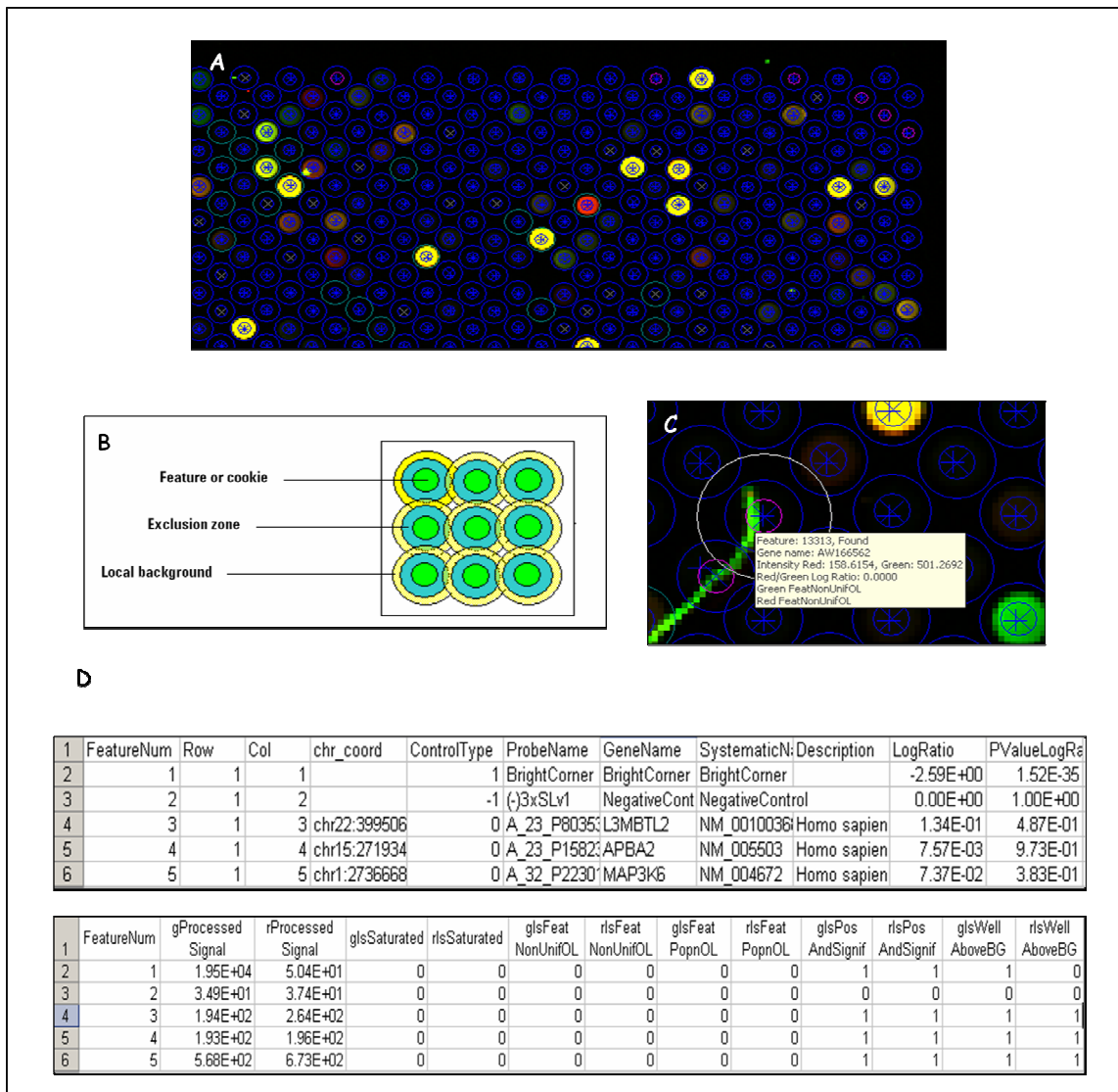


Fig.5. Resumen del proceso de cuantificación de la imagen. A) Identificación y localización exacta de las sondas en el microarray. B) Obtención de los valores de intensidad de fluorescencia de la muestra y del ruido de fondo para los dos canales mediante el algoritmo “Cookie Cutter”. C) Identificación de los puntos de baja calidad. D) Esquema de los datos más importantes generados por el programa “Feature Extraction 8.5 y 9.5”. En el archivo original el número de filas es aproximadamente 44.000. En la parte superior se muestra la localización de los puntos en el microarray, sus diferentes identificaciones, su valor de expresión definitivo para cada punto y el p-valor asociado a dicho punto. En la parte inferior se muestra la señal procesada para los dos canales en cada punto y los identificadores de puntos problemáticos.

9. Análisis de datos.

9.1. Control de calidad. “QC-Report”, “QC-Chart” y PCA.

En paralelo al proceso de extracción de datos, el programa “Feature Extraction” genera un archivo denominado “Quality Control Report” que se utiliza para el descarte de muestras que hubiesen sufrido fallos durante el proceso de hibridación y/o escaneado. Este tipo de archivo recoge de manera global los parámetros más importantes para cada microarray, permitiendo hacer una evaluación inicial del resultado del experimento. Para ello se realizó un análisis en conjunto de todos los parámetros y todos los arrays utilizando el programa “Quality Control Chart” (Agilent Technology. Palo Alto. EEUU) que nos proporcionó un criterio objetivo para el rechazo o la aceptación de cada uno de los microarrays. Estos parámetros fueron seleccionados en base a criterios que tenían en cuenta la intensidad de la fluorescencia de cada sonda y su relación con el ruido de fondo, la fluorescencia de los controles negativos, el número de puntos no uniformes y la relación entre puntos estimulados e inhibidos. Los límites, superior e inferior, aplicados a cada parámetro se especifican en la Tabla.6 y en la Fig.6. Las muestras debían encontrarse entre los límites establecidos en al menos 9 del total de los 12 parámetros analizados.

Las muestras se dividieron en dos grupos dependiendo del protocolo de hibridación y lavado. Las muestras escaneadas con el programa FE 8.5 (110 muestras) comparten el mismo sistema de hibridación y un sistema de lavado muy parecido para todas ellas. Las muestras escaneadas con el programa FE 9.5 (6 muestras) comparten entre si el mismo sistema de hibridación y lavado, pero este es diferente al del FE8.5. Los parámetros valorados en cada uno de los dos canales de manera independiente fueron los siguientes:

Parámetro	Microarrays escaneados con el FE 8.5				Microarrays escaneados con el FE 9.5			
	Limite Superior		Limite Inferior		Limite Superior		Limite Inferior	
	C. Rojo	C. verde	C. Rojo	C. verde	C. Rojo	C. verde	C. Rojo	C. verde
NegCtrlAveNetSig	70	70	—	—	70	70	—	—
NumFeatureNonUnifOL	150	150	—	—	150	150	—	—
NumPopnOL	Mediana + 2.6 DS (155.5)	Mediana + 2.6 DS (147.2)	—	—	Mediana + 2.6 DS (180.4)	Mediana + 2.6 DS (168.15)	—	—
NonCtrlMedPrntCVBGSubSig	25	25	0	0	25	25	0	0
NumFoundFeat	—		Mediana + 2.6 DS (39723)		—		Mediana + 2.6 DS (36067)	
NumPopnBGOL	Mediana + 2.6 DS (2172.8)	Mediana + 2.6 DS (1555.4)	—	—	Mediana + 2.6 DS (1565.6)	Mediana + 2.6 DS (3144.1)	—	—
Prop_down_vs_up_regulated	1.5		0.666		1.5		0.666	

Tabla.6. Esquema representativos de los parámetros investigados en el proceso del control de calidad de la hibridación de las muestras. Los parámetros “NegCtrlAveNetSig”, “NumFeatureNonUnifOL”, “NumPopnOL”, “NonCtrlMedPrntCVBGSubSig” y “NumPopnBGOL” fueron examinados de manera independiente para el canal rojo y para el canal verde, contabilizando por tanto un total de 10 de los 12 parámetros estudiados. Los parámetros “Num Found Feat” y “Prop_down_vs_up_regulated” fueron examinados en conjunto para los dos canales, completando los dos parámetros que faltaban para el total de 12 seleccionados.

“NegCtrlAveNetSig”: Media de la señal pura para los controles negativos.

“NumFeatureNonUnifOL”: Número de puntos no uniformes.

“NumPopnOL”: Número de puntos cuya expresión en relación con sus replicados están fuera del rango esperado.

“NonCtrlMedPrntCVBGSubSig”: Mediana del Coeficiente de Variación del “BGSubSig” para los puntos (no controles).

“NumFoundFeat”: Número de puntos descritos como presentes.

“NumPopnBGOL”: Número de puntos cuya expresión del ruido de fondo en relación con sus replicados están fuera del rango esperado.

“Prop down vs up regulated”: Cociente entre el número de punto estadísticamente inhibidos y estimulados.

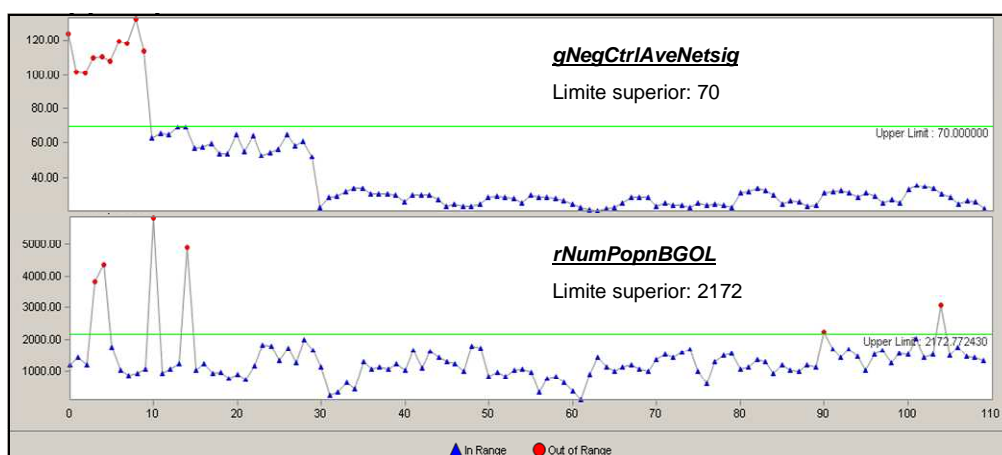


Fig.6. Visualización de dos de los parámetros analizados durante el proceso de control de calidad de la hibridación de las muestras.

Por último, para confirmar la validez del resto de muestras se realizó un Análisis de Componentes Principales o PCA (para más información sobre esta técnica ver Anexo1. Pág.181) que nos permitiese detectar un comportamiento anómalo de alguna de las muestras debido a sesgos que hubiesen pasado inadvertidos durante los controles previos (tipo de array, día de hibridación, componente tumoral de las muestras normales, etc.). Para este análisis se utilizó el programa “GeneSpring 7.3.” (Agilent Technologies. Palo Alto, USA). Las sondas utilizadas fueron las que pasaron el filtrado de sondas problemáticas expuesto a continuación en el apartado “filtrado de datos”.

9.2. Filtrado de datos.

En este caso se procedió al filtrado de sondas con problemas de hibridación o sin interés para el estudio en función de los criterios siguientes (Fig.7):

- ❖ Se excluyeron las sondas marcadas como “problemáticas” en más de un 10% de las muestras. Los criterios usados para describir una sonda como “problemática” son los referidos en la sección 8. en el apartado “identificación de puntos problemáticos”. Todas las sondas marcadas como aptas superaron estos controles de manera independiente para los dos canales de fluorescencia (Cy-5 y Cy-3).

Una vez eliminadas estas sondas, en las restantes se estableció una fluorescencia mínima de 10 cuentas en los dos canales (valor por defecto establecido por el programa “GeneSpring 7.3.”) para evitar tanto valores negativos de fluorescencia procesada como valores muy bajos que luego provoquen diferencias de expresión irreales.

- ❖ Se excluyeron aquellas sondas en las que en menos de 7 muestras no mostrasen un diferencia de expresión significativa (a un “P-valueLogRatio” < 0.01) entre la expresión de la muestra a estudio (Cy-5) y la expresión de la muestra de referencia (Cy-3). El hecho de seleccionar 7 muestras como criterio de filtrado fue debido a que era el número menor de muestras del que se podía esperar que formasen un grupo a nivel molecular. Revisando las características de los pacientes observamos que había un total de 7 muestras de tejido normal, 9 pacientes con inestabilidad en microsatélites, 10 pacientes con tumores de tipo histológico mucinoso y el grupo

más pequeño de estadios de Dukes estaba formado por 19 muestras (Dukes C y Dukes D). Por otro lado, se pensó que un grupo de menos de 7 muestras podría carecer de la estabilidad necesaria para ser reproducible, por lo que finalmente se tomó como referencia el número de tejidos normales utilizados.

Para estos filtrados anteriores se utilizó el programa “GeneSpring 7.3”.

La matriz final, en formato Excel, se importó al programa de tratamiento de datos BRB Array Tools 3.7 (<http://linus.nci.nih.gov/BRB-ArrayTools.html>) con el que se realizó el resto del análisis.

Durante el proceso de importación de los datos al programa “BRB Array Tools 3.7”. Se transformaron los datos numéricos (ratios) a datos en Log_2 , se promediaron aquellas sondas que representaban el mismo gen (el identificador usado fue el descrito en la columna Systematic Name) y se realizaron las anotaciones de genes correspondientes (términos GO, vías de señalización, etc.). Desafortunadamente, el objetivo de trabajar con genes únicos no se consiguió por completo, debido a que en este tipo de anotación hay genes que poseen diferentes identificadores. Finalmente, La matriz obtenida fue de 96 muestras y 14764 genes.

Para el establecimiento de los diferentes patrones de expresión se decidió trabajar con genes que mostrasen cambios claros en los niveles de expresión a lo largo de las muestras, de manera que fuesen fácilmente identificables. En base a esto, se llevo a cabo un último filtrado, por el cuál se seleccionaban aquellos genes que mostraban un cambio de al menos 2.5 veces con respecto a la mediana de expresión de ese gen, en al menos 7 de las muestras analizadas. Esto redujo el número de genes analizados a 1722.

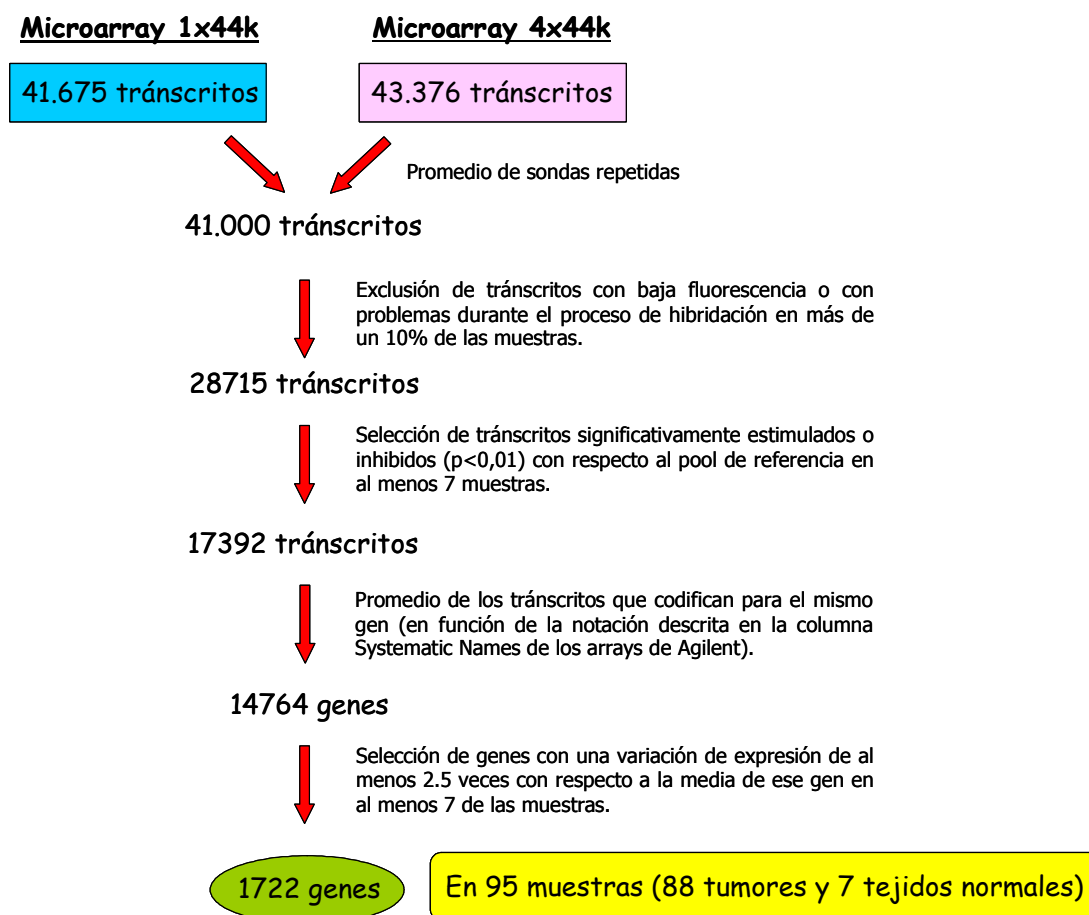


Fig.7. Proceso de filtrado de genes para conseguir la matriz final con la que llevar a cabo la clasificación molecular. Antes de llevar a cabo este proceso de filtrado, ya se habían eliminado los transcritos que codificaban para los controles internos del microarray.

9.3. Clasificación de los pacientes mediante métodos no supervisados.

Para la clasificación de las muestras en grupos homogéneos se utilizó la técnica de “clustering” jerárquico o HC (para más información, ver Anexo1. Pág.176). Para este propósito se seleccionó una matriz de genes más reducida, basándonos en el filtro en función de los niveles de expresión que acabamos de relatar en el apartado anterior.

En primer lugar se llevo a cabo una clasificación del conjunto total de las muestras (normales y tumorales). En este caso se usó como medida de la distancia, la correlación centrada de Pearson y como método de agrupación de elementos el “average linkage”.

En el caso de la clasificación tumoral, se descartaron las muestras de tejido colorrectal normal y se ejecutó el algoritmo de agrupación usando como distancia la correlación centrada de Pearson, como método de agrupación de elementos el “average linkage”, además de realizar un centrado de genes (para más información ver Anexo1. Pág.175).

En ambas clasificaciones, el número de agrupaciones seleccionadas mediante criterio visual del árbol, fue validado en primera instancia mediante técnicas matemáticas que tienen en cuenta dos parámetros [275]: la reproducibilidad del número y tamaño de los grupos encontrados (“R-Index”) y la reproducibilidad en la composición de los grupos (“D-Index”, que incluye omisiones y adiciones de muestras en los grupos) cuando se alteran aleatoriamente los valores de expresión de la matriz de genes en base a la variabilidad generada por las propias muestras (1000 perturbaciones) (Anexo1, Pág.178 y (<http://linus.nci.nih.gov/BRB-ArrayTools.html>)).

En el caso de la clasificación tumoral, y con el fin de confirmar la reproducibilidad de las agrupaciones obtenidas, se realizaron variaciones sobre los parámetros iniciales seleccionados para la realización de la clasificación y que afectaron a:

- ❖ Nº de genes: se variaron los criterios de filtrado de genes en base a su expresión para obtener matrices con distintos números de genes.
- ❖ Técnica de análisis utilizada: “clustering” jerárquico, “k-means”, “k-medians”, Análisis de Componentes Principales.
- ❖ Método de agrupación de elementos: “average linkage”, “complete linkage”, “single linkage”.
- ❖ Tipo de distancia seleccionado: correlación no centrada de Pearson, correlación de Spearman, kendal Tau, Euclídea.

Finalmente, se compararon los nuevos grupos obtenidos con las agrupaciones iniciales. Algunos de estas técnicas cuyos algoritmos no estaban presentes en el

programa BRB Array Tools fueron llevadas a cabo mediante la aplicación directa que posee este programa al programa de análisis de microarrays de Eisen [276] (<http://rana.lbl.gov/EisenSoftware.htm>).

9.4. Comparación de clases y análisis funcional.

Con la intención de caracterizar las diferencias en los perfiles de expresión obtenidos para el “clustering” jerárquico de muestras normales frente a tumorales, así como para la clasificación de los distintos subtipos tumorales se llevaron a cabo técnicas de comparación de clase y análisis funcional mediante el estudio de las KEGG (Kyoto Encyclopedia of Genes and Genomes) Pathways (www.genome.jp/kegg/pathway.html) [277].

Para la comparación de clases entre tejido normal y tumoral y para la comparación de clases entre grupos tumorales se calcularon dos parámetros:

- ❖ p-valor de cada gen: para cada gen se analiza, utilizando un t-test con un modelo de varianza aleatoria y el método de correcciones múltiples de Benjamini y Hochberg, si existen diferencias estadísticamente significativas a un nivel de significancia $\leq 0,001$.

- ❖ p-valor global: Tras obtener los genes que muestran una diferencia de expresión estadísticamente significativa y mediante el uso de 10000 permutaciones de los identificadores de las muestras, se analiza si la clasificación propuesta para el análisis es significativamente diferente a los perfiles obtenidos al azar.

Estas comparaciones de clase, se llevaron a cabo partiendo de los 1722 genes seleccionados inicialmente. Para la comparación entre normales frente a tumorales, se estableció además un “fold-change” igual o superior a dos veces. También cabe resaltar que en el caso de la comparación de los grupos tumorales, las comparaciones se realizaron dos a dos, de manera que cada uno de los grupos propuestos se comparó con el resto de muestras tumorales analizadas.

Para la realización del análisis funcional mediante el estudio de conjuntos de genes agrupados en función de la vía molecular a la que pertenecen (KEGG Pathways) e incluidos en la base de datos del Cancer Genome Anatomic Project (<http://cgap.nci.nih.gov/Pathways>) se utilizaron cuatro pruebas estadísticas diferentes denominadas “LS test”, “KS test”, “Goeman’s test” y “Gen Set Enrichment Análisis (GSEA)” (Para más información sobre estos test ver Anexo I. Pág.184).

En los análisis funcionales la lista completa de genes fue de 14764 genes y se estudiaron un total de 164 vías moleculares. Para los cuatro test se estableció un nivel de significancia de $p \leq 0.005$ (establecido por defecto en el software BRB Array Tools). Se decidió seleccionar como significativo cualquier conjunto de genes con un nivel de significancia $p \leq 0.005$ en dos de los cuatro test. Este criterio se debió al hecho de que el análisis que lleva a cabo el test de Goeman es diferente al de los otros tres. Mientras que el primero analiza la hipótesis de si existen o no genes diferencialmente expresados dentro del conjunto de genes (N) analizados, los otros tres métodos analizan la hipótesis de si el grado medio de expresión diferencial de ese conjunto de genes es mayor que el esperado si se seleccionase un conjunto de genes al azar.

10. Estudio de inestabilidad en secuencias microsatélites.

Se caracterizó el perfil de inestabilidad tumoral en todos los pacientes incluidos finalmente en el análisis de expresión génica global mediante microarrays. Debido a que en la literatura no se ha descrito un patrón de evolución tumoral diferente de los pacientes con perfil de baja estabilidad (MSI-L) con respecto al de aquellos con estabilidad en microsatélites (MSS) [278-280], en este trabajo solo se distinguió entre pacientes con alta inestabilidad o MSI-H (que pasaremos a denominar a lo largo del trabajo como MSI) y pacientes con baja inestabilidad o estables (que denominaremos a lo largo del trabajo como MSS).

A pesar de lo establecido durante el National Cancer Institute Workshop on MSI for Cancer Detection and Familial Predisposition [281] en el que se consensuó el estudio de cinco secuencias microsatélites (**Bat25**, **Bat26**, D2S123, D5S346 y D17S250) para poder definir los patrones de inestabilidad, hay estudios que afirman que el uso del microsatélite BAT26 por sí sólo bastaría para establecer esta diferencia. [282,283]. Teniendo en cuenta estos antecedentes y habiendo definido los MSI-H como

aqueellos con ganancias o pérdidas de estas repeticiones en al menos 2 microsatélites, se analizaron los 5 marcadores (HNPCC microsatellite instability kit. Roche) en un grupo de muestras iniciales. Una vez confirmado que Bat26 aparecía como un predictor fiable de las muestras MSI-H, se decidió estudiar solamente tres (Bat25, Bat26 y S5S346) de las 5 secuencias del panel de Bethesda y en el caso de encontrar inestabilidad en una sola de las tres secuencias, someter a dicha muestra al análisis de las otras dos secuencias restantes del panel inicial de 5 (Tabla.7). Pacientes con inestabilidad en menos de dos microsatélites fueron considerados como MSS.

Resumen del análisis de inestabilidad en secuencias microsatélites		
Método	"HNPCC microsatellite instability kit" (Roche. Nº Cat:2041901)	Home-made PCR
Nº muestras analizadas	12	80
Cebadores utilizados	"HNPCC microsatellite instability kit" (Roche. Nº Cat:2041901)	Bat 26-F: FAM-TGACTACTTTTGACTTCAGCC
		Bat 26-R: AACCATTCAACATTTTAAACCC
		Bat 25-F: HEX-TCGCCTCCAAGAATGTAAGT
		Bat-26-R: TCTGCATTTTAACTATGGCTC
		S5S346-F: TET-ACTCACTCTAGTGATAAATCGGG
		S5S346-R: AGCAGATAAGACAAGTATTACTAGTT
Protocolo de amplificación	"HNPCC microsatellite instability kit" (Roche. Nº Cat:2041901)	<p>94°C--30"</p> <p>95°C--3" ? 56°C--30" ? 72°C--5'</p> <p>72°C--30"</p> <p>1 paso ? 25 ciclos ? 1 paso</p>
Análisis del amplicón	Electroforesis capilar (ABI Prism 3.10. Applied Biosystems)	Electroforesis capilar (ABI Prism 3.10. Applied Biosystems)

Tabla.7. Esquema de los protocolos utilizados para el análisis de los perfiles de inestabilidad en los sujetos a estudio.

El análisis se llevó a cabo comparando para cada paciente, el perfil de microsatélites obtenido con el DNA tumoral frente al obtenido con el DNA de la mucosa del colon normal de ese mismo paciente. El método de análisis usado se basó en la amplificación de las secuencias microsatélites mediante la técnica de PCR ("Polimerase Chain Reaction") utilizando cebadores marcados. Los fragmentos amplificados fueron sometidos posteriormente a una electroforesis capilar mediante polímero (POP4) a 60°C en el secuenciador ABI Prism 310 (Applied Biosystems) utilizando el programa ABI Prism Collection. Los resultados se analizaron utilizando el programa "Gene Scan 3.1" (Applied Biosystem, California. EEUU) (Fig.8). Esta técnica nos permitió distinguir diferencias en el tamaño de los amplicones de hasta 1 base.

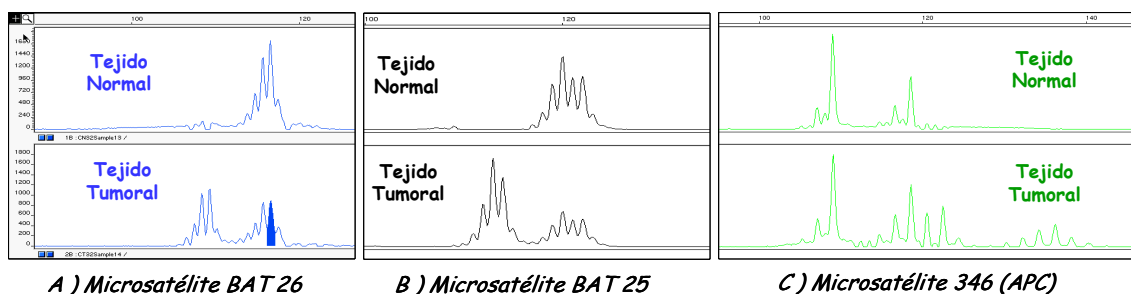


Fig.8. Resultado del análisis de inestabilidad a microsatélites en la muestra CT32.

11. Estudio de mutaciones en el gen KRAS.

El análisis de las mutaciones del codon 12 y 13 del gen KRAS se realizó siguiendo la técnica descrita por Hatzaki A et al [284] con pequeñas modificaciones y que de manera general conlleva la amplificación (PCR) del fragmento de DNA tumoral que contiene la secuencia, una incubación posterior del fragmento amplificado con enzimas de restricción y el análisis de los fragmentos mediante técnicas de electroforesis capilar.

Se utilizaron los oligonucleótidos mutados diseñados por Hatzaki (primer “sense” en el codon 12 y primer “antisense” en el codon 13), se diseñó la pareja de primer correspondiente para cada codon y un primer de cada pareja fue marcado con FAM-6 (Fig.10-12). Los primers mutados poseían una base alterada (missmatch) con respecto a la secuencia original, creando así un sitio de restricción artificial que nos permitió diferenciar las secuencias mutadas de las no mutadas. La electroforesis capilar se realizó mediante polímero (POP4) usando el secuenciador ABI Prism 3.10 y el programa de análisis “Gene Scan 3.10” (Applied Biosystems, California. EEUU) (Fig.9).

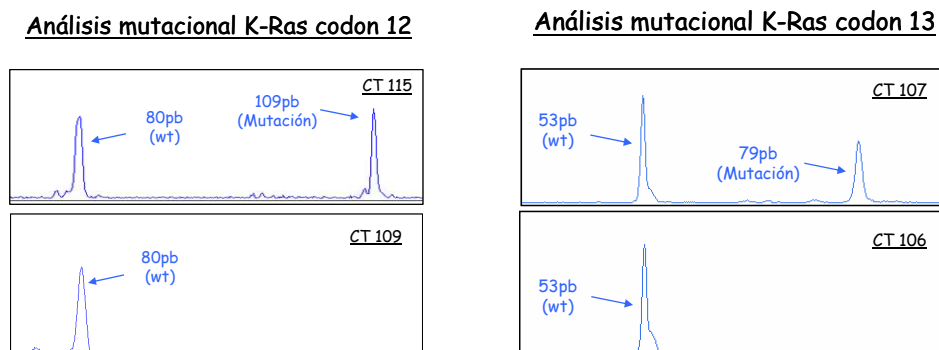


Fig.9. Resultados obtenidos en el análisis mutacional de KRAS en las muestras CT115, CT106, CT107 y CT109. Las muestras CT106 y CT 109 muestran un perfil salvaje para las mutaciones del codon 13 y 12 respectivamente mientras que las muestras CT115 y CT 107 muestran un perfil mutado para los codones 12 y 13 de KRAS respectivamente.

tamaños (29pb y 80pb) aunque solo uno (80pb) permanecerá marcado y podrá verse en el secuenciador. Si el gen está mutado, veremos una secuencia en la que se habrá perdido el sitio de restricción obteniéndose por tanto un fragmento del tamaño original (109pb). (Fig.11).

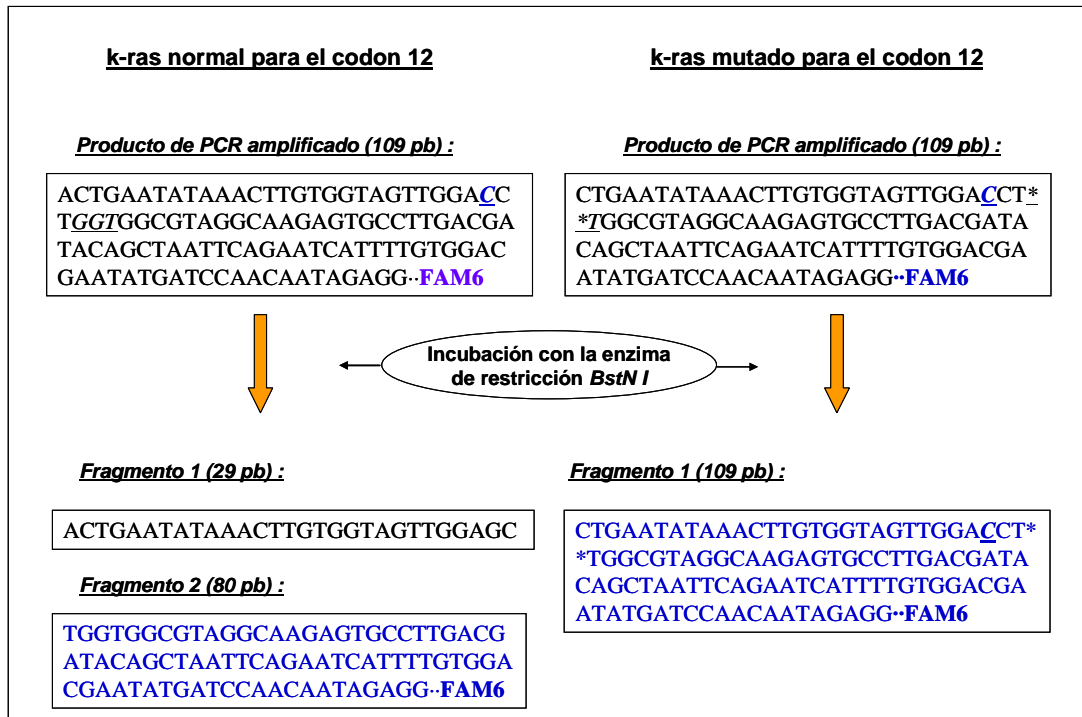


Fig.11. Proceso de detección de mutaciones en el codon 12. La presencia de una mutación viene definida por la aparición de un único amplicón (109pb) tras la digestión del fragmento de PCR con la enzima *BstNI*. La ausencia de la misma viene definida con la aparición de dos fragmentos más pequeños (29pb y 80pb) debido a la digestión del fragmento de PCR con la enzima de restricción *BstNI*. El símbolo “*” representa las dos bases del codon 12 que se han descrito que pueden sufrir mutaciones.

❖ Mutación en el codon 13 de KRAS:

Para detectar las mutaciones en este codon, se seleccionaron los siguientes cebadores:

G

Cebador sentido → (FAM-6) -TAACGCCTGCTGAAAATGACTG (22pb)

Cebador antisentido → GTATCGTCAAGGCACTCTTGCCTAGG (26pb)

C

En el cebador antisentido, se sustituyó una citosina (C) por una guanina (G) para crear un sitio de restricción para la enzima “HaeIII” (corta en las secuencias GG/CC). En el cebador sentido se sustituyó una guanina (G) por una citosina (C) para eliminar un sitio de restricción y así mantener el marcaje del fragmento con FAM-6 y la posibilidad de verlo en el secuenciador.

Cuando no se encuentren mutaciones en el codon 13 de KRAS, obtendremos dos fragmentos de diferentes tamaños (26pb y 53pb) aunque solo uno (53pb) permanecerá marcado y podrá verse en el secuenciador. Si el gen está mutado, veremos una secuencia en la que se habrá perdido el sitio de restricción obteniéndose por tanto un fragmento del tamaño original (79pb). (Fig.12).

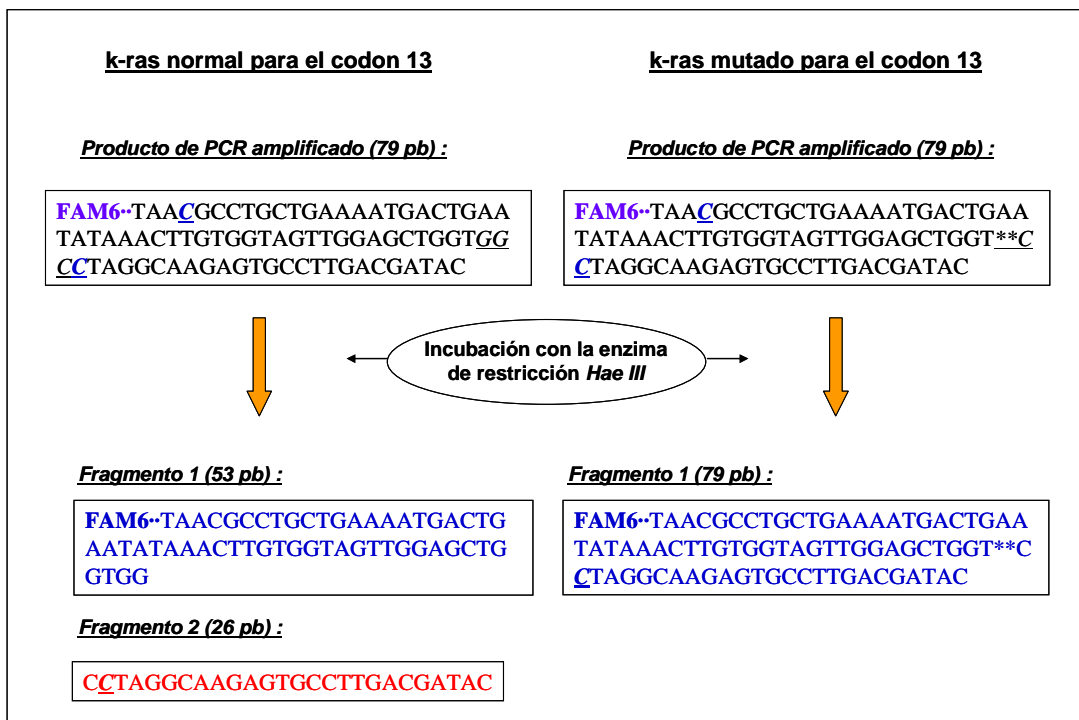


Fig.12. Proceso de detección de mutaciones en el codon 13. La presencia de una mutación viene definida por la aparición de un único amplicón (79pb) tras la digestión del fragmento de PCR con la enzima HaeIII. La ausencia de la misma viene definida con la aparición de dos fragmentos más pequeños (53pb y 26pb) debido a la digestión del fragmento de PCR con la enzima de restricción HaeIII. El símbolo “*” representa las dos bases del codon 12 que se han descrito que pueden sufrir mutaciones.

12. Estudio de asociación de las agrupaciones con los parámetros clínicos y patológicos.

Los estudios de asociación entre los parámetros clínicos, patológicos y las distintas clasificaciones (Dukes y clasificación molecular) se realizaron utilizando el programa estadístico *SPSS v13.0* (SPSS Inc. Chicago, USA).

La comparación de variables cualitativas se realizó mediante el análisis de tablas de contingencia utilizando el estadístico Chi-cuadrado o el test exacto de Fisher/likelihood ratio (cuando la frecuencia esperada de un 25% o más de las categorías era menor de 5) para calcular la probabilidad de la distribución. En todos los casos se aplicó el test de dos colas.

La asociación con las variables cuantitativas (caracterizadas mediante media y desviación estándar) y debido al escaso número de muestras en algunas de las categorías fue analizada mediante métodos no paramétricos para múltiples categorías (test de Kruskal-Wallis). Las comparaciones por pares cuando fueron necesarias se llevaron a cabo mediante el test de U-Mann-Whitney.

Para todos los contrastes de hipótesis, la hipótesis nula que establece que no existen diferencias entre los grupos fue rechazada cuando el error de tipo I o error- α fue menor de 0.05. No se llevaron a cabo correcciones para los p-valores.

13. Análisis inmunohistoquímico.

13.1. Construcción de los Tissue- Microarrays.

La técnica del “Tissue-MicroArray” permite incluir en único un bloque de parafina (bloque receptor) cilindros de tejido (parafinado) de múltiples pacientes (bloques donadores), pudiendo así realizar análisis inmunohistoquímico de un gran número de muestras a la vez (Fig.13).

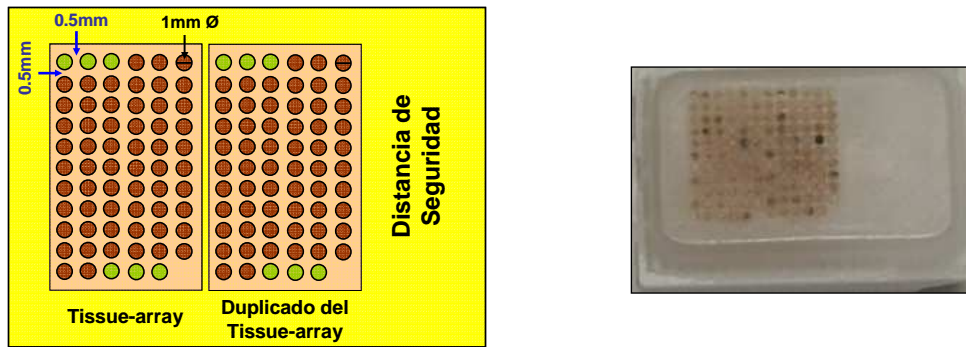


Fig.13. Esquema representativo de la distribución de los cilindros de tejido en el tissue-array. En el diagrama de la izquierda, los puntos verdes indican los tejidos control, mientras que los puntos rojos representan las muestras tumorales. A la derecha, resultado de uno de los TMA realizados.

Durante este trabajo se construyeron dos tissue-arrays (Fig.13) en los que se incluyó tejido tumoral de 116 pacientes para un total de 234 muestras. Además de introducir por cada paciente 2 cilindros de tejidos (1mm Ø), también se incorporaron al bloque varios tejidos controles con la finalidad de comprobar el correcto funcionamiento de la técnica (un total de 13 tejidos para 26 muestras) (Fig.13). Cada uno de los tissue-array fue posteriormente cortado en secciones de 3µm de grosor, cada una de las cuales fue montada sobre un portaobjetos y embebida completamente en parafina para su almacenamiento hasta su procesamiento inmunohistoquímico.

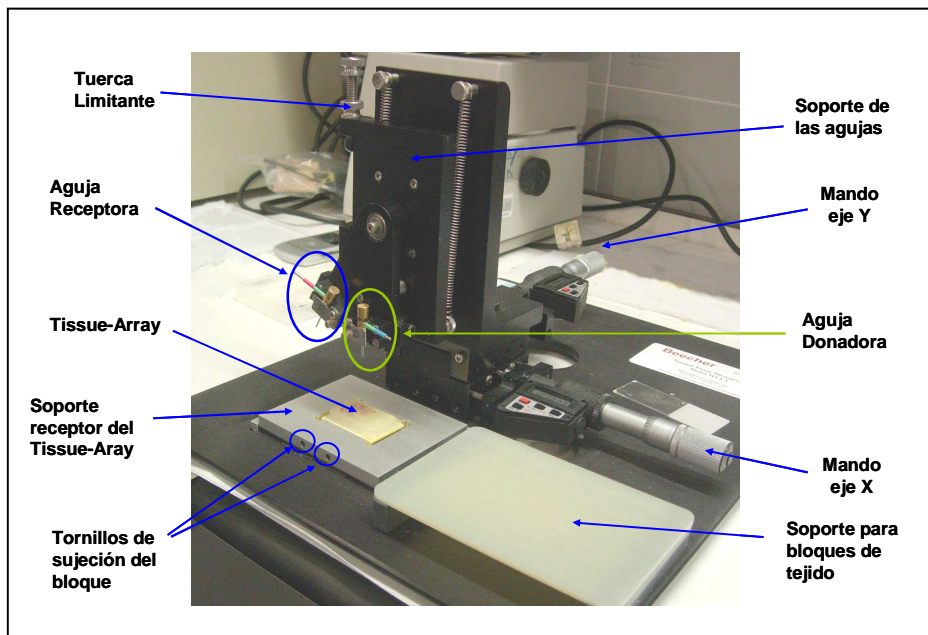


Fig.14. Esquema representativo de los componentes de un Tissue-MicroArrayer.

De manera resumida la construcción del TMA consta de dos pasos [285]:

- ❖ Selección de los tejidos (Fig.15A). De cada una de las muestras que se incluyó en este estudio se obtuvo un bloque de tejido parafinado. Para seleccionar la región del bloque más adecuada en cada uno de ellos se realizó, al microtomo, un corte de unos 4µm de grosor que se tiñó con hematoxilina–eosina y fue examinado y marcado por un patólogo (zona tumoral para los bloques a estudio y zona normal para los bloques control). La misma región seleccionada en la tinción fue seleccionada en el bloque de parafina.
- ❖ Inclusión de los tejidos en el bloque receptor (Fig.15B, C y D). Mediante una aguja se perfora parcialmente el bloque de parafina virgen (receptor) dejando un orificio (Fig.15B) en el que se incluirá un cilindro de tejido (Fig.15D) que previamente habíamos extraído (Fig.15C) también mediante aguja, de uno de los bloques donadores. Este paso se repetirá para cada uno de las muestras que se tengan que incluir. Este proceso se realizó utilizando un Tissue-Arrayer (Manual Tissue Microarrayer. N° modelo: MTA1675. Beecher Instrument Inc, Wisconsin, USA) (Fig.14).

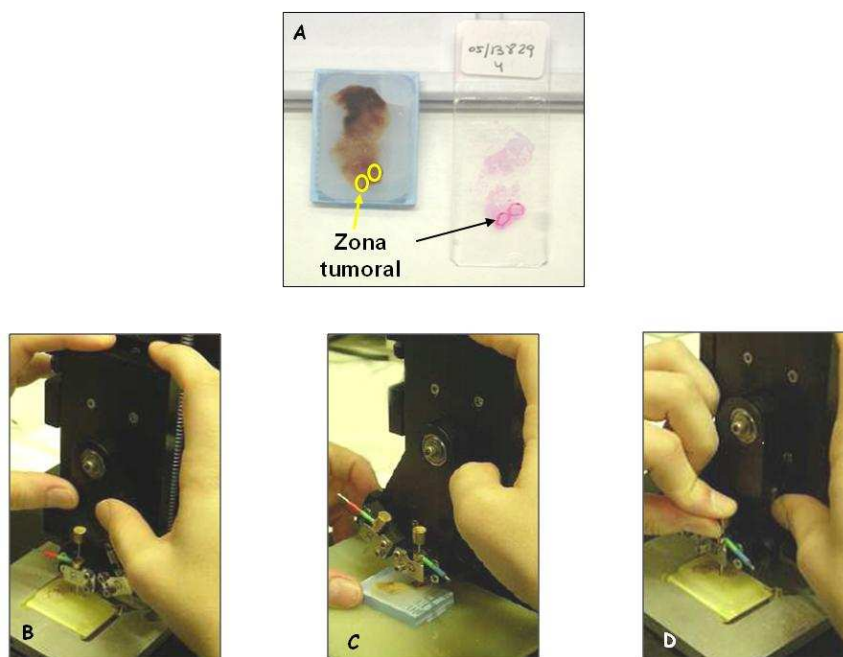


Fig.15. Proceso de construcción de un Tissue-MicroArray. A) Selección de la muestra. B) Perforación del bloque de parafina virgen. C) Extracción del cilindro de tejido del “bloque muestra”. D) Inserción y ajuste del cilindro de tejido en el bloque receptor.

13.2. Estudio inmunohistoquímico de las muestras

Tanto las técnicas inmunohistoquímicas como la técnica de hibridación in situ fluorescente (FISH) se realizaron sobre cortes de 3µm de grosor, obtenidos de los dos TMA previamente realizados, recogidos sobre cristal de inmunohistoquímica. En ambas técnicas el tejido se desparafinó con xilol (2 x 10') y posteriormente fue rehidratado en un gradiente de concentraciones decrecientes de etanol: etanol absoluto (2 x 5'), etanol al 96% (2 x 5'), etanol al 70% (2 x 5') y agua destilada.

13.2.1. Protocolo FISH para EGFR:

Consta de varios pasos que se describirán brevemente a continuación:

- ❖ **Desenmascaramiento del tejido:** mediante buffer citrato a pH:6 (Dako REAL Target Retrieval Solution 1x. Dako group. Glostrup, Denmark) en olla a presión durante 4'. Lavar 5' con Buffer 2 x SSC.
- ❖ **Rotura de membranas celulares:** mediante digestión con proteinasa K (Proteinase K. Dako group. Glostrup, Denmark) durante 10' a 37°C. Lavar con Buffer 2 x SSC durante 5'. Fijar con formol tamponado durante 10'.
- ❖ **Deshidratación:** con pasos sucesivos de 1' en un gradiente de alcohol (70%, 96% y 100%).
- ❖ **Hibridación de la sonda sobre el corte del TMA:** se añaden 6µl de la mezcla de sondas para el gen EGFR (Vysis LSI EGFR SpectrumOrange. Abbot laboratories. Illinois, EEUU) y el centrómero del cromosoma 7 (CEP 7 SpectrumGreen Probe. Abbot laboratories. Illinois, EEUU). Se sella con un cubre-objetos y se incuba 5' a 90°C para desnaturalizar el DNA. Se deja hibridar la sonda durante toda la noche en una cámara húmeda a 37°C.
- ❖ **Lavado de astringencia para eliminar las hibridaciones inespecíficas:** retirar el cubre-objetos mediante inmersión en solución de lavado (NP40 0,3% y 2 x SSC) a

temperatura ambiente. Posterior eliminación del exceso de sonda y de aquella sonda unida de manera inespecífica al tejido, mediante inmersión en la solución de lavado a 72°C durante 5'.

- ❖ Tinción de los núcleos celulares mediante incubación con DAPI (Abbot laboratories. Illinois, EEUU).

Se consideró la existencia de amplificación cuando el ratio de copias EGFR/Centrómero-7 era ≥ 2 . Se consideró como polisomía cuando como mínimo el 50% de los núcleos no superpuestos contenían al menos 3 copias del gen EGFR y del Centrómero-7.

13.2.2. Protocolo para tinción inmunohistoquímica (Ki-67, M-30 y β -catenina):

Se realizaron tinciones inmunohistoquímicas para la detección de las proteínas Ki-67 (proliferación), M-30 (apoptosis) y β -catenina (vía Wnt). A continuación se resumen los pasos más importantes:

- ❖ Desenmascaramiento de los antígenos (Tabla.8).

	Buffer	Método	Temperatura	Tiempo
Ki-67	Buffer Citrato. pH:6 (Dako Real Target Retrieval Solution 1x)	Presión	—	8'
M-30	React. ER-1. pH:6 (Bond Epitope Retrieval Solution)	Calor	100°C	15'
β-catenina	React. ER-1. pH:6 (Bond Epitope Retrieval Solution)	Calor	100°C	15'

Tabla.8. Esquema del proceso de desenmascaramiento de antígenos.

- ❖ Se produce el bloqueo de la peroxidasa endógena mediante H₂O₂ durante 10'.
- ❖ Incubación con el anticuerpo primario (Tabla.9).

	Anticuerpo	Tipo	Temperatura	Dilución	Tiempo
Ki-67	Ki-67 Antigen. Clone MIB1(Dako group)	Monoclonal	Ambiente	1:100	30'
M-30	M-30 Cytodeath (Roche Farma)	Monoclonal	Ambiente	1:50	60'
β-catenina	β -catenina prediluida (Master diagnostica)	Policlonal	Ambiente	—	60'

Tabla.9. Esquema del proceso de incubación con el anticuerpo primario.

❖ Incubación con el Anticuerpo secundario y revelado (Tabla.10).

	Método de detección	Enzima / Sustrato	Temperatura	Tiempo Enz / Sust	Contraste
Ki-67	Autostainer DAKO	HRP / DAB	Ambiente	30' / 5'	Hematoxilina
M-30	Bond Polymer Refine Detection	HRP / DAB	Ambiente	30' / 5'	Hematoxilina
β-catenina	Bond Polymer Refine Detection	HRP / DAB	Ambiente	30' / 5'	Hematoxilina

HRP: Peroxidasa de Rabano; DAB: Diamino-Bencidina

Tabla.10. Esquema del proceso de incubación con el anticuerpo secundario y de revelado.

Nota: entre cada uno de los pasos previos se realiza un lavado con la solución Dako Wash Buffer 1x (Dako group. Glostrup, Denmark).

❖ Se procede de nuevo a la deshidratación del tejido mediante gradiente de concentraciones crecientes de etanol (Etanol 70%, 96% y 100%) y al posterior pase en xilol necesario para el montaje permanente de la tinción.

La valoración tanto de la proteína Ki-67 como de la proteína M-30 fueron realizadas como porcentajes de células teñidas en relación al total de células epiteliales (Fig.16). En el caso de β -catenina se valoró la tinción de la proteína en el núcleo de las células epiteliales (Fig.16).

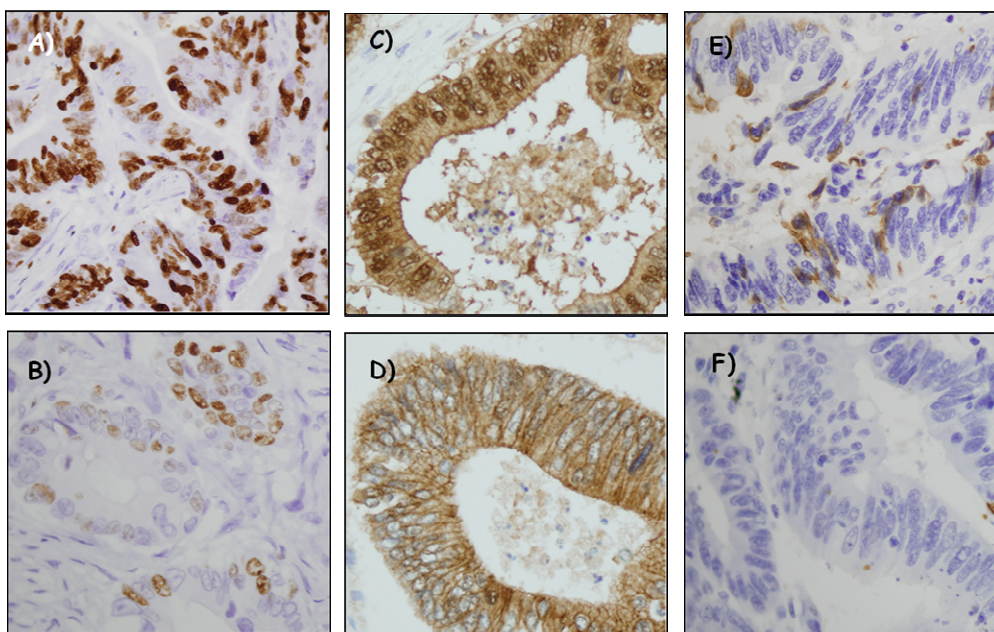


Fig.16. Tinciones con los anticuerpos para la proteína Ki-67 (A y B), β -catenina nuclear (C y D) y M-30 (E y F). Las secciones A, C y E muestran una alta tinción para las tres proteínas analizadas, mientras que las secciones B, D y F muestran niveles bajos o nulos de tinción para dichas proteínas en diversas muestras analizadas durante este estudio.

14. Validación externa de la clasificación molecular.

Se realizó una validación “in silico”, utilizando matrices de datos procedentes de estudios de microarrays en pacientes con cáncer colorrectal previamente publicados.

Los criterios para la selección de la matriz de datos externos fueron:

- ❖ La matriz debía incluir en su análisis al menos 60 pacientes.
- ❖ Los datos de expresión génica estuviesen públicamente disponibles.

De los dos estudios que cumplían este requisito ([252,260]) se seleccionó el realizado por el grupo de Eschrich [252] debido a la mayor similitud técnica del estudio con el nuestro y a que poseían de manera pública los datos de supervivencia de sus pacientes. El proceso de validación constó de tres fases:

- ❖ Fusión de los dos conjuntos de datos en una única matriz: este proceso consta de dos partes.

- Selección de la lista de genes comunes a los dos estudios. Esto conlleva una serie de transformaciones en ambas matrices de datos para poder promediar aquellos transcritos con un mismo identificador (se usó el “Unigene Cluster ID” como identificador único) y así poder comparar las listas de genes incluidos en ambos experimentos:
 - HCSC: se eliminaron los transcritos que no poseían este tipo de identificador y a la vez se promediaron aquellos que compartían la misma identificación.
 - Eschrich S et al: la matriz estaba formada inicialmente por 78 muestras (23 estadio B, 22 estadio C, 30 estadios D y 3 adenomas) y 32208 transcritos normalizados mediante la técnica de LOWESS. Tras deshacer la transformación inicial, con la que se recibieron los datos, del tipo $-1/x$ para los ratios de expresión (“x”) que mostraban inhibición, los valores de la matriz al completo fueron transformados a \log_2 . Se eliminaron los transcritos que no poseían ningún tipo de identificación (denominados “null”) y aquellos que poseían más de un 25% de valores missing (hacen referencia a los que el autor consideraba como puntos problemáticos o de baja calidad). De los transcritos restantes, aquellos con el mismo identificador (“Unigen Cluster ID”) fueron promediados y posteriormente aquellos con más de un 10% de datos missing fueron eliminados.
- Fusión de las matrices: se utilizó el programa “Distance Weighted Discrimination” (<https://genome.unc.edu/pubsup/dwd/>) [286] que utiliza técnicas multivariantes para identificar y eliminar los efectos propios de las diferentes plataformas, haciendo posible el análisis conjunto de diversas matrices. En este caso, al ser matrices similares y según recomendaciones del creador del programa se ajustó la matriz Eschrich S et al. a la nuestra (HCSC).
- ❖ Predicción de las muestras de la matriz externa: se crearon algoritmos de predicción utilizando aquellos genes significativamente diferentes entre los grupos de la clasificación molecular propuesta ($p < 0,001$) (muestras HCSC) para así poder asignar las muestras de la matriz externas (Eschrich S et al) a los grupos

correspondientes. Se elaboraron un total de cuatro predictores diferentes mediante cuatro métodos distintos :

- “Diagonal Linear Discriminant Analysis” (DLDA).
- “1-Nearest Neighbord” (1NN).
- “3-Nearest Neighbors” (3NN).
- “Nearest Centroid” (NC).

Para cada uno de ellos se realizó una validación cruzada utilizando la técnica de “Leave One Out Cross Validation” (LOOCV) y obteniéndose datos sobre la precisión, la tasa de error global de cada uno de ellos, además de sobre los coeficientes de sensibilidad, especificidad, valor predictivo positivo y valor predictivo negativo de cada uno de los grupos en función del predictor. El predictor obtenido se consideró fiable cuando la probabilidad de que un predictor obtenido al azar clasificase mejor, no superaba el 0,01% (tras 100 repeticiones).

Posteriormente dichos algoritmos fueron utilizados para clasificar los tumores de la matriz de Eschrich como pertenecientes a alguno de los cuatro grupos moleculares propuestos.

Basado en las mismas reglas utilizadas para el otro predictor, se creó también un algoritmo de predicción que distinguía únicamente tumores “de bajo estroma” del resto de grupos. Partiendo de los 461 genes restantes de los 1722 utilizados originalmente para establecer la clasificación, se seleccionaron 167 genes que diferenciaban ambos grupos ($p < 0.001$) y se creó un predictor basado en el método del “Nearest Centroid” que se usó exclusivamente en los análisis de supervivencia.

- ❖ Correlación entre la predicción realizada y el análisis no supervisado: el resultado del proceso de predicción de las muestras de Eschrich se comparó con el resultado obtenido de un análisis no supervisado (“clustering” jerárquico con centrado de genes, correlación centrada de Pearson y “average linkage”) de las muestras tumorales (muestras HCSC y Eschrich et al. excluyendo la muestras normales, los adenomas y las muestras del grupo 5) estudiándose cual era su coincidencia. Para el

análisis no supervisado (HC) se utilizaron los genes coincidentes en ambas matrices de los 1722 genes originales (obteniéndose al final una matriz de 461 genes).

A modo de resumen de este procedimiento se muestra la Fig.17.

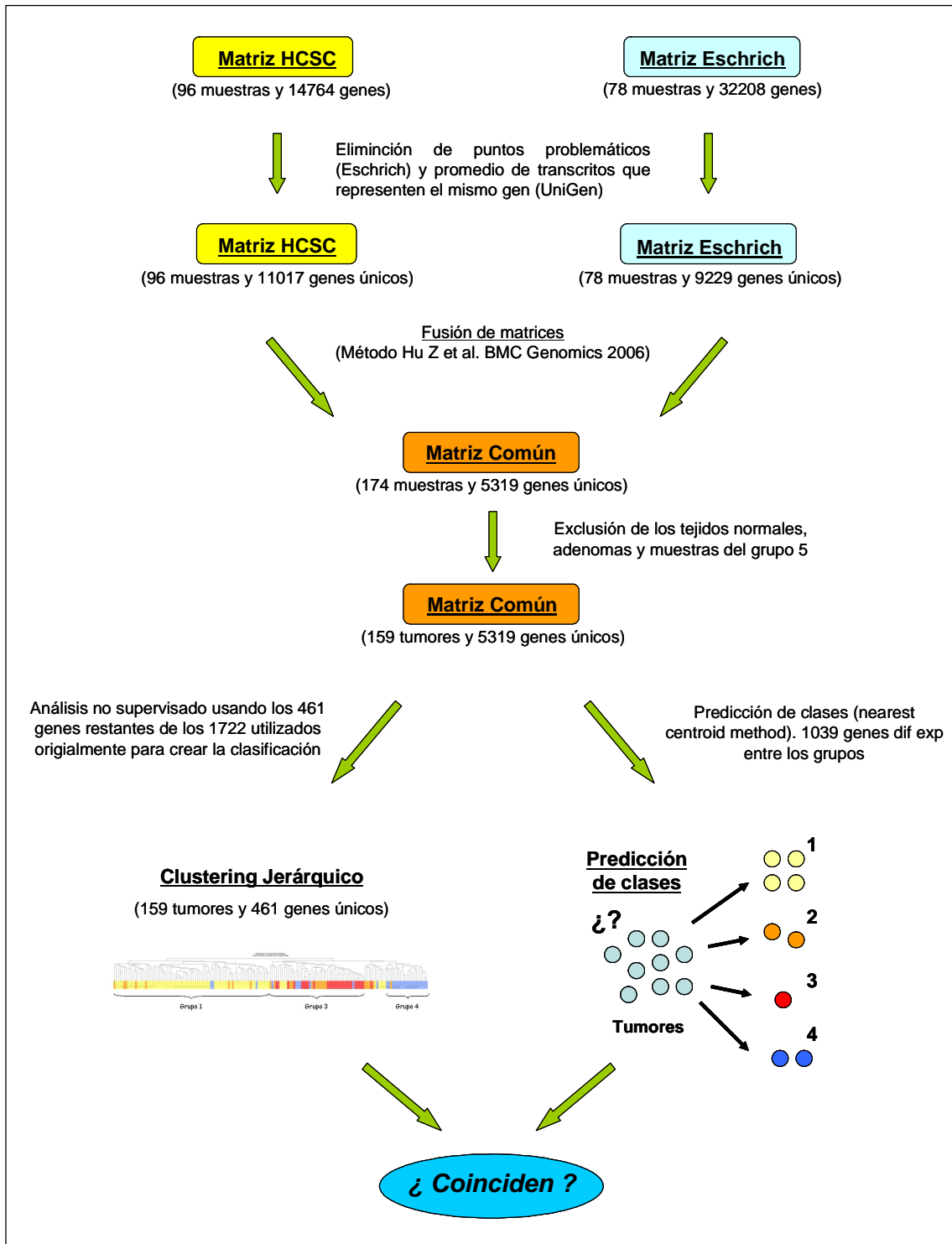


Fig.17. Esquema del proceso de fusión de matrices para la validación de la clasificación molecular.

15. Análisis de los perfiles de supervivencia.

Este análisis se llevo a cabo en dos fases:

- ❖ Aplicación de los predictores de supervivencia públicos a las muestras HCSC: ante la imposibilidad de aplicar de manera exacta los algoritmos de predicción publicados, debido al uso de diferentes plataformas, se llevaron a cabo análisis no supervisados (“clustering” jerárquico) usando los genes de estos predictores y comparando los resultados con los grupos establecidos en nuestra clasificación molecular. Los predictores utilizados fueron los pertenecientes al trabajo de Wang [253], Eschrich [252] y Lin [260] ya que en los tres casos eran publicaciones que usaban más de 60 pacientes en el trabajo y habían realizado validaciones cruzadas para confirmar su validez.
- ❖ Análisis de supervivencia: se llevó a cabo tanto para nuestra matriz de datos (datos HCSC) como para la matriz de validación (datos Eschrich). En el primer caso se realizaron análisis de supervivencia global y de período libre de recurrencia. En el caso de la matriz de datos externa (Eschrich), se contó exclusivamente con los datos de supervivencia global que fueron amablemente cedidos, al igual que los datos de expresión, por los autores del trabajo. Todos los paciente incluidos en el análisis contaron con seguimiento mínimo de 36 meses (salvo, claro está, en el caso de muerte previa debida al tumor).

Las funciones de supervivencia se llevaron a cabo usando el método actuarial. Para comparar dichas funciones de supervivencia se utilizó el test de la probabilidad exacta (likelihood exact test). Se estableció el modelo de regresión de riesgos proporcionales de Cox con el que se calcularon los “Hazard Ratios” (HR) y los correspondientes intervalos de confianza al 95%. En cada contraste de hipótesis, la tasa de proporcionalidad fue verificada. La visualización de las funciones de supervivencia se llevó a cabo mediante curvas de Kaplan-Meier. Para todos los contrastes de hipótesis, la hipótesis nula que establece que no existen diferencias entre los grupos fue rechazada cuando el error de tipo I o error- α fue menor de 0.05. El programa utilizado fue el SPSS v.13 (SPSS Inc, Chicago, USA).

Resultados.

1. Puesta a punto de la técnica y controles de calidad.

La aplicación de un cuidadoso diseño experimental en el que se estandaricen las fases del trabajo que todavía dependen del criterio del investigador es, en el caso de los microarrays, más esencial si cabe que en el resto de técnicas de biología molecular para la obtención de resultados fiables y libres de sesgos. Entre estas etapas no estandarizadas se encontraba el análisis histológico de los tumores, la medición de la calidad del RNA purificado que se iba a hibridar y los controles de calidad de la hibridación. Los resultados obtenidos tras la puesta a punto de las diferentes etapas y la aplicación de los controles de calidad se muestran a continuación:

El estudio comenzó con un total de 116 pacientes

- ❖ Análisis histológico de la composición de las muestras tumorales. La aplicación de este análisis conllevó que se descartaran un total de 20 pacientes (paciente CT4, 6, 7, 9, 25, 26, 43, 61, 65, 66, 67, 71, 77, 79, 80, 87, 88, 89, 111 y 117) de los 116 que inicialmente habían sido incluidos en el estudio, por poseer una contaminación con tejido normal superior al 20%.

El estudio histológico de las muestras finalmente incluidas en el trabajo (88 tumores), mostró (Tabla.1) que el porcentaje medio de células epiteliales tumorales en los tejidos analizados era aproximadamente del 80%. Por otro lado dentro de lo que denominamos componente tumoral (epitelio tumoral + estroma tumoral), el estroma malignizado representaba en media aproximadamente un 16%. Para verificar si la composición de la pieza tumoral analizada era un reflejo de la morfología del tumor completo y no consecuencia del criterio establecido durante el análisis histológico de las muestras (selección de al menos un 80% de componente tumoral), se analizó para cada tumor un fragmento distinto al utilizado para la obtención del RNA. Estas nuevas piezas pertenecían a zonas diferentes de los tumores que habían sido embebidas en parafina para su almacenamiento y su uso en la práctica clínica rutinaria. El porcentaje medio de estroma tumoral existente en las muestras parafinadas analizadas se estimó en un 15.3%, resultando casi idéntico al obtenido en el tejido congelado (16%) (Tabla.1). Además, el estudio de correlación entre el porcentaje de estroma de las dos regiones tumorales analizadas (tejido

congelado y parafinado) para cada uno de los pacientes, indicó que no existían diferencias significativas entre ambas áreas tumorales y que la correlación entre dichas regiones era elevada (ICC: 0.781; (IC95%: 0.681-0.852)), pudiendo por tanto asumir que la asociación entre el porcentaje de estroma tumoral y los grupos moleculares es real y no fruto del fragmento de tejido seleccionado.

La presencia de necrosis (considerada como 5% o más del componente tumoral) en los tumores no fue muy frecuente, observándose solamente en 14 de las 88 muestras estudiadas (16%) (Tabla.1). No se observaron diferencias en la calidad del RNA de las muestras con necrosis (Media RIN: 8.84) frente a las muestras que no presentaban necrosis (Media RIN: 8.96).

	% cél. epiteliales tumorales	% de células dentro del componente tumoral		
		Estroma. Muestra congelada	Estroma. Muestra parafinada	% necrosis. Muestra congelada
Media	79.6	16.6	15.3	2.4
Mediana	80	15	15	0
Desv std	10.7	10.7	9.2	6.0
Max	95	65	60	30
Min	35	5	5	0

Tabla.1. Resumen del análisis histológico realizado a las muestras incluidas en el estudio.

- ❖ Control de calidad del RNA extraído. Sólo 5 muestras (paciente CT13, 21, 46, 51 y 85) de las 96 analizadas fueron eliminadas por no alcanzar el criterio mínimo de calidad exigido a los RNAs extraídos ($RIN \geq 7.5$). La calidad de los RNAs restantes que se incluyeron en el estudio fue muy elevada obteniéndose un RIN medio de 8.9 (en una escala de 0-10) con un mínimo de 7.5 y un máximo de 10 (Fig.1 y Tabla.3).

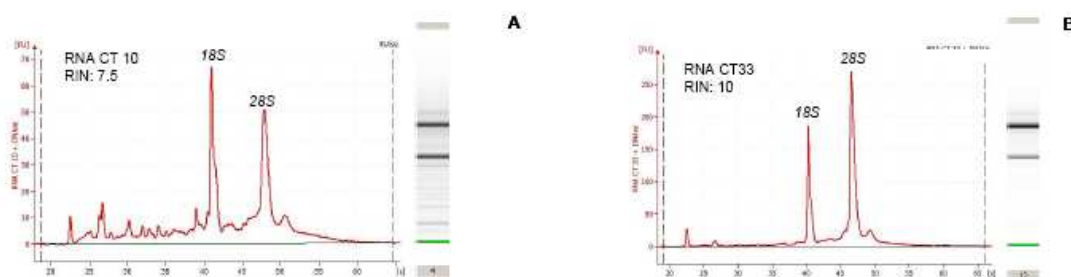


Fig.1. Representación del rango de calidad de los RNAs seleccionados para el estudio. A) Muestra CT 10 con el límite mínimo de calidad aceptado (RIN: 7.5) para el estudio. B) Muestra CT 33 con el RNA extraído intacto (RIN: 10, máxima calidad).

Durante el estudio se llevaron a cabo un total de 116 hibridaciones mediante doble marcaje (Cy5 y Cy3) que comprendían las 91 muestras tumorales restantes, a las que se sumaron 8 tejidos de mucosa normal de colon cuya calidad de RNA estaba dentro de los rangos establecidos. La distribución de las hibridaciones fue la siguiente: 95 hibridaciones Tumorales (Tumor-Cy5/"Pool"CN-Cy3) con un total de 91 tumores únicos y 4 réplicas de tumores cuya hibridación inicial no había sido adecuada; 8 tejidos de colon normal (Normal-Cy5/"Pool" CN-Cy3) que facilitaron el estudio de las alteraciones entre cada uno de los distintos grupos tumorales encontrados con respecto al tejido colorrectal normal, ya que la muestra de referencia se utilizó sólo para los procesos de normalización y de control de la técnica; 13 réplicas control ("Pool" CN-Cy5/"Pool" CN-Cy3) indicativas de la variabilidad propia de la técnica. En la Fig.2. se muestra una representación de los gráficos obtenidos (MA plots) para las muestras de cada uno de los 3 tipos de hibridación que se generaron durante este estudio y que acabamos de describir.

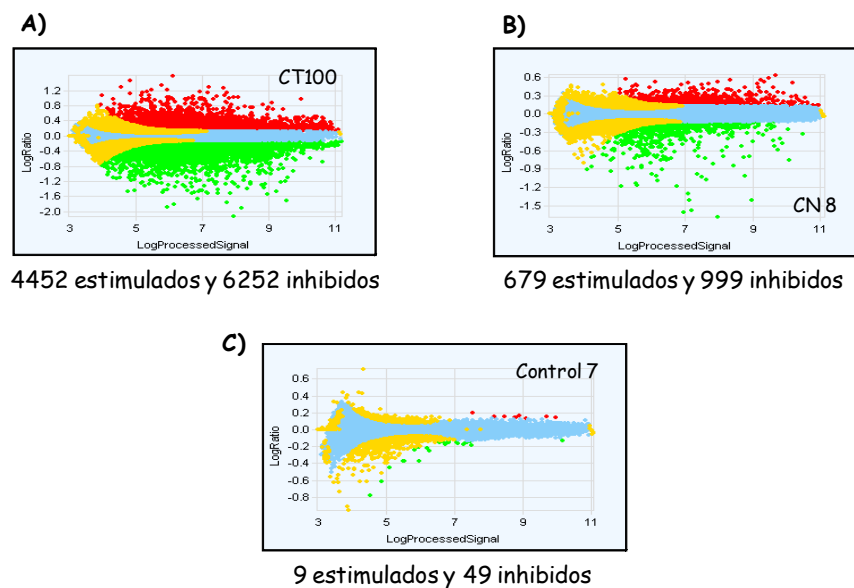


Fig.2. Representación mediante M-A plot de los genes diferencialmente expresados en los tres tipos de hibridaciones realizadas. A) Tejido tumoral vs. "Pool" CN; B) tejido normal vs. "Pool" CN; C) "Pool" CN vs. "Pool" CN. El color rojo indica genes sobreexpresado en la muestra analizada en relación a la muestra de referencia; el color verde indica genes inhibidos; el color azul indica los genes que se han utilizado para normalizar los valores de expresión; el color amarillo indica genes que no están diferencialmente expresados a un p-valor < 0.01. $M = \text{Log}_2 \text{Tumor} - \text{Log}_2 \text{"Pool" CN}$; $A = (\frac{1}{2}) \times (\text{Log}_2 \text{Tumor} + \text{Log}_2 \text{"Pool" CN})$.

La diferencia media observada en el número de transcritos (~5000 transcritos) en estas representaciones entre las muestras de tejido tumoral (media \pm DS: 7369 ± 2015) y las muestras de tejido normal (2212 ± 857), fueron reflejo de los numerosos cambios que se producen en el proceso de malignización tumoral.

Por otro lado, mediante el análisis de los MA-plots de la hibridaciones control (“Pool” CN-Cy5 / “Pool” CN-Cy3) se pudo comprobar la escasa variabilidad técnica de los microarrays (media + DS en las réplicas control (218 ± 326)) observándose que la media del número de genes diferencialmente expresados no superaba la que se esperaba encontrar por azar (p-valor = 0.01) si las muestras no fuesen realmente diferentes.

❖ Control de calidad de las hibridaciones (microarrays). Del total de 116 hibridaciones realizadas sólo 5 fueron eliminadas (nº CT19, CT32, CT33, CT47 y CT54). Para 3 de estas muestras eliminadas (nº CT19, CT32 y CT33), se volvió a realizar el proceso de marcaje e hibridación y al superar los controles de calidad fueron finalmente incluidas en el análisis. El comportamiento de la muestra CT102 y de su duplicado a lo largo del análisis, sirvieron para verificar que todas las réplicas, hibridadas en microarrays de formato 4x44k (CT19, CT32 y CT102duplicado) y con un protocolo diferente de lavado (CT19, CT32, CT33 y CT102duplicado), eran compatibles con el resto de muestras, hibridadas en formato 1x44k y con protocolos de lavado diferentes (entre las que se encontraba el CT102 original).

Los valores medios obtenidos para las muestras finalmente incluidas en el estudio se ajustaron bien a los controles establecidos (Tabla.2). El hecho de que puntualmente, determinadas muestras superasen alguno de estos límites marcados no fue de suma importancia ya que todas ellas cumplían con los controles establecidos en al menos el 75% de los parámetros analizados. Se puede concluir por tanto, que las hibridaciones de los microarrays considerados para posteriores estudios fueron buenas y que por tanto, los datos obtenidos son fiables.

Parámetro	Microarrays escaneados con el FE 8.5				Microarrays escaneados con el FE 9.5			
	Límite Superior		Límite Inferior		Límite Superior		Límite Inferior	
	C. Rojo	C. verde	C. Rojo	C. verde	C. Rojo	C. verde	C. Rojo	C. verde
NegCtrlAveNetSig	37.8 ± 5.7 (70)	38.7 ± 24.1 (70)	—	—	85.9 ± 2 (70)	28.2 ± 8.6 (70)	—	—
NumFeature NonUnifOL	22.3 ± 25.3 (150)	29.2 ± 26 (150)	—	—	7 ± 5.1 (150)	8.7 ± 5 (150)	—	—
NumPopnOL	78.1 ± 27 (155.5)	58.3 ± 28.5 (147.2)	—	—	156.8 ± 15.6 (180.4)	128.2 ± 28.8 (168.15)	—	—
NonCtrlMedPront CVBGSubSig	19 ± 3.9 (25)	18.3 ± 7.4 (25)	FR = 13 (0)	FR = 0 (0)	14.2 ± 1.4 (25)	13.7 ± 1.4 (25)	FR = 0 (0)	FR = 0 (0)
NumFoundFeat	—		42864 ± 913 (39723)		—		40036 ± 1412 (36067)	
NumPopnBGOL	1236 ± 432 (2172.8)	407 ± 525 (1555.4)	—	—	977 ± 186.9 (1565.6)	908 ± 655 (3144.1)	—	—
Prop_down_ vs_up_regulated Tumores	FR = 13. Cuartiles (0.7-0.8-0.9-1.1) (0.666-1.5)				FR = 0 (0.666-1.5)			

Tabla.2. Valores medios obtenidos (**Media ± DS**) para cada uno de los parámetros analizados durante los controles de calidad de las hibridaciones. Los datos fueron obtenidos a partir de los 110 microarrays que fueron considerados válidos para el análisis. En cada una de las celdas aparece entre paréntesis el límite superior o inferior para el cuál las muestras se consideran fuera de rango, según lo establecido en la sección 9.1 de materiales y métodos (Pág.69). En el caso del parámetro “Prop_down_vs_up_regulated Tumores”, solo fueron incluidos datos de los tejidos tumorales. FR: muestras fuera de rango.

Como excepción, la muestra 113 fue eliminada por la existencia de un número excesivamente elevado de puntos problemáticos (no uniformes) en el canal rojo (Cy5). La media para ese canal a lo largo del resto de muestras fue de 28.22 ± 56.08 , mientras que para el tumor CT113 el número de puntos no uniformes en el canal rojo fue de 8947. Finalmente, el estudio continuó con 110 microarrays hibridados de los cuales 89 correspondieron a un total de 88 tumores únicos y una réplica (CT102).

- ❖ Estudio de comportamientos anómalos de las muestras mediante Análisis de Componentes Principales o PCA. En las 110 muestras analizadas (89 tumores, 8 normales y 13 controles) no se observaron sesgos significativos debidos al día en que se realizó la hibridación, al tipo de chip utilizado o al método de lavado empleado (Fig.3.). Sin embargo se eliminó una muestra de tejido normal (CN61) cuya expresión era más parecida a la de los tejidos tumorales que a la de los propios tejidos normales. Tras llevar a cabo el análisis histológico de dicha muestra, se confirmó la existencia de contaminación con tejido tumoral (~ 30% del tejido). El resto de muestras normales fueron también sometidas a análisis histológico sin que se encontrasen células tumorales. Por tanto el número de tejidos de colon normal finalmente analizado fue de 7.

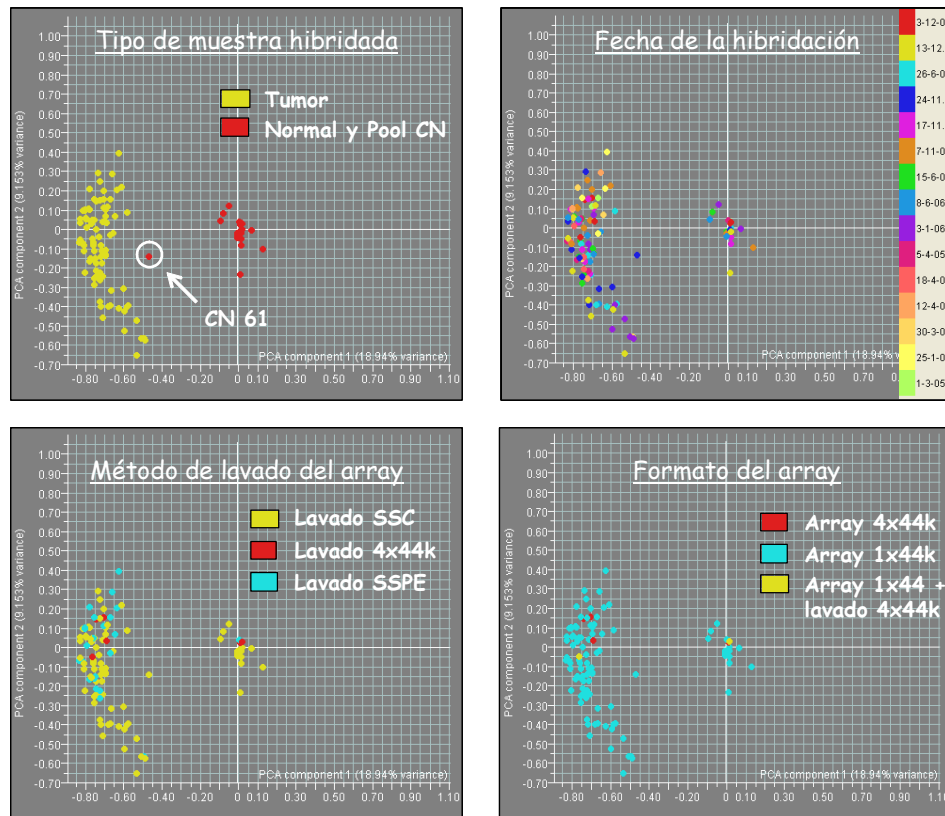


Fig.3. Evaluación mediante PCA de los 110 microarrays hibridados. De izquierda a derecha y de arriba abajo las muestras aparecen coloreadas en función del tipo de tejido, del día en el que se realizó la hibridación, del método de lavado utilizado y del formato de array hibridado.

El control de estos cuatro pasos fue por tanto esencial para no introducir sesgos artificiales de importancia durante la generación de los datos de análisis y por consiguiente para ser capaces de obtener e interpretar los resultados de manera fiable en el curso de su análisis posterior.

2. Características de los pacientes.

Del total de 116 pacientes con adenocarcinomas colorrectales esporádicos en estadios localizados o avanzados incluidos inicialmente, un total de 88 superaron los controles de calidad para ser finalmente incluidos en el estudio (Tabla.3).

De estos 88 pacientes ninguno, a excepción del paciente 119, recibió quimioterapia o radioterapia neoadyuvante. Aunque este tipo de tratamiento fue

considerado como un criterio de exclusión, este caso no fue descartado de inicio porque no se tuvo constancia de dicho tratamiento hasta una vez finalizado el análisis de datos. Sin embargo tras el mismo, se comprobó que este hecho no tuvo especial relevancia en el resultado final ya que esta muestra se asoció con el “cluster” 5 o “cluster” de muestras no clasificadas.

<u>88 Pacientes</u>			
		<u>Nº Pacientes</u>	<u>Porcentaje</u>
<u>Sexo</u>	Hombre	43	48.9%
	Mujer	45	51.1%
<u>Edad (años)</u>	Media (min-max)	70.9	(25-93)
<u>RIN</u>	Media (min-max)	8.9	(7.5-10)
<u>Estadio de Dukes</u>	A	24	27.3%
	B	26	29.5%
	C	19	21.6%
	D	19	21.6%
<u>MSS/MSI Status</u>	MSS	79	89.8%
	MSI	9	10.2%
<u>Location</u>	C. Ascendente	33	37.5%
	Sigmoide	32	36.4%
	Recto	14	15.9%
	C. Descendente	3	3.4%
	C. Transverso	6	6.8%
<u>Nodos analizados</u>	Menos de 12	28	31.8%
	Más de 12	60	68.2%
<u>Mut. K-ras</u>	Codon 12	28	31.8%
	Codon 13	6	6.8%
	Wild Type	54	61.4%
<u>Grado Histológico</u>	Indiferenciado	1	1.1%
	Pobrementemente dif.	5	5.7%
	Moderadamente dif.	51	58.0%
	Bien dif.	31	35.2%
<u>Tipo Histológico</u>	Convencional	78	88.6%
	Mucinoso	10	11.4%
<u>Margen Tumoral</u>	Expansivo	16	18.2%
	Infiltrante	61	69.3%
	Mixto	11	12.5%
<u>Invasión Vascular</u>	Si	33	37.5%
	No	55	62.5%
<u>Invasión Perineural</u>	Si	8	9.1%
	No	80	90.9%
<u>Infiltrado linfocitario</u>	Ausente	53	60.2%
	Bajo	15	17.0%
	Medio	17	19.3%
	Alto	3	3.4%

Tabla.3. Descripción de los parámetros clínicos y patológicos más característicos de los pacientes.

La edad media de diagnóstico del tumor en los pacientes seleccionados fue de 70.9 años, en consonancia con la edad media de diagnóstico de 68 años descrita en otros trabajos [5]. De entre todos los pacientes, destacó el paciente nº 27 con una edad de diagnóstico anormalmente temprana (25 años) para tratarse de un tumor colorrectal de tipo esporádico.

La relación de pacientes en función del sexo se acercó a la paridad (49% hombres y 51% mujeres) no observándose una mayor frecuencia de aparición en hombres como está descrito [3]. Al tratarse de un trabajo de clasificación se intentó seleccionar un número igualmente representativo de tumores en cada estadio. El estadio más numeroso fue el B de Dukes con un 29.5% de las muestras y los menos numerosos los estadios C y D con un 21.6% cada uno. El resultado de los análisis de inestabilidad en microsatélites mostró que un 10.2% de los tumores colorrectales estudiados presentaban alta inestabilidad, acercándose a los rangos (12-15%) descritos en la literatura [64]. En cuanto a la localización, los tumores analizados parecieron desarrollarse de manera algo más frecuente en el colon distal (55.7%) que en proximal (44.3%) en consonancia con lo propuesto en otros trabajos [287,288].

En función del tipo histológico, las muestras se agruparon en adenocarcinomas convencionales (88.6%) y mucinosos (11.4%). No existe un consenso sobre la frecuencia de tumores colorrectales de tipo mucinoso. Tradicionalmente se ha estimado entre un 10-20% de los tumores colorrectales [289], aunque en función del criterio patológico estos porcentajes varían del 3-40% [136,289,290]. El análisis del gen KRAS, detalló una frecuencia de mutaciones en nuestra muestra del 38,6%, lo que también se ajusta a lo descrito (30-45%) hasta el momento en la literatura [27,291].

Finalmente, los datos obtenidos para el Antígeno Carcinoembrionario (CEA) al diagnóstico no fueron analizados, ya que pese a ser el único marcador molecular aceptado para su uso en la clínica (excluyendo el recientemente aprobado KRAS) y pese a estar incluido su análisis en la práctica rutinaria en el hospital, solo contamos con la información en 37 de los 88 casos estudiados (42.05%).

3. Análisis de microarrays.

El estudio de los perfiles de expresión génica global se llevó a cabo a partir de 88 muestras tumorales (89 hibridaciones) y 7 muestras colorrectales normales que pasaron los diferentes controles de calidad. El número de genes incluidos en la matriz final fue de 14764 genes según los criterios de filtrado explicado en la sección 9.2 de materiales y métodos (Pág.71). Sin embargo, para el propósito de la clasificación y caracterización tumoral solo fueron utilizados aquellos genes que mostraban una mayor variación en la expresión a lo largo de las muestras, utilizando de esta forma un grupo más reducido de 1722 genes (ver también sección 9.2 de materiales y métodos. Pág.71).

3.1. Caracterización del patrón de expresión de muestras normales y tumorales.

No siendo la comparación entre tejido tumoral y tejido normal el objetivo central de este trabajo, consideramos que los resultados obtenidos de su análisis podían ser interesantes ya que nos permitirían conocer los genes y ruta biológicas afectadas en nuestros tumores en el proceso de malignización, nos permitiría verificar la capacidad discriminatoria de este tipo de técnicas y por último nos brindaría la oportunidad de descubrir nuevos genes cuyo estudio más en profundidad podrían ser importantes para el tratamiento de la enfermedad.

Por tanto, mediante técnicas no supervisadas (“clustering” o agrupamiento jerárquico) realizamos una agrupación de genes y muestras en las que se encontraban los 88 tumores y las 7 muestras de colón normal y los 1722 genes previamente filtrados. El resultado obtenido, como cabía esperar, separó por un lado los tejidos normales y por otro lado los tumorales y nos permitió identificar los patrones de expresión diferencial entre ambos tipos de muestras (Fig.4). La validación matemática mediante el “D-Index” y el “R-Index” (ver materiales y métodos Sección 9.3. Pág.73) de las agrupaciones obtenidas confirmaron que molecularmente el tejido normal y el tejido tumoral son completamente diferentes entre sí, siendo capaces de asignar correctamente las muestras a su grupo correspondiente en el 100% de los casos (Tabla.4). Este experimento confirmó también la solidez de esta técnica para detectar alteraciones basadas en perfiles globales de expresión y por tanto apoyando así su utilización en la creación de

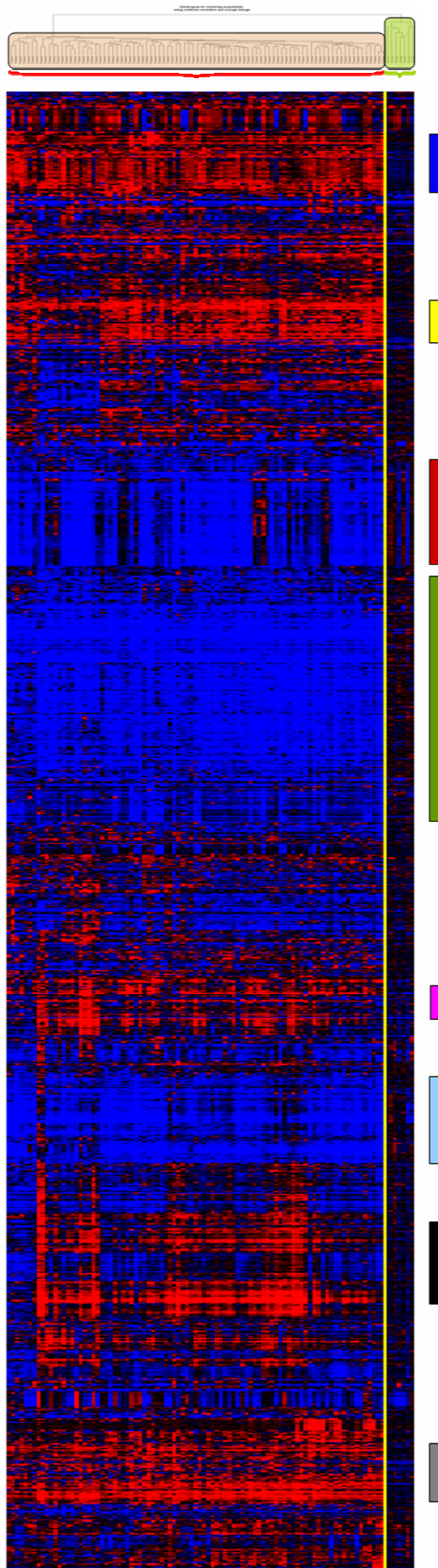
una clasificación molecular de tumores colorrectales en base a perfiles de expresión diferencial. La existencias de grandes diferencias entre tejido normal y tejido tumoral ya habían sido observadas cuando se analizaron los resultados de expresión mediante los gráficos MA-plot, en donde se observaba que la diferencia global entre ambos tejidos en cuanto al número medio de transcritos diferencialmente expresados (p -valor $< 0,01$) era de aproximadamente 5000 transcritos (Ver sección 1. de resultados. Pág.97).

Cluster	Nº Muestras	Robustez	Omisiones	Adiciones
Tejido Tumoral	88	1	0	0
Tejido Normal	7	1	0	0

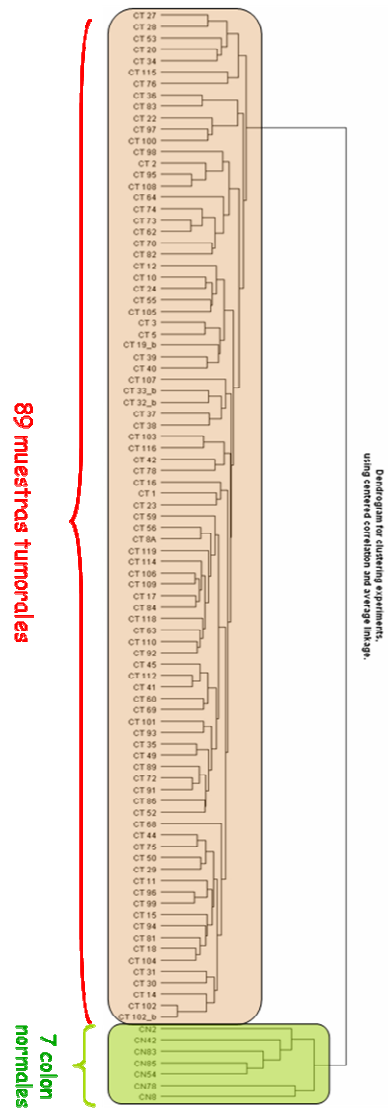
Tabla.4. Resumen del resultado de la validación matemática (tras 1000 perturbaciones) en función de los dos grupos obtenidos.

Fig.4. (En la página siguiente). Resultado del “clustering” jerárquico utilizando 1722 genes en las 88 muestras tumorales y los 7 tejidos normales. En la ejecución del mismo se utilizó como distancia la correlación centrada de Pearson y como método de medida de agrupación de elementos “average linkage”. A) Resultado del “clustering” jerárquico de las 96 muestras (horizontal). En vertical se muestra la matriz de expresión o “heatmap”, ordenada tras la realización de un “clustering” jerárquico para los 1722 genes. El color rojo en el heatmap indica sobreexpresión del gen en la muestra problema en relación a la muestra de referencia; el color azul indica inhibición; el negro ausencia de expresión diferencial. A la derecha de la matriz de expresión se muestran un grupo de barras rectangulares de diferentes colores que indican los clusters de genes encontrados. Los genes de cada uno de estos clusters, muestran en general un comportamiento diferente entre tejidos normales y tumorales y funciones similares en los procesos biológicos. Sin embargo no todos los genes de estos clusters muestran una expresión significativamente diferente y con una diferencia de expresión de 2 veces entre ambos grupos (criterio seleccionado en la comparación de clases posterior). B) Ampliación del “clustering” jerárquico de muestras.

A)



B)



3.1.1. Comparación de clases entre tejido tumoral vs. tejido normal.

Posteriormente y partiendo de los 1722 genes utilizados en el “clustering” o agrupamiento jerárquico se llevó a cabo una comparación de clases para identificar los genes que se encontraban diferencialmente expresados entre las dos clases y que, por tanto, presumiblemente podrían participar en la aparición y desarrollo del cáncer. De los 1722 genes de inicio, un total de 559 se encontraron diferencialmente expresados (p -valor < 0.001 B&H) y con una diferencia de expresión de al menos 2 veces entre ambos grupos (ver Anexo IV para lista detallada. Pág.221). Estos 559 genes se encontraban en su mayoría asociados a determinadas regiones del “heatmap” formando grupos o clusters de genes que parecían en muchos casos compartir una misma función biológica. La interpretación biológica de estos resultados se completó mediante un análisis de vías moleculares (KEGG pathways) que nos permitió comprender aún mejor las funciones alteradas en el proceso de malignización. Los resultados más importantes se muestran a continuación:

- ❖ Proliferación celular (“cluster” azul). Se observó la sobreexpresión en tumores de genes implicados en replicación del DNA como TOP2A o CDC6 [292,293], en ciclo celular como CDC2, Ciclina B1 o CDC6 [292,294] o en procesos relacionados con interacción entre los cromosomas y el huso acromático durante la división celular, así como en el proceso de citocinesis (Ej.: MAD2, MPS1, CENP-A, NUF2, BIRC5, NEK2A o CEP55) [295-299]. Todo ello en consonancia con la alteración observada de rutas moleculares (Tabla.5.) como ciclo celular (hsa04110), RNA polimerasa (transcripción) (hsa03020), así como metabolismo de purinas y pirimidinas (principalmente en la síntesis de estos nucleótidos) (hsa00230 y hsa00240).
- ❖ Desregulación de la vía Wnt (“cluster” gris). Observándose tanto la alteración de la vía (hsa04310) (Tabla.5) como y más en concreto la sobreexpresión de moléculas activadoras de dicha vía como Wnt5a, de genes diana de dicha vía como sp5, uPA, MMP7, SPP1 o Myc [300-302], de genes que conforman parte de la vía como LEF1 y ayudan a la transcripción de los genes diana, o como de genes pertenecientes a otras vías que interaccionan con Wnt como es el caso de la proteína Inhibina A de la ruta del TGF- β . La sobreexpresión de estos genes contrasta sin embargo con la sobreexpresión del gen SFRP4 que pueden inhibir la activación de esta vía.

- ❖ Activación del proceso inflamatorio (“cluster” rosa) e inhibición del sistema inmune (“cluster” azul claro y rojo). Caracterizados por la sobreexpresión de citoquinas y quimioquinas como CXCL1, CXCL2, CXCL3, CXCL12 o interleuquinas como IL11 e IL8 implicadas en angiogénesis e inflamación, así como por la alteración de la vía molecular “interacción citoquina-receptor de citoquina” (hsa04060) (Tabla.5) y caracterizados también por la inhibición de genes que codifican para inmunoglobulinas (14q32.33, 22q11.22 y 2p11.2), quimioquinas implicadas en la atracción de linfocitos como CCL13, CCL15 o CCL8, además de moléculas implicadas en la cascada del complemento como CFD, CFH o CFHR3.

- ❖ Invasividad y progresión tumoral (“cluster” gris y rosa). Mediada por la sobreexpresión de genes implicados en procesos como la degradación de la matriz extracelular (Ej. uPA [303], MMP11 [304] o MMP7 [305], MMP1 o MMP3), la inhibición de procesos de muerte celular por anoikis y la activación de procesos de transición epitelio-mesenquima (Ej. Claudina-1) [306]), la linfoangiogénesis (Ej. Podoplanina [307]) o la potenciación de los procesos de extravasación (Ej. TGFβI [308]). Además, genes como TIMP1, sobreexpresado en tumores, cuya función principal se creía relacionada con la inhibición de MMPs y por tanto con la supresión tumoral, parecen tener un papel importante en la progresión tumoral al estar también implicadas en la inhibición de la apoptosis, la activación de la proliferación o la activación de la angiogénesis [309].

- ❖ Alteración de la matriz extracelular (“cluster” negro). Se produce un aumento de los componentes de la MEC como colágenos (COL10A1, COL12A1, COL1A1, COL5A2), Nidógeno-2, Spondina-2 o SPARC, así como de genes (MMP14, THBS2, PDGFRB, SPARC, Spondina-2 o FAP) relacionados con funciones asociadas con el estroma, la remodelación de la MEC, la activación del estroma o las interacciones célula-matriz e implicados en la alteración de la vía de adhesión focal (hsa04510), importante en el desarrollo del tumor [310-315].

- ❖ Inhibición de las vías de detoxificación (“cluster” verde y azul claro). La inhibición de genes que podrían estar implicadas en el metabolismo xenobiótico (hsa00980) como ADH1C, ADH1A, EPHX2, GSTA5, NAT1, SULT1B1, UGT2B10/11/15/17

así como la inhibición de metalotioneinas (MT1B, MT1E, MT1F, MT1G, MT1H, MT1L, MT1M, MT1X o MT2A) y posiblemente la inhibición de las anhidrasas carbónicas (CA1, CA2, CA4 y CA12) estarían relacionadas con la acumulación de metales y de especies reactivas de oxígeno que podrían estar favoreciendo la aparición del tumor. Alguno de estos genes, podrían estar, no obstante, implicados también en el metabolismo de lípidos.

Finalmente estos perfiles de expresión también son capaces de reflejar las diferencias (“cluster” azul claro) en la composición tisular entre tumores y la mucosa de colon normal. En este último se observa la aparición de un componente muscular (músculo liso) asociado a la sobreexpresión de genes como MYH11, MYLK, ATP1A2, Desmina (DES) o Calponina-1 (CNN1) en contraposición al primero.

Identificador	Descripción de la vía molecular	LS p-valor	KS p-valor	E-T's GSA p-valor	Goeman's p-valor
hsa00040	Pentose and glucuronate interconversions	0.0000	0.0009	0.020	< 0.0000001
hsa00071	Fatty acid metabolism	0.0000	0.0075	0.010	< 0.0000001
hsa00120	Bile acid biosynthesis	0.0057	0.0035	0.010	< 0.0000001
hsa00150	Androgen and estrogen metabolism	0.0003	0.0055	0.090	< 0.0000001
hsa00190	Oxidative phosphorylation	0.0000	0.0000	0.140	0.0000002
hsa00230	Purine metabolism	0.0000	0.0000	0.010	< 0.0000001
hsa00240	Pyrimidine metabolism	0.0048	0.0034	0.010	< 0.0000001
hsa00362	Benzoate degradation via hydroxylation	0.0000	0.0037	0.020	< 0.0000001
hsa00500	Starch and sucrose metabolism	0.0007	0.0036	0.170	< 0.0000001
hsa00630	Glyoxylate and dicarboxylate metabolism	0.0026	0.1449	0.010	< 0.0000001
hsa00670	One carbon pool by folate	0.0007	0.0292	< 0.005	< 0.0000001
hsa00910	Nitrogen metabolism	0.0018	0.0461	< 0.005	< 0.0000001
hsa00980	Metabolism of xenobiotics by cytochrome P450	0.0084	0.0001	0.100	< 0.0000001
hsa03020	RNA polymerase	0.1971	0.0020	0.040	< 0.0000001
hsa03320	PPAR signaling pathway	0.0016	0.1983	0.120	< 0.0000001
hsa04020	Calcium signaling pathway	0.0001	0.0657	0.010	< 0.0000001
hsa04060	Cytokine-cytokine receptor interaction	0.0000	0.0000	0.180	< 0.0000001
hsa04110	Cell cycle	0.0001	0.0001	0.010	< 0.0000001
hsa04310	Wnt signaling pathway	0.0000	0.0000	0.230	< 0.0000001
hsa04510	Focal adhesion	0.0000	0.0000	0.460	0.0000043
hsa04740	Olfactory transduction	0.0000	0.0053	< 0.005	< 0.0000001
hsa04810	Regulation of actin cytoskeleton	0.0000	0.0000	0.280	< 0.0000001

Tabla.5. Análisis de vías moleculares (KEGG pathways) diferencialmente expresadas entre tejido normal y tejido tumoral. P-valor límite = 0.005. Para ser consideradas significativa, la ruta analizada tiene que ser significativa para al menos dos de los cuatro test realizados. En gris aparecen los p-valor no significativos.

3.2. Clasificación molecular de los tumores colorrectales.

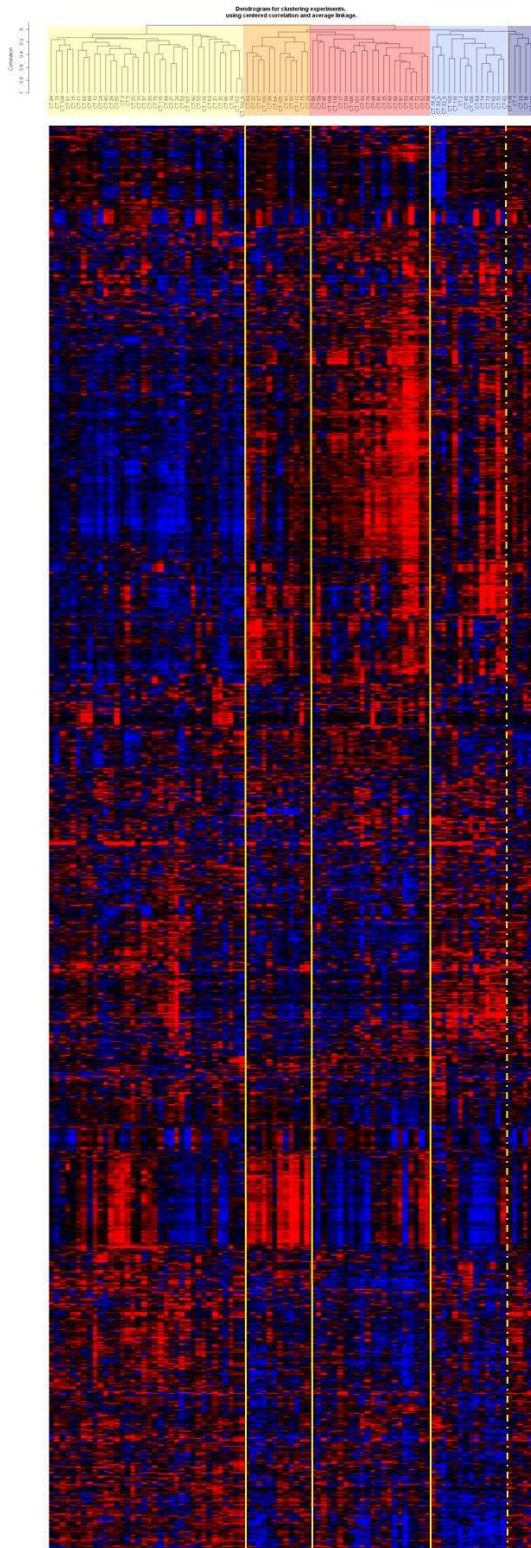
3.2.1. Clasificación de las muestras tumorales mediante métodos no supervisados.

Una vez establecido el contexto molecular que rige el desarrollo de los tumores colorrectales, nos centramos en el establecimiento de una clasificación molecular que distribuyese los tumores en grupos molecularmente homogéneos, utilizando los 1722 genes seleccionados y los 88 tumores incluidos en el análisis. Previamente al análisis se realizó el centrado de genes, de manera que el punto de referencia para determinar si un gen está o no alterado pasa a ser la mediana de los valores de expresión para ese gen a lo largo de los tumores, posibilitando una mejor discriminación de grupos similares. El proceso de clasificación se realizó mediante “clustering” jerárquico, utilizando como distancia la correlación centrada de Pearson y como método de agrupación de elementos el “average linkage”.

La exploración del dendrograma obtenido tras el análisis (Fig.5A y 5B) permitió diferenciar un total de 5 grupos o clusters en lo referente a la distribución de las muestras. El “cluster” 1 (amarillo) englobó un total de 36 muestras (incluyendo las dos hibridaciones de la muestras CT 102), el “cluster” 2 (naranja) un total de 12 muestras, el “cluster” 3 (rojo) un total de 22 muestras y el “cluster” 4 (azul claro) un total de 14 muestras. Las 5 muestras restantes formaron el denominado “cluster” 5 (malva), que debido al escaso tamaño muestral fue considerado por el momento como un grupo de muestras no clasificadas.

Del estudio de este dendrograma (Fig.5B) también se puede reseñar que no parece existir un agrupamiento de las muestras en función del estadio de Dukes al que pertenecen, si no que más bien, estos estadios se distribuyen heterogéneamente a lo largo de todo el árbol. Por otro lado, otras características como el tipo histológico tumoral (mucinoso) o la presencia de tumores con inestabilidad en microsatélites sí que parecen asociarse a esta clasificación molecular al agruparse mayoritariamente en el denominado grupo 4.

A)



B)

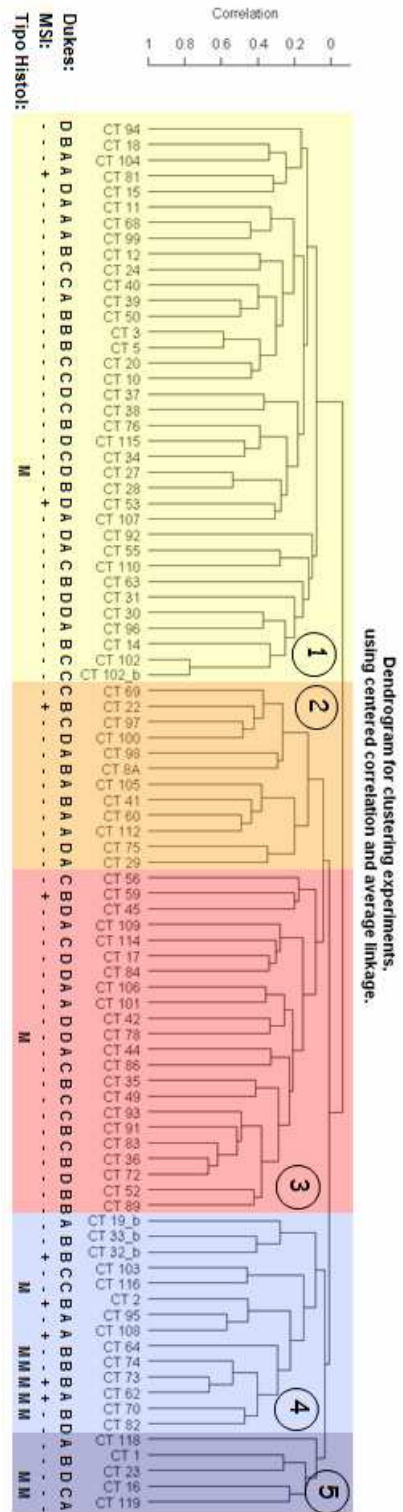


Fig.5. Pie de figura en la siguiente página.

Fig.5. Resultado del “clustering” jerárquico utilizando 1722 genes en las 88 muestras tumorales. En la ejecución del mismo se utilizó el centrado de genes, como distancia la correlación centrada de Pearson y como método de agrupación de elementos el “average linkage”. A) Resultado del “clustering” jerárquico de las 88 muestras (horizontal). En vertical se muestra la matriz de expresión o “heatmap”, ordenada tras la realización de un “clustering” jerárquico para los 1722 genes. El color rojo en el heatmap indica sobreexpresión del gen en la muestra en relación con la mediana de expresión de los tumores; el color azul indica inhibición; el negro ausencia de expresión diferencial. A la derecha de la matriz de expresión se muestran un grupo de barras rectangulares de diferentes colores que indican clusters de genes. Los genes de cada uno de estos clusters, muestran un comportamiento diferente entre los diferentes grupos moleculares definidos. Los genes que componen cada uno de los clusters estarían teóricamente implicados en funciones biológicas similares. B) Ampliación del “clustering” jerárquico de muestras. En este “clustering” jerárquico se describen 4 grupos (“cluster” 1 o amarillo; “cluster” 2 o naranja; “cluster” 3 o rojo; “cluster” 4 o azul claro). El “cluster” malva o “cluster” 5 engloba las muestras no clasificables. Debajo de cada muestra se describe alguna de sus características más importantes como el estadio de Dukes, el estatus MSI o el tipo histológico al que pertenecen. El símbolo “+” indica tumor MSI. La letra “M” indica tumor mucinoso.

Los resultados obtenidos de la validación matemática, tras la perturbación de la matriz (1000 veces), reflejaron que las agrupaciones propuestas no eran debidas al azar. Como se puede comprobar en la Tabla.6, los cuatro grupos moleculares definidos poseen una reproducibilidad en cuanto a la forma (“R-Index”) superior al 75% (en una escala de 0-100%), siendo además para los grupos 1, 3 y 4 cercana al 90%. En cuanto a la estabilidad en las muestras (“D-Index”) que conformaron cada uno de los 4 grupos, se observó que en ninguno de los casos, el número promedio de muestras que tras ser asignadas a un determinado grupo cambió de agrupación al alterar la matriz (omisiones) superó las 2.5 muestras por grupo. En cuanto al número promedio de muestras que tras ser incluidas en cualquiera de los otros tres grupos fueron incluidas en el grupo de referencia (adiciones) al alterar la matriz, fue incluso más bajo para los grupos 1, 3 y 4 ya que no superaron las 2.3 muestras por grupo (destacando una puntuación de 0.58 para el grupo 4), siendo sin embargo un poco más elevada para el grupo 2 (4.98). Este último dato reflejaría el hecho que este grupo 2 parecía tener unos criterios de selección peor definidos que el resto.

El grupo 5, obtuvo unos valores de “D-Index” más elevados (Om: 1.29 y Ad: 1.56) y de “R-Index” (65%) sensiblemente más bajos que los demás, teniendo en cuenta el número de muestras que lo componían. La falta de criterio para poder decidir si estos datos se debieron a una falta de tamaño muestral o a una falta de homogeneidad

biológica aconsejó como ya señalamos anteriormente que estas muestras fueran excluidas en los análisis posteriores.

Cluster	# Muestras	Robustez	Omisiones	Adiciones	Reproducibilidad Global	
Cluster 1	36	0.90	2.54	1.78	R-Index	0.888
Cluster 2	12	0.75	2.24	4.98	D-Index	9.838
Cluster 3	22	0.87	1.78	2.34		
Cluster 4	14	0.89	0.92	0.58		
Cluster 5	5	0.65	1.29	1.56		

Tabla.6. Resultado de la validación matemática (tras 1000 perturbaciones) para cada uno de los grupos moleculares obtenidos. A la derecha se muestra la reproducibilidad global de todos los grupos analizados.

Conscientes del efecto que tiene el número de genes seleccionados, la elección del algoritmo de clasificación, así como las medidas utilizadas en la obtención de los resultados de la clasificación, se comprobó la consistencia de dichos resultados, analizando de nuevo las mismas muestras pero utilizando diferentes combinaciones de los parámetros mencionados.

Los resultados obtenidos se muestran en la Tabla.7 y Fig.6. Se llevaron a cabo un total de 18 combinaciones diferentes de la seleccionada inicialmente, en las que se analizó la distribución de las muestras. Para 12 de ellas se encontró una concordancia en los grupos, en relación a la clasificación original, superior al 80% (alcanzando en 8 de ellos valores en torno al 90% o superiores). Una de las combinaciones (la nº 13) obtuvo un 71% de concordancia. Cuatro combinaciones (nº 5, 7, 8 y 17) mostraron una reproducibilidad parcial que impedía definir de manera fiable los cuatro grupos, coincidiendo con la elección de parámetros poco adecuados para la realización de clasificaciones (ver discusión). En cuanto a la representación gráfica del Análisis de Componentes Principales (Fig.6), se observó que la variabilidad representada por las tres primeras componentes (28% de variabilidad) era suficiente para diferenciar de manera consistente los grupos moleculares previamente propuestos. En general, la reproducibilidad de los grupos fue buena, indicando de nuevo la consistencia de la clasificación propuesta.

	Centrado de genes	Método de Agrupación	Distancia	Métodos de medir distancias	Repeticiones	Filtrado de genes	Nº genes	Grupo 1	Grupo 2	Grupo 3	Grupo 4	% Total de concordancia
Clasificador Seleccionado	Si	Clustering Jerárquico	Correlación centrada de Pearson	Average Linkage	–	2.5-fold en al menos 8% de las muestras	1722g	36/36	12/12	22/22	14/14	100%
Clasificador 1	Si	<i>Clustering Jerárquico Eisen</i>	Correlación centrada de Pearson	Average Linkage	–	2.5-fold en al menos 8% de las muestras	1722g	35/36	12/12	22/22	14/14	99%
Clasificador 2	Si	Clustering Jerárquico Eisen	<i>Spearman</i>	Average Linkage	–	2.5-fold en al menos 8% de las muestras	1722g	36/36	7/12	19/22	10/14	86%
Clasificador 3	Si	Clustering Jerárquico Eisen	<i>Kendal Tau</i>	Average Linkage	–	2.5-fold en al menos 8% de las muestras	1722g	36/36	7/12	18/22	10/14	85%
Clasificador 4	Si	Clustering Jerárquico Eisen	<i>Correlación no centrada de Pearson</i>	Average Linkage	–	2.5-fold en al menos 8% de las muestras	1722g	34/36	6/12	19/22	11/14	83%
Clasificador 5	Si	Clustering Jerárquico	<i>Euclidea</i>	Average Linkage	–	2.5-fold en al menos 8% de las muestras	1722g	No se pueden distinguir los 4 grupos				No disponible
Clasificador 6	Si	Clustering Jerárquico Eisen	Correlación centrada de Pearson	<i>Centroid Linkage</i>	–	2.5-fold en al menos 8% de las muestras	1722g	32/36	6/12	18/22	11/14	80%
Clasificador 7	Si	Clustering Jerárquico Eisen	Correlación centrada de Pearson	<i>Single Linkage</i>	–	2.5-fold en al menos 8% de las muestras	1722g	No se pueden distinguir los 4 grupos				No disponible
Clasificador 8	Si	Clustering Jerárquico Eisen	Correlación centrada de Pearson	<i>Complete Linkage</i>	–	2.5-fold en al menos 8% de las muestras	1722g	Se distingue el grupo 1 del resto				No disponible
Clasificador 9	Si	<i>K-Means Eisen</i>	Correlación centrada de Pearson	Average Linkage	200	2.5-fold en al menos 8% de las muestras	1722g	32/36	12/12	21/22	13/14	93%
Clasificador 10	Si	<i>K-Medians Eisen</i>	Correlación centrada de Pearson	Average Linkage	200	2.5-fold en al menos 8% de las muestras	1722g	31/36	12/12	19/22	13/14	89%
Clasificador 11	Si	Clustering Jerárquico	Correlación centrada de Pearson	Average Linkage	–	<i>2-fold en al menos 8% de las muestras</i>	3943g	36/36	11/12	21/22	6/14	88%
Clasificador 12	Si	Clustering Jerárquico	Correlación centrada de Pearson	Average Linkage	–	<i>3-fold en al menos 8% de las muestras</i>	995g	35/36	6/12	21/22	13/14	89%
Clasificador 13	Si	Clustering Jerárquico	Correlación centrada de Pearson	Average Linkage	–	<i>3.5-fold en al menos 8% de las muestras</i>	649g	24/36	5/12	18/22	13/14	71%
Clasificador 14	Si	Clustering Jerárquico	Correlación centrada de Pearson	Average Linkage	–	<i>2.5-fold en al menos 5% de las muestras</i>	2737g	36/36	12/12	18/22	11/14	92%
Clasificador 15	Si	Clustering Jerárquico	Correlación centrada de Pearson	Average Linkage	–	<i>2.5-fold en al menos 10% de las muestras</i>	1327g	35/36	6/12	21/22	13/14	89%
Clasificador 16	Si	Clustering Jerárquico	Correlación centrada de Pearson	Average Linkage	–	<i>2.5-fold en al menos 12% de las muestras</i>	1049g	35/36	9/12	18/22	13/14	89%
Clasificador 17	No	Clustering Jerárquico	Correlación centrada de Pearson	Average Linkage	–	2.5-fold en al menos 8% de las muestras	1722g	No se pueden distinguir los 4 grupos				No disponible

Tabla.7. (página anterior) Tabla resumen de las diferentes combinaciones realizadas para comprobar la estabilidad de nuestras agrupaciones. Cada fila muestra cada una de las combinaciones realizadas, especificándose los parámetros utilizados para su realización, el número de genes seleccionados y la concordancia con los grupos de la clasificación original (tanto en número como finalmente en porcentaje).

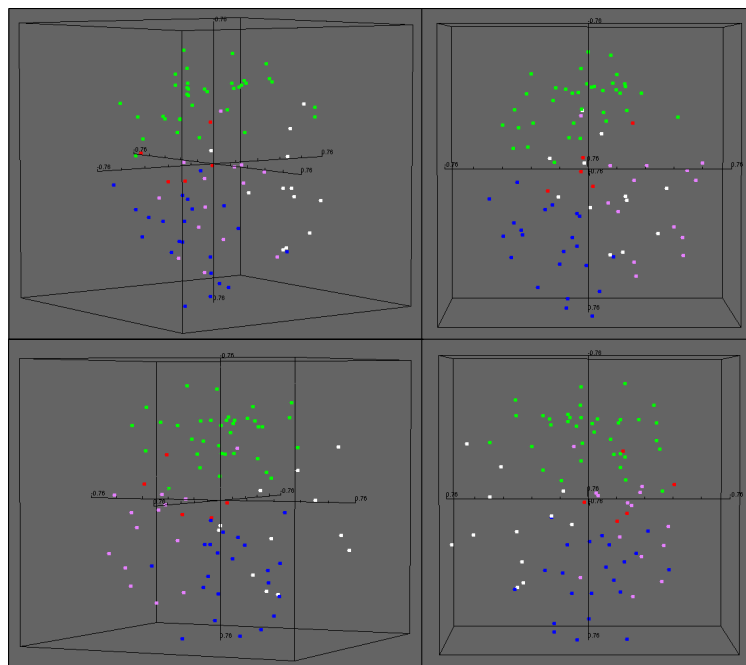


Fig.6. Representación gráfica de los tumores en función de la variabilidad representada por las tres primeras componentes. Muestras del “cluster” 1 en verde; muestras “cluster” 2 en blanco; muestras del “cluster” 3 en azul marino; muestras del “cluster” 4 en rosa; muestras del “cluster” 5 en rojo.

3.2.2. Comparación de subtipos y análisis funcional.

La caracterización de los genes más representativos y significativos de los grupos moleculares propuestos es complicada debido a la necesidad de comparar la expresión de dichos genes entre los cuatro grupos a la vez. En este caso, se comparó cada uno de los grupos contra el resto de muestras, seleccionándose aquellos genes con una diferencia estadísticamente significativa a un nivel p -valor $< 0,001$ utilizando además el método de corrección para comparaciones múltiples de B&H. El número de genes iniciales introducidos en la comparación fue de 1722 genes. Un total de 397 genes se encontraban diferencialmente expresados en el grupo 1 con respecto al resto de tumores, 101 en el caso del grupo 2, 300 genes en el caso del grupo 3 y 85 en el caso del grupo 4. A continuación se estudió cuál era la distribución en el “clustering” jerárquico de estos genes (ver Anexo IV. Pág 221 para un listado detallado y Fig.5) y en función

de ellos se analizaron los clusters o grupos de genes más destacados. Las regiones marcadas en la figura indican las zonas donde se agrupaban preferentemente los genes diferencialmente expresados, aunque hay que reseñar que en dichas áreas también se encontraban genes cuya expresión era igual a lo largo de todos los tumores.

Esta estrategia, aunque en algún caso puede resultar poco selectiva en cuanto a la discriminación de genes, permite establecer un retrato aproximado de las diferencias existentes entre los grupos moleculares propuestos. En cada “cluster” de genes se intentó encontrar una o varias funciones que permitiesen entender porque dichos elementos se agrupaban conjuntamente, tarea facilitada gracias a la realización de análisis de vías moleculares (KEGG pathways) que nos permitió entender mejor el marco biológico que caracterizaba estos subtipos. Los resultados más significativos fueron los que se exponen a continuación:

❖ “cluster” A (o verde). Genes del 270 a la 523.

Este “cluster” engloba genes relacionados principalmente con el estroma. La expresión de los genes que se describen a continuación está más elevada en las muestras tumorales que en las muestras normales señalando su importancia en el desarrollo tumoral y los niveles de expresión entre los subtipos tumorales muestras niveles elevados en el grupo 3 e inhibidos en el grupo 1 en comparación con el resto de tumores. Entre los más destacados, encontramos genes que parecen relacionados con los procesos de angiogénesis y una mayor densidad de microvasos como son ANTXR1, PDGFC o S100A4 [316-318], proliferación tumoral como GREM1 [319], asociadas con activación de fibroblastos como es el caso de FAP [120], procesos de adhesión y migración celular como fibronectina 1 [320] (<http://www.ncbi.nlm.nih.gov/gene/2335>) o remodelación de la matriz como MMP2, MMP9, MMP11, MMP14 y uPA [302,321]. Se ha descrito que la activación de fibroblastos, provoca además una mayor secreción de proteínas de MEC (matriz extracelular) como SPARC, Fibronectina 1 o Tenascina C [120] que también se encuentran alteradas y englobadas en este “cluster”. Finalmente, también encontramos sobreexpresados diferentes tipos de colágenos (COL10A1/ 12A1/ 1A1/ 1A2/ 3A1/ 4A2/ 5A2/ 6A1/ 6A2/ 8A1/ 8A2) y otras proteínas de MEC como lumican [322]. Estos datos concuerdan con la alteración de vías implicadas en

comunicación celular (hsa01430), adhesión focal (hsa04510) o interacción con receptores de la matriz extracelular (hsa04512) que se encontraron durante el análisis funcional de vías moleculares (Tabla.8).

Otra muestra de la importancia de estos genes en el desarrollo del tumor es que muchos de ellos como S100A4, FAP, Lumican, Periostina o SPP1 están relacionadas con invasión, metástasis y mal pronóstico [315,322-325].

Por otra parte, se ha descrito que proteínas como Fibulina 1, DACT3, PTGIS se encuentran inhibidas en tumores colorrectales con respecto al tejido normal y que esta inhibición podría estar favoreciendo la progresión tumoral. Sorprendentemente, teniendo en cuenta lo observado hasta este momento, la expresión de estos genes en el “cluster” 3 muestra niveles ligeramente superiores al de los tejidos normales. El resto de los grupos muestra inhibición con respecto al tejido normal siendo la más marcada la del grupo 1.

Por último, cabe destacar la presencia en este “cluster” de los genes THBS2 y TIMP1, no tanto porque se encuentren elevados en el grupo 3 e inhibidos en el grupo 1, si no porque en todos los grupos tumorales se encuentran estimuladas con respecto al tejido normal, a pesar de su conocida implicación en procesos de supresión tumoral en el caso de THBS2 [326] y debido a su más que probable implicación en la progresión tumoral a pesar de su teórico papel de inhibidor de MMP9, en el caso de TIMP1 [327].

❖ “cluster” B (o rojo). Genes del 524 al 664.

En este grupo de genes la situación frente al tejido normal varía dependiendo del gen en cuestión. Una parte importante de los genes englobados en este “cluster” están implicados en respuesta inmune e inflamación. Entre los genes implicados en respuesta inmune tenemos la granulicina (GNLY) y la granzima H (GZMH) que son liberados por células del sistema inmune como los linfocitos; genes que codifican para las cadenas del complejo mayor de histocompatibilidad tipo II (HLA-DMB, -DPB1, -DPB2, -DRB4, -DRB5); receptores de membrana que se unen a moléculas presentadas por el complejo MHC I como LILRB3. También encontramos genes que

codifican para receptores de inmunoglobulinas como FCGR1B o FCGR3A. En todo este grupo de genes se mantiene siempre una constante en cuanto a los niveles de expresión génica: aunque estadísticamente no son grupos con diferencias de expresión muy elevadas, las muestras del grupo 1 tienen claramente los niveles más bajos de los cuatro grupos, mientras que el grupo 2 tiene los niveles más altos (siendo estos iguales o superiores a los del tejido normal). Los grupos 3 y 4 tienen niveles más cercanos al grupo 2 que al grupo 1. Coincidiendo con esta aparente alteración del sistema inmune durante el desarrollo tumoral, el análisis de vías moleculares mostraba la desregulación de vías relacionadas como las líneas celulares hematopoyéticas (hsa04640), el procesamiento y presentación de antígenos (hsa04612) o las moléculas de adhesión celular (hsa04514) (Tabla.8).

También se observan la alteración de la vía molecular “citoquinas e interacción con sus receptores” (hsa04050) (Tabla.8) a la vez que la alteración de un gran número de genes implicados en respuesta inflamatoria entre las que destacan quimioquinas y citoquinas como CXCL10, IL-8, IL-1 β , CCL18, CCL2, CCL3, CCL4, CCL8, CX3CL1. En este caso el patrón de expresión sigue indicando una menor expresión de las muestras pertenecientes al “cluster” 1 en contraposición a las muestras del grupo 2 aunque en esta ocasión las muestras del grupo 4 muestran unos niveles similares y en algunos casos superiores a los del grupo 2. En comparación con el tejido normal, la expresión de los grupos 2 y 4 es casi siempre más elevada.

Otro grupo de genes que cabe destacar es el grupo de las metalotioneinas como MT1B, MT1E, MT1G, MT1H, MT1L, MT1M, MT1X, MT2A todas ellas codificadas en la región cromosómica 16q13 y que podrían estar implicadas por un lado en la detoxificación de elementos nocivos y en la resistencia de los tumores inestables (MSI) a QT [245]. En este caso la expresión elevada se da sobre todo en el grupo 4 y de nuevo el grupo 1 vuelve a mostrar el menor nivel expresión de los grupos de tumores. Con respecto al tejido normal, todos los grupos tumorales se encuentran inhibidos.

Finalmente aparecen genes con importancia en cáncer, inhibidos en el grupo 1, cuya expresión ha sido asociada a buen pronóstico como ICAM1 que

impide la diseminación de células tumorales [328] u Oncostatina (elevada sobre todo en el grupo 4) [329].

❖ “cluster” C (o azul oscuro). Genes del 1037 al 1102.

Este “cluster” de genes engloba, en general, genes estimulados en el grupo 4 frente al resto de grupos. Estos niveles de grupo 4 son similares a los encontrados en tejido de colon normal. La mayoría de los genes codifican en este caso para proteínas relacionadas con el componente mucinoso. Ya sea mediante su función, como en el caso de AGR2 que han sido relacionados con procesos de desregulación de las células caliciformes o goblets cells (secretoras de moco) [330,331], ya sea porque el gen en cuestión comparte una estructura típica de mucinas como FCGBP [332], asociado también con células caliciformes [333]. Genes como el TFF2 y principalmente el TFF1 han sido relacionados con procesos relacionados con el componente mucinoso del colon además de haber sido asociados con tumores tanto de tipo histológico serrado [334] como mucinoso [335]. También se engloban en este “cluster” genes relacionados con el ácido siálico (ST3GAL4, ST6GALNAC1) que ha sido asociado con progresión, metástasis y mal pronóstico [336,337] y con interacción con el gen MUC2 para producir apoptosis de células dendríticas [338]. También descrito como gen de mal pronóstico en este grupo destaca el gen REG4, que se asocia con metástasis [339] y en cáncer gástrico con resistencia a 5-FU [340].

No asociado a este “cluster” de genes, pero sí diferencialmente expresados en el grupo 4 destacan otra serie de genes. Un claro ejemplo es el del gen MUC1, característico de tumores mucinosos [341], cuyos niveles altos se han visto asociados con mal pronóstico [342,343]. Se ha descrito que este gen puede unirse a β -catenina evitando así la unión de este último a E-Cadherina y favoreciendo la transcripción de genes [344] o correlacionándolo con una mayor velocidad de desarrollo tumoral [345]. También destacan genes implicados en la respuesta inmune y respuesta inflamatoria como IL1 β o IL1R2 o genes implicados en la cascada del complemento como CD55 que parece que puede evitar precisamente la muerte de la célula en la que se expresa cuando es activada esta cascada del complemento, etc. Otros genes son los relacionados con Serpina (SERPINA1, SERPINB5), que en general parecen correlacionar con una mayor sobreexpresión en tumores MSI [346]. Todos estos

genes están inhibidos en todos los tumores a excepción de los del grupo 4 que muestran unos niveles similares, o incluso superiores en algunos casos, a los del tejido normal.

❖ “cluster” D (o amarillo). Genes del 1247 al 1352.

Engloba genes relacionados con respuesta inmune y más en concreto con las inmunoglobulinas. Así, se observa un aumento en el grupo 2 frente al resto de los grupos de la expresión de las regiones cromosómicas 14q32, 16p11, y 15q11, correspondientes a genes que codifican para las cadenas pesadas de las inmunoglobulinas y de las regiones 22q11, 2p11, correspondientes a genes que codifican para las cadenas ligeras (Lambda y kappa respectivamente) de las inmunoglobulinas. Entre los genes destacan el IGHA1, IGKC, IGKV1D-8, IGHV1-69 o IGLL1. En concordancia con esto, también se encuentran expresados genes como POU2AF1 que parece favorecer la transcripción de genes de las inmunoglobulinas y la recombinación de las regiones V (D) J de algunas inmunoglobulinas [347], además de haberse relacionado su inhibición con fallos en la respuesta inmune [348]. También destaca la expresión de receptores del TNF (TNFRSF17) importante en respuesta inmune.

Genes asociados con progresión tumoral como laminina (LAMC2) [310,349] están disminuidos en este grupo 2 frente al resto de tumores. Cabe destacar que la mayoría de los genes reseñados en este “cluster” tienen, en el grupo 2, niveles de expresión similares a los de los tejidos normales. Este “cluster” de genes muestra de nuevo, como ya se viese en el “cluster” rojo, la asociación de los tumores del grupo 2 con vías moleculares relacionadas con el sistema inmune.

❖ “cluster” E (o azul claro). Del 1428 al 1528.

Engloba a un grupo de genes cuyo nivel de expresión está elevado en las muestras del grupo 1 con respecto al conjunto de los otros tres grupos y principalmente con respecto a las muestras de los grupos 2 y 3. En relación con el tejido colorrectal normal, la expresión de estos genes en los tumores se encuentra inhibida, exceptuando en las muestras del grupo 1 que muestran niveles similares a los del

tejido normal. En este caso destacan genes que han sido implicados en procesos del metabolismo xenobiótico o del metabolismo de lípidos como la UDP glucuronosiltransferasa 2 (UGT2B7 y UGT2B28) o la apolipoproteína B (APOBEC1). Coincidiendo con esto, también se encuentran estimulados en el grupo 1, aunque no de manera diferencial, genes que codifican para el citocromo P450 (CYP3A4, CYP3A5 y CYP3A7). Por otra parte, este “cluster” engloba además genes relacionados con adhesión celular e interacción célula-célula como Cadherina 17 (CDH17) o nectina 3 (PVRL3).

Kegg Pathway	Pathway description	LS permutation p-value	KS permutation p-value	Efron-Tibshirani's GSA test p-value	Goeman's global test p-value
hsa00190	Oxidative phosphorylation	0.00001	0.00001	0.14	< 0.0000001
hsa00230	Purine metabolism	0.00001	0.00001	0.38	< 0.0000001
hsa01430	Cell Communication	0.00001	0.0023208	< 0.005	< 0.0000001
hsa04060	Cytokine-cytokine receptor interaction	0.00001	0.00001	0.15	< 0.0000001
hsa04310	Wnt signaling pathway	0.00001	0.00001	0.51	< 0.0000001
hsa04360	Axon guidance	0.00001	0.00001	0.01	< 0.0000001
hsa04510	Focal adhesion	0.00001	0.00001	0.01	< 0.0000001
hsa04512	ECM-receptor interaction	0.00001	0.00001	< 0.005	< 0.0000001
hsa04514	Cell adhesion molecules (CAMs)	0.00001	0.00001	0.15	< 0.0000001
hsa04610	Complement and coagulation cascades	0.00001	0.0046577	0.01	< 0.0000001
hsa04670	Leukocyte transendothelial migration	0.00001	0.0012397	0.07	< 0.0000001
hsa04810	Regulation of actin cytoskeleton	0.00001	0.00001	0.05	< 0.0000001
hsa00532	Chondroitin sulfate biosynthesis	0.0001609	0.0699698	0.04	< 0.0000001
hsa04350	TGF-beta signaling pathway	0.0001835	0.0629803	0.01	< 0.0000001
hsa04940	Type I diabetes mellitus	0.0006505	0.0003245	0.3	0.0000033
hsa05130	NA	0.0032635	0.1833333	0.12	< 0.0000001
hsa05131	NA	0.0032635	0.1833333	0.12	< 0.0000001
hsa04640	Hematopoietic cell lineage	0.0039121	0.0003785	0.26	< 0.0000001
hsa04612	Antigen processing and presentation	0.0046006	0.0071049	0.39	0.0000083
hsa00530	Aminosugars metabolism	0.063982	0.0031385	0.06	< 0.0000001

Tabla.8. Resultado del análisis de vías moleculares en el que se muestran aquellas diferencialmente expresadas entre los grupos moleculares propuestos (con el correspondiente nivel de significación para los test estadísticos realizados).

Por último, es interesante resaltar la aparición en el análisis de vías moleculares de alteraciones en la vía Wnt (hsa04310) y en la vía del TGF β (hsa04350) dos de las rutas más importantes implicadas cáncer y que no se agrupan de manera específica a un “cluster” de genes en concreto.

En el caso de la vía Wnt, tumores del grupo 3 mostraban sobreexpresión con respecto al resto de tumores de genes diana de la vía como SPP1, MMP14 o S100A4, pero también de genes inhibidores de la vía como DACT3, SFRP2 y SFRP4. También

se observaba una sobreexpresión en los tumores del grupo 3 en contraposición con los del grupo 1 para genes implicados en la vía del TGF β entre los que se encontraban genes como INHBA, HTRA1, THBS2, COMP o Decorina.

A la vista de los genes que caracterizaban los cuatro subtipos, el grupo 1 fue denominado como subtipo “de bajo estroma”, el grupo 2 como subtipo “relacionado con respuesta inmune”, el grupo 3 como subtipo “de alto estroma” y el grupo 4 como subtipo “MSI-Mucinoso”

3.2.3. Estudio de asociación entre la clasificación molecular y los parámetros clínicos y patológicos.

Se llevaron a cabo estudios de asociación entre los grupos moleculares propuestos y los datos clínicos y patológicos que fueron recogidos y analizados a lo largo de este estudio. Los primeros análisis mostraron que no existía asociación de los grupos moleculares con parámetros tales como la edad, el sexo de los pacientes o la calidad del RNA hibridado (RIN).

El estudio de correlación entre la clasificación propuesta en este trabajo y los grupos histológicos de la clasificación de Dukes (Tabla.9.) revelaron que no había asociación entre ambas variables (p-valor: 0.646), confirmando lo que ya se había podido observar durante la realización del “clustering” jerárquico, en el que los estadios de Dukes mostraban una distribución heterogénea a lo largo de las agrupaciones de la clasificación molecular propuesta (Fig.5B).

Por el contrario, parámetros como el tipo histológico y la cantidad de estroma tumoral si mostraron asociación con la clasificación propuesta (p-valor = 0.001 y p-valor < 0.001 respectivamente). En el primer caso se observó que los tumores de tipo mucinoso se agrupaban mayoritariamente en el denominado grupo 4 o “MSI-Mucinoso”. En el caso del estroma tumoral, se observó que las muestras del grupo 1 o “de bajo estroma” contenían de manera estadísticamente significativa una menor proporción de contenido estromal que el resto de grupos moleculares (Tabla.9). Las muestras del grupo 3 o “de alto estroma” contenían una mayor proporción de estroma tumoral que las muestras de los grupos 2 (“relacionadas con el sistema inmune”) y 4

("MSI-Mucinoso"), aunque estas diferencias no llegaron a ser significativas. Cabe recordar que la cantidad de estroma de las piezas tumorales analizadas era representativa de la cantidad de estroma del tumor completo de cada paciente (ver resultados sección 1. Pág.94). Como cabía esperar, la cantidad de estroma de las muestras parafinadas para cada grupo mostró unos valores similares a los observados en el tejido congelado, siendo las muestras del grupo 1 las que menor porcentaje de estroma contenían y las del grupo 3 las que mayor, aunque en este caso con porcentajes muy similares a los grupos 2 y 4. En el caso del tejido parafinado también se encontró asociación entre el estroma y la clasificación molecular (p-valor: 0.047), aunque solamente existieron diferencias significativas entre el porcentaje de estroma del grupo 3 y del grupo 1 (p-valor: 0.005).

Otros dos parámetros, localización tumoral e invasión vascular, aunque no mostraron asociación con los nuevos subgrupos, si mostraron un nivel de significancia estadística (p-valor: 0.103 y p-valor: 0.168 respectivamente) relativamente cercano a lo que podemos llamar tendencia estadística ($0.05 < p\text{-valor} < 0.1$). En el caso de la localización tumoral se observó una mayor tendencia de los grupos 2 y 3 a correlacionar con tumores distales, mientras que el grupo 4 correlacionaba principalmente con tumores proximales y el grupo 1 no mostró diferencias en cuanto a la localización. En el caso de la invasión vascular se observó una asociación del grupo 4 con una menor invasión vascular mientras que el grupo 1 se asociaba con mayores niveles de invasión de vasos.

El resto de variable histopatológicas, que habían sido previamente descritas en diferentes estudios como asociadas con progresión tumoral y con mal pronóstico, no mostraron asociación alguna con la clasificación molecular propuesta (Tabla.9).

Variable Independiente		Variable de clasificación				p-global	Comparaciones por pares
		Grupo 1 (35)	Grupo 2 (12)	Grupo 3 (22)	Grupo 4 (14)		
Dukes	Estadio A	9	5	4	4	0.646 ^{LR}	No significativas
	Estadio B	9	3	6	7		
	Estadio C	8	2	6	2		
	Estadio D	9	2	6	1		
Localización	C. Derecho	18	3	7	9	0.103 ^x	No significativas
	C. Izquierdo	17	9	15	5		
Grado de diferenciación	Alto grado	1	2	2	1	0.462 ^{LR}	No significativas
	Bajo grado	34	10	20	13		
Margen Tumoral	Infiltrante	27	9	15	7	0.325 ^{LR}	No significativas
	Mixto	4	1	4	1		
	Expasivo	4	2	3	6		
Infiltrado Linfocitario	Inexistente	24	6	12	7	0.452 ^{LR}	No significativas
	Bajo	3	3	6	2		
	Medio	6	3	3	5		
	Alto	2	0	1	0		
Inv. Vascular	Si	17	5	8	2	0.168 ^x	No significativas
	No	18	7	14	12		
Invasión Perineural	Si	3	1	1	1	0.944 ^{LR}	No significativas
	No	32	11	21	13		
Extensión Tumoral (T)	T1	1	0	0	0	0.356 ^{LR}	No significativas
	T2	8	5	5	5		
	T3	21	6	17	7		
	T4	5	1	0	2		
Inv. Ganglionar (N)	N0	21	8	12	11	0.458 ^{LR}	No significativas
	N1	8	2	6	3		
	N2	6	2	4	0		
Metástasis (M)	M0	26	10	16	13	0.376 ^{LR}	No significativas
	M1	9	2	6	1		
Tipo Histológico	Convencional	34	12	21	8	0.001 ^{LR}	G4 ≠ G1 (0.001) G2 ≠ G4 (0.017) G4 ≠ G3 (0.008)
	Mucinoso	1	0	1	6		
Porcent. Estroma Congelado	Media	11.2	18.8	23.4	18.9	< 0.001 ^{KW}	G1 ≠ G4 (0.013) G1 ≠ G2 (0.020) G1 ≠ G3 (< 0.001) 3vs4 tendencia (0.088)
	Desv Stnd	6.0	10.3	10.5	14.8		
Porcent. Estroma Parafina	Media	12.6	17.1	18.9	16.8	0.047 ^{KW}	G1 ≠ G3 (0.005)
	Desv Stnd	6.0	9.9	8.6	14.2		
Sexo	Hombre	20	5	8	9	0.282 ^x	No significativas
	Mujer	15	7	14	5		
Edad	Media	70.5	72.3	72.6	71.0	0.865 ^{KW}	No significativas
	Desv Stnd	13.3	10.5	9.6	11.6		
RIN	Media	9.0	8.7	8.9	9.1	0.438 ^{KW}	No significativas
	Desv Stnd	0.6	0.8	0.7	0.8		

Tabla.9. Resumen de los estudios de asociación entre la clasificación molecular y los distintos parámetros clínicos y patológicos disponibles. El número de muestras totales analizadas es de 83. χ : test Chi-Cuadro; LR: test "Likelihood Ratio"; KW: test de Kruskal Wallis. Grupo 1: bajo estroma; Grupo2: relacionado con respuesta inmune; Grupo 3: alto estroma; Grupo 4: MSI-Mucinoso.

En cuanto a los marcadores moleculares, dos fueron los analizados en este trabajo: la inestabilidad en microsatélites (MSI) y la presencia de mutaciones en el gen KRAS (Tabla.10). La presencia de inestabilidad en microsatélites, una de las vías alternativas de desarrollo del cáncer colorrectal, se correlacionó con la nueva

clasificación (p-valor: 0.039), agrupándose las muestras MSI de manera mayoritaria en el grupo 4.

En el caso de las mutaciones en el gen KRAS no se encontró asociación con los subtipos propuestos, aunque el grupo 1 mostraba mayor proporción que el grupo 3.

Variable Independiente		Variable de clasificación				p-global	Comparaciones por pares
		Grupo 1 (35)	Grupo 2 (12)	Grupo 3 (22)	Grupo 4 (14)		
Inestabilidad	MSS	33	11	21	9	0.039 ^{LR}	G4 ≠ G1 (0.015) G4 ≠ G3 (0.024)
	MSI	2	1	1	5		
Mutación K-Ras	Si	17	5	5	4	0.217 ^x	No significativas
	No	18	7	17	10		

Tabla.10. Resumen de los estudios de asociación entre la clasificación molecular y los marcadores moleculares disponibles. El número de muestras totales analizadas es de 83. x : test Chi-Cuadro; LR: test "Likelihood Ratio".

3.2.4. Análisis inmunohistoquímico (Ki-67, M-30, β-catenina) y por FISH (EGFR) y correlación con la clasificación molecular.

Cuatro fueron los parámetros estudiados por inmunohistoquímica o por FISH: proliferación, apoptosis, activación de la vía Wnt y amplificación del receptor EGFR (Tabla.11).

El análisis inmunohistoquímico para los procesos de proliferación (Ki-67) y de apoptosis (M-30) no registró asociación con los grupos moleculares (p-valor = 0.341 y p-valor = 0.825 respectivamente). Por el contrario, la expresión de β-catenina nuclear (Fig.7), uno de los principales síntomas de activación de la vía canónica de Wnt y cuya desregulación tiene gran importancia en el desarrollo de tumores colorrectales, correlacionaba con la clasificación propuesta (p-valor = 0.021). Prácticamente no se encontró expresión nuclear de esta proteína en los grupos 2 y 4 (<20% de las muestras), mientras que en los grupos 1 y 3 la expresión de β-catenina nuclear superaba el 40% de las muestras. Existía, por tanto, asociación significativa entre el parámetro y la clasificación, siendo estadísticamente diferente entre los grupos 2 y 4 frente al grupo 3 y el grupo 1 frente al grupo 4. Para el grupo 1 frente al grupo 2 no existían diferencias significativas, quizás fruto del bajo tamaño muestral del grupo 2.

El análisis del EGFR, realizado mediante FISH, no mostró amplificación del gen en ninguno de los tumores, si no polisomía del cromosoma 17. Sin embargo la presencia de esta polisomía tampoco se asoció con ninguno de los grupos propuestos (p-valor = 0.78).

Variable Independiente		Variable de clasificación				p-global	Comparaciones por pares
		Grupo 1 (35)	Grupo 2 (12)	Grupo 3 (22)	Grupo 4 (14)		
β-Catenina nuclear (IHQ)	Negativa (<1)	21	9	11	12	0.021 ^{LR}	G1 ≠ G4 (0.020) G2 ≠ G3 (0.047) G3 ≠ G4 (0.014)
	Baja (1 ≥ X < 2)	4	0	6	1		
	Media (x ≥ 2)	10	2	5	0		
Ki-67 (IHQ)	Media	54.2	64.0	53.4	55.0	0.341 ^{KW}	No significativas
	Desv Stnd	19.6	25.2	13.0	20.6		
Apoptosis (IHQ)	Media	8.5	7.8	8.5	6.6	0.825 ^{KW}	No significativas
	Desv Stnd	8.6	6.2	6.1	4.9		
EGFR (FISH)	Normal	28	10	14	11	0.785 ^{LR}	No significativas
	Polisomía	5	2	5	2		

Tabla.11. Resumen de los estudios de asociación entre la clasificación molecular y los análisis inmunohistoquímicos realizados. El número de muestras totales analizadas es de 83 en todas las comparaciones a excepción de β-Catenina que es de 81 y de EGFR que fue de 77, debido a que 2 y 6 muestra respectivamente fueron no informativas. LR: test "Likelihood Ratio"; KW: test de Kruskal Wallis.

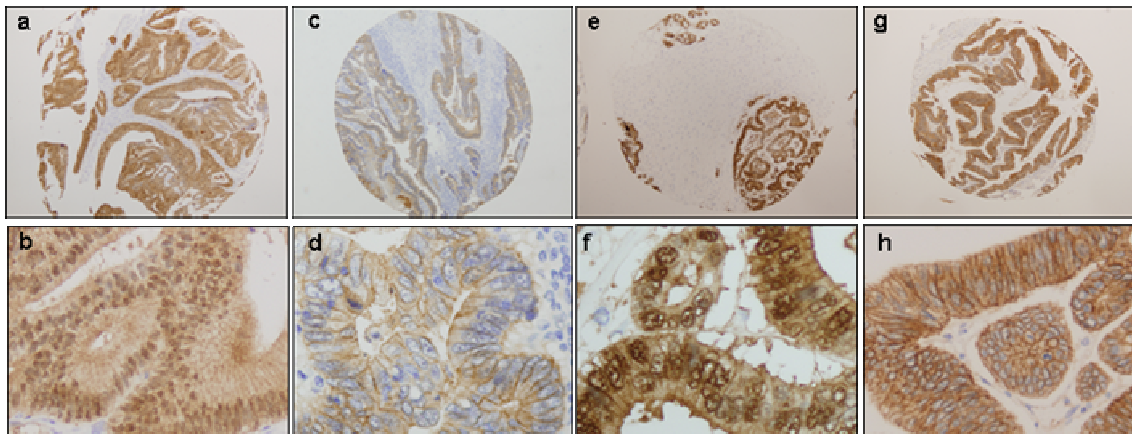


Fig.7. Inmunohistoquímica con el anticuerpo contra β-catenina realizado en los Tissue-arrays. Aparecen representados tumores pertenecientes a cada uno de los cuatro grupos moleculares definidos. a y b) Grupo 1, muestra CT5; c y d) Grupo 2, muestra CT105; e y f) Grupo 3, muestra CT42; g y h) Grupo 4, muestra CT103. Tanto en las secciones correspondientes a las muestras del grupo 1 como en las de las muestras correspondientes al grupo 3 se puede observar la presencia de β-catenina nuclear en contraposición con las de las secciones de las muestras de los grupos 2 y 4 en donde se observa la presencia de β-catenina en la membrana celular. Secciones a, c, e y g obtenidas mediante un objetivo de 100 aumentos. Secciones b, d, f y h obtenidas mediante un objetivo de 400 aumentos.

3.2.5. Estudios de asociación entre la clasificación histológica (Dukes), los parámetros analizados y los perfiles de expresión.

La clasificación basada en criterios clínicos y patológicos (Dukes, TNM, etc.) es la que se utiliza actualmente en la clínica para establecer el diagnóstico de los pacientes. Debido a su importancia, estudiamos también su asociación con los parámetros patológicos y moleculares recogidos a lo largo del estudio (Tablas. 12, 13 y 14).

Los primeros análisis mostraron que no había asociación ni con el sexo (p-valor = 0.614), ni con la calidad del RNA o RIN (p-valor = 0.823) (Tabla.12). Sin embargo en el caso de la edad, aunque las diferencias no eran significativas si se acercaban a la tendencia, mostrando una menor edad de diagnóstico cuando más avanzado era el tumor (Tabla.12). Esto podría ser reflejo de la dificultad que conlleva el diagnóstico temprano de los tumores, siendo capaces de detectar a edades tempranas solo aquellos tumores muy agresivos que posiblemente muestren ya claros síntomas de su presencia.

Posteriormente se pudo observar que existía una correlación entre alguno de los marcadores patológicos asociados a pronóstico y la progresión tumoral (Tabla.12). La aparición de infiltrado linfocitario cuya presencia está asociada con buen pronóstico, se observó mayoritariamente en tumores en estadio tempranos (Dukes A), y fue disminuyendo hasta alcanzar sus niveles más bajos en tumores en estadios tardíos (Dukes D), siendo esta diferencia estadísticamente significativa (p-valor < 0.001) y coincidiendo con lo observado por otros grupos [350].

Lo contrario ocurrió para el margen tumoral y la invasión vascular, ambos asociados a mal pronóstico y que en nuestro caso también se asociaban con la clasificación de Dukes (p-valor = 0.001 y p-valor = 0.001 respectivamente) (Tabla.12). En el caso del margen tumoral, se ha descrito que la presencia de un borde de tipo infiltrante en contraposición al borde expansivo tiene peor pronóstico. En nuestras muestras se observó que a medida que progresaba el tumor se producía una disminución de tumores con borde expansivo y un aumento de tumores con borde infiltrante, hasta llegar al estadio D donde todos ellos poseían borde infiltrante. En el caso de la invasión vascular, descrita como una de las vías para producir metástasis, se observó que el

número de tumores con invasión de vasos iba aumentando según progresaba el tumor hasta alcanzar la mayor proporción en tumores en estadio D.

Variable Independiente		Variable de clasificación				p-global	Comparaciones por pares
		Dukes A (22)	Dukes B (25)	Dukes C (18)	Dukes D (18)		
Localización	C. Derecho	9	12	7	9	0.876 ^x	No significativas
	C. Izquierdo	13	13	11	9		
Grado de diferenciación	Alto grado	0	1	2	3	0.116 ^{LR}	No significativas
	Bajo grado	22	24	16	15		
Margen Tumoral	Infiltrante	9	17	14	18	0.001 ^{LR}	A ≠ D (<0.001) B ≠ D (0.007) C ≠ D (0.049) A vs C tendencia (0.053)
	Mixto	5	3	2	0		
	Expasivo	8	5	2	0		
Infiltrado Linfocitario	Inexistente	8	11	12	18	< 0.001 ^{LR}	A ≠ D (<0.001) B ≠ D (<0.001) C ≠ D (0.009) A vs C tendencia (0.085)
	Bajo	4	6	4	0		
	Medio	9	6	2	0		
	Alto	1	2	0	0		
Inv. Vascular	Si	2	10	7	13	0.001 ^x	A ≠ B (0.015) A vs C tendencia (0.053) A ≠ D (<0.001) B ≠ D (0.037) C ≠ D (0.044)
	No	20	15	11	5		
Invasión Perineural	Si	1	1	1	3	0.378 ^{LR}	No significativas
	No	21	24	17	15		
Tipo Histológico	Convencional	21	21	17	16	0.531 ^{LR}	No significativas
	Mucinoso	1	4	1	2		
Porcent. Estroma Congelado	Media	13.9	16.6	15.5	22.1	0.2 ^{KW}	No significativas
	Desv Stnd	8.0	11.4	8.2	10.9		
Porcent. Estroma Parafina	Media	11.8	15.0	16.4	20.3	0.019 ^{KW}	A ≠ C (0.034) A ≠ D (0.004) B vs D tendencia (0.078)
	Desv Stnd	7.8	8.4	7.2	9.3		
Sexo	Hombre	11	13	7	11	0.614 ^x	No significativas
	Mujer	11	12	11	7		
Edad	Media	74.5	72.9	71.4	65.7	0.111 ^{KW}	No significativas
	Desv Stnd	8.5	11.0	10.9	14.7		
RIN	Media	9.0	9.0	8.8	9.1	0.823 ^{KW}	No significativas
	Desv Stnd	0.7	0.7	0.8	0.6		

Tabla.12. Resumen de los estudios de asociación entre la clasificación clínica (Dukes) y los distintos parámetros clínicos y patológicos disponibles. El número de muestras totales analizadas es de 83. x : test Chi-Cuadro; LR: test "Likelihood Ratio"; KW: test de Kruskal Wallis.

Finalmente, del resto de parámetros histopatológicos analizados, solo la cantidad de estroma tumoral parecía correlacionar con la clasificación de Dukes (Tabla.12). El estudio del porcentaje de estroma de la pieza congelada no mostró diferencias estadísticamente significativas (p-valor = 0.20) aunque si se observó un aumento de la cantidad de estroma con la progresión tumoral. El análisis de la proporción de estroma de la pieza parafinada reflejó la misma tendencia, aunque en este caso de manera más

clara ya que sí existían diferencias estadísticamente significativas (p-valor = 0.019). Esto es interesante, ya que la clasificación molecular propuesta también muestra asociación con la cantidad de estroma, y sin embargo no hay asociación entre la clasificación de Dukes y la clasificación molecular.

Variable Independiente		Variable de clasificación				p-global	Comparaciones por pares
		Dukes A (22)	Dukes B (25)	Dukes C (18)	Dukes D (18)		
Inestabilidad	MSS	19	20	18	17	0.082 ^{LR}	No significativas
	MSI	3	5	0	1		
Mutación K-Ras	Si	8	12	5	6	0.562 [?]	No significativas
	No	14	13	13	12		

Tabla.13. Resumen de los estudios de asociación entre la clasificación clínica (Dukes) y los marcadores moleculares disponibles. El número de muestras totales analizadas es de 83. x : test Chi-Cuadro; LR: test "Likelihood Ratio".

En el caso de los parámetros moleculares estudiados (KRAS e inestabilidad) (Tabla.13) no se observó asociación con la clasificación de Dukes (p-valor = 0.562 y p-valor = 0.082 respectivamente). Sin embargo, en el caso de la inestabilidad en microsatélites (MSI) se observó una tendencia (p-valor = 0.082) hacia la presencia de este tipo de tumores en estadios tempranos (Dukes A y B).

Variable Independiente		Variable de clasificación				p-global	Comparaciones por pares
		Dukes A (22)	Dukes B (25)	Dukes C (18)	Dukes D (18)		
β-Catenina nuclear (IHQ)	Negativa (<1)	15	15	10	13	0.983 ^{LR}	No significativas
	Baja (1 >= X < 2)	3	3	3	2		
	Media (x >= 2)	4	6	4	3		
Ki-67 (IHQ)	Media	65.7	57.6	49.6	46.3	0.004 ^{KW}	A ≠ C (0.001) A ≠ D (0.002) B vs D tendencia (0.084)
	Desv Stnd	17.5	19.6	14.5	19.2		
Apoptosis (IHQ)	Media	6.5	8.9	10.2	6.7	0.398 ^{KW}	No significativas
	Desv Stnd	6.1	6.9	9.8	4.3		
EGFR (FISH)	Normal	17	22	11	13	0.415 ^{LR}	No significativas
	Polisomía	4	2	4	4		

Tabla.14. Resumen de los estudios de asociación entre la clasificación clínica (Dukes) y las inmunohistoquímicas realizadas. El número de muestras totales analizadas es de 83 en todas las comparaciones a excepción de β-catenina que es de 81 y de EGFR que fue de 77, debido a que 2 y 6 muestra respectivamente fueron no informativas. LR: test "Likelihood Ratio"; KW: test de Kruskal Wallis.

Los datos obtenidos del análisis inmunohistoquímico (Tabla.14) reflejaron que no había correlación entre la clasificación clínica y la presencia de β-catenina nuclear, la

apoptosis celular o la presencia de polisomía en el cromosoma de EGFR. Sin embargo y contrariamente a lo descrito en la literatura [351,352] se observó una correlación entre estadios tempranos de la progresión tumoral (Dukes A) y un mayor índice de proliferación (Ki-67).

Finalmente y desde un punto de vista molecular, se confirmó lo observado durante la realización del “clustering” jerárquico (Fig.5B). No existía asociación entre los perfiles de expresión (clasificación molecular) y la clasificación de Dukes (clasificación histológica) ($p = 0.646$) (Tabla.9). A esto hay que añadir que la realización de un análisis de comparación de clases reveló que no existían genes diferencialmente expresados (utilizando un p -valor < 0.001 y B&H) entre ambas clasificaciones, confirmando lo descrito previamente.

3.2.6. Validación de la clasificación molecular mediante una matriz de datos externa.

La validación de una nueva clasificación mediante el uso de datos externos independientes es un aspecto esencial a la hora de demostrar la existencia y universalidad de dicha clasificación. Lo ideal es llevar a cabo la validación a través de la realización de experimentos en nuevas muestras. Sin embargo, a menudo la falta de disponibilidad de un número suficientemente elevado de nuevas muestras y el alto coste que estos experimentos suelen tener, hace que los laboratorios se planteen esta validación usando matrices de datos ya publicadas por otros laboratorios. La dificultad de esta segunda opción es como ya se ha comentado anteriormente la disparidad entre grupos sobre los criterios de selección de muestras, de purificación de las mismas, de las plataformas utilizadas y de los procesos de pre-análisis. Siendo conscientes de estos problemas se utilizó la matriz de datos publicada por el grupo de Eschrich [252] para este propósito.

Este trabajo contaba con más de 60 pacientes, había llevado a cabo una validación externa de sus resultados en otra serie de datos y poseía datos de supervivencia que podían complementar nuestro estudio. Además, presentaba la ventaja de tener un diseño experimental similar al utilizado en nuestro caso, puesto que se llevó a cabo con pacientes en diferentes estadios clínicos (B, C y D de Dukes), mediante

hibridación doble (cDNA) y los datos fueron normalizados mediante la técnica de LOWESS.

Para llevar a cabo esta validación era necesario poder aplicar de manera precisa los algoritmos resultantes de un estudio en el otro y viceversa. Esto implicaba eliminar previamente sesgos sistemáticos existentes en ambas matrices debido al tipo de plataforma utilizada y contar con el mismo conjunto de genes. Así, tras promediar los valores de expresión de los genes repetidos bajo un identificador único y eliminar aquellos con un número demasiado elevado de datos perdidos (esto último sólo en la matriz de Eschrich) se obtuvieron los siguientes valores para las dos matrices:

- ❖ Matriz HCSC: 96 muestras y 11017 genes únicos analizados en este trabajo.
- ❖ Matriz Eschrich: 78 muestras y 9229 genes únicos utilizados por el grupo de Eschrich.

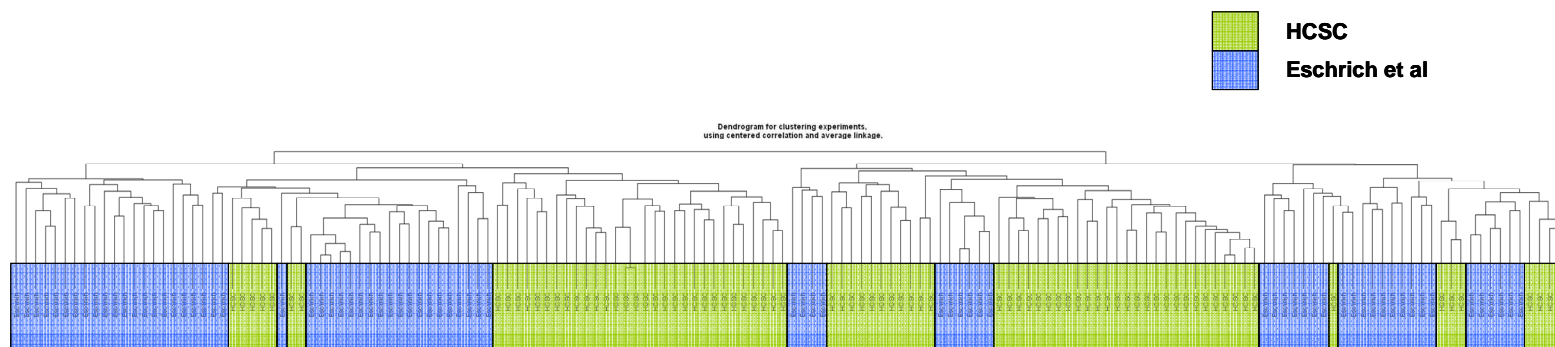
Tras combinar los dos conjuntos mediante el método DWD [286] se obtuvo una matriz de expresión con 5319 genes y 174 muestras.

Se llevó a cabo un análisis no supervisado de todas las muestras tumorales (se eliminaron los tejidos normales, los adenomas y las 5 muestras no clasificadas) con la intención de ver si se distribuían en los cuatro grupos esperados (159 muestras en total). Solo los 461 genes, de los 1722 utilizados originalmente para construir el “cluster”, que se encontraron en la matriz combinada fueron utilizados.

Lo primero que se observó tras la realización del “clustering” jerárquico fue la validez del método DWD para eliminar los sesgos debido al uso de microarrays de plataformas diferentes (Fig.8). Aunque es cierto que no consiguió eliminarlos completamente, puesto que las muestras de ambas matrices aunque entremezcladas siguen apareciendo en bloques perfectamente distinguibles), si permitió eliminar los más notables y revelar los principales patrones de expresión de las muestras.

A) Clustering Jerárquico. Muestras tumorales HCSC y Schrich et al. Usando el DWD.

(Centrado de genes, Correlación Centrada de Pearson y Average linkage)



B) Clustering Jerárquico. Muestras tumorales HCSC y Schrich et al. No usando el DWD.

(Centrado de genes, Correlación Centrada de Pearson y Average linkage)

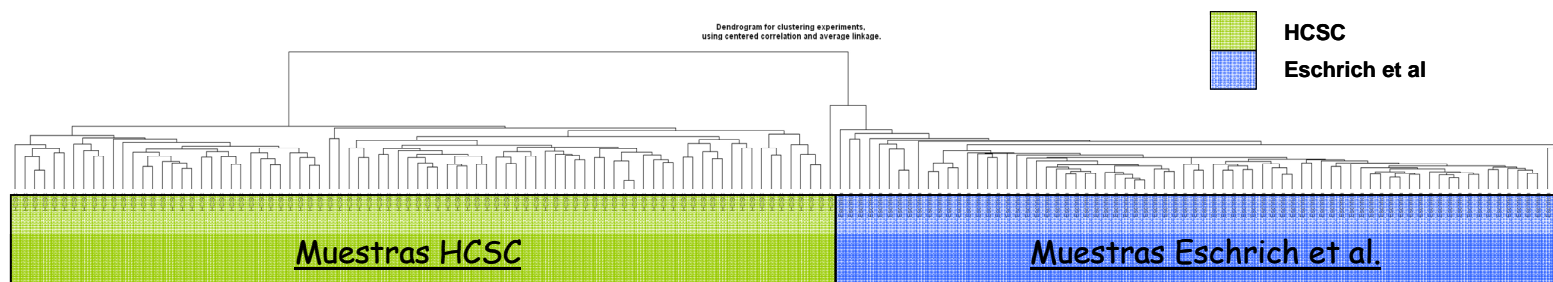


Fig.8. Resultado del “clustering” jerárquico utilizando las muestras (un total de 159) de las dos matrices (HCSC y Eschrich) y los 461 genes comunes a ambas series de datos de los 1722 originales que se usaron para establecer la clasificación molecular. A) “clustering” jerárquico tras utilizar el método DWD para fusión de matrices de datos. B) “clustering” jerárquico sin usar el programa DWD.

Un análisis más en profundidad del árbol jerárquico permitió identificar tres grupos o clusters de muestras bien definidos (Fig.10). En el “cluster” o grupo de la izquierda aparecen agrupadas 31 de las 36 muestras de la matriz HCSC definidas como grupo 1 o “de bajo estroma”. El grupo central engloba las 22 muestras de la matriz HCSC previamente definidas como grupo 3 o “de alto estroma”. El grupo de la derecha engloba 8 de las 14 muestras de la matriz HCSC definidas como grupo 4 o “MSI-Mucinoso”. Las muestras del grupo 2 no se agrupan en ningún “cluster” aparte, aunque en su mayoría (9 de 12) se engloban en el grupo central. El hecho de que las muestras de la matriz de validación (Eschrich) no conformen clusters independientes si no que se distribuyen asociándose con los grupos moleculares propuestos en este estudio debido a la existencia de perfiles de expresión similares entre las muestras, pareció confirmar la aplicabilidad de nuestra clasificación molecular en su matriz de datos.

Para comprobar si las muestras de la matriz de Eschrich en este “cluster” conjunto se agrupaban con los grupos propuestos en nuestra clasificación debido a perfiles comunes o sólo debido al azar, se llevo a cabo una predicción de clases en las muestras de la matriz de validación en base a nuestros 4 grupos moleculares, con la intención de ver cuál era la coincidencia con el análisis no supervisado que se acababa de realizar.

Para el predictor se usaron todos aquellos genes que mostraban una diferencia significativa entre grupos ($p < 0.001$) (Tabla.15). Los cuatro métodos de clasificación disponibles (DLDA, 1NN, 3NN y NC) arrojaron resultados estadísticamente significativos ($p < 0.01$) a la hora establecer un predictor que fuese capaz de separar las muestras de la matriz HCSC en sus grupos correspondientes (100 repeticiones). La tasa de error de clasificación optimizada (mediante la técnica del “Leave One Out Cross Validation” o LOOCV) en cualquiera de las cuatro técnicas solo varió entre el 15-19% a pesar de no contarse con la matriz de genes más óptima (Tabla.15). Durante la combinación de ambas matrices de datos (Eschrich y HCSC), se observó que genes cuya diferencia de expresión entre los grupos moleculares propuestos utilizando los datos HCSC eran muy significativos y que por tanto los hacía importantes para la clasificación, no se encontraban incluidos en la matriz de validación, por lo que no pudieron ser usado para crear el predictor. De todas maneras, los cuatro predictores

parecían funcionar correctamente, aunque los métodos del NC y el 3NN parecían los más adecuados al tener la menor tasa de error optimizada (Tabla.15).

	DLDA	NN1	NN3	NC
% medio de clasificaciones correctas	81	81	85	85
Tasa de Error Optimizado	0.19	0.19	0.15	0.16
Nivel óptimo de significancia	0.0005	0.001	0.0005	0.001
Genes utilizados	903	1039	903	1039

Tabla.15. Descripción global de los predictores utilizados para la clasificación de muestras externas según la clasificación molecular propuesta. El cálculo de estos parámetros se llevó a cabo usando, como técnica de validación cruzada, el LCOOV. La fiabilidad de estos predictores se testó mediante el uso de 100 permutaciones.

Se calcularon los coeficientes de sensibilidad y especificidad obtenidos para los predictores, así como los coeficientes del Valor Predictivo Positivo y Valor Predictivo Negativo para cada grupo mediante la técnica de LOOCV (Tabla.16).

Los resultados (Tabla.16.) reflejaron una especificidad superior al 85% para los cuatro grupos independientemente de la técnica de predicción utilizada, alcanzando en la mayoría de los casos valores superiores al 90%. La sensibilidad para los grupos 1, 3 y 4 fue siempre superior al 70% con cualquiera de los métodos, siendo sin embargo más baja para el grupo 2, lo que se corresponde con la reproducibilidad encontrada para el “cluster” de muestras HCSC original (Tabla.16). El VPP y el VPN fueron calculados mediante LOOCV para nuevas muestras. El VPP para los grupos 1, 3 y 4 fue siempre superior al 70% independientemente del método usado, mientras que para el grupos 2 rondaba una precisión del 54-64% (Tabla.16). El VPN de los predictores fue sumamente elevado, superando siempre el 90% de precisión (Tabla.16). Los coeficientes de VPP y VPN iguales o superiores a los valores de sensibilidad, confirman que no parece existir sobreajuste (“overfitting”) en los métodos utilizados.

Aunque los resultados fueron similares para los cuatro casos, el método final de predicción de muestras seleccionado fue el “Nearest Centroid” ya que poseía de manera global los valores de Sensibilidad, Especificidad, VPP y VPN más elevados (Tabla.16),

además de poseer una de las tasas de error optimizadas más bajas de las técnicas utilizadas (Tabla.15).

Diagonal Linear Discriminat Analysis.

Grupo	Sensibilidad	Especificidad	VPP	VPN
1	0.92	0.94	0.92	0.94
2	0.67	0.90	0.53	0.94
3	0.77	0.94	0.81	0.92
4	0.71	0.97	0.83	0.94

1-Nearest Neighbor.

Grupo	Sensibilidad	Especificidad	VPP	VPN
1	0.94	0.94	0.92	0.96
2	0.58	0.93	0.58	0.93
3	0.77	0.89	0.71	0.92
4	0.71	0.99	0.91	0.95

3-Nearest Neighbor.

Grupo	Sensibilidad	Especificidad	VPP	VPN
1	0.94	0.90	0.87	0.96
2	0.50	0.93	0.55	0.92
3	0.86	0.87	0.70	0.95
4	0.71	0.96	0.77	0.94

Nearest Centroid.

Grupo	Sensibilidad	Especificidad	VPP	VPN
1	0.94	0.94	0.92	0.96
2	0.75	0.93	0.64	0.96
3	0.82	0.94	0.82	0.94
4	0.71	0.99	0.91	0.95

Tabla.16. Valores de sensibilidad, especificidad, VPP y VPN de los predictores creados para cada uno de los grupos moleculares propuestos, en función de los diferentes algoritmos disponibles (DLDA, 1NN, 3NN, NC).

Cuando se aplicó este predictor a la matriz de Eschrich et al., el 55% se clasificó como grupo 1 o “de bajo estroma”, el 16% como grupo 2 o “relacionado con respuesta inmune”, el 8% como grupo 3 o “de alto estroma” y el 21% como grupo 4 o “MSI-Mucinoso” según los perfiles moleculares propuestos anteriormente. La clasificación de las muestras de Eschrich según los estadios de Dukes y los grupos moleculares se muestra en la Tabla.17.

	Dukes B	Dukes C	Dukes D
Grupo 1	9	12	20
Grupo 2	7	1	4
Grupo 3	3	1	2
Grupo 4	4	8	4

Fig.17. Distribución de las muestras de Eschrich en función de su estadio de Dukes y del estadio molecular estimado. Grupo 1: “de bajo estroma”; Grupo 2: “relacionado con respuesta inmune”; Grupo 3: “de alto estroma”; Grupo 4: “MSI-Mucinoso”.

Tras comparar el resultado del “clustering” jerárquico con la predicción realizada, pudimos establecer en la matriz de Eschrich la existencia de 3 de los 4 subtipos moleculares identificados en HCSC (grupos 1, 3 y 4) (Fig.10). Treinta y siete de las 41 muestras de Eschrich definidas como grupo 1 o “bajo estroma” se agrupaban junto a las muestras HCSC del grupo 1. Cinco de las 6 muestras de la matriz de Eschrich definidas como grupo 3 o “de alto estroma” se dispusieron a lo largo del

“cluster” central en donde también se agrupaban las muestras HCSC definidas como grupo 3. En el caso de las muestras de la matriz de Eschrich clasificadas como grupo 4 o “MSI-Mucinoso”, un total de 13 de las 16 se asociaban con las muestras HCSC del grupo 4. Solo las muestras de Eschrich definidas como grupo 2 o “relacionado con sistema inmune” no clasificaban en ninguno de los “cluster” de manera preferencial. Como era de prever, los genes más significativos de los diferentes grupos moleculares mostraban el mismo perfil de expresión para las muestras de las dos matrices de datos, a excepción de los diferencialmente expresados en el grupo 2, cuyo perfil era un poco más confuso (Fig.9).

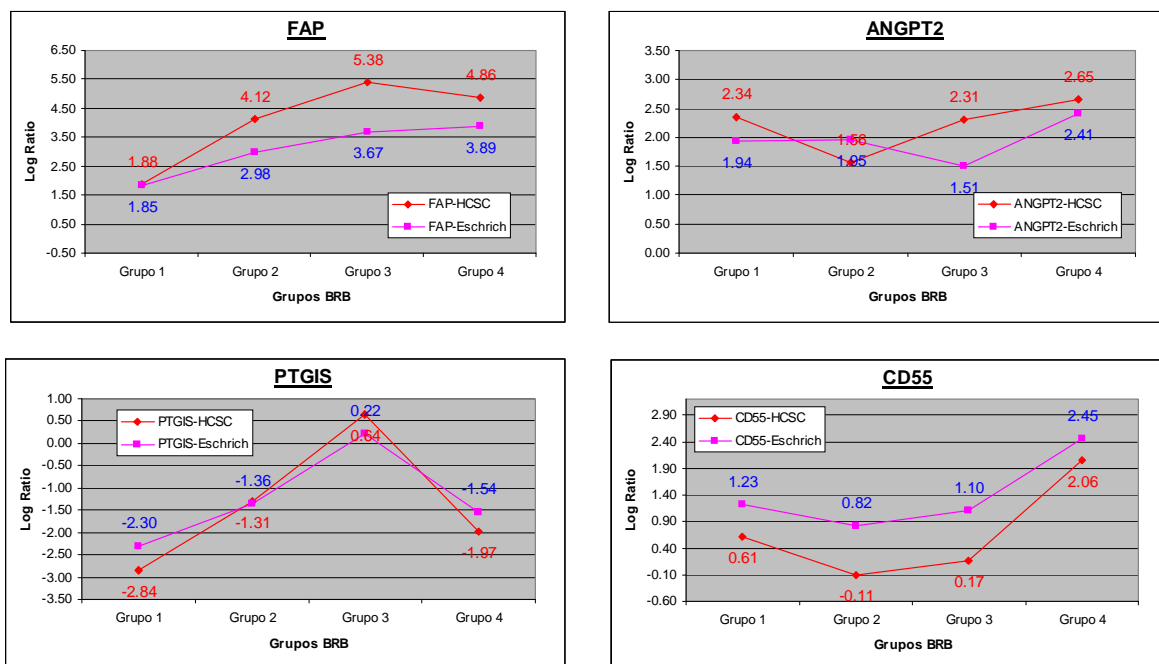


Fig.9. Comparación de los perfiles de expresión entre la matriz de datos de Eschrich (azul) y la matriz HCSC (rojo) para alguno de los genes importantes en para el establecimiento de la clasificación molecular. De derecha a izquierda y de arriba-abajo se muestra un ejemplo de uno de los genes diferencialmente expresado en el grupo 1, en el grupo 2, en el grupo 3 y en el grupo 4.

Los resultados obtenidos parecen confirmar la existencia, en conjuntos de datos independientes (Eschrich), de perfiles similares a los de los clusters definidos en este trabajo. Sin embargo, también hay que destacar ciertas diferencias entre ambas matrices de datos en lo que se refiere a la plataforma utilizada o a los criterios de selección histológica utilizados, además, de la falta de criterios externos para confirmar la validación (tumores MSI, mucinosos, % estroma, etc. en la matriz de Eschrich).

"Clustering" Jerárquico. Muestras tumorales HCSC y Eschrich et al.

(Centrado de genes, Correlación Centrada de Pearson y "Average linkage")

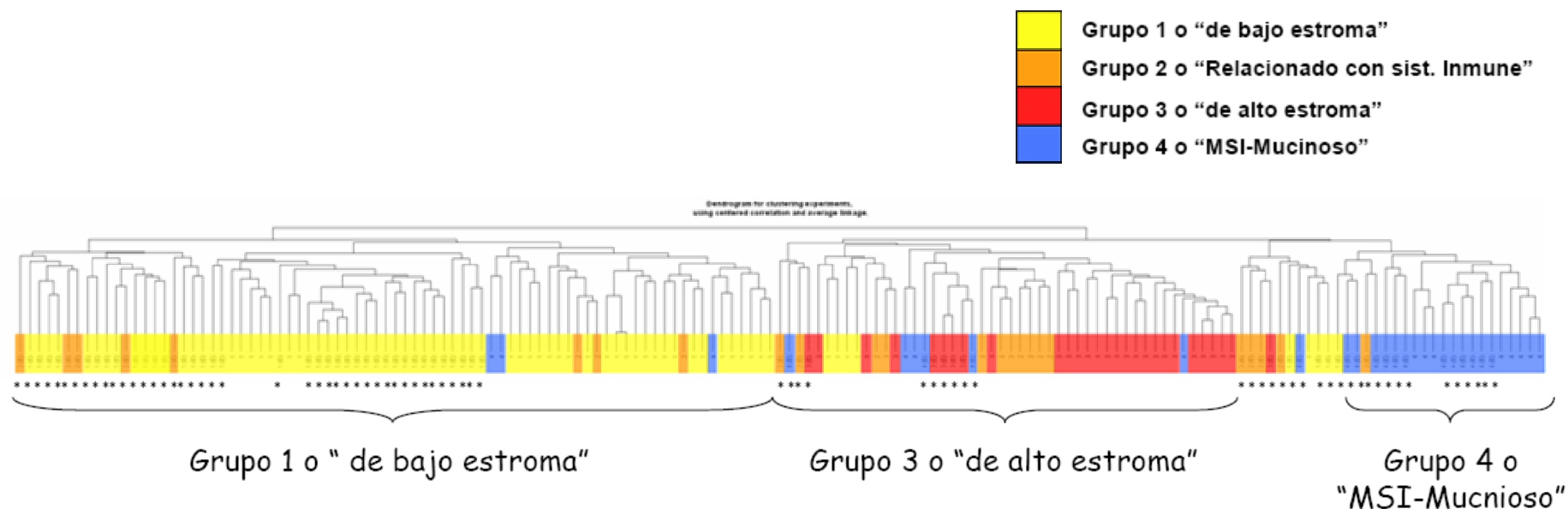


Fig.10. Resultado del "clustering" jerárquico conjunto de las muestras tumorales de la matriz HCSC (excluyendo las muestras no clasificadas del grupo 5) y de la matriz de Eschrich (definidas con un * en el árbol jerárquico). Para ello se usaron los 461 genes, de los 1722 originales, que eran comunes para ambos conjuntos de datos. Se utilizó la correlación centrada de Pearson, el "average linkage" y el centrado de genes. El código de colores en función de los subtipos moleculares se muestra en la parte superior derecha de la figura. La asignación del subtipo molecular a las muestras de Eschrich se llevo a cabo utilizando el método del "Nearest Centroid" (predicción de clases).

3.2.7. Análisis de supervivencia.

Con el objetivo de estudiar si existía asociación entre la clasificación molecular propuesta y la evolución de los pacientes (supervivencia global o período libre de enfermedad) se utilizaron dos enfoques diferentes:

- ❖ Se utilizaron los predictores de supervivencia o de período libre de enfermedad propuestos en otros trabajos para comprobar si estos correlacionaban con nuestros grupos propuestos.
- ❖ Se llevaron a cabo estudios de supervivencia propiamente dichos para las dos matrices de datos (HCSC y Eschrich) en función de la clasificación molecular.

3.2.7.1. Aplicación de predictores pronóstico.

Se utilizaron los predictores publicados en el propio trabajo de Eschrich [252] además de los de Wang [253] y Lin [260] (New Zeland). Ante la imposibilidad de aplicar de manera exacta el predictor propuesto por cada uno de estos laboratorios, al no coincidir completamente las matrices de genes (uso de diferentes plataformas, falta de una notación común para los genes, criterios diferentes de filtrado por “flags”) y al carecer del algoritmo matemático concreto con el que habían sido definidos, realizamos una aproximación usando los genes incluidos en estos tres predictores como matrices de expresión para la realización de análisis no supervisados. Si bien es cierto que la construcción de predictores es más exacta a la hora de clasificar muestras, los mismos patrones de expresión utilizados para su construcción, pueden ser detectados mediante análisis no supervisado (“clustering” jerárquico), logrando con esta última técnica obtener una idea aproximada de cuál podría ser el comportamiento de cada muestra si realmente hubiésemos aplicado el predictor tal y como fue publicado.

Al ser plataformas diferentes, el número de genes de cada uno de los predictores utilizados disminuyó sensiblemente (Tabla.18), pudiéndose utilizar 11 de 23 genes para el predictor de Wang, 17 de 43 en el caso del predictor de Eschrich y 17 de 22 en el caso del predictor de Lin.

	Predictor Wang Y	Perdeictor Eschrich S	Predictor Lin YH (N.Zeland)
Nº muestras usadas	72	78	149
Precisión Predictor	78	90	77
Sensibilidad	72	93	53
Especificidad	83	84	88
Nº Genes propuestos	23	43	22
Nº genes con flags	8	7	2
Nº genes no pasan el filtro del expresión diferencial	2	0	1
Nº genes sin coincidencia	2	19	2
Nº genes usados	11	17	17

Tabla.18. Resumen de los aspectos más importantes de los clasificadores publicados por otros grupos y usados en este trabajo. En las primeras cinco filas se hace una descripción de los predictores publicados. Las cuatro últimas resumen la adaptación de esos predictores a nuestra matriz de datos, indicando las causas de exclusión de los genes para los diferentes predictores y el número final de genes utilizados.

Los resultados obtenidos tras la realización del “clustering” jerárquico en la matriz de datos HCSC, utilizando los genes incluidos en los predictores publicados por los diferentes laboratorios, se muestran en la Fig.11. La reproducibilidad de los “cluster” depende del predictor usado, siendo el más reproducible desde un punto de vista matemático el realizado con el predictor de Eschrich y el menos reproducible el realizado con el predictor de Lin. En los tres casos analizados, el dendrograma dividía las muestras en dos grupos principales, que teniendo en cuenta lo expuesto en los artículos originales, cabría esperar que un grupo englobase las muestras de buen pronóstico y otro las de mal pronóstico. En los tres casos se observó la separación de la mayoría de las muestras del grupo 1 o “de bajo estroma” en una rama del árbol jerárquico, mientras que la mayoría de las muestras de los grupos 2, 3 y 4 se encuadraron en la rama opuesta. Esto coincidiría con las grandes diferencias de expresión encontradas entre este grupo y el resto en nuestra matriz de datos. En el predictor propuesto por Lin et al. se observó también una tendencia a clasificar un número elevado de tumores del grupo 4 (un total de 8 de las 14 existentes), entre los que se encontraban la mayoría de tumores MSI (4 de 5), pero no así los tumores mucinosos (solo 1 de 6 que además resultó ser a la vez MSI).

Estos datos indicaban por tanto la existencia de una posible asociación entre los grupos moleculares inicialmente propuestos y la evolución de los pacientes.

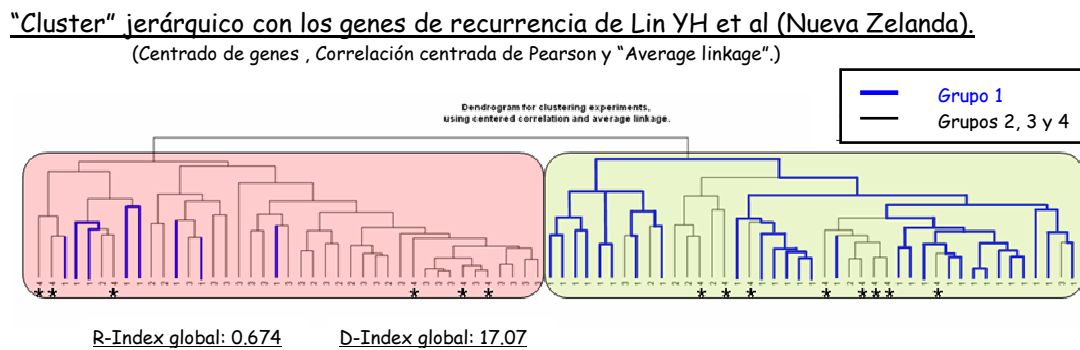
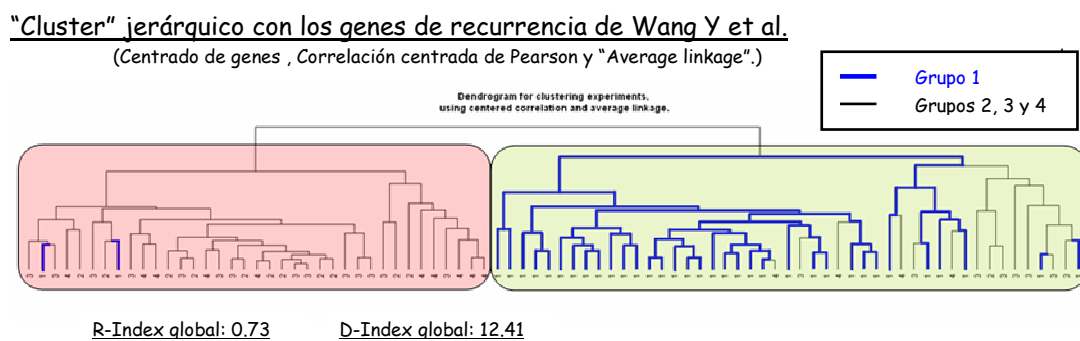
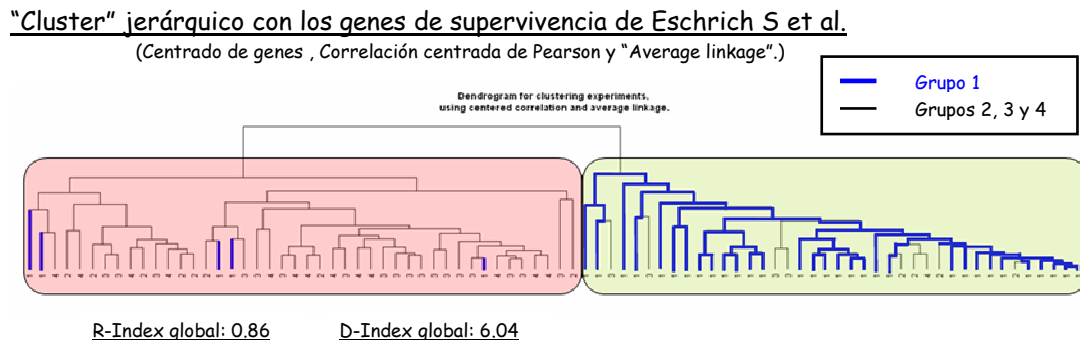


Fig.11. Resultado del "clustering" jerárquico de las muestras de la matriz HCSC utilizando los genes referidos en los predictores de Eschrich, Wang y Lin (de arriba hacia abajo). Las muestras del "cluster" 1 aparecen marcadas en azul, mientras que las muestras del resto de grupos aparecen en negro. En el "clustering" realizado con el predictor de Lin, las muestras del grupo 4 aparecen marcadas con un "*". Debajo de cada "cluster" aparece información sobre la reproducibilidad global de los mismos ("D-Index" y "R-Index"). Grupo 1: "de bajo estroma"; Grupo2: "relacionado con respuesta inmune"; Grupo 3: "de alto estroma"; Grupo 4: "MSI-Mucinoso"

3.2.7.2. Análisis de supervivencia en pacientes HCSC.

Para llevar a cabo este análisis, se excluyeron las muestras que contaban con menos de 36 meses de seguimiento no habiendo fallecido debido al tumor (13 pacientes), que fallecieron durante el ingreso hospitalario en el que se realizó la cirugía (7 pacientes) y que fueron considerados como no clasificables (grupo N°5) tras el establecimiento de la clasificación molecular (5 pacientes). Finalmente se eliminó otro paciente (n° 27) debido a que desarrolló otro tumor primario de colon posterior al analizado y para el que además estaba recibiendo varias líneas de quimioterapia. Siguiendo estas reglas generales se excluyeron 26 de los 88 pacientes inicialmente incluidos en el estudio. 23 de los 62 pacientes restantes recibieron tratamiento adyuvante o paliativo con diferentes esquemas de QT o RTx. Todos los esquemas tenían en común el tratamiento con 5-FU, aunque en función de cada paciente se aplicaron otros fármacos o tratamientos que incluían oxaliplatino, irinotecan, anti-EGFR, anti-VEGF o radiofrecuencia.

Inicialmente se realizó el análisis de supervivencia (supervivencia global y período libre de enfermedad, excluyendo en este último caso los estadios D) en aquellos pacientes que no habían recibido tratamiento complementario a la cirugía (39 pacientes), de manera que la población estudiada fuese lo más homogénea posible (Tabla.19. A, B y C). La finalidad era ver si existía asociación entre nuestra clasificación molecular y la supervivencia. Debido al bajo número de muertes (1 de 39) y al bajo número de recidivas (4 de 39) el test de regresión de Cox no fue capaz ni de encontrar diferencias significativas para el pronóstico de los grupos, ni de crear un modelo convergente que explicase el riesgo asociados a cada grupo. Esto ocurrió tanto para el análisis de supervivencia global como para el de período libre de enfermedad, no pudiéndose extraer conclusiones en ninguno de los dos casos.

Posteriormente, realizamos el estudio independientemente del tratamiento, utilizando los 62 pacientes que habían pasado los criterios de selección (Tabla.19. D, E y F). Cuando realizamos el análisis de supervivencia global (10 muertes de 62 pacientes) el test de Cox no fue capaz, de nuevo, ni de encontrar diferencias significativas en el pronóstico de los grupos, ni de crear un modelo convergente que explicase el riesgo asociado a cada grupo. Cuando se analizó el período libre de

recurrencia (10 recidivas de un total de 50 pacientes, una vez excluidos los D) entre los diferentes grupos tampoco se encontraron diferencias estadísticamente significativas (p-valor: 0.406).

		Distribución muestras sin tratamiento (39)			
		Grupo 1	Grupo 2	Grupo 3	Grupo 4
A)	Dukes A	8	4	3	4
	Dukes B	6	3	3	3
	Dukes C	1	2	2	0
	Dukes D	0	0	0	0
	Total	15	9	8	7

		Distribución todas las muestras (62)			
		Grupo 1	Grupo 2	Grupo 3	Grupo 4
D)	Dukes A	8	4	3	4
	Dukes B	9	3	3	5
	Dukes C	5	2	4	0
	Dukes D	6	2	3	1
	Total	28	11	13	10

		Distribución muestras sin tratamiento (39)			
		Grupo 1	Grupo 2	Grupo 3	Grupo 4
B)	Dukes A	1	0	0	0
	Dukes B	0	1	0	0
	Dukes C	1	0	1	—
	Dukes D	—	—	—	—
	Total	2	1	1	0

		Distribución todas las muestras (50)			
		Grupo 1	Grupo 2	Grupo 3	Grupo 4
E)	Dukes A	1	0	0	0
	Dukes B	2	1	0	1
	Dukes C	4	0	1	—
	Dukes D	—	—	—	—
	Total	7	1	1	1

		Distribución muestras sin tratamiento (39)			
		Grupo 1	Grupo 2	Grupo 3	Grupo 4
C)	Dukes A	0	0	0	0
	Dukes B	0	0	0	0
	Dukes C	0	0	1	—
	Dukes D	—	—	—	—
	Total	0	0	1	0

		Distribución todas las muestras (62)			
		Grupo 1	Grupo 2	Grupo 3	Grupo 4
F)	Dukes A	0	0	0	0
	Dukes B	2	0	0	0
	Dukes C	1	0	1	—
	Dukes D	4	0	1	1
	Total	7	0	2	1

Tabla.19. Distribución de las muestras en función de la clasificación molecular propuesta y los estadios de Dukes. A) Distribución de todas las muestras que no recibieron tratamiento complementario a la cirugía. B) Distribución de las muestras sin tratamiento complementario y que recidivaron durante el estudio. C) Distribución de las muestras sin tratamiento complementario y que fallecieron durante el estudio. D) Distribución de todas las muestras independientemente de si recibieron o no tratamiento complementario a la cirugía. E) Distribución de todas las muestras independientemente del tratamiento complementario y que recidivaron durante el estudio. F) Distribución de todas las muestras independientemente del tratamiento complementario y que fallecieron durante el estudio.

Posteriormente se comprobó si la clasificación molecular era capaz de aportar un mayor poder pronóstico a los estadios de Dukes. Debido al escaso tamaño muestral se tuvieron en cuenta de manera conjunta los estadios B y C (grupos más homogéneos) y se realizó el análisis de supervivencia de nuestros grupos moleculares. Los resultados obtenidos se correspondieron con lo expuesto anteriormente. Así, cuando se eliminaron los pacientes que habían recibido QT o RTx en estadios B y C de Dukes (se analizaron un total de 20 muestras) el análisis de Cox no encontró diferencias significativas entre los grupos moleculares, ni fue capaz de crear un modelo que explicase el riesgo asociado a cada grupo por lo que no se pudieron extraer conclusiones ni para la SG (1 muerte), ni para el PLE (3 recidivas). Cuando se analizaron las muestras en estadios B y C de Dukes, independientemente de si habían recibido otro tratamiento diferente a la cirugía (31 pacientes) el análisis de SG (4 muertes) tampoco fue capaz de encontrar diferencias significativas, ni de encontrar un modelo convergente, por lo que no

podimos extraer conclusiones, mientras que el análisis del PLE (9 recidivas) dio un resultado no significativo (p-valor: 0.570). Todos estos análisis en los que se incluyen muestras con tratamientos complementarios a la cirugía, deben ser vistos con cautela debido a que no estamos teniendo en cuenta el posible efecto que este tratamiento complementario puede tener en la supervivencia.

Además de estos análisis, también se estudió si en nuestro conjunto de muestras se observaba el efecto pronóstico que se asocia a la clasificación de Dukes. Al igual que en el caso de la clasificación molecular, cuando se analizaba la función de supervivencia para los estadios de Dukes teniendo en cuenta aquellos pacientes que no habían recibido tratamiento complementario a la cirugía no se encontraron diferencias ni para el PLE (p-valor = 0.155. Excluido el estadio D) ni para la SG (p-valor = 0.128) siendo imposible crear un modelo convergente que explicase el riesgo de recurrencia o muerte de cada grupo en cada una de las respectivas situaciones, debido al bajo número de eventos.

Cuando se incluyeron los pacientes que habían recibido tratamiento complementario se observaron diferencias estadísticamente significativas tanto para la SG (p-valor = 0.002) como para el PLE (p-valor = 0.039. Excluidos los estadios D). En el primer caso el aumento en el número de muertes correlacionaba con la progresión tumoral. Aunque no pudo establecerse un modelo de asociación de riesgo, ya que el número de eventos no fue suficientemente elevado, se observó una diferencia elevada en el porcentaje de muertes entre el estadio A y B con respecto al D (0%, 9% y 33%). En el caso del estudio del período libre de enfermedad se observó que el riesgo de recaída también era mayor en función del estadio de Dukes (mayor cuanto más avanzado era el estadio). Esta diferencia era estadísticamente significativa (p-valor = 0.034) en el caso del estadio C en comparación con el estadio A asociándose un riesgo de recaída 10 veces mayor en caso del estadio C.

3.2.7.3. Análisis de supervivencia de la matriz de validación (Eschrich).

Al contar con los datos de expresión y el tiempo de supervivencia de los pacientes incluidos en el trabajo publicado por Eschrich S et al. [252] (75 muestras de las cuales 48 murieron por tumor antes de los 36 meses de seguimiento), realizamos un

análisis de supervivencia global de estos pacientes en base a nuestros grupos moleculares.

En primer lugar y teniendo en cuenta lo observado al aplicar los predictores de supervivencia de Wang [253], Eschrich [252] y Lin [260] en nuestra matriz de datos quisimos comprobar si un predictor (“Nearest Centroid”) que diferenciase el grupo “de bajo estroma” del resto de grupos, era capaz, en las muestras de la matriz de Eschrich, de encontrar diferencias significativas en cuanto a supervivencia (Fig.12). El predictor creado a partir de 167 genes (Grupo “de bajo estroma”: Sensibilidad = 0.896; Especificidad = 0.972; VPP = 0.977; VPN = 0.875) mostraba de manera estadísticamente significativa (p -valor = 0.030) diferencias entre ambos grupos (HR = 1.86 (1.07-3.22)), indicando una menor supervivencia para el “resto de grupos” en relación con el grupo “de bajo estroma”.

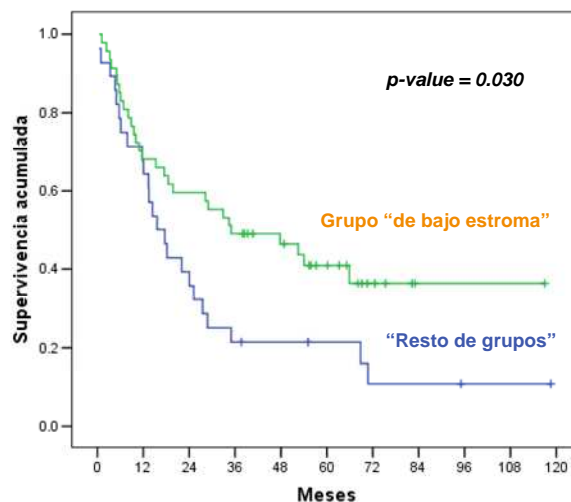


Fig.12. Representación de la función de supervivencia para las muestras de la matriz de Eschrich clasificadas en función de los grupos moleculares (grupo “de bajo estroma” y “resto de grupos”, usando 167genes y el método del “Nearest Centroid”), mediante curvas de Kaplan-Meier. El test de Cox da un riesgo de muerte 1,86 veces mayor para los pacientes del “resto de grupos” en comparación con las muestras del grupo “de bajo estroma” ($p = 0.030$).

Una vez comprobado que existía asociación, quisimos comprobar por separado, la relación de cada uno de los 4 grupos molecularmente definidos, con la supervivencia. Para ello se utilizaron los resultados obtenidos previamente en el proceso de predicción de clases realizado en la etapa de validación de la clasificación (Sección 3.2.6. Pág.128).

Los resultados mostraron que teniendo en cuenta los cuatro grupos moleculares como entidades individuales independientes de los estadios de Dukes, no existían diferencias globales significativas en cuanto a la supervivencia (p -valor = 0.121). A pesar de ello, analizando los grupos por pares se observaba que el grupo 1 o “de bajo estroma” sí se asociaba con una supervivencia mayor (2.39 veces) y estadísticamente significativa (p -valor = 0.012) cuando se le comparaba con el grupo 4 o grupo “MSI-Mucinoso” (Fig.13).

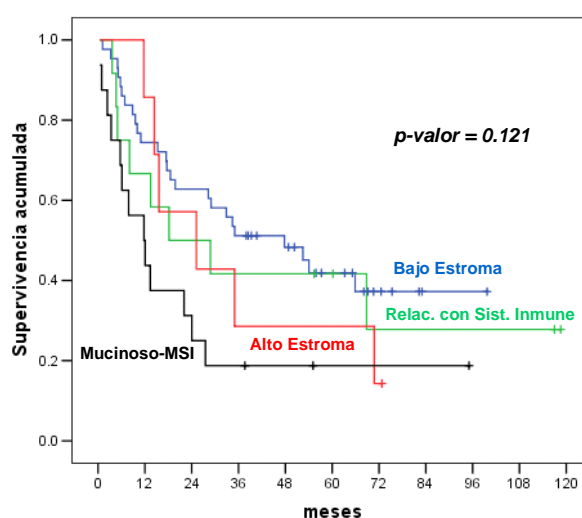


Fig.13. Representación de la función de supervivencia para todas las muestras tumorales de la matriz de Eschrich en función de la clasificación molecular (4 grupos), mediante curvas de Kaplan-Meier. Grupo 1: “de bajo estroma”; Grupo2: “relacionado con respuesta inmune”; Grupo 3: “de alto estroma”; Grupo 4: “MSI-Mucinoso”.

A la vista de estos resultados, analizamos de nuevo si existía asociación entre la supervivencia y nuestra clasificación molecular, pero esta vez usando un conjunto de muestras más homogéneas (estadios B y C). A pesar de las diferencias, decidimos combinar ambos subtipos (B y C), ya que los dos engloban estadios localizados, en este caso concreto no mostraban diferencias significativas en cuanto a supervivencia y el número de muestras total permitiría llevar a cabo estos análisis de supervivencia.

Cuando se usó de nuevo el predictor de 167 genes que diferenciaba el grupo “de bajo estroma” del “resto de grupos”, pero sólo en estadios B y C de Dukes, se observó con mayor claridad que los pacientes del “resto de grupos” en comparación con el grupo “de bajo estroma” tenían peor pronóstico (p -valor = 0.001 y HR = 3.907 (1.64-9.31)) (Fig.14).

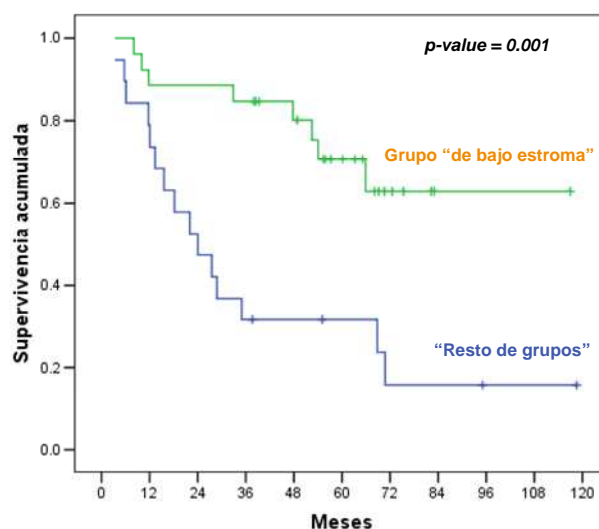


Fig.14. Representación de la función de supervivencia para las muestras de la matriz de Eschrich (sólo estadios B y C) clasificadas en función de los grupos moleculares (grupo “de bajo estroma” y “resto de grupos”, usando 167genes y el método del “Nearest Centroid”), mediante curva de Kaplan-Meier. El test de Cox da un riesgo de muerte 3.907 veces mayor para los pacientes del “resto de grupos” en comparación con las muestras del grupo “de bajo estroma” ($p = 0.001$).

Cuando se analizaban en los estadios B y C los 4 grupos moleculares de manera individual, el test de regresión de Cox mostró una diferencia significativa (p-valor: 0.005) en la función de supervivencia global de los diferentes grupos moleculares coincidiendo con lo reflejado por las curvas de Kaplan-Meier. Fig.15. y Tabla.20.

El grupo 1 mostró un riesgo relativo de muerte 5.17 veces menor que el grupo 3 (IC95% = 1.45-18.48, con un p-valor = 0.012) y 5.66 veces menor que con respecto al grupo 4 (IC95% = 1.98-16.23, con un p-valor = 0.001). El grupo 2 mostró una diferencia en el riesgo relativo de muerte 1.93 veces superior al del grupo 1, aunque esta diferencia no fue significativa (p-valor = 0.31). En relación con el grupo 2 cabe destacar también el hecho de que parecía poseer un riesgo relativo de muerte 2.94 veces menor que el grupo 4 con un p-valor cercano a la significancia (p-valor = 0.077). Con respecto al grupo 3 y a pesar de que su riesgo relativo de muerte (el de grupo 2) es 2.68 veces menor, esta diferencia es no significativa (p-valor = 0.165), quizás debido al bajo número de muestras sobre todo en el grupo 3. Cuando se realizó la comparación del grupo 3 vs 4, se encontró que el riesgo relativo de muerte era prácticamente igual (HR = 1.10), no encontrándose diferencias significativas (p-valor = 0.88).

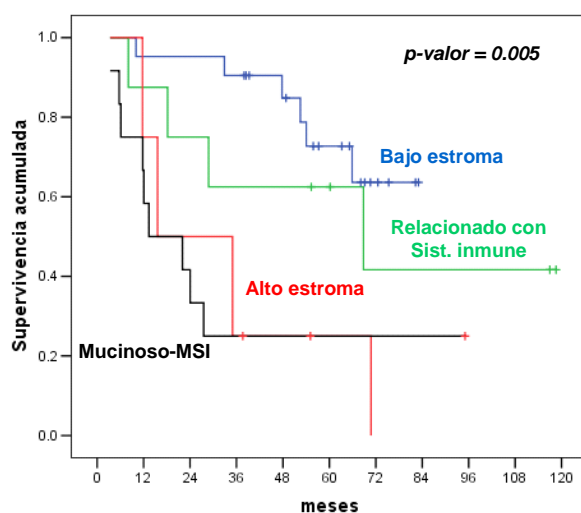


Fig.15. Representación de la función de supervivencia para las muestras tumorales de la matriz de Eschrich diagnosticadas como estadios de Dukes B o C, en función de la clasificación molecular (4 grupos), mediante curvas de Kaplan-Meier.

Modelo de Cox	Hazard Ratio (95% IC)	p-valor
Alto estroma vs. Bajo Estroma	5.17 (1.45-18.48)	0.012
Alto estroma vs. Relac. con Sist. Inmune	2.68 (0.66-10.78)	0.216
Mucinoso-MSI vs Alto estroma	1.10 (0.33-3.58)	0.88
Mucinoso-MSI vs. Bajo Estroma	5.66 (1.98-16.23)	0.001
Mucinoso-MSI vs. Relac. con Sist. Inmune	2.94 (0.89-9.70)	0.077
Relac.con Sist. Inmune vs. Bajo Estroma	1.93 (0.54-6.85)	0.310

Tabla.20. Comparación del riesgo relativo de muerte entre los 4 subtipos moleculares propuestos, utilizando las muestras de la matriz de Eschrich pertenecientes a los estadios B y C de Dukes. IC: intervalo de confianza.

Estos resultados apuntan al hecho de que la nueva clasificación molecular es capaz de aportar nueva información en los que a pronóstico y supervivencia se refiere, al ser combinada con la clasificación de Dukes. Sin embargo hay que recordar que este pronóstico no ha sido todavía corroborado en nuestros propios pacientes y ni en el contexto de un análisis multivariable. Para ello, una muestra más grande y homogénea de pacientes debería ser usada.

Discusión.

1. Introducción.

El cáncer en general, y el colorrectal en particular, se han revelado como una de las enfermedades más importantes del siglo XX y XXI, cuya incidencia ha ido aumentando en paralelo con el desarrollo de la sociedad y el aumento de la esperanza de vida. En los últimos años el manejo clínico de los pacientes ha experimentado una evolución importante, introduciéndose grandes avances en lo que a supervivencia y calidad de vida del paciente se refiere. Los conocimientos clínicos, patológicos y moleculares adquiridos han permitido crear clasificaciones pronósticas y predictivas de gran valor para el tratamiento del paciente que nos acercan un poco más hacia el objetivo de una medicina personalizada en la que el abordaje de la enfermedad y los procedimientos a seguir estarán en función de las características individuales de cada paciente y de su tumor. Pese a todo, cuando tratamos con pacientes clasificados dentro de un mismo subtipo (Dukes, TNM, etc.), sigue existiendo una gran heterogeneidad que impide rentabilizar al máximo las posibilidades terapéuticas disponibles. El estudio de los tumores a nivel molecular mediante el uso de técnicas de análisis global parece que se está revelando como una pieza clave para descifrar los mecanismos de desarrollo del cáncer y alcanzar el objetivo de una medicina personalizada en el área de la oncología. Los frutos obtenidos mediante las técnicas de análisis masivos como los microarrays de expresión en tumores de mama son un buen reflejo de las posibilidades que esta técnica puede tener en tumores como el colorrectal. El establecimiento de una clasificación de tumores colorrectales en base a sus perfiles de expresión génica y el conocimiento de las rutas moleculares implicadas en los procesos de malignización tal y como se han planteado en nuestro trabajo podrían suponer un avance para el futuro tratamiento de los pacientes.

2. Clasificación de tumores colorrectales.

La clasificación de tumores colorrectales mediante perfiles de expresión, al contrario que ocurre en tumores de otros tejidos, es un área poco explorada y en general, los trabajos realizados no han sido llevados a cabo como verdaderos experimentos no supervisados. El establecimiento de una clasificación molecular independiente de otros parámetros puede aportar nueva información útil para el manejo clínico de los pacientes y el tratamiento de los tumores colorrectales.

Los resultados de nuestro laboratorio presentan un modelo nuevo de clasificación molecular mediante microarrays que agrupa los tumores en 4 subtipos asociados además a parámetros biológicos y cuyas características moleculares están presentes desde el inicio del desarrollo tumoral. Los 4 grupos descritos se denominaron como subtipo “de bajo estroma”, subtipo “relacionado con sistema inmune”, subtipo “de alto estroma” y subtipo “MSI-mucinoso”.

La autenticidad de los 4 grupos moleculares propuestos y obtenidos a partir de 1722 genes en 88 tumores, la confirmamos por un lado mediante criterios estadísticos y por otro lado mediante la obtención de resultados similares tras la variación de los parámetros originales de agrupación. Cerca de un 80% de las combinaciones que analizamos (14 de las 18) mostraron una alta reproducibilidad. Variaciones del número de genes utilizados durante la clasificación (Min: 649 genes utilizados. Max: 3943 genes), el uso de algoritmos de clasificación diferentes como el k-means, o el uso de diferentes medidas de distancia (Pearson, Kendal-Tau, Spearman), no alteraron los patrones de reconocimiento de los cuatro grupos, demostrando su consistencia. Aún así, uno de los principales problemas de este tipo de estudios es que en muchas ocasiones solo se aportan los mejores resultados obtenidos, obviando o no aclarando aquellos que no arrojan resultados tan buenos como nos gustaría [185]. Durante este proceso también se encontraron combinaciones (cuatro en total) que no mostraron reproducibilidad con los subtipos moleculares originalmente descritos. Esto creemos que es debido al hecho de haber utilizado parámetros cuyas características no siempre se ajustan al objetivo de obtener una clasificación basada en perfiles de expresión similares. Las cuatro combinaciones que no mostraban esta reproducibilidad se comentan a continuación:

- ❖ Diferencias totales frente a perfiles similares (distancia euclídea vs. Correlación de Pearson): la distancia euclídea al contrario que el resto de distancias utilizadas en nuestro trabajo, hace hincapié en las diferencias absolutas existentes entre dos perfiles de expresión y no tiene en cuenta si a pesar de que existan ciertas diferencias cuantitativas de expresión entre ambos perfiles, en general la forma de dichos perfiles es similar. Por ello, en la mayoría de los casos en los que se busca establecer una nueva clasificación, se utiliza como distancia, medidas como la correlación de Pearson o similares [207,218,353].

- ❖ Métodos de agrupación de elementos (“average linkage” vs “complete and single linkage”): Los cuatro métodos utilizados a lo largo de las combinaciones son teóricamente válidos, pero tanto el “average linkage” como el “centroid linkage” son mucho más representativos de las muestras que conforman las agrupaciones, ya que ambos métodos de agrupación, en contraposición con los otros dos, están influidos por todos los miembros (muestras) que componen los clusters o grupos. El “single linkage” y el “complete linkage” serán útiles, solo en el caso de que los miembros de cada “cluster” tengan una gran variabilidad intergrupala, acompañada de una variabilidad intragrupal muy pequeña.
- ❖ Centrado vs. No Centrado de genes: el centrado de genes permite eliminar el efecto que la muestra de referencia ejerce sobre los niveles de expresión, pudiendo identificar diferencias, que cabría esperar que fuesen pequeñas, entre los perfiles de los subtipos tumorales encontrados. En nuestro caso el hecho de clasificar solo tumores, hizo necesario eliminar el efecto que el “pool” de tejidos normales tuvo sobre estas, al ser usado como referencia.

Este cuidadoso proceso de análisis ha permitido establecer la primera clasificación de adenocarcinomas colorrectales basada exclusivamente en datos de expresión. Otros trabajos se plantearon anteriormente retos parecidos, aunque con resultados poco concluyentes. En 2003, Frederiksen et al. [248] presentaron un trabajo de clasificación que analizaba muestras que incluían todos los estadios de Dukes, además de muestras de tejido normal (sólo 5 tejidos de cada subtipo). Como resultado principal observaron que mientras que sí que eran capaces de diferenciar de manera más o menos precisa estadios B y C, los estadios A y D no se agrupaban cada uno en un “cluster” independiente y que además los estadios D clasificaban en su mayoría con las muestras de tejido normal. Estos sorprendentes resultados parecían, sin embargo, estar asociados con problemas en los procedimientos utilizados durante los experimentos y en concreto con los que se refieren al número de muestras analizadas y a los análisis histológicos realizados.

En los datos presentados por nuestro laboratorio para la clasificación de tumores colorrectales, se paliaron en gran medida este tipo de deficiencias incluyendo por un lado un número cercano a los 20 pacientes para cada uno de los estadios analizados y por otro

llevando a cabo un exhaustivo análisis histológico. Gracias al criterio histológico utilizado (tejidos con al menos un 80% de componente tumoral) y realizado sobre la misma pieza que se iba a hibridar se evitó una elevada contaminación con tejido normal. Sin embargo, en el caso del trabajo de Frederiksen et al. el número de muestras analizadas era demasiado pequeño (25 muestras en total) como para poder corroborar la existencia de diferentes subtipos y no se había llevado a cabo un estudio histológico del tejido que se hibridó. Que los tejidos tumorales en estadio D tuviesen un alto porcentaje de contaminación con tejido normal parece la causa más probable de que muestras normales y los tumores en ese estadio se agrupen conjuntamente. Trabajos como el de Birkenkamp-Demtroder [235] et al. sufrieron inconvenientes parecidos a los observados en el trabajo del grupo de Frederiksen.

El hecho de trabajar con un tamaño muestral pequeño como ocurrió en los dos casos anteriores, solo es aceptable, aunque no completamente aconsejable, en determinadas situaciones. En nuestro laboratorio y previamente a la realización de la clasificación tumoral, llevamos a cabo una caracterización de las vías afectadas en el desarrollo de los tumores colorrectales mediante el estudio de tejidos tumorales (88 muestras) frente a tejidos normales (7 muestras). El bajo tamaño muestral de nuestro grupo de tejidos normales, se vio sin duda compensado por el exhaustivo control histológico realizado y por las notables diferencias que existen entre estos dos fenotipos. Nosotros y otros grupos [234,236] hemos comprobado la validez de esta aproximación cuando se comparan tejidos normales y tumorales.

Otros factores que han influido en los resultados de los trabajos de clasificación publicados hasta ahora, hacen referencia a los criterios de inclusión de muestras. Lo ideal en un trabajo experimental es contar con una serie consecutiva de pacientes, reflejo de la población enferma en general. Desafortunadamente, esto no es posible ante determinadas circunstancias. El trabajo propuesto en nuestro caso consistía en crear una clasificación molecular, ver si esta correlacionaba con la clasificación de Dukes, caracterizarla y definir si estos patrones se establecían desde el principio del desarrollo tumoral. Esto implicaba analizar un porcentaje de tumores similar para cada uno de los cuatro estadios de Dukes, incluido el estadio inicial o estadio A. Conscientes de este hecho se intentó obtener una población de muestras libre de sesgos, obteniéndose una serie de pacientes cuyos parámetros analizados (edad, MSI, KRAS, localización, etc.) parecen mostrar frecuencias en su mayoría coincidentes con las frecuencias descritas en la literatura para la población

general con cáncer colorrectal. En el caso de trabajos como el de Birkenkamp-Demtroder [235] o el de Frederiksen [248] llama la atención, el hecho de realizar una selección sesgada de muestras (sólo tumores colorrectales del lado izquierdo) cuando el objetivo principal es el establecimiento de una clasificación de pacientes. En el caso de Frederiksen su decisión de no seleccionar tumores de colon derecho con la intención de no incluir tumores de tipo hereditarios, no justifica el sesgo introducido al eliminar posibles subtipos tumorales como los tumores MSI esporádicos que se desarrollan preferentemente en el colon derecho.

Por último, el hecho de utilizar técnicas no supervisadas (“clustering” jerárquico) partiendo de criterios supervisados (teniendo en cuenta los subtipos de la clasificación histológica, trabajando con los subtipos en forma de “pool” o utilizando listas de genes obtenidas de comparación de clases) como sucede en los trabajos de Friederichs [354] y sobre todo en el de Birkenkamp-Demtroder [235,354], impide la obtención de datos concluyentes.

De la clasificación que proponemos en este estudio, el dato más destacado además de los cuatro subgrupos moleculares definidos, es su falta de correlación con la clasificación de Dukes (p-valor: 0.646). La falta de asociación entre los perfiles moleculares y la clasificación de Dukes, se vio confirmada al no observarse genes diferencialmente expresados en los análisis de comparación de clases entre los estadios de Dukes. Esto confirmaría a nivel molecular, y por vez primera utilizando microarrays, la heterogeneidad de los estadios de Dukes y además establecería que los perfiles moleculares de los tumores estarían presentes desde el inicio del tumor (estadio A). Aunque, parece que somos los primeros en confirmar la ausencia de correlación entre ambas clasificaciones mediante esta tecnología, otros trabajos de microarrays también aportaron datos en este sentido. Frederiksen et al. [248] describieron que tumores en estadio A aparecían dispersos a lo largo de los distintos grupos moleculares. También Bertucci et al. [251] observaron que mediante perfiles de expresión era posible diferenciar, de manera más precisa que a través de la clasificación patológica, los pacientes que han desarrollado o desarrollaran metástasis en un corto espacio de tiempo de aquellos que aún con metástasis linfáticas no llegarán a realizar metástasis en órganos distantes. Los trabajos de Kwong [249] y de de Friederichs [354], aunque más dirigidos a la comparación de clases entre estadios de Dukes que a la clasificación, señalaron también la dificultad de encontrar genes que diferenciasen

los estadios de Dukes (en concreto B, C y D) y no consiguieron agrupar las muestras en base a dichos estadios mediante análisis no supervisados, reflejando la heterogeneidad molecular de la clasificación de Dukes. Posteriormente el grupo de Groene publicó un trabajo [250] de comparación de clases entre estadios II y III (clasificación UICC), que aunque consiguió crear un predictor que diferenciaba ambos estadios, volvió a evidenciar la falta de homogeneidad biológica desde un punto de vista molecular como se desprendió del hecho de que el análisis de vías moleculares no mostrase ninguna diferencia entre estadios. Podemos considerar por tanto, que el trabajo realizado en nuestro laboratorio es el primer estudio que clasifica y caracteriza en subtipos moleculares los tumores colorrectales a partir de perfiles globales de expresión.

2.1. Validación de la clasificación molecular.

La finalidad de los trabajos de clasificación de tumores usando perfiles de expresión, consiste en crear grupos homogéneos de pacientes en base a la alteración de sus vías moleculares y analizar su posible valor pronóstico y/o predictivo para su posterior utilización en la clínica. Sin embargo, para que la universalidad de la nueva clasificación sea aceptada debe de ser reproducible en otras poblaciones de pacientes.

Los resultados de la validación de nuestra clasificación muestran la presencia, y por tanto reproducibilidad, de perfiles similares a los grupos “de alto estroma”, “de bajo estroma” y “MSI-Mucinoso” en la matriz de muestras independientes pertenecientes al grupo de Eschrich. Desafortunadamente, el perfil de las muestras pertenecientes al grupo “relacionado con sistema inmune” no parece altamente reproducible y en la nueva matriz de datos, las muestras clasificadas como grupo 2 quedaban distribuidas heterogéneamente entre los otros tres grupos. A pesar de ello, esta situación es esperable cuando existen diferencias entre las matrices de datos que se comparan y además el predictor de ese determinado grupo en concreto, no es altamente robusto. Trabajos como el de Sorlie et al. [355] para la validación de la clasificación molecular del cáncer de mama, utilizando métodos muy similares a los aplicados en nuestro estudio, encontraron las mismas dificultades para diferenciar todos los grupos propuestos en su clasificación. A pesar de ello, ese trabajo supuso una prueba más de la existencia y universalidad de dicha clasificación en tumores mamarios. En nuestro caso, y a pesar de que la reproducibilidad no sea total, este proceso de validación también confirma la universalidad de al menos, una

gran parte de nuestra clasificación. En comparación con los métodos utilizados por Sorlie et al., nuestro estudio tiene la ventaja de haber contado con el uso del programa DWD [286] que elimina sesgos importantes debido al uso de diferentes plataformas. Este tipo de herramientas son indispensables en este tipo de proyectos y comienzan a ser utilizadas por diferentes grupos [356].

No obstante de los resultados obtenidos nos llamo la atención la forma en la que se distribuyeron las muestras de la matriz de validación en los nuevos grupos moleculares. El porcentaje de muestras clasificadas como grupo “de alto estroma” en la matriz de validación (Eschrich) es sensiblemente inferior al descrito en nuestra matriz de datos (HCSC) (8% frente al 26%). Este hecho puede ser fruto exclusivamente de la composición muestral de ambas matrices, aunque, no debemos excluir la posibilidad de que el distinto criterio histológico establecido por Eschrich (mayor % de células epiteliales) para la selección de muestras en relación con el establecido en nuestro trabajo, sea el causante de las distintas proporciones encontradas.

2.2. Alteración de los perfiles de expresión en el proceso tumoral (Tejido Normal vs Tejido Tumoral).

Los análisis supervisados y no supervisados llevados a cabo en nuestro laboratorio entre tejido tumoral y tejido normal mostraron un retrato bastante completo de las alteraciones sufridas en la malignización del tejido, permitiendo posteriormente contextualizar e interpretar mejor los patrones de los nuevos subtipos moleculares definidos.

El bajo tamaño muestral del grupo de mucosa normal, no pareció ser un problema, como ya se comentó anteriormente y como se desprende del hecho de observar una amplia concordancia entre nuestros resultados y los mostrados en el meta-análisis publicado por el grupo de Chan [357]. En él, se comparaban estudios que investigaban diferencias entre tumores colorrectales y mucosa normal de colon y se publicó una lista de 38 genes (22 sobreexpresados en tumores y 16 inhibidos) presentes en al menos 5 de los 23 estudios (tumor vs. normal) analizados. De estos 38 genes, 35 coincidían con lo observado en nuestra matriz de datos (92%). Solo 3 genes de los 38 propuestos (8%) en el meta-análisis o

bien no mostraban ningún tipo de diferencia o bien no aparecían representados en nuestra matriz de datos (ver lista completa en el Anexo III).

Entre las alteraciones observadas en la transformación cancerígena, describimos un aumento de la expresión de genes implicados en proliferación, invasión y progresión tumoral, de genes de la MEC y de la vía Wnt e inhibición de genes implicados en la respuesta inmune y en el metabolismo de lípidos, coincidiendo con lo encontrado por grupos como el de Notterman [240], Birkenkamp-Demtroder [235] o Bertucci [251]. Profundizando un poco más fuimos capaces de precisar alguna de estas funciones.

Así en el caso de la proliferación, además de genes implicados en ciclo celular y procesos de síntesis y replicación de DNA, también observamos la alteración de genes implicados en la segregación de cromosomas y en procesos de citocinesis.

A raíz de alguno de los genes implicados en el metabolismo de lípidos, encontramos un grupo más amplio de genes (implicados o no en dicho metabolismo de lípidos) como son ADH1C, GSTA5, SULT1B1, EPHX2 o metalotioneinas, etc. que también están involucrados en los procesos de las rutas xenobióticas y de detoxificación de sustancias nocivas. En la mayoría de los casos estas rutas han sido relacionadas con los procesos de metabolismo de fármacos, sin embargo poco se sabe del papel que su inhibición puede jugar en el desarrollo del tumor. La acumulación de sustancias nocivas y carcinogénicas podría favorecer la aparición de mutaciones y por tanto del desarrollo del tumor, haciendo de esta ruta una vía interesante de análisis.

Por último también observamos la capacidad de esta técnica para diferenciar la composición celular de los distintos tejidos. En este caso encontramos sobreexpresados en tejido normal, genes (MYH11, MYLK, ATP1A2, Desmina (DES) o Calponina-1 (CNN1)) relacionados con el músculo liso, existente en dicho tejido normal, pero ausente en tumores, coincidiendo con lo encontrado en los trabajos de Alon [232] y Notterman [240].

2.3. Subtipos moleculares propuestos.

Teniendo en cuenta la alteración de los perfiles del tejido normal, durante el desarrollo tumoral, los perfiles de los cuatro subgrupos moleculares obtenidos fueron caracterizados y asociados tanto con parámetros clínicos y patológicos como moleculares:

2.3.1. Estroma y clasificación. Subtipos “de bajo estroma” y “de alto estroma”.

La clasificación molecular propuesta se asocia de manera significativa con la cantidad de estroma que rodea al tumor como se comprueba por el hecho de que dos de los cuatro grupos definidos se caractericen por ser grupos “de alto estroma” y “de bajo estroma”. Según lo observado en nuestro trabajo mediante los perfiles de expresión, el subtipo 1 o “de bajo estroma” posee niveles más bajos de genes asociados con estroma y que en muchos casos se han asociado, además, con malignización y mal pronóstico como ocurre con SPP1, FAP, MMP9, lumican, etc. De manera inversa, el perfil de expresión del grupo 3 o “de alto estroma” muestra niveles elevados de los genes anteriormente mencionados. Confirmando la asociación de estos grupos con el estroma se observa la alteración de vías implicadas en comunicación celular (hsa01430), adhesión focal (hsa04510) o interacción con receptores de la matriz extracelular (hsa04512).

La importancia del componente estromal en el desarrollo tumoral parece confirmarse tanto por los datos obtenidos en trabajos de microarrays como los de Lenz [217] y Finak [218] llevados a cabo en linfomas difusos de células B largas y en tumores de mama respectivamente, así como por trabajos de análisis puramente histológicos como los del grupo de Mesker [358,359] en cáncer colorrectal. En 2007 y 2009, Mesker et al. [358,359] publicaron sendos trabajos en los que observaban asociación entre la cantidad de estroma del tumor en la zona invasiva y la supervivencia (supervivencia global, período libre de enfermedad y período libre de metástasis), de manera que pacientes con una cantidad de estroma superior al 50% en dicha región del tumor tienen un peor pronóstico que aquellos pacientes con una proporción de estroma inferior al 50%. Estos resultados, que parecían independientes de la clasificación de Dukes, dejan entrever que existe una relación entre agresividad tumoral y cantidad de estroma repercutiendo de forma directa en la supervivencia de los pacientes. Dichos resultados parecerían coincidir con el perfil molecular encontrado en los grupos “de alto y bajo estroma”.

Conviene destacar sin embargo, que existe una diferencia importante entre el estudio de Mesker y el nuestro. Así, mientras que Mesker lleva a cabo su análisis solamente en la región invasiva del tumor, en nuestro laboratorio examinamos los perfiles de expresión del frente invasivo y del centro del tumor al mismo tiempo. Esta puede ser una de las razones por la que la cantidad de estroma encontrada en sus muestras es más elevada que en las nuestras. Se ha descrito que los perfiles de expresión cuando se analiza el frente tumoral difieren de los perfiles de expresión del centro del tumor [360,361]. Sin embargo, alguno de estos autores [360] también observa que la heterogeneidad de estas regiones entre sí en un determinado tumor, es en general menor que las diferencias en el perfil de expresión de dicho tumor con respecto a otros. Teniendo en cuenta todo esto, parecería verosímil que nuestro grupo de tumores “de alto estroma”, que está asociado con genes de mal pronóstico y los tumores de alto estroma definidos por Mesker y que están asociados a menor supervivencia, pudiesen estar representado un mismo tipo de tumores.

La asociación encontrada en nuestro estudio entre los estadios de Dukes tardíos y el aumento de la cantidad de estroma, también podría estar incidiendo sobre el hecho de que los tumores para llegar a producir invasión y metástasis, generan una mayor cantidad de estroma, ya sea consecuencia de procesos de transformación epitelio-meséquima, de la necesidad de reclutar determinados tipos celulares como fibroblastos que parece que facilitan los procesos de invasión y metástasis tumoral o de otros mecanismos desconocidos. En esta misma línea apuntaba ya un trabajo de Mesker publicado en 2007 [359] en el que además de observarse que la clasificación creada en función del estroma correlacionaba con supervivencia, también observaba que la cantidad de estroma tumoral aumenta con los estadios (I-III. Clasificación AJCC). En dicho trabajo, la relación de la supervivencia con el estroma parecía independiente de la relación entre la supervivencia y la clasificación de Dukes. Esto podría hacer pensar que el aumento de la cantidad de estroma es necesaria en todos los tumores para que se lleven a cabo los procesos de invasión y metástasis, pero que además en determinados tumores se produce un aumento extra de dicho componente estromal, quizás mediante la atracción hacia la zona tumoral de tipos celulares específicos como fibroblastos o macrófagos, lo que aumentaría la agresividad de estos tumores en comparación con el resto.

Nuevos estudios, permitirán confirmar si el aumento de la cantidad de estroma independientemente del estadio de Dukes correlaciona con algún tipo de célula estromal concreta que aumente la agresividad. Estos permitirán además establecer el efecto, si lo hay, que la presencia de β -catenina nuclear y la tendencia mostrada por la de invasión vascular, tienen sobre el grupo “de bajo estroma” de nuestra clasificación molecular.

2.3.2. Subtipo “relacionado con sistema inmune”.

Desde un punto de vista molecular, su asociación con los procesos de respuesta inmune (inmunoglobulinas (región 14q32, 16p11, 15q11, 22q11 y 2p11), genes que favorecen la transcripción de estas, proteínas del complejo MHC-II, receptores de moléculas presentadas por el complejo MHC-I, moléculas liberadas por linfocitos o receptores de inmunoglobulinas) y de respuesta inflamatoria (CXCL10, CX3CL1, CCL2, CCL3, CCL4, CCL8), así como con la inhibición de genes de mal pronóstico como LAMC2, hacen de él un subtipo muy interesante.

Este interés viene reforzado por el hecho de que firmas genéticas de estroma relacionadas con sistema inmune (GZMA, CD247, HLA-F, etc.) y asociadas a buen pronóstico en otros tumores [218], muestren sus niveles más elevados en las muestras de nuestro subtipo tipo 2 (1.5 fold change con respecto al resto de grupos). En la misma dirección señalan los resultados del trabajo publicados por el grupo de Lin [260] en cáncer colorrectal. Genes como CXCL9, CXCL10, INDO o GZMB, asociados a respuesta inmune y descritos en ese trabajo como parte de un predictor de buen pronóstico (matriz alemana), muestran los niveles más elevados en las muestras de nuestro subtipo 2. Coincidiendo con todo esto, la baja presencia de β -catenina nuclear en este subgrupo también podría estar apuntando hacia un tumor de tipo menos agresivo.

El hecho de ser el grupo menos robusto (tanto en la matriz de experimentación como en la de validación) desde un punto de vista estadístico y de presentar diversos transcritos en las regiones cromosómicas de las inmunoglobulinas que podrían estar codificando para una misma inmunoglobulina, aumentando de este modo la homogeneidad del grupo, hace necesario ser prudentes con las conclusiones obtenidas. El aumento del tamaño muestral podría ayudar a definir de manera más precisa el perfil de este subtipo, despejando las dudas existentes sobre la robustez del mismo.

2.3.3. Subtipo “MSI-Mucinoso”.

Este subtipo engloba tumores MSI y tumores mucinosos. Aunque ambos tipos se encuentran relacionados entre sí, ya que muchos tumores MSI son de tipo mucinoso, sin embargo también poseen características diferentes.

Los tumores mucinosos, en la literatura, se han asociado a personas más jóvenes, con estadios más avanzados, mayor diseminación peritoneal y con una menor tasa de supervivencia que los pacientes no mucinosos [19,362]. Este peor pronóstico de los tumores mucinosos coincidiría con los perfiles moleculares para este subtipo encontrados en nuestro trabajo, en donde genes como MUC1, REG-IV, CD55 han sido asociados con peor pronóstico.

Por otro lado, los tumores MSI forman la denominada vía del fenotipo mutador [55] y a lo largo de los años, diversos trabajos los han definido como tumores con localización preferentemente proximal, de fenotipo mucinoso [363,364] y con mejor supervivencia en comparación con tumores MSS [20,365,366]. Estas características clínicas y patológicas coinciden, en parte, con lo encontrado en nuestro trabajo, en donde este subtipo se asoció (aunque no significativamente) con una localización preferentemente proximal (en concreto todos los tumores MSI de este grupo se desarrollaron en la región proximal) y con un fenotipo mucinoso (dos de los 5 tumores MSI fueron descritos además como tumores mucinosos). La presencia de infiltrado linfocitario descrita en este tipo de tumores [367,368] y una de las posibles causas de la mejor supervivencia de los tumores MSI [369], no se ha observado en el grupo 4, aunque cuando se analiza este parámetro en tumores MSI frente a MSS en nuestra serie de datos sí que se observa un cierto aumento de la cantidad de infiltrado linfocitario (aunque no significativo). Posiblemente la ausencia de asociación con este parámetro se deba, al menos en parte, a la presencia de tumores de fenotipo exclusivamente mucinoso asociados en la literatura a falta de infiltrado linfocitario.

Este subtipo “MSI-Mucinoso”, presenta también ausencia de β -catenina nuclear, coincidiendo con datos recientemente publicados [370]. Esto podría implicar una menor activación de la ruta canónica de la vía Wnt, conduciendo ello, a un tumor posiblemente

menos agresivo. En esta misma dirección, apuntaría el hecho de que observásemos una menor tendencia de este subtipo a desarrollar invasión vascular.

Este subtipo también cuenta con una frecuencia baja de mutaciones en KRAS ajustándose a lo descrito en la literatura para tumores MSI [371]. Es difícil valorar la implicación que esto puede tener en la evolución del tumor ya que en general no se ha llegado a asociar la presencia de mutaciones con un peor pronóstico [26].

De nuevo, desde el punto de vista de los perfiles de expresión, nos encontramos que este subtipo “MSI-Mucinoso”, se caracterizaba por sobreexpresar genes relacionados tanto con el componente mucinosos (Ej. TFF1 o MUC1), como con tumores inestables (Ej. SERPINA1 o SERPINB5). Además, también se observan genes relacionados con inflamación como IL-8 y en general con mal pronóstico [343,372], además de genes con resistencia a quimioterapia (REG4 y metalotioneinas). Los resultados de los trabajos de microarrays presentados por Kim [243] o Di Pietro [244], que diferencian tumores MSI de tumores MSS mediante perfiles de expresión, observan la desregulación de genes como MUC1 o las metalotioneinas, coincidiendo con lo observado en nuestro estudio.

El hecho de que tumores MSI y tumores mucinosos se nos agrupen conjuntamente en un mismo subtipo, podría deberse a su posible origen común. Los grupos de Jass [73] y Ogino [72] clasifican de manera general los tumores colorrectales en cuatro grupos, definiendo dos de ellos como tumores con fenotipo “CIMP+ y MSI” y tumores con fenotipo “CIMP+ y MSS” (Tabla.7. de Introducción). Ambos grupos tendrían su origen en los pólipos serrados caracterizados por su elevado componente mucinoso. En función de algunas de las características propuestas en estos trabajos (infiltrado linfocitario, localización, inestabilidad a microsatélites, carácter mucinoso) nuestro grupo de tumores MSI podría coincidir con el grupo “CIMP+ y MSI”, mientras nuestros tumores mucinosos parece que se asociarían con el grupo “CIMP+ y MSS”. La buena supervivencia propuesta en los trabajos de estos dos autores para el grupo “CIMP+ y MSI” coincidiría con el buen pronóstico al que siempre han sido asociados los MSI. La mala supervivencia asociada al grupo “CIMP+ y MSS” por Jass y Ogino, también estaría de acuerdo con la peor supervivencia descrita para los tumores de tipo mucinoso. El trabajo presentado por Leopoldo S et al. [290] en el que se observan diferencias tanto en supervivencia global como en tiempo a la recurrencia entre tumores Mucinosos-MSI y tumores Mucinosos-MSS

apoyan la teoría sostenida por los trabajos de Jass y Ogino. Por tanto los pólipos serrados como origen común podría explicar el hecho de que tumores MSI y tumores mucinosos se agrupen conjuntamente.

Sin embargo hay que resaltar, que en este trabajo no hemos analizado el perfil metilador de los tumores, algo necesario para poder confirmar si nuestro grupo “MSI-Mucinoso” se ajusta a los dos grupos propuestos en los trabajos previos y que sería interesante incluir en próximos proyectos. Es posible además, que la presencia de un tamaño muestral mayor nos permita ser capaces de diferenciar estos dos grupos a través de los perfiles de expresión.

2.4. Supervivencia asociada a los subtipos moleculares.

Aunque el estudio del posible valor pronóstico de la clasificación molecular, es uno de los aspectos más interesante de este tipo de estudios, en nuestro caso se ha visto limitado por diversos factores. La necesidad de seleccionar pacientes pertenecientes a todos los estadios de Dukes (incluido los estadios A que presentan una tasa muy baja de eventos o los estadios D con una gran variedad en cuanto a esquemas y duración de los tratamientos de QT), la limitación de tener que usar tejidos congelados, disponibles en menor número, así como la pérdida en el seguimiento de algunos pacientes hizo muy difícil la obtención de conclusiones en este estudio. De hecho, el bajo número de eventos detallado (16% de muertes y 20% de recaídas) en nuestra matriz de datos contrasta con la proporción de los mismos detallada en otros trabajos de cáncer colorrectal y que se encuentra alrededor del 45% (min: 32% y max: 62%) [252,253,257-260].

Como era de esperar ante estos datos, los análisis llevados a cabo en nuestras muestras no mostraron diferencias significativas en cuanto a supervivencia global ni en cuanto a tiempo libre de recaída para ninguno de los subgrupos y en la mayoría de los casos el método de Cox no fue capaz de crear un modelo de riesgo asociado al estudio debido al bajo número de eventos. Por el contrario, el análisis de supervivencia en función de los estadios de Dukes y, a pesar también del bajo número de eventos, si mostró diferencias en cuanto a supervivencia (al menos entre estadios tempranos y tardíos), demostrando el gran poder discriminatorio que se le asocia a esta clasificación. Sin embargo, si es posible (como

discutimos a continuación) que nuestra clasificación molecular, combinada con la clasificación de Dukes aporte nueva información pronóstica al diagnóstico de los pacientes.

El uso de predictores y matrices externas publicados por otros grupos para aplicarlos a los estudios propios es una herramienta útil para obtener una mayor información de los datos propios, aunque su aplicación suele ser difícil [355]. La aplicación de los predictores de Wang [253], Eschrich [252] y Lin [260], nos sirvió para darnos cuenta, de que existe un punto de coincidencia entre los perfiles de supervivencia encontrados en estos trabajos y los subtipos tumorales propuestos en nuestro estudio. En base a estos resultados la supervivencia para el grupo “de bajo estroma” o grupo 1 sería diferente a la del resto de tumores. Debido a que los predictores en general incluyen siempre un número muy bajo de genes, y de los que en algún caso no se conoce ni siquiera la función, es a menudo difícil establecer mediante una simple comparación, si los genes y vías moleculares que determinan una clasificación coinciden con aquellos que determinan el predictor de supervivencia. De los genes de los predictores de Wang, Eschrich y Lin que se encontraban en nuestra matriz, observamos que un pequeño grupo mostraba alteración de sus patrones de expresión correlacionándose tanto con el predictor de supervivencia como con los genes implicados en el establecimiento de los subtipos moleculares. Entre ellos cabe reseñar genes como Apolipoproteína C1, SPP1, Cadherina 17, HEYL o LILRB3, asociados con interacción epitelio-estroma [373] [374], procesos de inflamación [375], angiogénesis [376], progresión tumoral [377,378]. Estas funciones coincidirían con las descritas durante el análisis de los subtipos de nuestra clasificación molecular.

Los datos aportados a través de la comparación realizada con los predictores publicados, indicaba que sí que podía existir una relación entre la clasificación molecular y la supervivencia. Para analizar un poco más a fondo esta cuestión se utilizaron las muestras de la matriz de Eschrich con sus correspondientes datos de supervivencia. Al analizar la función de supervivencia usando el predictor “de bajo estroma” y de manera independiente de los estadios de Dukes, se vio que había diferencias entre las muestras del grupo “de bajo estroma” y las demás en conjunto. Esta diferencia dejaba de ser significativa cuando se analizaban los cuatro grupos de la clasificación molecular por separado. La asociación entre nuestra clasificación y la supervivencia, sin tener en cuenta los estadios de Dukes, debería revisarse en futuros estudios, ya que es posible que el bajo número de muestras impida que las diferencias entre subtipos sean estadísticamente significativas.

El análisis de supervivencia de un grupo más homogéneo de muestras (estadios B y C) confirmó el mejor pronóstico de los tumores "de bajo estroma", además de revelar que los grupos "de alto estroma" y "MSI-Mucinoso" podrían tener una menor supervivencia. De este análisis se extrae por tanto, la conclusión de que la clasificación molecular parece presentar poder pronóstico, al menos, dentro de cada estadio. Las diferencias en supervivencia entre los muestras que han sido pronosticadas como "de alto estroma" y de bajo estroma" parecen coincidir tanto en lo observado de manera general en los análisis de comparación de clases (genes diferencialmente expresados que se asociaban a mal pronóstico), como con lo publicado por el grupo de Mesker [358]. Sin embargo, el hecho de que el grupo "MSI-Mucinoso" se asocie con mala supervivencia es llamativo, puesto que en la literatura ha sido descrita la asociación entre los tumores MSI y una mejor supervivencia [20]. La presencia de tumores mucinosos, asociados a mal pronóstico, apoyados por el perfil de genes que caracteriza este grupo "MSI-mucinoso" y que también se asocian con mal pronóstico, pueden explicar en parte este hecho. Por otro lado, quizás las diferencias en el criterio histológico para la selección de muestras observadas entre el laboratorio de Eschrich y el nuestro tengan reflejo en el proceso de asignación de muestras, produciéndose una sobreestimación del número de muestras del grupo 4. El predictor creado con nuestra matriz, contaba con valores de precisión, especificidad, sensibilidad, VPP y VPN elevados pero sin embargo no recoge el hecho de que las muestras de Eschrich contengan una menor cantidad de estroma. Esto estaría incidiendo en los análisis de supervivencia de este grupo en concreto y podría explicar también en parte la asociación del grupo "MSI-Mucinoso" con mal pronóstico.

Más comprensible parece el hecho de que el grupo de bajo estroma posea un mejor pronóstico que el resto de pacientes y que de este modo este componente tisular esté jugando un papel importante en la clasificación y pronóstico de los tumores colorrectales. Importante es también el hecho, como ocurre en otro tipo de tumores (Masagué J. NEJM 2007, Fan C. NEJM 2006) de que distintos perfiles de expresión estén reconociendo una misma característica. En nuestro caso tanto los tres predictores pronóstico publicados, como nuestro predictor del grupo "de bajo estroma" y nuestro predictor de la clasificación molecular en general están reconociendo un mismo grupo de pacientes que presentan tumores con niveles de expresión del componente estromal disminuido, lo que subraya la relevancia de nuestros hallazgos. Esto también reafirma el hecho de que no son tanto los

genes individuales que componen cada predictor, si no el ser capaces de reconocer determinadas actividades y rutas biológicas alteradas en el tumor las que servirán para mejorar el manejo clínico de los pacientes.

Sin embargo hay que resaltar que estos hallazgos no han sido confirmados mediante los datos de supervivencia de nuestros propios pacientes en el contexto de un análisis multivariable. La realización de trabajos con un mayor tamaño muestral de muestras más homogéneas y a ser posible se trabaje a nivel prospectivo es, sin duda, el camino a seguir para verificar los hallazgos, en lo que a supervivencia se refiere, realizados en este trabajo.

Conclusiones

1. La obtención de resultados fiables mediante la técnica de microarrays implica la necesidad de estandarizar todos los procesos. El establecimiento de un criterio histológico de selección de muestras, el control de calidad de los RNAs hibridados así como de los procesos de hibridación y el estudio del comportamiento anómalo de las muestras mediante PCA son claves para la obtención de resultados fiables.
2. Es posible caracterizar y diferenciar, en base a alteraciones en sus perfiles de expresión, muestras tumorales colorrectales de muestras de mucosa de colon normal, identificando incluso diferencias en la composición celular de los tejidos. La expresión de genes relacionados con músculo liso en el tejido normal y su ausencia en el tejido tumoral así lo atestiguan.
3. El análisis, mediante perfiles de expresión, del tejido malignizado en comparación con el tejido de colon normal confirma la desregulación de determinadas rutas moleculares, destacando la sobreexpresión de aquellas relacionadas con proliferación, invasión y progresión tumoral, de genes de la MEC y de la vía Wnt y la inhibición de otras vías como son la de respuesta inmune y el metabolismo de lípidos. Este análisis también señala la posible implicación que la alteración de procesos como la segregación de cromosomas, los procesos de citocinesis o las rutas xenobióticas pueden tener en el cáncer colorrectal.
4. Los tumores colorrectales se clasifican, en base a sus perfiles de expresión, en cuatro subtipos que se asocian además con distintas características biológicas y clínicas. Estos son el subtipo “de bajo estroma”, el subtipo “relacionado con sistema inmune”, el subtipo “de alto estroma” y el subtipo “MSI-mucinoso”.
5. No existe asociación entre la clasificación molecular propuesta y la clasificación histológica de Dukes, observándose además, que los subtipos moleculares se establecen desde el inicio del desarrollo tumoral (tumores en estadio A).
6. El estroma adyacente al tumor juega un papel importante en el establecimiento de los subtipos que conforman la clasificación molecular propuesta y principalmente en los que se refiere a los grupos “de bajo estroma” y “de alto estroma”. Ambos subtipos se

caracterizan por la alteración en la expresión de genes que codifican tanto para proteínas del estroma y la matriz extracelular, como para aquellas implicadas en sus procesos de remodelación.

7. El subtipo “MSI-Mucinoso” engloba tumores con inestabilidad a microsatélites y tumores con un fenotipo histológico mucinoso, caracterizados por la expresión de genes relacionados con células caliciformes y con componente mucinoso como son AGR2, TFF1 o MUC1 entre otros.
8. Existe un grupo de muestras asociado, en base a su expresión génica, con una mayor respuesta inmune, caracterizada principalmente por genes relacionados con inmunoglobulinas, respuesta de linfocitos o respuesta del complejo mayor de histocompatibilidad.
9. Los perfiles que definen los subgrupos de nuestra clasificación se encuentran presentes al menos en una población independiente de pacientes. En dicha población, los subtipos de nuestra clasificación se asocian con distinta supervivencia, siendo los pacientes del subtipo “de bajo estroma” los que poseen mejor pronóstico y los subtipos “de alto estroma” y “MSI-Mucinoso” los que presentan un peor pronóstico.

Anexo I.
Análisis de microarrays.

1. Análisis de la imagen.

El análisis de este tipo de experimentos comienza con la extracción de los datos a partir de la imagen obtenida tras los procesos de hibridación y escaneado de los microarrays. Durante este proceso, se obtiene información tanto de los niveles de expresión de cada uno de los genes incorporados en el microarray, así como de los posibles errores que hayan podido ocurrir durante los procesos de hibridación y lavado.

1.1. Puntos problemáticos.

Hacen referencia a aquellos puntos del microarray que han sufrido algún problema durante los procesos de hibridación o lavado. Existen una serie de parámetros que identifican diversas alteraciones que pueden influir en los niveles de expresión detectados para cada gen y que son necesarios tener en cuenta para evitar la aparición de resultados poco fiables. Los criterios seleccionados en este trabajo y mencionados en la sección 8 de Materiales y métodos (Pág.67) se detallan a continuación:

- ❖ IsSaturated (Saturación): se considera que un punto está saturado cuando más del 50% de los píxeles para ese punto superan el límite máximo de detección de fluorescencia del escáner (65535 cuentas de fluorescencia).
- ❖ Feature NonUnifOL (No uniformidad): se considera que un punto no es uniforme cuando la varianza en la intensidad de los píxeles medida para ese punto supera la varianza esperada por el intervalo de confianza calculado a partir de una distribución Chi-cuadrado.
- ❖ IsFeatPopOL: sólo se aplica en aquellos puntos que poseen sondas repetidas a lo largo del array. Identifica si la señal de fluorescencia para un determinado punto es diferente al intervalo de confianza creado por el rango intercuartílico (1.42 veces el IQR) de los replicados de ese punto.
- ❖ IsPosAndSig (Positividad y significancia): se considera que un punto supera este criterio si la media de la señal de fluorescencia en ese punto es mayor y estadísticamente significativa con respecto al ruido de fondo ajustado (BGUsed).

-
- ❖ IsWellAbove: Indica que la señal, una vez restado el ruido de fondo (BGSubSig), es mayor que el posible error cometido en la medición de dicho ruido ($2.6 \times BG_StdDesv$) y que además cumple el criterio anterior (IsPosAndSig). En el caso de que los puntos no cumplan estos dos últimos criterios su valor de fluorescencia será sustituido por el límite mínimo de detección (“SurrogatedUsed”).

1.2. Procesamiento de la fluorescencia.

Este tratamiento permite corregir el dato bruto de la fluorescencia de manera que se elimine el efecto que distintos factores como el ruido de fondo, la incorporación de distintos fluorocromos, etc. pueden tener sobre la medida de la fluorescencia, evitando así la aparición de sesgos en la matriz de datos. Este proceso consta de 3 pasos:

- ❖ Señal menos el ruido de fondo (BGSubSignal): a la señal media de fluorescencia se le resta la señal producida por el ruido de fondo una vez ha sido ajustado (“BGUsed”). Este ajuste se realiza mediante un algoritmo que tiene en cuenta el ruido de fondo a nivel local (para ese punto), a nivel global (en el microarray) o en forma de gradiente espacial que pueda estar afectando la expresión del punto.
- ❖ Señal normalizada (DyeNormSig): partiendo del “BGSubSignal” se evalúan posibles sesgos que se hayan podido introducir en las medidas, debido a la diferente incorporación de los fluorocromos, la distinta cantidad de RNA de partida, etc. y los corrige mediante el método LOWESS. Este algoritmo asume que dicho sesgo es dependiente de intensidad y calcula diversos factores de normalización para diferentes rangos de intensidades, obteniendo una normalización mucho más precisa. Los puntos usados para el cálculo de los factores de normalización son aquellos marcados como consistentes por el método “Rank Consistency Probes”.
- ❖ Señal procesada (ProcessedSignal): en el caso de que el punto cumpla los criterios de “IsWellAbove” e “IsPosAndSig”, la señal procesada será la señal normalizada. En caso contrario la señal procesada será igual al límite mínimo de detección (“SurrogatedUsed”).

2. Análisis de datos.

2.1. Análisis no supervisados.

Los métodos no supervisados se basan en técnicas que agrupan u organizan los datos a partir de las propiedades intrínsecas de estos, sin referencias a informaciones adicionales o externas (ej.: supervivencia, criterios patológicos, etc.), siendo los más utilizados el “clustering” jerárquico, el “k-means” y el PCA.

La agrupación de las muestras o genes por perfiles de expresión similares se expresa en función de la distancia existente entre ellos. Por ello, previo a la ejecución de estas técnicas no supervisadas es necesario conocer las diferentes distancias existentes, los distintos métodos de agrupación de elementos y algún otro parámetro como el “centrado de genes”.

2.1.1. Distancias.

Hay muchas formas de medir la distancia entre dos sujetos, dependiendo de lo que el observador quiera estudiar y en función del tipo de datos con los que se trabaje. Los tipos de distancias más comunes son [178,379]:

- ❖ Euclídea: mide la distancia absoluta en el espacio existente entre dos elementos (genes o muestras) definidos en forma de vectores. Es una medida sensible a la magnitud y a la dirección de dichos vectores. Suele ser una medida muy útil cuando estamos interesados en medir la dimensión del cambio. Esta distancia se expresa de la forma:

$$d_{rs}^2 = \sum_{j=1}^n (x_{rj} - x_{sj})^2$$

Donde “r” y “s” son los genes (o las muestras), $x_j = 1 \dots n$ indica cada una de las dimensiones del vector para cada gen (o cada muestra).

- ❖ Manhattan: mide la distancia que deberíamos recorrer entre dos puntos si estuviésemos en un espacio (como la ciudad de Manhattan) estructurado en forma de cuadrículas

rectangulares regulares. Los resultados que proporciona son similares a los de la distancia euclídea (Fig.2). Esta distancia se expresa de la forma:

$$d_{rs} = \sum_{j=1}^n |x_{rj} - x_{sj}|$$

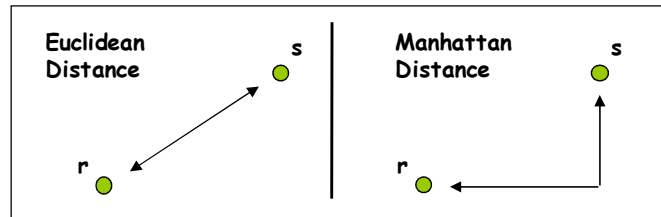
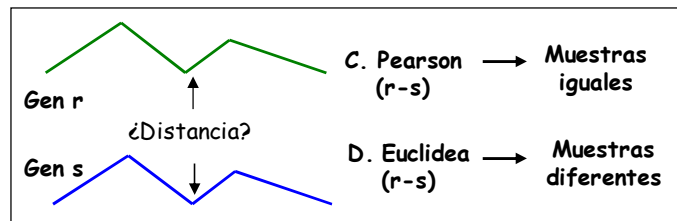


Fig.2. Representación gráfica de la diferencia en la medición de la distancia entre dos genes o muestras “r” y “s” cuando se utiliza la distancia Euclídea o la distancia Manhattan.

- ❖ Correlación centrada de Pearson: es un tipo de correlación paramétrica (se asume normalidad en la distribución de los datos) que calcula la similitud existente entre dos vectores. Es, al contrario que la distancia Euclídea, insensible a la magnitud de los vectores y muy sensible a la forma o perfil del vector de expresión de las muestras. Sus valores van desde +1 (correlación perfecta) hasta -1 (anticorrelación perfecta) pasando por el 0 (no hay correlación) (Fig.3). Esta distancia se expresa de la forma:

$$d_{rs} = \frac{\sum_{j=1}^n (x_{rj} - \bar{x}_r)(x_{sj} - \bar{x}_s)}{\sqrt{\sum_{j=1}^n (x_{rj} - \bar{x}_r)^2} \sqrt{\sum_{j=1}^n (x_{sj} - \bar{x}_s)^2}}$$

A)



B)

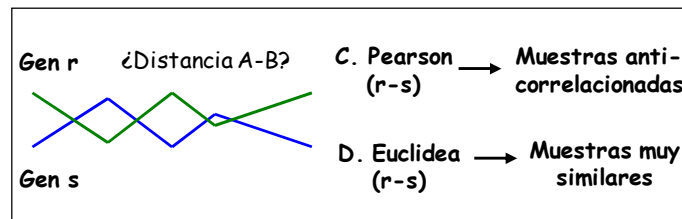


Fig.3. Los gráficos A y B muestran las diferencias a la hora de considerar la similitud o diferencia entre dos genes dependiendo de la medida seleccionada (Correlación centrada de Pearson o distancia Euclídea).

- ❖ Correlación no centrada de Pearson: es una variación de la anterior medida en la que además de detectar patrones parecidos en la expresión del vector, también detecta si hay una diferencia absoluta en la expresión de los vectores de las dos muestras o genes. Esta distancia se expresa de la forma:

$$d_{rs} = \frac{\sum_{j=1}^n (x_{rj}x_{sj})}{\sqrt{\sum_{j=1}^n (x_{rj} - \bar{x}_r)^2} \sqrt{\sum_{j=1}^n (x_{sj} - \bar{x}_s)^2}}$$

- ❖ Correlación de Spearman: es una medida de asociación no paramétrica (no asume ninguna hipótesis sobre la distribución de los datos) que organiza la matriz en rangos y establece la asociación de las muestras o genes en función de estos rangos. Da una medida similar a la obtenida con la correlación de Pearson aunque menos específicamente interpretable. Los valores siguen la escala de la correlación de Pearson (-1 → 0 → +1). Esta distancia se expresa de la forma:

$$r_s = \frac{\sum_i (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_i (R_i - \bar{R})^2} \sqrt{\sum_i (S_i - \bar{S})^2}}$$

Donde R_i es el rango para los valores “ x_{ri} ” del vector de la muestra o gen “r” y S_i es el rango para los valores “ x_{si} ” de la muestra o gen “s”.

2.1.2. Métodos de agrupación de elementos.

Una vez seleccionada la distancia que se va a utilizar hay que definir el algoritmo que nos permita estudiar la distancia entre un nodo o grupo de muestras y una muestra individual u otro nodo (Fig.4.). Cuatro son las formas más comunes [178,379]:

- ❖ Enlace simple (“single linkage”): la distancia entre dos grupos “r” y “s” es calculada como la distancia mínima entre un miembro del grupo “r” y otro del grupo “s”. Suele crear agrupaciones alargadas en forma de cadena, lo que a veces hace difícil su interpretación. Se expresa de la forma:

$$d(r,s) = \min (d(x_{ri}, x_{sj}))$$

Donde “r” y “s” son dos grupos y “ x_{ri} ” e “ x_{sj} ” son miembros del grupo “r” y “s” respectivamente.

- ❖ Enlace completo (“complete linkage”): la distancia entre dos grupos “r” y “s” es calculada como la distancia máxima entre un miembro del grupo “r” y otro del grupo “s”. Crea agrupaciones de aspecto muy compacto y bien definido que en ocasiones distorsionan las diferencias reales entre ellos. Se expresa de la forma:

$$d(r,s) = \max (d(x_{ri}, x_{sj}))$$

- ❖ Enlace promedio (“average linkage”): la distancia entre dos grupos “r” y “s” es calculada como la media del sumatorio de las distancias de cada uno de los elementos de la agrupación “r” a cada uno de los miembros de la agrupación “s”.

$$d(r,s) = (1/n_r n_s) \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} d(x_{ri}, x_{sj})$$

Donde “n” indica el número de muestras en la agrupación.

- ❖ Enlace del centroide (“centroid linkage”): la distancia entre dos grupos “r” y “s” es calculada como la distancia existente entre el centroide del grupo “r” y el centroide del grupo “s”. Dependiendo del tipo de medida seleccionada el uso del centroide puede no ser muy representativo de las agrupaciones, dando a veces resultados difícilmente interpretables. Se expresa de la forma:

$$d(r,s) = d(\bar{x}_{ri}, \bar{x}_{sj})$$

Donde,

$$\bar{x}_r = (1/n_r) \sum_{i=1}^{n_r} x_{ri}$$

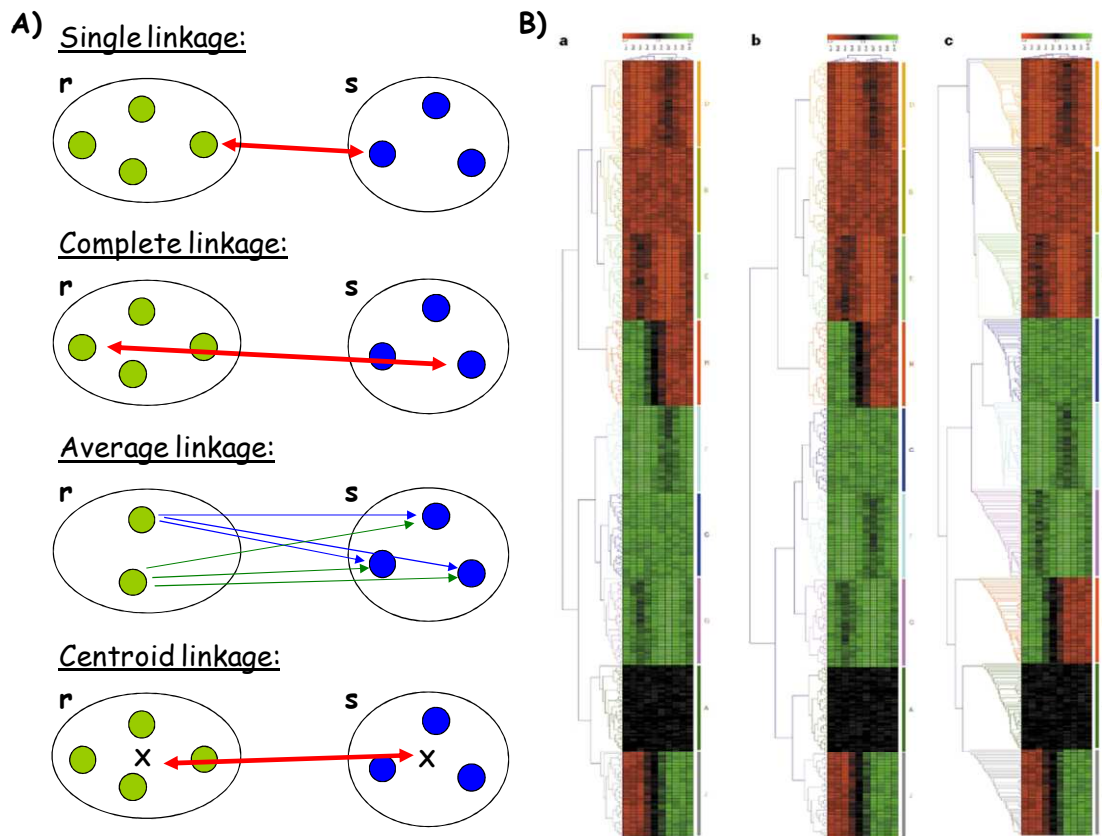


Fig.4. Representación de los métodos de medida más comunes [178] entre dos grupos “r” y “s”. A) representación gráfica de los cuatro métodos más comunes de medida de distancia. B) Ejemplo de “clustering” jerárquico de una matriz de datos utilizando “average linkage” (a), “complete linkage” (b) y “single linkage” (c).

2.1.3. Centrado de genes.

Es una transformación de la matriz de datos que permite reducir la influencia del perfil de expresión de la muestra de referencia que en muchos casos puede no ser de interés, dándonos la posibilidad de comparar en igualdad de condiciones genes que se diferencian mucho en relación a la muestra de referencia con aquellos que se diferencian poco [178].

El hecho de centrar o no los genes, tiene gran influencia sobre el resultado obtenido cuando tratamos de medir distancias entre muestras o entre genes (métodos de “clustering”), pero no así cuando realizamos comparación.

Para realizar este ajuste se calcula la media de expresión de cada gen a lo largo del conjunto de muestras. Luego el valor de expresión de ese gen en cada muestra es dividido

por la media anteriormente calculada, de tal forma que se consigue ajustar la media del gen a lo largo de las muestras a una media igual a 1. Otra forma de centrar los genes es restando a cada uno de los datos de expresión para ese gen la media de todos los valores. En este caso la media de expresión de ese gen a lo largo de las muestras se habrá ajustado a cero.

2.1.4. Métodos de análisis no supervisados

2.1.4.1. “Clustering” jerárquico.

Es un método exploratorio de análisis de datos [276] que no evalúa ninguna hipótesis a priori, si no que se limita a agrupar los sujetos a estudio en función de su similitud de manera que resulte más fácil el trabajo con dichos datos y la obtención de conclusiones. Esta técnica suele ser “aglomerativa” de manera que cada sujeto a estudio empieza formando una agrupación de una sola muestra. En un primer paso se calcula la distancia de cada muestra al resto y aquellas muestras más parecidas entre sí se unen para formar una nueva agrupación (estas agrupaciones variarán dependiendo del tipo de distancia y del método de agrupación de elementos seleccionados). El segundo paso calcula las distancias entre las nuevas agrupaciones, y aquellas más parecidas se unirán formando nuevos grupos. Este proceso se repetirá hasta que todos los sujetos formen una única agrupación o raíz. El resultado obtenido aparece en forma de árbol o dendrograma, de manera que cuanto más cortas sean las ramas que unen dos sujetos mayor será la similitud entre ellos y cuánto más cerca esté el nodo de la raíz, mayor será la heterogeneidad entre los sujetos que forman dicho nodo (ver fig.5.). Esta técnica plantea, sin embargo, ciertos problemas:

- ❖ El resultado del “cluster” es afectado por el número de genes que se están estudiando a la vez, de manera que un número excesivamente alto de estos puede enmascarar patrones de expresión interesantes y un número demasiado bajo puede crear perfiles que no se ajustan a la realidad.
- ❖ A medida que la agrupación aumenta de tamaño puede que el vector que representa a dicha agrupación no represente de forma fiel a ninguno de los sujetos que formaban el

grupo de forma que los patrones de expresión de esos genes se vuelven cada vez menos relevantes (eso ocurre por ejemplo cuando se utiliza el “centroid linkage”).

- ❖ Una mala asignación de grupos en etapas tempranas del proceso no podrá ser corregida dando lugar a estimaciones erróneas.

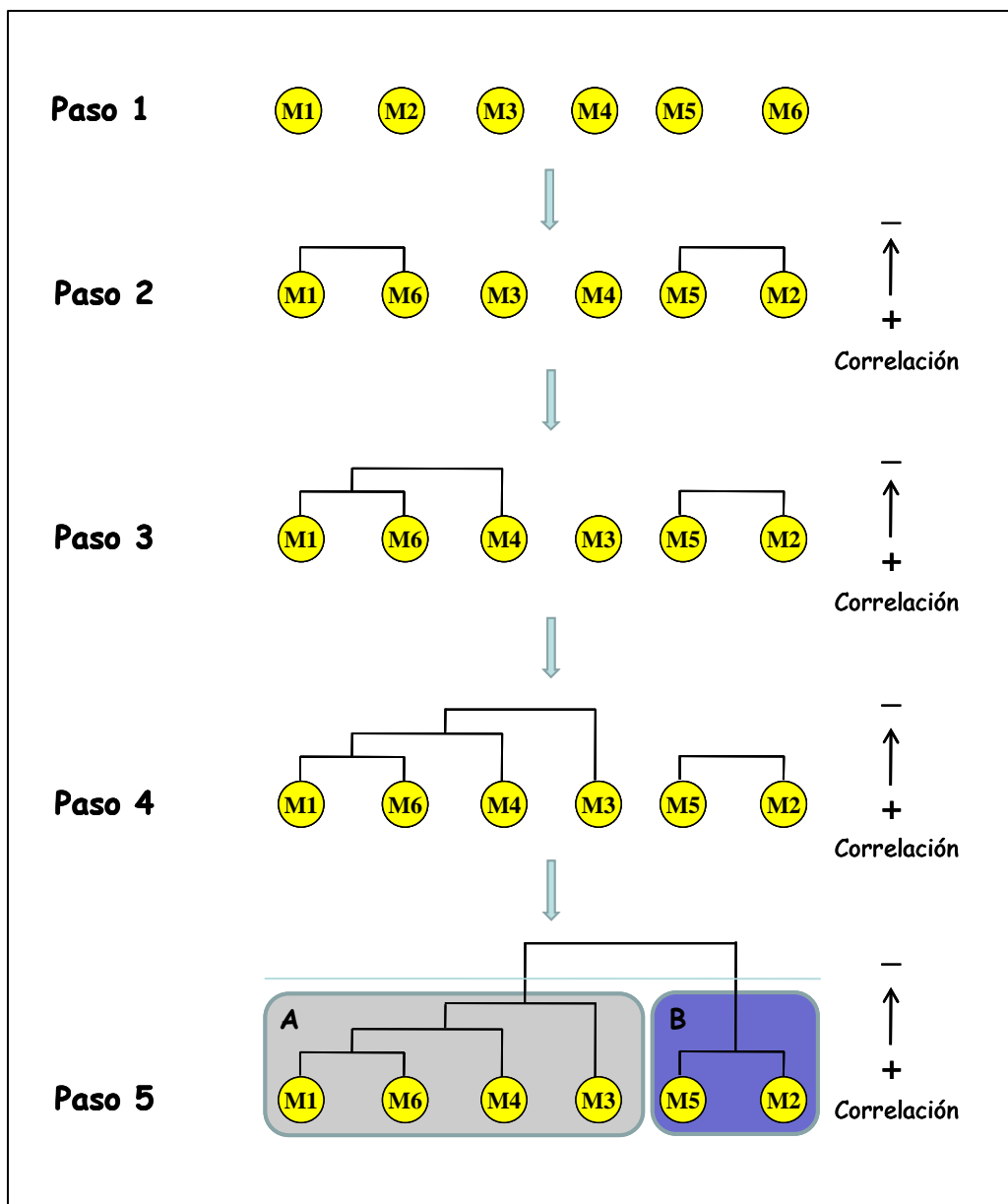


Fig.5. Esquema representativo del proceso de agrupación jerárquica (o “clustering” jerárquico) con la obtención de dos grupos o patrones de expresión destacados. M = muestra.

Otro aspecto importante es la selección del número de grupos dentro del “cluster” y la validación de dichos grupos. Aunque se han propuesto diversas técnicas no hay ninguna que se haya mostrado ampliamente efectiva a la hora de identificar dicho número exacto

de agrupaciones. Para establecer si la elección de grupos ha sido la correcta hay diversos mecanismos que permiten validar en cierta medida la elección realizada. Entre ellos destaca:

- ❖ Validación matemática: aunque esta técnica tiene que ser tomada con cautela, existen determinados algoritmos matemáticos capaces de medir la homogeneidad y reproducibilidad de las agrupaciones, lo que nos da cierta evidencia de que pueden tener un significado biológico y que no han sido obtenidas al azar.

Una de estas técnicas es la denominada “Gaussian random noise” (aplicada durante el desarrollo de esta tesis) que genera nuevas matrices de datos resultado de alterar la matriz original con datos de una distribución creada a partir del ruido de fondo experimental y las compara con la matriz original para ver si continúan reflejando los mismos patrones, obteniéndose así unos índices de confianza en las agrupaciones obtenidas. Para el cálculo de la distribución generada a partir del ruido de fondo se calcula la varianza del ratio ($\log_2(\text{ratio})$) para cada gen a lo largo de todas las muestras. La mediana de todos esos valores se utilizará como la varianza experimental estimada de una distribución normal de media cero.

Los parámetros indicativos de la consistencia de las agrupaciones son:

- Reproducibilidad del “cluster” (“R-Index”): es una estimación de la reproducibilidad de las agrupaciones, comparando el número de muestras que forma cada “cluster” en la matriz “perturbada” con respecto al número de muestras que forma cada grupo en la matriz original de datos. Una $R = 1$ significa reproducibilidad perfecta mientras que una $R = 0$ significa que no hay reproducibilidad.
- Medida de la discrepancia (“D-Index”): para el cálculo de este factor se comparan los miembros que conforman un “cluster” obtenido a partir de la matriz original con aquellos que conforman el “cluster” más parecido generado a partir de la matriz de datos “perturbada”. Aquellas muestras que aparezcan en la matriz original y no en la perturbada serán consideradas como “omisiones” y aquellas que aparezcan en la matriz perturbada y no en la original serán consideradas como

“adiciones”. Esto se realizará para cada uno de los clusters obtenidos de manera independiente.

Es importante reseñar que una vez confirmada la reproducibilidad y robustez de los grupos es fundamental encontrar el significado biológico o la posible aplicación clínica y además verificar la universalidad de dicha clasificación mediante su estudio en conjuntos de muestras independientes.

2.1.4.2. *K-means.*

Se trata de una técnica de análisis no supervisada cuya principal característica es que exige la selección previa del número de agrupaciones que se quieren obtener [380]. Esta decisión marca completamente los resultados que se obtienen, por lo que si no se tiene una certeza previa de cuál puede ser el número final de agrupaciones, se puede llegar a conclusiones erróneas. Una vez seleccionado el número de grupos, el algoritmo asigna las muestras de manera aleatoria a cada uno de los grupos. Tras estos se calcula el centroide de cada uno de los grupos y se reasigna cada muestra a un grupo según sea su proximidad a dichos centroides. Como en la técnica de “clustering” jerárquico hay que definir el tipo de distancia que se va a utilizar (euclídea, correlación de Pearson, etc.). La función de enlace que se aplicará viene definida por defecto y es el “centroide linkage”. Este proceso de asignaciones, cálculos y reasignaciones es iterativo y se repite (al menos idealmente) hasta que se encuentra una distribución de las muestras en grupos que no varía por más repeticiones que se realicen (Fig.6.). Sin embargo esto no suele ocurrir ya que habitualmente hay muestras que pueden pertenecer a dos grupos a la vez con lo que nunca se alcanza una solución final. La forma más correcta, en este caso, de obtener una confirmación del resultado obtenido es repetir la técnica un número determinado de veces, variando los parámetros de selección y viendo si los distintos resultados son similares.

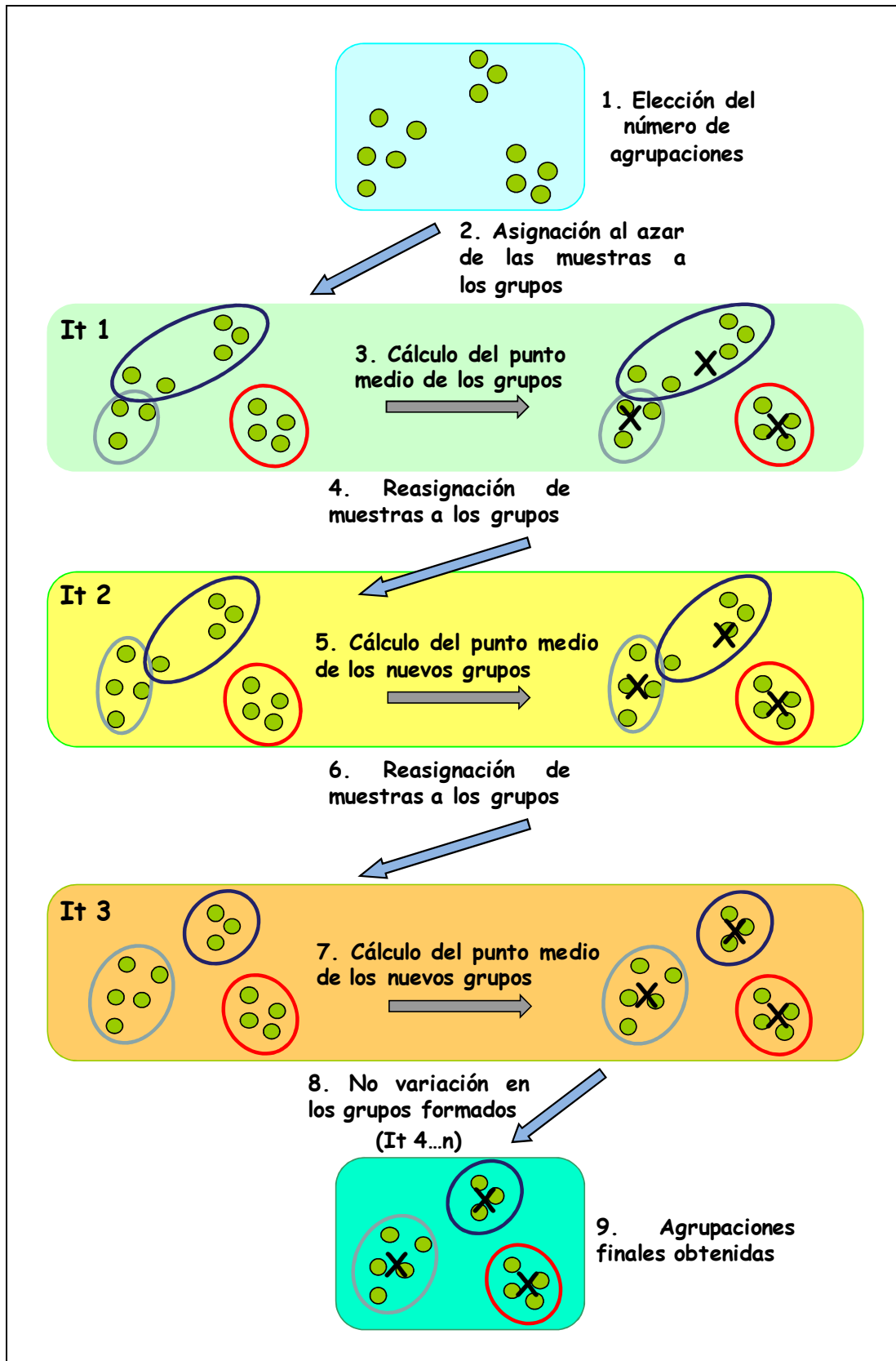


Fig.6. Esquema representativos del proceso de iteración realizado en la técnica “k-means”. A diferencia de otras técnicas como el “clustering” jerárquico, el primer paso del algoritmo especifica el número de grupos que obtendremos al final. El resultado ideal se alcanza cuando a partir de una determinada iteración el resultado que se alcanza es siempre el mismo (It 4...n en el esquema). It = iteración.

2.1.4.3. *Análisis de Componentes Principales (PCA).*

Es una técnica matemática que permite reducir la dimensionalidad de los datos mediante la convergencia de información redundante (variables correlacionadas). Para ello transformará las variables iniciales en variables no correlacionadas (convergencia de la información en un menor número de variables) reduciendo la dimensionalidad sin perder la variabilidad de los datos (Fig.7.). Este análisis nos permite de manera aproximada ver cuál es el comportamiento de las muestras [381] y la forma en la que podrían agruparse con lo que puede ser de gran ayuda en conjunción con otros métodos como el “k-means”.

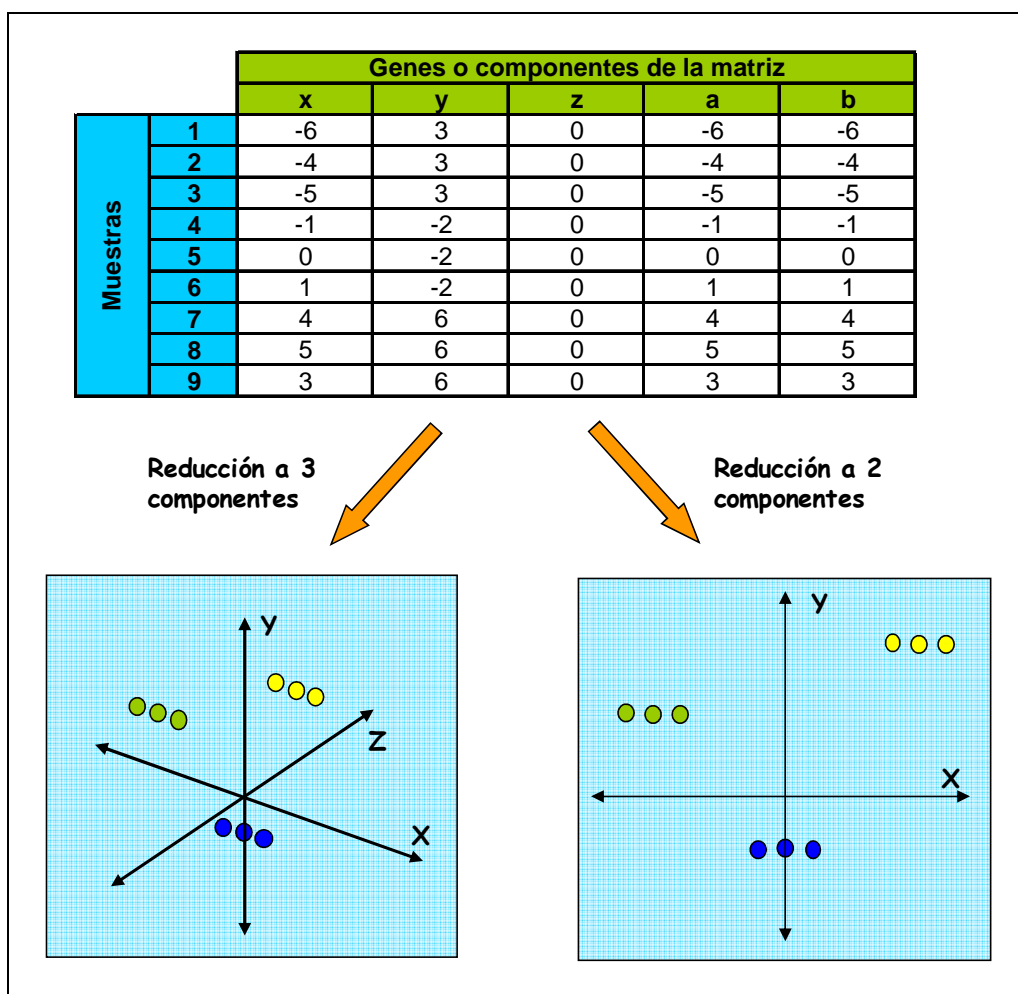


Fig.7. Ejemplo básico de la técnica de Análisis de Componentes Principales. La matriz inicial consta de 9 muestras en las que se han analizado 5 genes. En este caso, cada muestra estaría representada por un vector de 5 componentes. Un análisis de PCA realizado sobre las muestras nos permite reducir la matriz original de 5 componentes a 3 o 2, al eliminar los datos redundantes (las componentes “a” y “b” son iguales a la componente “x” y la componente “z” es igual en todas las muestras con los que no aportan más información y las tres (a, b y z) pueden ser eliminadas.

2.2. Métodos supervisados.

Este tipo de técnicas de análisis parten de un conocimiento previo de las muestras que se van a analizar (Ej.: tejido normal- tejido tumoral, enfermo-sano, respondedor- no respondedor, etc.) y normalmente, son dos los objetivos que se persiguen con su aplicación:

- ❖ Comparación de clases o comparación de grupos experimentales.
- ❖ Predicción de clases o clasificación de nuevas muestras de función de una clasificación ya conocida.

2.2.1. Comparación de clases.

En el caso de estudios de microarrays, estas técnicas se utilizan para seleccionar los perfiles de expresión génica diferencial entre grupos muestrales para los que ya se conoce alguna característica que los diferencia. Así, en nuestro caso podría ser interesante conocer qué genes se expresan de manera estadísticamente diferente entre el fenotipo normal y el fenotipo malignizado o bien entre los diferentes subtipos moleculares que se han definido. Los métodos de análisis más utilizados son:

- ❖ T-Test: analiza las diferencias entre dos grupos experimentales. Se computa mediante el cálculo del parámetro “t” según la formula:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(S_1^2/n_1) + (S_2^2/n_2)}}$$

Donde \bar{X}_1 y \bar{X}_2 son la media de los dos grupos, S_1 y S_2 son la desviación de cada una de los conjuntos muestrales y n_1 y n_2 son el tamaño muestral de cada uno de los grupos. Valores de t cercanos a 0 indicaran que no hay diferencias entre los grupos y valores elevados indicaran que si hay diferencias estadísticamente significativas entre ambos conjuntos. El valor de “t” será considerado grande o pequeño en

función de la significancia estadística que queramos obtener y del tamaño muestral (grados de libertad) que estemos manejando.

- ❖ ANOVA o F-test: son generalizaciones del t-test cuando se comparan más de dos grupos experimentales.

Nº genes analizados	p-valor seleccionado	Incidencia de falsos positivos	Probabilidad de obtener 1 falso positivo
1	0.05	1/20	5.0%
2	0.05	1/10	10.0%
20	0.05	1	64.0%
100	0.05	5	99.4%

Tabla.1. Representación del efecto que tiene la comparación de varias variables en la probabilidad de obtener falsos positivos.

2.2.1.1. Método de corrección para comparaciones múltiples.

El análisis de un elevado número de variables exige además que se apliquen sobre estas técnicas estadísticas métodos de corrección múltiples que tengan en cuenta el número de variables analizadas (Tabla.1). Las más comunes son:

- ❖ Método de Bonferroni: esta técnica corrige el p-valor calculado para las distintas comparaciones, multiplicándolo por el número de comparaciones realizadas. Cuando el número de variables analizadas es muy elevado, esta corrección es demasiado restrictiva, imposibilitando la obtención de diferencias significativas y obteniendo falsos negativos.
- ❖ Método de “Benjamini and Hochberg”: esta técnica está especialmente diseñada para análisis en los que se analizan miles de variables como son los experimentos de microarrays. De manera resumida, esta técnica ordena de menor a mayor los p-valores obtenidos para las distintas variables. El p-valor más elevado permanece con su valor original y se analiza si es significativo o no. Para el siguiente p-valor más elevado se realiza la corrección de la siguiente manera: $p\text{-valor corregido} = (p\text{-valor} * n) / a$, siendo “n” el número total de variables analizadas y “a” la posición que ese p-valor posee en el ranking ordenado de más significativo (p-valor menor) a menos significativo (p-valor mayor) (Tabla.2).

Gen	p-valor ordenado	Nº total de variables	Posición en el ranking	p-valor corregido	¿Significativo?
A	0.1	1000	500	$(0.1 * 1000) / 500 = 0.2$	No
B	0.075	1000	100	$(0.075 * 1000) / 100 = 0.75$	No
C	0.0006	1000	10	$(0.0006 * 1000) / 10 = 0.06$	No
D	0.00003	1000	1	$(0.00003 * 1000) / 1 = 0.03$	Si

Tabla.2. Efecto de la técnica de Benjamini and Hochberg sobre el p-valor en procesos corrección para comparaciones múltiples.

2.2.1.2. Análisis funcional.

Este tipo de análisis, perteneciente al campo de la comparación de clases, examina si conjuntos de genes, normalmente pertenecientes a una misma vía molecular (Ej. Genes de la vía Wnt) o a una misma localización celular (Ej. Genes de matriz extracelular), se encuentran diferencialmente expresados entre dos o más grupos. Tienen como ventaja principal, en relación a una comparación de clases tradicional (utilizando genes a nivel individual), que proporciona una imagen más global de los procesos que han sido alterados entre los grupos que se están estudiando. Una descripción de los test estadísticos que se han utilizado para llevar a cabo este tipo de comparaciones se muestra a continuación:

- ❖ “LS y KS test”: en ambos test y para cada conjunto de genes, se analiza el p-valor de cada gen individual que forma parte de dicho conjunto. Posteriormente se calculan los estadísticos:

$$LS = \sum_{i=1}^N (-\log(p_i)) / N$$

$$KS = \max_{i=1}^N (i/N - p_i)$$

donde “N” es el número de genes que componen el conjunto de genes que se está analizando y “p” es el p-valor. En el caso del Test KS, $p_1 \leq p_2 \leq \dots \leq p_N$ y p_i siempre hace referencia al p-valor más pequeño del test univariante.

Posteriormente se calcula la significancia estadística de conjuntos de genes del mismo tamaño, seleccionados al azar (100.000 repeticiones) y se calculan los correspondientes estadísticos. El p-valor final para estos dos test indica la proporción de conjuntos de

genes seleccionados al azar con un estadístico (LS o KS) superior al de la lista a analizar.

- ❖ “Goeman’s global test”: comprueba para cada conjunto de genes, si su perfil de expresión está significativamente relacionado con el parámetro que se está analizando (Ej. Tejido normal vs. tejido tumoral). Este test está descrito en los trabajos de Goeman de 2004 y 2005 [382,383].
- ❖ “Efron Tibshirani’s Gene Set Enrichment Análisis” (GSEA): brevemente esta técnica consta de los siguientes pasos [384]: 1) se calcula cuál es la correlación del conjunto de genes (N) que se está analizando en relación al parámetro estudiado. 2) se ordenan los genes de este conjunto (N) dentro de la lista completa de genes con la que se inició el estudio, en función de la correlación con el parámetro a estudio. 3) Se calcula el nivel de enriquecimiento de nuestro conjunto de genes dentro de la lista completa de genes. Este enriquecimiento será mayor cuanto más genes del conjunto (N) analizado estén en los extremos de la lista (correlación directa o inversa con el parámetro a estudio). 4) se compara el nivel de enriquecimiento encontrado en el conjunto de genes a estudio con los enriquecimientos obtenidos para conjuntos de genes del mismo tamaño elegidos al azar y se analiza si existe una diferencia estadísticamente significativa.

2.2.2. Predicción de clases.

Es una técnica de análisis supervisada que permite, una vez que se conocen las diferencias entre 2 o más grupos muestrales, crear un algoritmo capaz de clasificar nuevas muestras en función de la similitud encontrada con los grupos previamente definidos. Los algoritmos capaces de crear este tipo de predictores son muy numerosos: “Support Vector Machines”, “Neural Networks”, “3-Nearest Neighbor”, “Decisión Trees”, etc. No es nuestro propósito hacer una revisión sobre los mismos, por lo que a continuación solo se detallaran brevemente aquellos que se han usado durante el presente trabajo (<http://linus.nci.nih.gov/BRB-ArrayTools.html>):

- ❖ “1-Nearest Neighbor” (1NN) / “3-Nearest Neighbor” (3NN): Se basa en determinar que perfil de expresión (1NN o la muestra más cercana) o que perfiles (3NN o las 3

muestra más cercanas) de las muestras que conforman la matriz de original es más parecido al de la nueva muestra que se está analizando. La nueva muestra será por tanto clasificada con dicho perfil/es. Utiliza la distancia Euclídea para llevar a cabo los cálculos de la predicción. Se puede utilizar cuando hay más de dos grupos en el clasificador.

- ❖ “Nearest Centroid” (NC): se puede usar para construir predictores cuando se trabaja con más de 2 grupos de muestras. Para cada uno de los grupos definidos se calcula el centroide de expresión. Posteriormente se calcula la distancia de la nueva muestra que se quiere clasificar con respecto a los centroides de cada grupo. La muestra será clasificada en el grupo que se encuentre a menor distancia de ella. Se utilizará la distancia euclídea.
- ❖ “Support Vector Machine” (SVM): esta técnica esta implementada para crear clasificadores que diferencien entre dos grupos de muestras. La idea principal se centra en encontrar un plano que separe en el espacio los dos grupos de muestras, maximizando la distancia entre ellos.
- ❖ “Compound Covariate Predictor” (CCP): solo es válido cuando se trata de predicciones entre dos grupos. Establece una combinación lineal de los Log (ratios) de los genes incluidos en el clasificador (en la que cada gen tiene un peso específico dentro del predictor). El resultado que se obtiene es de signo diferente (+ o -) para cada uno de los dos grupos. Esta combinación lineal será aplicada a los datos de expresión de la nueva muestra obteniéndose un resultado que se asociara con uno de los dos grupos.
- ❖ “Diagonal Linear Discriminant Analysis” (DLDA): es similar al CCP, aunque ignora la correlación existente entre los genes para evitar el sobreajuste del clasificador. Sirve para crear predictores para más de dos grupos de muestras.

El hecho de ser capaces de crear un predictor, no indica que este sea efectivo a la hora de clasificar nuevas muestras. Para validarlo es necesario usar una matriz de muestras independientes o matriz de validación de la que ya se conoce la distribución de las muestras y ver cual es la tasa de error del clasificador. En situaciones en las que no se

dispone de una matriz de validación independiente, se pueden utilizar técnicas de validación cruzada en las que se utilizan como matriz de validación las propias muestras de la matriz de entrenamiento. Las dos técnicas más comunes se exponen a continuación.

2.2.2.1. Técnicas de validación cruzada.

- ❖ **“K-fold validation”**: es un método de validación que se aplica sobre la propia matriz de experimentación (ME) para estimar la precisión del clasificador creado. En este caso, la ME se divide “k” veces, obteniéndose “k” grupos de tamaño similar. El clasificador se creará un total de “k” veces, excluyendo en cada una de ellas uno de los grupos de muestras que formarán la matriz de validación (MV). En cada ocasión, las muestras que permanecen en la ME crearán un predictor que será testado sobre la MV. La suma de los errores, dará el error global de la técnica. Este método es efectivo cuando se utiliza un tamaño muestral grande. El número de genes del predictor final será el creado con un predictor en el que se incluyan todas las muestras.

- ❖ **“Leave One Out Cross-Validation” (LOOCV)**: es otro método de validación que se aplica sobre la propia matriz de experimentación para estimar la precisión del clasificador creado. Este método es muy útil cuando se trabaja con tamaños muestrales bajos. En este caso para un conjunto muestral de “n” muestras se van a realizar “n-1” predictores a partir de “n-1” muestras. La muestra restante en cada caso será usada como MV para computar la tasa de error global.

Finalmente, conviene calcular si el predictor creado a pesar de la precisión obtenida, clasifica mejor que un clasificador creado al azar. Esto se puede comprobar mediante el cálculo del parámetro:

- ❖ **“Permutation p-value for cross-validated misclassification error rate”**: es un indicador de la probabilidad de obtener un clasificador al azar que tenga una tasa de error menor a la encontrada en el clasificador que se está analizando.

Anexo II.
Predictores de supervivencia publicados.

1. Coincidencia entre los predictores de supervivencia y nuestra matriz de datos.

Predictor Eschrich	
ID	NAME
NM_016133	INSIG2
NM_016227	C1orf9
NM_016491	MRPL37
NM_003262	TLOC1
NM_002270	TNPO1
NM_015340	LARS2
NM_012328	DNAJB9
NM_003569	STX7
NM_173574	ZNF683
NM_153607	LOC153222
NM_014571	HEYL
NM_001645	APOC1
NM_016651	DACT1
NM_000582	SPP1
NM_020755	SERINC1
NM_001542	IGSF3
NM_002094	GSPT1

Predictor Wang	
ID	NAME
NM_003405	YWHAH
NM_004063	CDH17
NM_015655	ZNF337
NM_001747	CAPG
NM_014387	LAT
NM_001269	RCC1
NM_001645	APOC1
NM_001730	KLF5
NM_001271	CHD2
NM_012345	NUFIP1
NM_006864	LILRB3

Predictor Lin (N. Zeland)	
ID	NAME
NM_012087	GTF3C5
NM_002717	PPP2R2A
NM_032547	SCOC
NM_002813	PSMD9
NM_014016	SACM1L
NM_016836	RBMS1
NM_015642	ZBTB20
NM_021188	ZNF410
NM_012484	HMMR
NM_003753	EIF3S7
NM_203505	G3BP2
NM_003101	SOAT1
NM_080425	GNAS
NM_002802	PSMC1
NM_001001396	ATP2B4
NM_018492	PBK
NM_003347	UBE2L3

Tabla.1. Listado de genes pertenecientes a los predictores publicados en los trabajos de Eschrich S [252], Wang Y [253] y Lin YH [260] que estaban presentes en el conjunto de 1722 genes utilizado para la construcción de nuestra clasificación molecular.

Anexo III.

Listado genes meta-análisis de Chan et al. [357].

1. Listado de genes publicado en el estudio de Chan SK et al.

Sobre-expresión		Inhibición	
Gen	Dif Exp	Gen	Dif Exp
TGFβI	Si	CA2	Si
MYC	Si	CEACAM1	Si
SPARC	Si	HSD11B2	Si
GDF15	Si	SLC26A2	Si
CXCL1	Si	FCGBP	Si
TOP2A	Si	ACADS	Si
UBE2C	Si	CKB	Si
INHBA	Si	CLU	Si
MMP11	Si	CES2	Si
IFITM1	T	CA1	Si
CDC25B	T	GPA33	Si
HMBG1	T	KRT20	Si
IFITM2	T	SELENBP1	Si
COL1A2	T	CA12	Si
CKS2	T	MALL	T
CDH3	T	FABP1	X
SLC12A2	T		
GSE1L	T		
COL3A1	T		
COL4A1	T		
HNRPA1	X		
CDK10	X		

Tabla.1- Listado de genes publicado en el estudio de Chan SK [357] cuya expresión diferencial aparecía descrita en al menos 5 de los 23 trabajos analizados. A la derecha de cada gen se muestra si dicho gen se encontraba en la lista de 1722 genes examinados con expresión diferencial entre tejido tumoral y tejido normal (Si), si no estaban entre los 1722 seleccionados, pero mostraban el mismo tipo de diferencias indicadas en el artículo de Chan (Tendencia o T) o si dicho gen no se encontraba en la matriz de datos utilizada (X).

Bibliografía

1. **Nystrom M, Mutanen M.** *Diet and epigenetics in colon cancer.* World J Gastroenterol 2009;15:257-63.
2. **Weitz J, Koch M, et al.** *Colorectal cancer.* Lancet 2005;365:153-65.
3. **López-Abente G.** *La situación del cáncer en España.* Ministerio de sanidad y consumo centro de publicaciones 2005;ISBN: 84-7670-673-1.
4. **Gatalica Z, Torlakovic E.** *Pathology of the hereditary colorectal carcinoma.* Fam Cancer 2008;7:15-26.
5. **Grady WM, Carethers JM.** *Genomic and epigenetic instability in colorectal cancer pathogenesis.* Gastroenterology 2008;135:1079-99.
6. **Meyerhardt JA, Niedzwiecki D, et al.** *Association of dietary patterns with cancer recurrence and survival in patients with stage III colon cancer.* Jama 2007;298:754-64.
7. **Dukes C.** *The classification of cancer of the rectum.* J Pathol Bacteriol 1932;35:323-31.
8. **Greene F, Page D, et al.** *AJCC Cancer Staging Manual.* 6th ed. New York: Springer 2002.
9. **Nelson H, Petrelli N, et al.** *Guidelines 2000 for colon and rectal cancer surgery.* J Natl Cancer Inst 2001;93:583-96.
10. **Quirke P, Williams GT, et al.** *The future of the TNM staging system in colorectal cancer: time for a debate?* Lancet Oncol 2007;8:651-7.
11. **Sobin L, Wittekind C.** *UICC TNM Classification of Malignant Tumors.* Tumors. 6th ed. New York: John Wiley & Sons, Inc 2002.
12. **Mutch MG.** *Molecular profiling and risk stratification of adenocarcinoma of the colon.* J Surg Oncol 2007;96:693-703.
13. **Saltz LB.** *Biomarkers in colorectal cancer: added value or just added expense?* Expert Rev Mol Diagn 2008;8:231-3.
14. **Zlobec I, Lugli A.** *Prognostic and predictive factors in colorectal cancer.* J Clin Pathol 2008;61:561-9.
15. **Duffy MJ, van Dalen A, et al.** *Tumour markers in colorectal cancer: European Group on Tumour Markers (EGTM) guidelines for clinical use.* Eur J Cancer 2007;43:1348-60.
16. **Compton C, Fenoglio-Preiser CM, et al.** *American Joint Committee on Cancer Prognostic Factors Consensus Conference: Colorectal Working Group.* Cancer 2000;88:1739-57.

17. **Broders A.** *The grading of carcinoma.* . Minn Med. 8:726-30. 1925.
18. **Fujita S, Shimoda T, et al.** *Prospective evaluation of prognostic factors in patients with colorectal cancer undergoing curative resection.* J Surg Oncol 2003;84:127-31.
19. **Compton CC.** *Colorectal carcinoma: diagnostic, prognostic, and molecular features.* Mod Pathol 2003;16:376-88.
20. **Popat S, Hubner R, et al.** *Systematic review of microsatellite instability and colorectal cancer prognosis.* J Clin Oncol 2005;23:609-18.
21. **Ribic CM, Sargent DJ, et al.** *Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer.* N Engl J Med 2003;349:247-57.
22. **Wiratkapun S, Kraemer M, et al.** *High preoperative serum carcinoembryonic antigen predicts metastatic recurrence in potentially curative colonic cancer: results of a five-year study.* Dis Colon Rectum 2001;44:231-5.
23. **Duffy MJ, van Dalen A, et al.** *Clinical utility of biochemical markers in colorectal cancer: European Group on Tumour Markers (EGTM) guidelines.* Eur J Cancer 2003;39:718-27.
24. **Jen J, Kim H, et al.** *Allelic loss of chromosome 18q and prognosis in colorectal cancer.* N Engl J Med 1994;331:213-21.
25. **Munro AJ, Lain S, et al.** *P53 abnormalities and outcomes in colorectal cancer: a systematic review.* Br J Cancer 2005;92:434-44.
26. **Andreyev HJ, Norman AR, et al.** *Kirsten ras mutations in patients with colorectal cancer: the 'RASCAL II' study.* Br J Cancer 2001;85:692-6.
27. **Gnanasampanthan G, Elsaleh H, et al.** *Ki-ras mutation type and the survival benefit from adjuvant chemotherapy in Dukes' C colorectal cancer.* J Pathol 2001;195:543-8.
28. **Karapetis CS, Khambata-Ford S, et al.** *K-ras mutations and benefit from cetuximab in advanced colorectal cancer.* N Engl J Med 2008;359:1757-65.
29. **Allen WL, Johnston PG.** *Role of genomic markers in colorectal cancer treatment.* J Clin Oncol 2005;23:4545-52.
30. **Galizia G, Lieto E, et al.** *Prognostic significance of epidermal growth factor receptor expression in colon cancer patients undergoing curative surgery.* Ann Surg Oncol 2006;13:823-35.
31. **Cohen SJ, Punt CJ, et al.** *Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer.* J Clin Oncol 2008;26:3213-21.

-
32. **Fearon ER, Vogelstein B.** *A genetic model for colorectal tumorigenesis.* Cell 1990;61:759-67.
 33. **Bodmer W, Bishop T, et al.** *Genetic steps in colorectal cancer.* Nat Genet 1994;6:217-9.
 34. **Hanahan D, Weinberg RA.** *The hallmarks of cancer.* Cell 2000;100:57-70.
 35. **Kinzler KW, Vogelstein B.** *Lessons from hereditary colorectal cancer.* Cell 1996;87:159-70.
 36. **Nagase H, Nakamura Y.** *Mutations of the APC (adenomatous polyposis coli) gene.* Hum Mutat 1993;2:425-34.
 37. **Vogelstein B, Fearon ER, et al.** *Genetic alterations during colorectal-tumor development.* N Engl J Med 1988;319:525-32.
 38. **van Es JH, Giles RH, et al.** *The many faces of the tumor suppressor gene APC.* Exp Cell Res 2001;264:126-34.
 39. **Aoki K, Taketo MM.** *Adenomatous polyposis coli (APC): a multi-functional tumor suppressor gene.* J Cell Sci 2007;120:3327-35.
 40. **Rusan NM, Peifer M.** *Original CIN: reviewing roles for APC in chromosome instability.* J Cell Biol 2008;181:719-26.
 41. **Esteller M, Herman JG.** *Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours.* J Pathol 2002;196:1-7.
 42. **Chung DC.** *The genetic basis of colorectal cancer: insights into critical pathways of tumorigenesis.* Gastroenterology 2000;119:854-65.
 43. **Luongo C, Dove WF.** *Somatic genetic events linked to the Apc locus in intestinal adenomas of the Min mouse.* Genes Chromosomes Cancer 1996;17:194-8.
 44. **Velcich A, Corner G, et al.** *Altered phenotype of HT29 colonic adenocarcinoma cells following expression of the DCC gene.* Oncogene 1999;18:2599-606.
 45. **Chen YQ, Hsieh JT, et al.** *Induction of apoptosis and G2/M cell cycle arrest by DCC.* Oncogene 1999;18:2747-54.
 46. **Chuong CM, Jiang TX, et al.** *cDCC (chicken homologue to a gene deleted in colorectal carcinoma) is an epithelial adhesion molecule expressed in the basal cells and involved in epithelial-mesenchymal interaction.* Dev Biol 1994;164:383-97.
 47. **Mehlen P, Rabizadeh S, et al.** *The DCC gene product induces apoptosis by a mechanism requiring receptor proteolysis.* Nature 1998;395:801-4.

48. **Mehlen P, Fearon ER.** *Role of the dependence receptor DCC in colorectal cancer pathogenesis.* J Clin Oncol 2004;22:3420-8.
49. **Yin JJ, Selander K, et al.** *TGF-beta signaling blockade inhibits PTHrP secretion by breast cancer cells and bone metastases development.* J Clin Invest 1999;103:197-206.
50. **Akhurst RJ, Derynck R.** *TGF-beta signaling in cancer--a double-edged sword.* Trends Cell Biol 2001;11:S44-51.
51. **Torre-Amione G, Beauchamp RD, et al.** *A highly immunogenic tumor transfected with a murine transforming growth factor type beta 1 cDNA escapes immune surveillance.* Proc Natl Acad Sci U S A 1990;87:1486-90.
52. **Hickman ES, Moroni MC, et al.** *The role of p53 and pRB in apoptosis and cancer.* Curr Opin Genet Dev 2002;12:60-6.
53. **Kastan MB.** *Wild-type p53: tumors can't stand it.* Cell 2007;128:837-40.
54. **Hahn WC, Weinberg RA.** *Rules for making human tumor cells.* N Engl J Med 2002;347:1593-603.
55. **Aaltonen LA, Peltomaki P, et al.** *Clues to the pathogenesis of familial colorectal cancer.* Science 1993;260:812-6.
56. **Thibodeau SN, Bren G, et al.** *Microsatellite instability in cancer of the proximal colon.* Science 1993;260:816-9.
57. **Ionov Y, Peinado MA, et al.** *Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis.* Nature 1993;363:558-61.
58. **Thibodeau SN, French AJ, et al.** *Microsatellite instability in colorectal cancer: different mutator phenotypes and the principal involvement of hMLH1.* Cancer Res 1998;58:1713-8.
59. **Dionigi G, Bianchi V, et al.** *Differences between familial and sporadic forms of colorectal cancer with DNA microsatellite instability.* Surg Oncol 2007;16 Suppl 1:S37-42.
60. **Duval A, Hamelin R.** *Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability.* Cancer Res 2002;62:2447-54.
61. **Eshleman JR, Casey G, et al.** *Chromosome number and structure both are markedly stable in RER colorectal cancers and are not destabilized by mutation of p53.* Oncogene 1998;17:719-25.
62. **Olschwang S, Hamelin R, et al.** *Alternative genetic pathways in colorectal carcinogenesis.* Proc Natl Acad Sci U S A 1997;94:12122-7.

-
63. **Lanza G, Ferracin M, et al.** *mRNA/microRNA gene expression profile in microsatellite unstable colorectal cancer.* Mol Cancer 2007;6:54.
 64. **Soreide K, Janssen EA, et al.** *Microsatellite instability in colorectal cancer.* Br J Surg 2006;93:395-406.
 65. **Jass JR.** *Serrated adenoma of the colorectum and the DNA-methylator phenotype.* Nat Clin Pract Oncol 2005;2:398-405.
 66. **Dong SM, Lee EJ, et al.** *Progressive methylation during the serrated neoplasia pathway of the colorectum.* Mod Pathol 2005;18:170-8.
 67. **Hawkins N, Norrie M, et al.** *CpG island methylation in sporadic colorectal cancers and its relationship to microsatellite instability.* Gastroenterology 2002;122:1376-87.
 68. **Kambara T, Simms LA, et al.** *BRAF mutation is associated with DNA methylation in serrated polyps and cancers of the colorectum.* Gut 2004;53:1137-44.
 69. **Nagasaka T, Sasamoto H, et al.** *Colorectal cancer with mutation in BRAF, KRAS, and wild-type with respect to both oncogenes showing different patterns of DNA methylation.* J Clin Oncol 2004;22:4584-94.
 70. **Kawasaki T, Nosho K, et al.** *Correlation of beta-catenin localization with cyclooxygenase-2 expression and CpG island methylator phenotype (CIMP) in colorectal cancer.* Neoplasia 2007;9:569-77.
 71. **Ward RL, Cheong K, et al.** *Adverse prognostic effect of methylation in colorectal cancer is reversed by microsatellite instability.* J Clin Oncol 2003;21:3729-36.
 72. **Ogino S, Kawasaki T, et al.** *Evaluation of markers for CpG island methylator phenotype (CIMP) in colorectal cancer by a large population-based sample.* J Mol Diagn 2007;9:305-14.
 73. **Jass JR.** *Classification of colorectal cancer based on correlation of clinical, morphological and molecular features.* Histopathology 2007;50:113-30.
 74. **van Es JH, Barker N, et al.** *You Wnt some, you lose some: oncogenes in the Wnt signaling pathway.* Curr Opin Genet Dev 2003;13:28-33.
 75. **van Amerongen R, Mikels A, et al.** *Alternative wnt signaling is initiated by distinct receptors.* Sci Signal 2008;1:re9.
 76. **Caldwell GM, Jones CE, et al.** *Reorganisation of Wnt-response pathways in colorectal tumorigenesis.* Br J Cancer 2008;98:1437-42.
 77. **Behrens J.** *The role of the Wnt signalling pathway in colorectal tumorigenesis.* Biochem Soc Trans 2005;33:672-5.

-
78. **Polakis P.** *Wnt signaling and cancer.* Genes Dev 2000;14:1837-51.
79. **Morin PJ, Sparks AB, et al.** *Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC.* Science 1997;275:1787-90.
80. **He TC, Sparks AB, et al.** *Identification of c-MYC as a target of the APC pathway.* Science 1998;281:1509-12.
81. **Tetsu O, McCormick F.** *Beta-catenin regulates expression of cyclin D1 in colon carcinoma cells.* Nature 1999;398:422-6.
82. **Koh TJ, Bulitta CJ, et al.** *Gastrin is a target of the beta-catenin/TCF-4 growth-signaling pathway in a model of intestinal polyposis.* J Clin Invest 2000;106:533-9.
83. **Wilson CL, Heppner KJ, et al.** *Intestinal tumorigenesis is suppressed in mice lacking the metalloproteinase matrilysin.* Proc Natl Acad Sci U S A 1997;94:1402-7.
84. **Conacci-Sorrell ME, Ben-Yedidia T, et al.** *Nr-CAM is a target gene of the beta-catenin/LEF-1 pathway in melanoma and colon cancer and its expression enhances motility and confers tumorigenesis.* Genes Dev 2002;16:2058-72.
85. **Buo L, Meling GI, et al.** *Antigen levels of urokinase plasminogen activator and its receptor at the tumor-host interface of colorectal adenocarcinomas are related to tumor aggressiveness.* Hum Pathol 1995;26:1133-8.
86. **Zhang T, Otevrel T, et al.** *Evidence that APC regulates survivin expression: a possible mechanism contributing to the stem cell origin of colon cancer.* Cancer Res 2001;61:8664-7.
87. **Fuchs SY, Ougolkov AV, et al.** *Oncogenic beta-catenin signaling networks in colorectal cancer.* Cell Cycle 2005;4:1522-39.
88. **Easwaran V, Lee SH, et al.** *beta-Catenin regulates vascular endothelial growth factor expression in colon cancer.* Cancer Res 2003;63:3145-53.
89. **Boon EM, van der Neut R, et al.** *Wnt signaling regulates expression of the receptor tyrosine kinase met in colorectal cancer.* Cancer Res 2002;62:5126-8.
90. **Castellone MD, Teramoto H, et al.** *Cyclooxygenase-2 and colorectal cancer chemoprevention: the beta-catenin connection.* Cancer Res 2006;66:11085-8.
91. **Tian X, Du H, et al.** *Smad4 restoration leads to a suppression of Wnt/beta-catenin signaling activity and migration capacity in human colon carcinoma cells.* Biochem Biophys Res Commun 2009;380:478-83.
92. **Labbe E, Lock L, et al.** *Transcriptional cooperation between the transforming growth factor-beta and Wnt pathways in mammary and intestinal tumorigenesis.* Cancer Res 2007;67:75-84.

-
93. **van den Brink GR.** *Linking pathways in colorectal cancer.* Nat Genet 2004;36:1038-9.
 94. **Mueller MM, Fusenig NE.** *Friends or foes - bipolar effects of the tumour stroma in cancer.* Nat Rev Cancer 2004;4:839-49.
 95. **Dvorak HF.** *Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing.* N Engl J Med 1986;315:1650-9.
 96. **Shekhar MP, Pauley R, et al.** *Host microenvironment in breast cancer development: extracellular matrix-stromal cell contribution to neoplastic phenotype of epithelial cells in the breast.* Breast Cancer Res 2003;5:130-5.
 97. **Ronnov-Jessen L, Petersen OW, et al.** *Cellular changes involved in conversion of normal to malignant breast: importance of the stromal reaction.* Physiol Rev 1996;76:69-125.
 98. **Elenbaas B, Weinberg RA.** *Heterotypic signaling between epithelial tumor cells and fibroblasts in carcinoma formation.* Exp Cell Res 2001;264:169-84.
 99. **Lorusso G, Ruegg C.** *The tumor microenvironment and its contribution to tumor evolution toward metastasis.* Histochem Cell Biol 2008;130:1091-103.
 100. **Weinberg R.** *The Biology of cancer.* Garland Science, Taylor & Francis Group. 2007.
 101. **Galon J, Fridman WH, et al.** *The adaptive immunologic microenvironment in colorectal cancer: a novel perspective.* Cancer Res 2007;67:1883-6.
 102. **Rivoltini L, Carrabba M, et al.** *Immunity to cancer: attack and escape in T lymphocyte-tumor cell interaction.* Immunol Rev 2002;188:97-113.
 103. **Sanchez-Perez L, Kottke T, et al.** *Potent selection of antigen loss variants of B16 melanoma following inflammatory killing of melanocytes in vivo.* Cancer Res 2005;65:2009-17.
 104. **Drake CG, Jaffee E, et al.** *Mechanisms of immune evasion by tumors.* Adv Immunol 2006;90:51-81.
 105. **Cabrera T, Collado A, et al.** *High frequency of altered HLA class I phenotypes in invasive colorectal carcinomas.* Tissue Antigens 1998;52:114-23.
 106. **Maleno I, Cabrera CM, et al.** *Distribution of HLA class I altered phenotypes in colorectal carcinomas: high frequency of HLA haplotype loss associated with loss of heterozygosity in chromosome region 6p21.* Immunogenetics 2004;56:244-53.
 107. **Fukushima Y, Oshika Y, et al.** *Increased expression of human histocompatibility leukocyte antigen-G in colorectal cancer cells.* Int J Mol Med 1998;2:349-51.

-
108. **Medema JP, de Jong J, et al.** *Blockade of the granzyme B/perforin pathway through overexpression of the serine protease inhibitor PI-9/SPI-6 constitutes a mechanism for immune escape by tumors.* Proc Natl Acad Sci U S A 2001;98:11515-20.
 109. **Van Geelen CM, de Vries EG, et al.** *Lessons from TRAIL-resistance mechanisms in colorectal cancer cells: paving the road to patient-tailored therapy.* Drug Resist Updat 2004;7:345-58.
 110. **Durrant LG, Chapman MA, et al.** *Enhanced expression of the complement regulatory protein CD55 predicts a poor prognosis in colorectal cancer patients.* Cancer Immunol Immunother 2003;52:638-42.
 111. **Igney FH, Krammer PH.** *Immune escape of tumors: apoptosis resistance and tumor counterattack.* J Leukoc Biol 2002;71:907-20.
 112. **Sakaguchi S, Sakaguchi N, et al.** *Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases.* J Immunol 1995;155:1151-64.
 113. **Wolf AM, Wolf D, et al.** *Increase of regulatory T cells in the peripheral blood of cancer patients.* Clin Cancer Res 2003;9:606-12.
 114. **Ghiringhelli F, Larmonier N, et al.** *CD4+CD25+ regulatory T cells suppress tumor immunity but are sensitive to cyclophosphamide which allows immunotherapy of established tumors to be curative.* Eur J Immunol 2004;34:336-44.
 115. **O'Hara RJ, Greenman J, et al.** *Advanced colorectal cancer is associated with impaired interleukin 12 and enhanced interleukin 10 production.* Clin Cancer Res 1998;4:1943-8.
 116. **Kaminska B, Wesolowska A, et al.** *TGF beta signalling and its role in tumour pathogenesis.* Acta Biochim Pol 2005;52:329-37.
 117. **O'Connell J, Bennett MW, et al.** *Altered mechanisms of apoptosis in colon cancer: Fas resistance and counterattack in the tumor-immune conflict.* Ann N Y Acad Sci 2000;910:178-92; discussion 93-5.
 118. **Rabinovich GA, Baum LG, et al.** *Galectins and their ligands: amplifiers, silencers or tuners of the inflammatory response?* Trends Immunol 2002;23:313-20.
 119. **Pellegrini P, Berghella AM, et al.** *Disregulation in TH1 and TH2 subsets of CD4+ T cells in peripheral blood of colorectal cancer patients and involvement in cancer establishment and progression.* Cancer Immunol Immunother 1996;42:1-8.
 120. **Kalluri R, Zeisberg M.** *Fibroblasts in cancer.* Nat Rev Cancer 2006;6:392-401.

121. **Li H, Fan X, et al.** *Tumor microenvironment: the role of the tumor stroma in cancer.* J Cell Biochem 2007;101:805-15.
122. **Orimo A, Gupta PB, et al.** *Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion.* Cell 2005;121:335-48.
123. **Stetler-Stevenson WG, Aznavoorian S, et al.** *Tumor cell interactions with the extracellular matrix during invasion and metastasis.* Annu Rev Cell Biol 1993;9:541-73.
124. **Sternlicht MD, Lochter A, et al.** *The stromal proteinase MMP3/stromelysin-1 promotes mammary carcinogenesis.* Cell 1999;98:137-46.
125. **Li G, Satyamoorthy K, et al.** *Function and regulation of melanoma-stromal fibroblast interactions: when seeds meet soil.* Oncogene 2003;22:3162-71.
126. **Thiery JP.** *Epithelial-mesenchymal transitions in development and pathologies.* Curr Opin Cell Biol 2003;15:740-6.
127. **Lee JM, Dedhar S, et al.** *The epithelial-mesenchymal transition: new insights in signaling, development, and disease.* J Cell Biol 2006;172:973-81.
128. **Bates RC, Mercurio AM.** *The epithelial-mesenchymal transition (EMT) and colorectal cancer progression.* Cancer Biol Ther 2005;4:365-70.
129. **Huang D, Du X.** *Crosstalk between tumor cells and microenvironment via Wnt pathway in colorectal cancer dissemination.* World J Gastroenterol 2008;14:1823-7.
130. **Kornfeld D, Ekbohm A, et al.** *Is there an excess risk for colorectal cancer in patients with ulcerative colitis and concomitant primary sclerosing cholangitis? A population based study.* Gut 1997;41:522-5.
131. **Krishnamoorthy S, Honn KV.** *Inflammation and disease progression.* Cancer Metastasis Rev 2006;25:481-91.
132. **Robinson SC, Coussens LM.** *Soluble mediators of inflammation during tumor development.* Adv Cancer Res 2005;93:159-87.
133. **Mantovani A, Schioppa T, et al.** *Role of tumor-associated macrophages in tumor progression and invasion.* Cancer Metastasis Rev 2006;25:315-22.
134. **Wakabayashi O, Yamazaki K, et al.** *CD4+ T cells in cancer stroma, not CD8+ T cells in cancer cell nests, are associated with favorable prognosis in human non-small cell lung cancers.* Cancer Sci 2003;94:1003-9.
135. **Leffers N, Gooden MJ, et al.** *Prognostic significance of tumor-infiltrating T-lymphocytes in primary and metastatic lesions of advanced stage ovarian cancer.* Cancer Immunol Immunother 2009;58:449-59.

-
136. **Nonomura N, Takayama H, et al.** *Decreased number of mast cells infiltrating into needle biopsy specimens leads to a better prognosis of prostate cancer.* Br J Cancer 2007;97:952-6.
137. **Taskinen M, Karjalainen-Lindsberg ML, et al.** *Prognostic influence of tumor-infiltrating mast cells in patients with follicular lymphoma treated with rituximab and CHOP.* Blood 2008;111:4664-7.
138. **Mantovani A, Allavena P, et al.** *Cancer-related inflammation.* Nature 2008;454:436-44.
139. **Allavena P, Sica A, et al.** *The Yin-Yang of tumor-associated macrophages in neoplastic progression and immune surveillance.* Immunol Rev 2008;222:155-61.
140. **Le Bitoux MA, Stamenkovic I.** *Tumor-host interactions: the role of inflammation.* Histochem Cell Biol 2008;130:1079-90.
141. **Carmeliet P, Jain RK.** *Angiogenesis in cancer and other diseases.* Nature 2000;407:249-57.
142. **van Kempen LC, de Visser KE, et al.** *Inflammation, proteases and cancer.* Eur J Cancer 2006;42:728-34.
143. **Apte RN, Voronov E.** *Is interleukin-1 a good or bad 'guy' in tumor immunobiology and immunotherapy?* Immunol Rev 2008;222:222-41.
144. **Balkwill F.** *Cancer and the chemokine network.* Nat Rev Cancer 2004;4:540-50.
145. **Hiratsuka S, Watanabe A, et al.** *Tumour-mediated upregulation of chemoattractants and recruitment of myeloid cells predetermines lung metastasis.* Nat Cell Biol 2006;8:1369-75.
146. **Kaplan RN, Riba RD, et al.** *VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche.* Nature 2005;438:820-7.
147. **Pupa SM, Menard S, et al.** *New insights into the role of extracellular matrix during tumor onset and progression.* J Cell Physiol 2002;192:259-67.
148. **Kalluri R.** *Basement membranes: structure, assembly and role in tumour angiogenesis.* Nat Rev Cancer 2003;3:422-33.
149. **Montgomery AM, Reisfeld RA, et al.** *Integrin alpha v beta 3 rescues melanoma cells from apoptosis in three-dimensional dermal collagen.* Proc Natl Acad Sci U S A 1994;91:8856-60.
150. **Koshikawa N, Giannelli G, et al.** *Role of cell surface metalloprotease MT1-MMP in epithelial cell migration over laminin-5.* J Cell Biol 2000;148:615-24.

151. **Kuo CJ, LaMontagne KR, Jr., et al.** *Oligomerization-dependent regulation of motility and morphogenesis by the collagen XVIII NC1/endostatin domain.* J Cell Biol 2001;152:1233-46.
152. **Imai K, Hiramatsu A, et al.** *Degradation of decorin by matrix metalloproteinases: identification of the cleavage sites, kinetic analyses and transforming growth factor-beta1 release.* Biochem J 1997;322 (Pt 3):809-14.
153. **Park JE, Keller GA, et al.** *The vascular endothelial growth factor (VEGF) isoforms: differential deposition into the subepithelial extracellular matrix and bioactivity of extracellular matrix-bound VEGF.* Mol Biol Cell 1993;4:1317-26.
154. **Giancotti FG, Ruoslahti E.** *Integrin signaling.* Science 1999;285:1028-32.
155. **Longley DB, Harkin DP, et al.** *5-fluorouracil: mechanisms of action and clinical strategies.* Nat Rev Cancer 2003;3:330-8.
156. **Roche Farma S.** *Xeloda (Capecitabina). Monografía de producto. Tratamiento activado por la TP para tumores sólidos.* 2002.
157. **Casado E, Pfeiffer P, et al.** *UFT (tegafur-uracil) in rectal cancer.* Ann Oncol 2008;19:1371-8.
158. **DeLap RJ.** *The effect of leucovorin on the therapeutic index of fluorouracil in cancer patients.* Yale J Biol Med 1988;61:23-34.
159. **Graham J, Mushin M, et al.** *Oxaliplatin.* Nat Rev Drug Discov 2004;3:11-2.
160. **Muhsin M, Graham J, et al.** *Gefitinib.* Nat Rev Drug Discov 2003;2:515-6.
161. **Minna JD, Dowell J.** *Erlotinib hydrochloride.* Nat Rev Drug Discov 2005;Suppl:S14-5.
162. **Johnston SR, Leary A.** *Lapatinib: a novel EGFR/HER2 tyrosine kinase inhibitor for cancer.* Drugs Today (Barc) 2006;42:441-53.
163. **Goldberg RM.** *Cetuximab.* Nat Rev Drug Discov 2005;Suppl:S10-1.
164. **Carteni G, Fiorentino R, et al.** *Panitumumab a novel drug in cancer treatment.* Ann Oncol 2007;18 Suppl 6:vi16-21.
165. **Ellis LM.** *Bevacizumab.* Nat Rev Drug Discov 2005;Suppl:S8-9.
166. **Schena M, Shalon D, et al.** *Quantitative monitoring of gene expression patterns with a complementary DNA microarray.* Science 1995;270:467-70.
167. **DeRisi J, Penland L, et al.** *Use of a cDNA microarray to analyse gene expression patterns in human cancer.* Nat Genet 1996;14:457-60.

-
168. **Venter JC, Adams MD, et al.** *The sequence of the human genome.* Science 2001;291:1304-51.
169. **Lander ES, Linton LM, et al.** *Initial sequencing and analysis of the human genome.* Nature 2001;409:860-921.
170. **van de Vijver MJ, He YD, et al.** *A gene-expression signature as a predictor of survival in breast cancer.* N Engl J Med 2002;347:1999-2009.
171. **Wang Y, Broderick P, et al.** *Common 5p15.33 and 6p21.33 variants influence lung cancer risk.* Nat Genet 2008;40:1407-9.
172. **Holst F, Stahl PR, et al.** *Estrogen receptor alpha (ESR1) gene amplification is frequent in breast cancer.* Nat Genet 2007;39:655-60.
173. **Schetter AJ, Leung SY, et al.** *MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma.* Jama 2008;299:425-36.
174. **Martinez R, Martin-Subero JI, et al.** *A microarray-based DNA methylation study of glioblastoma multiforme.* Epigenetics 2009;4:255-64.
175. **Yi C, Zheng T, et al.** *Cancer-related transcriptional targets of the circadian gene NPAS2 identified by genome-wide ChIP-on-chip analysis.* Cancer Letters 2009;Mayo 2008 (Epub ahead of print).
176. **Lockhart DJ, Dong H, et al.** *Expression monitoring by hybridization to high-density oligonucleotide arrays.* Nat Biotechnol 1996;14:1675-80.
177. **Quackenbush J.** *Microarray analysis and tumor classification.* N Engl J Med 2006;354:2463-72.
178. **Quackenbush J.** *Computational analysis of microarray data.* Nat Rev Genet 2001;2:418-27.
179. **Fan J, Ren Y.** *Statistical analysis of DNA microarray data in cancer research.* Clin Cancer Res 2006;12:4469-73.
180. **Quackenbush J.** *Microarray data normalization and transformation.* Nat Genet 2002;32 Suppl:496-501.
181. **Simon R.** *Roadmap for developing and validating therapeutically relevant genomic classifiers.* J Clin Oncol 2005;23:7332-41.
182. **Schena M.** *Microarray analysis.* New York: Wiley-Liss 2003.
183. **Michiels S, Koscielny S, et al.** *Prediction of cancer outcome with microarrays: a multiple random validation strategy.* Lancet 2005;365:488-92.

-
184. **Ioannidis JP.** *Microarrays and molecular research: noise discovery?* Lancet 2005;365:454-5.
185. **Ioannidis JP.** *Is molecular profiling ready for use in clinical decision making?* Oncologist 2007;12:301-11.
186. **Biganzoli E, Lama N, et al.** *Prediction of cancer outcome with microarrays.* Lancet 2005;365:1683; author reply 84-5.
187. **Chuaqui RF, Bonner RF, et al.** *Post-analysis follow-up and validation of microarray experiments.* Nat Genet 2002;32 Suppl:509-14.
188. **Workman P, Johnston PG.** *Genomic profiling of cancer: what next?* J Clin Oncol 2005;23:7253-6.
189. **Michiels S, Koscielny S, et al.** *Interpretation of microarray data in cancer.* Br J Cancer 2007;96:1155-8.
190. **Simon R, Radmacher MD, et al.** *Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification.* J Natl Cancer Inst 2003;95:14-8.
191. **Larkin JE, Frank BC, et al.** *Independence and reproducibility across microarray platforms.* Nat Methods 2005;2:337-44.
192. **Irizarry RA, Warren D, et al.** *Multiple-laboratory comparison of microarray platforms.* Nat Methods 2005;2:345-50.
193. **Bammler T, Beyer RP, et al.** *Standardizing global gene expression analysis between laboratories and across platforms.* Nat Methods 2005;2:351-6.
194. **Shi L, Reid LH, et al.** *The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements.* Nat Biotechnol 2006;24:1151-61.
195. **Canales RD, Luo Y, et al.** *Evaluation of DNA microarray results with quantitative gene expression platforms.* Nat Biotechnol 2006;24:1115-22.
196. **Arikawa E, Sun Y, et al.** *Cross-platform comparison of SYBR Green real-time PCR with TaqMan PCR, microarrays and other gene expression measurement technologies evaluated in the MicroArray Quality Control (MAQC) study.* BMC Genomics 2008;9:328.
197. **Fan C, Oh DS, et al.** *Concordance among gene-expression-based predictors for breast cancer.* N Engl J Med 2006;355:560-9.
198. **Yu JX, Sieuwerts AM, et al.** *Pathway analysis of gene signatures predicting metastasis of node-negative primary breast cancer.* BMC Cancer 2007;7:182.

-
199. **Brazma A, Hingamp P, et al.** *Minimum information about a microarray experiment (MIAME)-toward standards for microarray data.* Nat Genet 2001;29:365-71.
200. **van't Veer LJ, Bernards R.** *Enabling personalized cancer medicine through analysis of gene-expression patterns.* Nature 2008;452:564-70.
201. **Sotiriou C, Piccart MJ.** *Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care?* Nat Rev Cancer 2007;7:545-53.
202. **Perou CM, Sorlie T, et al.** *Molecular portraits of human breast tumours.* Nature 2000;406:747-52.
203. **Chung CH, Bernard PS, et al.** *Molecular portraits and the family tree of cancer.* Nat Genet 2002;32 Suppl:533-40.
204. **Ramaswamy S, Tamayo P, et al.** *Multiclass cancer diagnosis using tumor gene expression signatures.* Proc Natl Acad Sci U S A 2001;98:15149-54.
205. **Sorlie T, Perou CM, et al.** *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proc Natl Acad Sci U S A 2001;98:10869-74.
206. **Garber ME, Troyanskaya OG, et al.** *Diversity of gene expression in adenocarcinoma of the lung.* Proc Natl Acad Sci U S A 2001;98:13784-9.
207. **Alizadeh AA, Eisen MB, et al.** *Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling.* Nature 2000;403:503-11.
208. **Bhattacharjee A, Richards WG, et al.** *Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses.* Proc Natl Acad Sci U S A 2001;98:13790-5.
209. **Pomeroy SL, Tamayo P, et al.** *Prediction of central nervous system embryonal tumour outcome based on gene expression.* Nature 2002;415:436-42.
210. **Takahashi M, Rhodes DR, et al.** *Gene expression profiling of clear cell renal cell carcinoma: gene identification and prognostic classification.* Proc Natl Acad Sci U S A 2001;98:9754-9.
211. **Rosenwald A, Wright G, et al.** *The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma.* N Engl J Med 2002;346:1937-47.
212. **Beer DG, Kardia SL, et al.** *Gene-expression profiles predict survival of patients with lung adenocarcinoma.* Nat Med 2002;8:816-24.
213. **van 't Veer LJ, Dai H, et al.** *Gene expression profiling predicts clinical outcome of breast cancer.* Nature 2002;415:530-6.

-
214. **Paik S, Shak S, et al.** *A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer.* N Engl J Med 2004;351:2817-26.
215. **Wang Y, Klijn JG, et al.** *Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer.* Lancet 2005;365:671-9.
216. **Ma XJ, Hilsenbeck SG, et al.** *The HOXB13:IL17BR expression index is a prognostic factor in early-stage breast cancer.* J Clin Oncol 2006;24:4611-9.
217. **Lenz G, Wright G, et al.** *Stromal gene signatures in large-B-cell lymphomas.* N Engl J Med 2008;359:2313-23.
218. **Finak G, Bertos N, et al.** *Stromal gene expression predicts clinical outcome in breast cancer.* Nat Med 2008;14:518-27.
219. **Potti A, Dressman HK, et al.** *Genomic signatures to guide the use of chemotherapeutics.* Nat Med 2006;12:1294-300.
220. **Lee JK, Havaleshko DM, et al.** *A strategy for predicting the chemosensitivity of human cancers and its application to drug discovery.* Proc Natl Acad Sci U S A 2007;104:13086-91.
221. **Kihara C, Tsunoda T, et al.** *Prediction of sensitivity of esophageal tumors to adjuvant chemotherapy by cDNA microarray analysis of gene-expression profiles.* Cancer Res 2001;61:6474-9.
222. **Takata R, Katagiri T, et al.** *Predicting response to methotrexate, vinblastine, doxorubicin, and cisplatin neoadjuvant chemotherapy for bladder cancers through genome-wide gene expression profiling.* Clin Cancer Res 2005;11:2625-36.
223. **Jansen MP, Foekens JA, et al.** *Molecular classification of tamoxifen-resistant breast carcinomas by gene expression profiling.* J Clin Oncol 2005;23:732-40.
224. **Folgueira MA, Carraro DM, et al.** *Gene expression profile associated with response to doxorubicin-based therapy in breast cancer.* Clin Cancer Res 2005;11:7434-43.
225. **Iwao-Koizumi K, Matoba R, et al.** *Prediction of docetaxel response in human breast cancer by gene expression profiling.* J Clin Oncol 2005;23:422-31.
226. **Hess KR, Anderson K, et al.** *Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer.* J Clin Oncol 2006;24:4236-44.
227. **Berns K, Horlings HM, et al.** *A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer.* Cancer Cell 2007;12:395-402.

-
228. **Swanton C, Marani M, et al.** *Regulators of mitotic arrest and ceramide metabolism are determinants of sensitivity to paclitaxel and other chemotherapeutic drugs.* *Cancer Cell* 2007;11:498-512.
229. **Griffith OL, Melck A, et al.** *Meta-analysis and meta-review of thyroid cancer gene expression profiling studies identifies important diagnostic biomarkers.* *J Clin Oncol* 2006;24:5043-51.
230. **Yasui W, Oue N, et al.** *Search for new biomarkers of gastric cancer through serial analysis of gene expression and its clinical implications.* *Cancer Sci* 2004;95:385-92.
231. **Nannini M, Pantaleo MA, et al.** *Gene expression profiling in colorectal cancer using microarray technologies: results and perspectives.* *Cancer Treat Rev* 2009;35:201-9.
232. **Alon U, Barkai N, et al.** *Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays.* *Proc Natl Acad Sci U S A* 1999;96:6745-50.
233. **Williams NS, Gaynor RB, et al.** *Identification and validation of genes involved in the pathogenesis of colorectal cancer using cDNA microarrays and RNA interference.* *Clin Cancer Res* 2003;9:931-46.
234. **Kitahara O, Furukawa Y, et al.** *Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia.* *Cancer Res* 2001;61:3544-9.
235. **Birkenkamp-Demtroder K, Christensen LL, et al.** *Gene expression in colorectal cancer.* *Cancer Res* 2002;62:4352-63.
236. **Croner RS, Foertsch T, et al.** *Common denominator genes that distinguish colorectal carcinoma from normal mucosa.* *Int J Colorectal Dis* 2005;20:353-62.
237. **Komori T, Takemasa I, et al.** *Gene expression of colorectal cancer: preoperative genetic diagnosis using endoscopic biopsies.* *Int J Oncol* 2008;32:367-75.
238. **Bianchini M, Levy E, et al.** *Comparative study of gene expression by cDNA microarray in human colorectal cancer tissues and normal mucosa.* *Int J Oncol* 2006;29:83-94.
239. **Lin YM, Furukawa Y, et al.** *Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas.* *Oncogene* 2002;21:4120-8.
240. **Notterman DA, Alon U, et al.** *Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays.* *Cancer Res* 2001;61:3124-30.

-
241. **Birkenkamp-Demtroder K, Olesen SH, et al.** *Differential gene expression in colon cancer of the caecum versus the sigmoid and rectosigmoid.* Gut 2005;54:374-84.
242. **Komuro K, Tada M, et al.** *Right- and left-sided colorectal cancers display distinct expression profiles and the anatomical stratification allows a high accuracy prediction of lymph node metastasis.* J Surg Res 2005;124:216-24.
243. **Kim H, Nam SW, et al.** *Different gene expression profiles between microsatellite instability-high and microsatellite stable colorectal carcinomas.* Oncogene 2004;23:6218-25.
244. **di Pietro M, Sabates Bellver J, et al.** *Defective DNA mismatch repair determines a characteristic transcriptional profile in proximal colon cancers.* Gastroenterology 2005;129:1047-59.
245. **Giacomini CP, Leung SY, et al.** *A gene expression signature of genetic instability in colon cancer.* Cancer Res 2005;65:9200-5.
246. **Watanabe T, Kobunai T, et al.** *Distal colorectal cancers with microsatellite instability (MSI) display distinct gene expression profiles that are different from proximal MSI cancers.* Cancer Res 2006;66:9804-8.
247. **Laiho P, Kokko A, et al.** *Serrated carcinomas form a subclass of colorectal cancer with distinct molecular basis.* Oncogene 2007;26:312-20.
248. **Frederiksen CM, Knudsen S, et al.** *Classification of Dukes' B and C colorectal cancers using expression arrays.* J Cancer Res Clin Oncol 2003;129:263-71.
249. **Kwong KY, Bloom GC, et al.** *Synchronous global assessment of gene and protein expression in colorectal cancer progression.* Genomics 2005;86:142-58.
250. **Groene J, Mansmann U, et al.** *Transcriptional census of 36 microdissected colorectal cancers yields a gene signature to distinguish UICC II and III.* Int J Cancer 2006;119:1829-36.
251. **Bertucci F, Salas S, et al.** *Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters.* Oncogene 2004;23:1377-91.
252. **Eschrich S, Yang I, et al.** *Molecular staging for survival prediction of colorectal cancer patients.* J Clin Oncol 2005;23:3526-35.
253. **Wang Y, Jatkoe T, et al.** *Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer.* J Clin Oncol 2004;22:1564-71.
254. **Nadal C, Maurel J, et al.** *Is there a genetic signature for liver metastasis in colorectal cancer?* World J Gastroenterol 2007;13:5832-44.

-
255. **Yamasaki M, Takemasa I, et al.** *The gene expression profile represents the molecular nature of liver metastasis in colorectal cancer.* Int J Oncol 2007;30:129-38.
256. **Koehler A, Bataille F, et al.** *Gene expression profiling of colorectal cancer and metastases divides tumours according to their clinicopathological stage.* J Pathol 2004;204:65-74.
257. **Arango D, Laiho P, et al.** *Gene-expression profiling predicts recurrence in Dukes' C colorectal cancer.* Gastroenterology 2005;129:874-84.
258. **Barrier A, Boelle PY, et al.** *Stage II colon cancer prognosis prediction by tumor gene expression profiling.* J Clin Oncol 2006;24:4685-91.
259. **Barrier A, Roser F, et al.** *Prognosis of stage II colon cancer by non-neoplastic mucosa gene expression profiling.* Oncogene 2007;26:2642-8.
260. **Lin YH, Friederichs J, et al.** *Multiple gene expression classifiers from different array platforms predict poor prognosis of colorectal cancer.* Clin Cancer Res 2007;13:498-507.
261. **Ki DH, Jeung HC, et al.** *Whole genome analysis for liver metastasis gene signatures in colorectal cancer.* Int J Cancer 2007;121:2005-12.
262. **Yanagawa R, Furukawa Y, et al.** *Genome-wide screening of genes showing altered expression in liver metastases of human colorectal cancers by cDNA microarray.* Neoplasia 2001;3:395-401.
263. **Nguyen DX, Massague J.** *Genetic determinants of cancer metastasis.* Nat Rev Genet 2007;8:341-52.
264. **Salazar R, van Laar R, et al.** *Independent Validation of a Genomic Profile for the Prognosis of Colorectal Cancer (CRC) Following Radical Surgery.* ASCO 2008.
265. **Mariadason JM, Arango D, et al.** *Gene expression profiling-based prediction of response of colon carcinoma cells to 5-fluorouracil and camptothecin.* Cancer Res 2003;63:8791-812.
266. **Watanabe T, Komuro Y, et al.** *Prediction of sensitivity of rectal cancer cells in response to preoperative radiotherapy by DNA microarray analysis of gene expression profiles.* Cancer Res 2006;66:3370-4.
267. **Ghadimi BM, Grade M, et al.** *Effectiveness of gene expression profiling for response prediction of rectal adenocarcinomas to preoperative chemoradiotherapy.* J Clin Oncol 2005;23:1826-38.
268. **Kim IJ, Lim SB, et al.** *Microarray gene expression profiling for predicting complete response to preoperative chemoradiotherapy in patients with advanced rectal cancer.* Dis Colon Rectum 2007;50:1342-53.

-
269. **Del Rio M, Molina F, et al.** *Gene expression signature in advanced colorectal cancer patients select drugs and response for the use of leucovorin, fluorouracil, and irinotecan.* J Clin Oncol 2007;25:773-80.
270. **Matsuyama R, Togo S, et al.** *Predicting 5-fluorouracil chemosensitivity of liver metastases from colorectal cancer using primary tumor specimens: three-gene expression model predicts clinical response.* Int J Cancer 2006;119:406-13.
271. **Graudens E, Boulanger V, et al.** *Deciphering cellular states of innate tumor drug responses.* Genome Biol 2006;7:R19.
272. **Imbeaud S, Graudens E, et al.** *Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces.* Nucleic Acids Res 2005;33:e56.
273. **Strand C, Enell J, et al.** *RNA quality in frozen breast cancer samples and the influence on gene expression analysis--a comparison of three evaluation methods using microcapillary electrophoresis traces.* BMC Mol Biol 2007;8:38.
274. **Van Gelder RN, von Zastrow ME, et al.** *Amplified RNA synthesized from limited quantities of heterogeneous cDNA.* Proc Natl Acad Sci U S A 1990;87:1663-7.
275. **McShane LM, Radmacher MD, et al.** *Methods for assessing reproducibility of clustering patterns observed in analyses of microarray data.* Bioinformatics 2002;18:1462-9.
276. **Eisen MB, Spellman PT, et al.** *Cluster analysis and display of genome-wide expression patterns.* Proc Natl Acad Sci U S A 1998;95:14863-8.
277. **Kanehisa M, Araki M, et al.** *KEGG for linking genomes to life and the environment.* Nucleic Acids Res 2008;36:D480-4.
278. **Pawlik TM, Raut CP, et al.** *Colorectal carcinogenesis: MSI-H versus MSI-L.* Dis Markers 2004;20:199-206.
279. **Yearsley M, Hampel H, et al.** *Histologic features distinguish microsatellite-high from microsatellite-low and microsatellite-stable colorectal carcinomas, but do not differentiate germline mutations from methylation of the MLH1 promoter.* Hum Pathol 2006;37:831-8.
280. **Halford S, Sasieni P, et al.** *Low-level microsatellite instability occurs in most colorectal cancers and is a nonrandomly distributed quantitative trait.* Cancer Res 2002;62:53-7.
281. **Boland CR, Thibodeau SN, et al.** *A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer.* Cancer Res 1998;58:5248-57.

-
282. **Hoang JM, Cottu PH, et al.** *BAT-26, an indicator of the replication error phenotype in colorectal cancers and cell lines.* *Cancer Res* 1997;57:300-3.
283. **Iacopetta B, Hamelin R.** *Rapid and nonisotopic SSCP-based analysis of the BAT-26 mononucleotide repeat for identification of the replication error phenotype in human cancers.* *Hum Mutat* 1998;12:355-60.
284. **Hatzaki A, Razi E, et al.** *A modified mutagenic PCR-RFLP method for K-ras codon 12 and 13 mutations detection in NSCLC patients.* *Mol Cell Probes* 2001;15:243-7.
285. **Kononen J, Bubendorf L, et al.** *Tissue microarrays for high-throughput molecular profiling of tumor specimens.* *Nat Med* 1998;4:844-7.
286. **Hu Z, Fan C, et al.** *The molecular portraits of breast tumors are conserved across microarray platforms.* *BMC Genomics* 2006;7:96.
287. **Gomez D, Dalal Z, et al.** *Anatomical distribution of colorectal cancer over a 10 year period in a district general hospital: is there a true "rightward shift"?* *Postgrad Med J* 2004;80:667-9.
288. **Qing SH, Rao KY, et al.** *Racial differences in the anatomical distribution of colorectal cancer: a study of differences between American and Chinese patients.* *World J Gastroenterol* 2003;9:721-5.
289. **Okuyama T, Oya M, et al.** *Budding (sprouting) as a useful prognostic marker in colorectal mucinous carcinoma.* *Jpn J Clin Oncol* 2002;32:412-6.
290. **Leopoldo S, Lorena B, et al.** *Two subtypes of mucinous adenocarcinoma of the colorectum: clinicopathological and genetic features.* *Ann Surg Oncol* 2008;15:1429-39.
291. **Mak T, Laloo F, et al.** *Molecular stool screening for colorectal cancer.* *Br J Surg* 2004;91:790-800.
292. **Borlado LR, Mendez J.** *CDC6: from DNA replication to cell cycle checkpoints and oncogenesis.* *Carcinogenesis* 2008;29:237-43.
293. **Staley BE, Samowitz WS, et al.** *Expression of DNA topoisomerase I and DNA topoisomerase II-alpha in carcinoma of the colon.* *Mod Pathol* 1999;12:356-61.
294. **Schwartz GK, Shah MA.** *Targeting the cell cycle: a new approach to cancer therapy.* *J Clin Oncol* 2005;23:9408-21.
295. **Musacchio A, Salmon ED.** *The spindle-assembly checkpoint in space and time.* *Nat Rev Mol Cell Biol* 2007;8:379-93.
296. **Liu D, Ding X, et al.** *Human NUF2 interacts with centromere-associated protein E and is essential for a stable spindle microtubule-kinetochore attachment.* *J Biol Chem* 2007;282:21415-24.

-
297. **Fukasawa K.** *Oncogenes and tumour suppressors take on centrosomes.* Nat Rev Cancer 2007;7:911-24.
298. **Malumbres M, Barbacid M.** *Cell cycle kinases in cancer.* Curr Opin Genet Dev 2007;17:60-5.
299. **Zhao WM, Seki A, et al.** *Cep55, a microtubule-bundling protein, associates with centralspindlin to control the midbody integrity and cell abscission during cytokinesis.* Mol Biol Cell 2006;17:3881-96.
300. **Takahashi M, Nakamura Y, et al.** *Identification of SP5 as a downstream gene of the beta-catenin/Tcf pathway and its enhanced expression in human colon cancer.* Int J Oncol 2005;27:1483-7.
301. **Hiendlmeyer E, Regus S, et al.** *Beta-catenin up-regulates the expression of the urokinase plasminogen activator in human colorectal tumors.* Cancer Res 2004;64:1209-14.
302. **Zucker S, Vacirca J.** *Role of matrix metalloproteinases (MMPs) in colorectal cancer.* Cancer Metastasis Rev 2004;23:101-17.
303. **Berger DH.** *Plasmin/plasminogen system in colorectal cancer.* World J Surg 2002;26:767-71.
304. **Skoglund J, Emterling A, et al.** *Clinicopathological significance of stromelysin-3 expression in colorectal cancer.* Oncology 2004;67:67-72.
305. **Remy L, Trespeuch C, et al.** *Matrilysin 1 influences colon carcinoma cell migration by cleavage of the laminin-5 beta3 chain.* Cancer Res 2006;66:11228-37.
306. **Smith JJ, Deane NG, et al.** *Regulation of metastasis in colorectal adenocarcinoma: a collision between development and tumor biology.* Surgery 2008;144:353-66.
307. **Lu Y, Yang Q, et al.** *Expression analysis of lymphangiogenic factors in human colorectal cancer with quantitative RT-PCR.* Cancer Invest 2007;25:393-6.
308. **Ma C, Rong Y, et al.** *Extracellular matrix protein betaig-h3/TGFBI promotes metastasis of colon cancer by enhancing cell extravasation.* Genes Dev 2008;22:308-21.
309. **Moller Sorensen N, Vejgaard Sorensen I, et al.** *Biology and potential clinical implications of tissue inhibitor of metalloproteinases-1 in colorectal cancer treatment.* Scand J Gastroenterol 2008;43:774-86.
310. **Hlubek F, Spaderna S, et al.** *Beta-catenin activates a coordinated expression of the proinvasive factors laminin-5 gamma2 chain and MT1-MMP in colorectal carcinomas.* Int J Cancer 2004;108:321-6.

-
311. **Tomii Y, Kamochi J, et al.** *Human thrombospondin 2 inhibits proliferation of microvascular endothelial cells.* Int J Oncol 2002;20:339-42.
312. **Lindmark G, Sundberg C, et al.** *Stromal expression of platelet-derived growth factor beta-receptor and platelet-derived growth factor B-chain in colorectal cancer.* Lab Invest 1993;69:682-9.
313. **Podhajcer OL, Benedetti LG, et al.** *The role of the matricellular protein SPARC in the dynamic interaction between the tumor and the host.* Cancer Metastasis Rev 2008;27:691-705.
314. **Jia W, Li H, et al.** *The extracellular matrix protein mindin serves as an integrin ligand and is critical for inflammatory cell recruitment.* Blood 2005;106:3854-9.
315. **Henry LR, Lee HO, et al.** *Clinical implications of fibroblast activation protein in patients with colon cancer.* Clin Cancer Res 2007;13:1736-41.
316. **Cho YG, Kim CJ, et al.** *Overexpression of S100A4 is closely associated with progression of colorectal cancer.* World J Gastroenterol 2005;11:4852-6.
317. **Crawford Y, Kasman I, et al.** *PDGF-C mediates the angiogenic and tumorigenic properties of fibroblasts associated with tumors refractory to anti-VEGF treatment.* Cancer Cell 2009;15:21-34.
318. **Rmali KA, Watkins G, et al.** *Tumour endothelial marker 8 (TEM-8) in human colon cancer and its association with tumour progression.* Eur J Surg Oncol 2004;30:948-53.
319. **Sneddon JB, Zhen HH, et al.** *Bone morphogenetic protein antagonist gremlin 1 is widely expressed by cancer-associated stromal cells and can promote tumor cell proliferation.* Proc Natl Acad Sci U S A 2006;103:14842-7.
320. **Ding J, Li D, et al.** *Fibronectin promotes invasiveness and focal adhesion kinase tyrosine phosphorylation of human colon cancer cell.* Hepatogastroenterology 2008;55:2072-6.
321. **Kim TD, Song KS, et al.** *Activity and expression of urokinase-type plasminogen activator and matrix metalloproteinases in human colorectal cancer.* BMC Cancer 2006;6:211.
322. **Seya T, Tanaka N, et al.** *Lumican expression in advanced colorectal cancer with nodal metastasis correlates with poor prognosis.* Oncol Rep 2006;16:1225-30.
323. **Stein U, Arlt F, et al.** *The metastasis-associated gene S100A4 is a novel target of beta-catenin/T-cell factor signaling in colon cancer.* Gastroenterology 2006;131:1486-500.
324. **Hemandas AK, Salto-Tellez M, et al.** *Metastasis-associated protein S100A4--a potential prognostic marker for colorectal cancer.* J Surg Oncol 2006;93:498-503.

325. **Ben QW, Zhao Z, et al.** *Circulating levels of periostin may help identify patients with more aggressive colorectal cancer.* Int J Oncol 2009;34:821-8.
326. **Tokunaga T, Nakamura M, et al.** *Thrombospondin 2 expression is correlated with inhibition of angiogenesis and metastasis of colon cancer.* Br J Cancer 1999;79:354-9.
327. **Hornebeck W, Lambert E, et al.** *Beneficial and detrimental influences of tissue inhibitor of metalloproteinase-1 (TIMP-1) in tumor progression.* Biochimie 2005;87:377-83.
328. **Maeda K, Kang SM, et al.** *Expression of intercellular adhesion molecule-1 and prognosis in colorectal cancer.* Oncol Rep 2002;9:511-4.
329. **Deng G, Kakar S, et al.** *Unique methylation pattern of oncostatin m receptor gene in cancers of colorectum and other digestive organs.* Clin Cancer Res 2009;15:1519-26.
330. **Park SW, Zhen G, et al.** *The protein disulfide isomerase AGR2 is essential for production of intestinal mucus.* Proc Natl Acad Sci U S A 2009;106:6950-5.
331. **Zheng W, Rosenstiel P, et al.** *Evaluation of AGR2 and AGR3 as candidate genes for inflammatory bowel disease.* Genes Immun 2006;7:11-8.
332. **Harada N, Iijima S, et al.** *Human IgGFc binding protein (FcgammaBP) in colonic epithelial cells exhibits mucin-like structure.* J Biol Chem 1997;272:15232-41.
333. **Kobayashi K, Hamada Y, et al.** *The molecular configuration and ultrastructural locations of an IgG Fc binding site in human colonic epithelium.* J Immunol 1991;146:68-74.
334. **Mochizuka A, Uehara T, et al.** *Hyperplastic polyps and sessile serrated 'adenomas' of the colon and rectum display gastric pyloric differentiation.* Histochem Cell Biol 2007;128:445-55.
335. **Kim DH, Kim JW, et al.** *Expression of mucin core proteins, trefoil factors, APC and p21 in subsets of colorectal polyps and cancers suggests a distinct pathway of pathogenesis of mucinous carcinoma of the colorectum.* Int J Oncol 2005;27:957-64.
336. **Inagaki Y, Tang W, et al.** *Sialoglycoconjugate expression in primary colorectal cancer and metastatic lymph node tissues.* Hepatogastroenterology 2007;54:53-7.
337. **Sebzda T, Saleh Y, et al.** *Total and lipid-bound plasma sialic acid as diagnostic markers in colorectal cancer patients: correlation with cathepsin B expression in progression to Dukes stage.* J Exp Ther Oncol 2006;5:223-9.

338. **Ishida A, Ohta M, et al.** *Mucin-induced apoptosis of monocyte-derived dendritic cells during maturation.* Proteomics 2008;8:3342-9.
339. **Oue N, Kuniyasu H, et al.** *Serum concentration of Reg IV in patients with colorectal cancer: overexpression and high serum levels of Reg IV are associated with liver metastasis.* Oncology 2007;72:371-80.
340. **Mitani Y, Oue N, et al.** *Reg IV is a serum biomarker for gastric cancer patients and predicts response to 5-fluorouracil-based chemotherapy.* Oncogene 2007;26:4383-93.
341. **Perez RO, Bresciani BH, et al.** *Mucinous colorectal adenocarcinoma: influence of mucin expression (Muc1, 2 and 5) on clinico-pathological features and prognosis.* Int J Colorectal Dis 2008;23:757-65.
342. **You JF, Hsieh LL, et al.** *Inverse effects of mucin on survival of matched hereditary nonpolyposis colorectal cancer and sporadic colorectal cancer patients.* Clin Cancer Res 2006;12:4244-50.
343. **Duncan TJ, Watson NF, et al.** *The role of MUC1 and MUC3 in the biology and prognosis of colorectal cancer.* World J Surg Oncol 2007;5:31.
344. **Niv Y.** *MUC1 and colorectal cancer pathophysiology considerations.* World J Gastroenterol 2008;14:2139-41.
345. **Baldus SE, Monig SP, et al.** *MUC1 and nuclear beta-catenin are coexpressed at the invasion front of colorectal carcinomas and are both correlated with tumor prognosis.* Clin Cancer Res 2004;10:2790-6.
346. **Bettstetter M, Woenckhaus M, et al.** *Elevated nuclear maspin expression is associated with microsatellite instability and high tumour grade in colorectal cancer.* J Pathol 2005;205:606-14.
347. **Casellas R, Jankovic M, et al.** *OcaB is required for normal transcription and V(D)J recombination of a subset of immunoglobulin kappa genes.* Cell 2002;110:575-85.
348. **Teitell MA.** *OCA-B regulation of B-cell development and function.* Trends Immunol 2003;24:546-53.
349. **Shinto E, Tsuda H, et al.** *Prognostic implication of laminin-5 gamma 2 chain expression in the invasive front of colorectal cancers, disclosed by area-specific four-point tissue microarrays.* Lab Invest 2005;85:257-66.
350. **Ropponen K, Eskelinen M, et al.** *Prognostic value of tumour-infiltrating lymphocytes (TILs) in colorectal cancer.* The Journal of Pathology 1997;182:318-24.

-
351. **Valera V, Yokoyama N, et al.** *Clinical significance of Ki-67 proliferation index in disease progression and prognosis of patients with resected colorectal carcinoma.* Br J Surg 2005;92:1002-7.
352. **Oshima C, Iriya K, et al.** *Ki-67 as a prognostic marker in colorectal cancer but not in gastric cancer.* Neoplasma 2005; 52(5):420-4.
353. **Bittner M, Meltzer P, et al.** *Molecular classification of cutaneous malignant melanoma by gene expression profiling.* Nature 2000;406:536-40.
354. **Friederichs J, Rosenberg R, et al.** *Gene expression profiles of different clinical stages of colorectal carcinoma: toward a molecular genetic understanding of tumor progression.* Int J Colorectal Dis 2005;20:391-402.
355. **Sorlie T, Tibshirani R, et al.** *Repeated observation of breast tumor subtypes in independent gene expression data sets.* Proc Natl Acad Sci U S A 2003;100:8418-23.
356. **Garman KS, Acharya CR, et al.** *A genomic approach to colon cancer risk stratification yields biologic insights into therapeutic opportunities.* Proc Natl Acad Sci U S A 2008;105:19432-7.
357. **Chan SK, Griffith OL, et al.** *Meta-analysis of colorectal cancer gene expression profiling studies identifies consistently reported candidate biomarkers.* Cancer Epidemiol Biomarkers Prev 2008;17:543-52.
358. **Mesker WE, Liefers GJ, et al.** *Presence of a high amount of stroma and downregulation of SMAD4 predict for worse survival for stage I-II colon cancer patients.* Cell Oncol 2009;31:169-78.
359. **Mesker WE, Junggeburst JM, et al.** *The carcinoma-stromal ratio of colon carcinoma is an independent factor for survival compared to lymph node status and tumor stage.* Cell Oncol 2007;29:387-98.
360. **Hlubek F, Brabletz T, et al.** *Heterogeneous expression of Wnt/beta-catenin target genes within colorectal cancer.* Int J Cancer 2007;121:1941-8.
361. **Cianchi F, Cuzzocrea S, et al.** *Heterogeneous expression of cyclooxygenase-2 and inducible nitric oxide synthase within colorectal tumors: Correlation with tumor angiogenesis.* Dig Liver Dis 2009.
362. **Kanemitsu Y, Kato T, et al.** *Survival after curative resection for mucinous adenocarcinoma of the colorectum.* Dis Colon Rectum 2003;46:160-7.
363. **Gafa R, Maestri I, et al.** *Sporadic colorectal adenocarcinomas with high-frequency microsatellite instability.* Cancer 2000;89:2025-37.
364. **Jass JR, Do KA, et al.** *Morphology of sporadic colorectal cancer with DNA replication errors.* Gut 1998;42:673-9.

365. **Gryfe R, Kim H, et al.** *Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer.* N Engl J Med 2000;342:69-77.
366. **Samowitz WS, Curtin K, et al.** *Microsatellite instability in sporadic colon cancer is associated with an improved prognosis at the population level.* Cancer Epidemiol Biomarkers Prev 2001;10:917-23.
367. **Buckowitz A, Knaebel HP, et al.** *Microsatellite instability in colorectal cancer is associated with local lymphocyte infiltration and low frequency of distant metastases.* Br J Cancer 2005;92:1746-53.
368. **Phillips SM, Banerjea A, et al.** *Tumour-infiltrating lymphocytes in colorectal cancer with microsatellite instability are activated and cytotoxic.* Br J Surg 2004;91:469-75.
369. **Chang EY, Dorsey PB, et al.** *Combination of microsatellite instability and lymphocytic infiltrate as a prognostic indicator in colon cancer.* Arch Surg 2009;144:511-5.
370. **Ortega P, Moran A, et al.** *Differential Wnt pathway gene expression and E-cadherin truncation in sporadic colorectal cancers with and without microsatellite instability.* Clin Cancer Res 2008;14:995-1001.
371. **Samowitz WS, Holden JA, et al.** *Inverse relationship between microsatellite instability and K-ras and p53 gene alterations in colon cancer.* Am J Pathol 2001;158:1517-24.
372. **Terada H, Urano T, et al.** *Association of interleukin-8 and plasminogen activator system in the progression of colorectal cancer.* Eur Surg Res 2005;37:166-72.
373. **Ricci F, Kern SE, et al.** *Stromal responses to carcinomas of the pancreas: juxtatumoral gene expression conforms to the infiltrating pattern and not the biologic subtype.* Cancer Biol Ther 2005;4:302-7.
374. **Fischer A, Steidl C, et al.** *Combined loss of Hey1 and HeyL causes congenital heart defects because of impaired epithelial to mesenchymal transition.* Circ Res 2007;100:856-63.
375. **Diao H, Kon S, et al.** *Osteopontin as a mediator of NKT cell function in T cell-mediated liver diseases.* Immunity 2004;21:539-50.
376. **Parker BS, Argani P, et al.** *Alterations in vascular gene expression in invasive breast carcinoma.* Cancer Res 2004;64:7857-66.
377. **Irby RB, McCarthy SM, et al.** *Osteopontin regulates multiple functions contributing to human colon cancer development and progression.* Clin Exp Metastasis 2004;21:515-23.
378. **Wai PY, Kuo PC.** *Osteopontin: regulation in tumor metastasis.* Cancer Metastasis Rev 2008;27:103-18.

379. **Gollub J, Sherlock G.** *Clustering microarray data.* Methods Enzymol 2006;411:194-213.
380. **Everitt B.** *Cluster analysis.* Wiley (New York) 1974.
381. **Chan YH.** *Biostatistics 302. Principal component and factor analysis.* Singapore Med J 2004;45:558-65, quiz 66.
382. **Goeman JJ, Oosting J, et al.** *Testing association of a pathway with survival using gene expression data.* Bioinformatics 2005;21:1950-7.
383. **Goeman JJ, van de Geer SA, et al.** *A global test for groups of genes: testing association with a clinical outcome.* Bioinformatics 2004;20:93-9.
384. **Subramanian A, Tamayo P, et al.** *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A 2005;102:15545-50.

Anexo IV.
Listado matriz de expresión HCSC.

Tabla.1. Listado de los 1722 genes utilizados a lo largo del estudio para la realización del “clustering” jerárquico y las comparaciones de clases.

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
THC2560976		THC2560976		1					5	
BC022881	Hs.652926	LOC644450		2					6	
NM_203414	Hs.417029	C17orf81	17p13.1	3					871	
NM_178537	Hs.148074	B4GALNT4	11p15.5	4					886	
NR_001564	Hs.529901	XIST	Xq13.2	5					8	
BC020847	Hs.644600	LOC644246	17q21.31	6					417	
NM_001828	Hs.889	CLC	19q13.2	7					421	
NM_173549	Hs.171455	C8orf47	8q22.2	8					414	
ENST00000302057	Hs.282089	IRX2	5p15.33	9					7	
NM_002344	Hs.434481	LTK	15q15.1	10					789	NvsT
NM_002220	Hs.2722	ITPKA	15q15.1	11	1				790	NvsT
THC2536817		THC2536817	2q31.1	12					802	
NM_172113	Hs.472877	EYA2	20q13.12	13					807	NvsT
NM_032391	Hs.116467	PRAC	17q21.32	14					797	
NM_006361	Hs.66731	HOXB13	17q21.32	15					798	
NM_001884	Hs.591758	HAPLN1	5q14.3	16					649	NvsT
NM_002054	Hs.516494	GCG	2q24.2	17					629	NvsT
NM_001819	Hs.516874	CHGB	20p12.3	18					613	NvsT
NM_017852	Hs.369279	NLRP2	19q13.42	19					803	
NM_001012978	Hs.47209	NGFRAP1L1	Xq22.1	20					804	
NM_006198	Hs.80296	PCP4	21q22.2	21					811	
NM_005987	Hs.46320	SPRR1A	1q21.3	22					812	
NM_005408	Hs.414629	CCL13	17q12	23					1222	NvsT
NM_001911	Hs.421724	CTSG	14q12	24					621	NvsT
NM_024025	Hs.8719	DUSP26	8p12	25				4	622	NvsT
NM_018286	Hs.173233	TMEM100	17q22	26					623	NvsT
ENST00000370548		ENST00000370548	1p22.3	27			3		1243	NvsT
NM_006741	Hs.505662	PPP1R1A	12q13.2	28					810	
NM_181789	Hs.526441	GLDN	15q21.2	29					805	NvsT
NM_003381	Hs.53973	VIP	6q25.2	30					829	NvsT
NM_001721	Hs.495731	BMX	Xp22.2	31					830	NvsT
NM_005398	Hs.303090	PPP1R3C	10q23.32	32	1				1154	
NM_033014	Hs.109439	OGN	9q22.31	33					1155	NvsT
NM_004750	Hs.114948	CRLF1	19p13.11	34			3		1300	
NM_004181	Hs.518731	UCHL1	4p13	35	1		3		836	
NM_000898	Hs.654473	MAOB	Xp11.3	36					837	
NM_003012	Hs.213424	SFRP1	8p11.21	37					838	
NM_006307	Hs.15154	SRPX	Xp11.4	38					839	
NM_198289	Hs.249129	CIDEA	18p11.21	39					809	
NM_007177	Hs.506357	FAM107A	3p14.2	40					624	NvsT
AF348994	Hs.513626	MT1JP	16q13	41					831	NvsT

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_153370	Hs.25391	PI16	6p21.2	42					832	NvsT
NM_000261	Hs.436037	MYOC	1q24.3	43					833	NvsT
NM_024730	Hs.115497	FLJ22655	12p12.3	44					834	
NM_138290	Hs.411488	RP1B9	7q21.12	45					814	NvsT
NM_002606	Hs.473927	PDE9A	21q22.3	46					650	NvsT
AK000158	Hs.210995	CA12	15q22.2	47					823	NvsT
NM_004816	Hs.118003	C9orf61	9q21.11	48					824	NvsT
NM_003759	Hs.5462	SLC4A4	4q13.3	49					640	NvsT
NM_000067	Hs.155097	CA2	8q21.2	50					641	NvsT
NM_006846	Hs.331555	SPINK5	5q33.1	51					625	NvsT
NM_014312	Hs.112377	VSIG2	11q24.2	52					639	NvsT
AK124946	Hs.162868	DPF3	14q24.2	53					753	NvsT
NM_017682	Hs.435611	BEST2	19p13.13	54					638	NvsT
FNST00000371483		GPR120	10q23.33	55					825	
NM_006615	Hs.498021	CAPN9	1q42.2	56					826	NvsT
THC2721785		THC2721785	1p33	57					813	
THC2559929		THC2559929	14q32.33	58					642	NvsT
NM_002614	Hs.444751	PDZK1	1q21.1	59					821	
NM_152672	Hs.651164	OSTalpha	3q29	60					815	NvsT
NM_000862	Hs.364941	HSD3B1	1p12	61					816	
NM_005123	Hs.282735	NR1H4	12q23.1	62					818	NvsT
NM_173625	Hs.439154	C17orf78	17q12	63					820	
NM_176782	Hs.554944	C1orf179	1p32.3	64					644	NvsT
NM_001169	Hs.176658	AQP8	16p12.1	65					646	NvsT
NM_001275	Hs.150793	CHGA	14q32.12	66					648	NvsT
NM_004827	Hs.480218	ABCG2	4q22.1	67					645	NvsT
NM_004160	Hs.169249	PYY	17q21.31	68					647	NvsT
NM_005478	Hs.251380	INSL5	1p31.3	69					626	NvsT
NM_017716	Hs.272789	MS4A12	11q12.2	70					632	NvsT
A_32_P157213		A_32_P157213	19q13.31	71					643	NvsT
NM_033553	Hs.778	GUCA2A	1p34.2	72					633	NvsT
NM_012128	Hs.567422	CLCA4	1p22.3	73					635	NvsT
AI659667	Hs.516105	AI659667	2p13.1	74					1165	NvsT
NM_139250	Hs.559536	CTAG1A	Xq28	75					827	
NM_000870	Hs.483773	HTR4	5q33.1	76					828	
A_32_P171043		A_32_P171043	18q21.32	77					835	NvsT
NM_002196	Hs.89584	INSM1	20p11.23	78					840	
THC2686967		THC2686967	17q25.1	79					841	
NM_017878	Hs.272805	HRASLS2	11q12.3	80					843	
NM_020973	Hs.653107	GBA3	4p15.31	81					817	
NM_006919	Hs.227948	SERP1NB3	18q21.33	82					842	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
A_32_P166653		A_32_P166653	1q41	83					845	NvsT
CB959193	Hs.97644	CB959193	11q12.3	84					844	NvsT
NM_006998	Hs.116428	SCGN	6p22.2	85					846	NvsT
NM_021637	Hs.45140	TMEM35	Xq22.1	86					847	NvsT
NM_000717	Hs.89485	CA4	17q23.1	87					634	NvsT
NM_001738	Hs.23118	CA1	8q21.2	88					630	NvsT
NM_007102	Hs.32966	GUCA2B	1p34.2	89					631	NvsT
AK124400	Hs.652438	LOC553137	6q21	90					796	
NM_022843	Hs.391781	PCDH20	13q21.31	91					564	
NM_000112	Hs.302738	SLC26A2	5q33.1	92					674	NvsT
NM_000111	Hs.1650	SLC26A3	7q31.1	93					627	NvsT
NM_178859	Hs.534533	OSTbeta	15q22.31	94					673	NvsT
NM_002010	Hs.111	FGF9	13q12.11	95					606	NvsT
NM_014033	Hs.655369	METTL7A	12q13.13	96				4	607	NvsT
NM_145259	Hs.352338	ACVR1C	2q24.1	97					568	NvsT
AF435956	Hs.591566	FAM82A	2p22.2	98					652	NvsT
NM_173833	Hs.591833	SCARA5	8p21.1	99					618	NvsT
NM_000667	Hs.654433	ADH1A	4q23	100					619	NvsT
NM_000669	Hs.654537	ADH1C	4q23	101					620	NvsT
BC033223	Hs.372492	SLC45A4	8q24.3	102					1511	
NM_002885	Hs.148178	RAP1GAP	1p36.12	103					1510	
NM_001012506	Hs.476399	CCDC66	3p14.3	104					1512	
BX119852	Hs.258100	BX119852	2p25.2	105					1525	
AA479896	Hs.587119	AA479896	20p12.3	106					1526	
NM_013347	Hs.659349	RPA4	Xq21.33	107					1513	
CV575560	Hs.659326	CV575560	17p11.2	108					1514	
NM_020857	Hs.23876	VPS18	15q15.1	109					1515	
NM_182527	Hs.643608	CABP7	22q12.2	110					1516	
THC2579650		THC2579650	15q26.1	111					1517	
THC2609820		THC2609820	8p21.1	112					1518	
THC2541992		THC2541992	1q32.1	113					1519	
NM_005853	Hs.435730	IRX5	16q12.2	114					1520	
CB162722	Hs.667977	CB162722	3p25.2	115					1521	
NM_018047	Hs.202023	RBM22	5q33.1	116					1522	
NM_014931	Hs.515610	SAPS1	19q13.42	117					1523	
NM_024512	Hs.657345	LRRC2	3p21.31	118					1524	
NM_014308	Hs.278901	PIK3R5	17p13.1	119					1527	
THC2672701		THC2672701	17q25.3	120					1528	
ENST00000314088	Hs.622290	HIST1H2AC	6p22.1	121					425	
NM_005319	Hs.7644	HIST1H1C	6p22.1	122					426	NvsT
A_23_P170713		A_23_P170713	6p22.1	123					427	NvsT

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_006980	Hs.532216	MTERF	7q21.2	124					212	
CR933606	Hs.653099	C20orf80	20q11.1	125					213	
NM_002252	Hs.414489	KCNS3	2p24.2	126					210	
NM_000313	Hs.64016	PROS1	3q11.2	127					211	
CR627475	Hs.125038	FAM92A1	8q22.1	128					273	NvsT
NM_018476	Hs.334370	BEX1	Xq22.1	129					1014	
NM_032621	Hs.398989	BFX2	Xq22.1	130					1015	
NM_025176	Hs.631508	RP4-691N24.1	20p11.21	131					1016	
ENST00000372602	Hs.496530	TMSL8	Xq22.2	132					1017	
NM_052969	Hs.647900	RPL39L	3q27.3	133					1004	
NM_020685	Hs.47166	C3orf14	3p14.2	134					1006	
NM_014766	Hs.520740	SCRN1	7p15.1	135					1003	
NM_001454	Hs.651204	FOXJ1	17q25.1	136					1001	
NM_024532	Hs.602792	SPAG16	2q34	137					1002	
NR_002166	Hs.446620	SEDLP	19q13.43	138					1009	
NM_021632	Hs.407694	ZNF350	19q13.33	139					1010	
NM_198542	Hs.579576	ZNF773	19q13.43	140				4	1011	
NM_005773	Hs.596242	ZNF256	19q13.43	141					1012	
NM_023926	Hs.235390	ZSCAN18	19q13.43	142			3		1013	
NM_022804	Hs.564847	SNURF	15q11.2	143					1532	
NM_004052	Hs.144873	BNIP3	10q26.3	144					1005	
NM_016073	Hs.513954	HDGFRP3	15q25.2	145			3		1259	
THC2644897		THC2644897	12p12.2	146					412	
NM_000484	Hs.693586	APP	21q21.3	147					1530	
NM_152718	Hs.60640	VWCE	11q12.2	148					13	
NM_025135	Hs.436636	FHOD3	18q12.2	149					1605	
NM_016848	Hs.656806	SHC3	9q22.1	150					1603	
NM_014344	Hs.39384	FJX1	11p13	151					1634	NvsT
NM_182898	Hs.437075	CREB5	7p15.1	152					1604	
NM_003102	Hs.2420	SOD3	4p15.2	153					1651	
NM_007283	Hs.277035	MGLL	3q21.3	154					594	NvsT
NM_020182	Hs.517155	TMEPAI	20q13.31	155			3		1619	NvsT
NM_003043	Hs.529488	SLC6A6	3p25.1	156			3		1620	NvsT
NM_000358	Hs.369397	TGFBI	5q31.2	157					1618	NvsT
ENST00000382327	Hs.156727	ANKH	5p15.2	158					1613	
NM_005978	Hs.516484	S100A2	1q21.3	159					1593	NvsT
NM_198129	Hs.436367	LAMA3	18q11.2	160					1027	
NM_018891	Hs.591484	LAMC2	1q25.3	161		2			1028	
NM_005562	Hs.591484	LAMC2	1q25.3	162		2			1029	
NM_002423	Hs.2256	MMP7	11q22.2	163					1617	NvsT
NM_014220	Hs.351316	TM4SF1	3q25.1	164					1592	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_020130	Hs.591849	C8orf4	8p11.21	165					1103	
NM_018948	Hs.605445	ERRFI1	1p36.23	166					1101	
NM_003897	Hs.655404	IER3	6p21.33	167					1104	
AL832183	Hs.436426	LOC284454	19p13.12	168					1105	
NM_006186	Hs.563344	NR4A2	2q24.1	169					1115	
NM_005252	Hs.25647	FOS	14q24.3	170					1114	
NM_004024	Hs.460	ATF3	1q32.3	171					1116	
NM_002135	Hs.524430	NR4A1	12q13.13	172					1117	
NM_001964	Hs.326035	EGR1	5q31.2	173					1118	
NM_005213	Hs.518198	CSTA	3q21.1	174					1097	
NM_000032	Hs.522666	ALAS2	Xp11.21	175					1107	
NM_000518	Hs.654454	HBB	11p15.4	176					1108	
NM_000519	Hs.523443	HBD	11p15.4	177					1109	
A_24_P147849		A_24_P147849		178					1110	
NM_000517	Hs.654744	HBA2	16p13.3	179					1111	
NM_000558	Hs.449630	HBA1	16p13.3	180					1112	
NM_006512	Hs.654493	SAA4	11p15.1	181					119	
NM_030754	Hs.654517	SAA2	11p15.1	182					120	
NM_000331	Hs.632144	SAA1	11p15.1	183					121	
NM_031419	Hs.319171	NFKBIZ	3q12.3	184					1036	
NM_005204	Hs.432453	MAP3K8	10p11.23	185					1037	
NM_005746	Hs.489615	PBFF1	7q22.2	186					1038	
A_24_P592591		A_24_P592591	10p11.21	187					1039	
NM_004591	Hs.75498	CCL20	2q36.3	188			3		1032	
NM_001511	Hs.789	CXCL1	4q13.3	189					1033	NvsT
NM_002089	Hs.590921	CXCL2	4q13.3	190					1034	NvsT
NM_002090	Hs.89690	CXCL3	4q13.3	191					1035	NvsT
NM_032895	Hs.597755	MGC14376	17p13.3	192					1152	NvsT
AY358804	Hs.505141	AY358804	12p11.21	193					424	
NM_002160	Hs.143250	TNC	9q33.1	194	1		3		1252	
NM_002658	Hs.77274	PLAU	10q22.2	195	1				1629	NvsT
NM_006474	Hs.468675	PDPN	1p36.21	196	1				1630	NvsT
NM_003392	Hs.643085	WNT5A	3p14.3	197					1046	NvsT
NM_002422	Hs.375129	MMP3	11q22.2	198					1047	NvsT
NM_002421	Hs.83169	MMP1	11q22.2	199					1048	NvsT
NM_002994	Hs.89714	CXCL5	4q13.3	200					1044	
NM_002425	Hs.2258	MMP10	11q22.2	201					1042	
NM_000640	Hs.336046	IL13RA2	Xq23	202					1043	
NM_000600	Hs.654458	IL6	7p15.3	203	1				1058	
NM_000963	Hs.196384	PTGS2	1q31.1	204					1059	
NM_006528	Hs.438231	TFPI2	7q21.3	205					1040	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_000641	Hs.467304	IL11	19q13.42	206					1041	NvsT
NM_000854	Hs.654462	GSTT2	22q11.23	207					1531	
AA554330	Hs.600066	AA554330	18q21.1	208					992	NvsT
NM_002200	Hs.521181	IRF5	7q32.1	209					20	
U52054	Hs.561411	U52054	1p11.2	210					1430	
NM_002436	Hs.496984	MPP1	Xq28	211					996	
NM_006646	Hs.635221	WASF3	13q12.13	212			3		997	
NM_005410	Hs.275775	SEPP1	5p12	213					653	NvsT
NM_003059	Hs.310591	SLC22A4	5q31.1	214					999	
THC2668815		THC2668815	9q34.11	215					1000	
NM_003462	Hs.406050	DNALI1	1p34.3	216					1718	
THC2623335		THC2623335	16q12.1	217					1719	
BC031342	Hs.445239	BC031342	7p15.2	218					1133	
NM_003063	Hs.334629	SLN	11q22.3	219	1			4	1030	
U10991	Hs.502266	C11orf41	11p13	220					1031	
BC040542	Hs.650822	KIAA0802	18p11.22	221					980	
NM_001013642	Hs.355747	LOC388610	1p36.11	222					981	
NM_053001	Hs.253247	OSR2	8q22.2	223	1				977	
NM_004282	Hs.55220	BAG2	6p12.1	224	1				978	
NM_005855	Hs.471783	RAMP1	2q37.3	225	1				979	
NM_002155	Hs.654614	HSPA6	1q23.3	226					1098	
NM_005345	Hs.520028	HSPA1A	6p21.33	227					1099	
NM_005527	Hs.690634	HSPA1L	6p21.33	228					1100	
NM_001174	Hs.435291	ARHGAP6	Xp22.2	229					1140	
NM_053277	Hs.693698	CLIC6	21q22.12	230					1141	
NM_021012	Hs.200629	KCNJ12	17p11.2	231					993	
NM_004962	Hs.2171	GDF10	10q11.22	232					994	
NM_000129	Hs.335513	F13A1	6p25.1	233			3		1280	NvsT
BX097190	Hs.7413	BX097190	3q23	234					995	
NM_133631	Hs.13640	ROBO1	3p12.3	235	1				1624	
NM_016315	Hs.470887	GULP1	2q32.2	236			3		1129	
NM_002091	Hs.153444	GRP	18q21.32	237					1130	
A_32_P131449		A_32_P131449	5q33.1	238	1		3		1131	
NM_182943	Hs.477866	PLOD2	3q24	239					1022	
BC044246	Hs.591341	KIAA1913	6q23.1	240					10	
NM_001873	Hs.75360	CPE	4q32.3	241			3		1244	
NM_007029	Hs.521651	STMN2	8q21.13	242					608	NvsT
S82024		STMN2	8q21.13	243					609	NvsT
NM_006744	Hs.50223	RBP4	10q23.33	244					1127	
AK124396	Hs.633447	AK124396	4q31.22	245					610	NvsT
BC092429	Hs.660885	BC092429	17p13.1	246					611	NvsT

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_183376	Hs.6093	ARRDC4	15q26.3	247					1238	NvsT
NM_006472	Hs.533977	TXNIP	1q21.1	248					1239	NvsT
NM_006274	Hs.50002	CCL19	9p13.3	249					1160	
NM_004867	Hs.17109	ITM2A	Xq21.1	250					431	NvsT
NM_001928	Hs.155597	CFD	19p13.3	251					1241	NvsT
NM_004167	Hs.656286	CCL15	17q12	252					1240	NvsT
NM_003278	Hs.476092	CLEC3B	3p21.31	253					615	NvsT
NM_001932	Hs.396566	MPP3	17q21.31	254	1				1639	
NM_207380	Hs.32433	FLJ43339	15q15.1	255			3		1132	
NM_003417	Hs.515634	ZNF264	19q13.43	256					1136	
NM_017870	Hs.118552	TMEM132A	11q12.2	257					1631	NvsT
NM_000152	Hs.1437	GAA	17q25.3	258	1				1137	
NM_015326	Hs.497575	SRGAP2	1q32.1	259	1				1637	
NM_006885	Hs.652666	ATBF1	16q22.3	260					1138	
AK129956	Hs.170131	AK129956	19p13.3	261	1				1139	
NM_003507	Hs.173859	FZD7	2q33.1	262					1135	
NM_032780	Hs.564188	TMEM25	11q23.3	263			3		1214	
NM_014380	Hs.448588	NGFRAP1	Xq22.2	264					1245	
NM_016269	Hs.555947	LEF1	4q25	265					1626	NvsT
NM_000076	Hs.106070	CDKN1C	11p15.4	266	1		3		1248	
NM_001546	Hs.519601	ID4	6p22.3	267					1249	
NM_000237	Hs.180878	LPL	8p21.3	268			3		1246	
NM_002514	Hs.235935	NOV	8q24.12	269					1145	
AJ710526	Hs.595068	AJ710526	10q23.33	270			3		1215	
NM_016606	Hs.416090	REEP2	5q31.2	271					1158	NvsT
NM_194272	Hs.436518	RBPMS2	15q22.31	272					1167	NvsT
NM_000702	Hs.34114	ATP1A2	1q23.2	273			3		1168	NvsT
NM_003248	Hs.211426	THBS4	5q14.1	274			3		1301	
NM_000743	Hs.89605	CHRNA3	15q25.1	275	1		3		1169	NvsT
NM_002247	Hs.144795	KCNMA1	10q22.3	276	1				1166	NvsT
NM_015424	Hs.432379	CHRD12	11q13.4	277					422	
NM_032105	Hs.444403	PPP1R12B	1q32.1	278			3		1188	NvsT
AL137734	Hs.437437	DKFZp586C0721	5q33.1	279			3		1171	
NM_004137	Hs.484099	KCNMB1	5q35.1	280	1		3		1189	NvsT
AL833294	Hs.655519	SYNPO2	4q26	281			3		1190	NvsT
NM_001615	Hs.516105	ACTG2	2p13.1	282	1		3		1191	NvsT
NM_001299	Hs.465929	CNN1	19p13.2	283	1		3		1192	NvsT
NM_053025	Hs.556600	MYLK	3q21.1	284	1		3		1193	NvsT
NM_002474	Hs.460109	MYH11	16p13.11	285	1		3		1194	NvsT
NM_022844	Hs.460109	MYH11	16p13.11	286	1		3		1195	NvsT
NM_001927	Hs.594952	DES	2q35	287			3		1196	NvsT

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_145728	Hs.207106	DMN	15q26.3	288			3		1197	NvsT
NM_203339	Hs.436657	CLU	8p21.1	289	1		3		1159	NvsT
NM_022370	Hs.435621	ROBO3	11q24.2	290	1				1465	
NM_002145	Hs.514289	HOXB2	17q21.32	291	1				1424	
NM_006622	Hs.398157	PLK2	5q11.2	292	1				1425	
NM_080927	Hs.203691	DCBLD2	3q12.1	293	1		3		1426	
NM_014945	Hs.49688	ABLI1	5q33.1	294					1427	
NM_018043	Hs.503074	TMEM16A	11q13.3	295	1				1428	
NM_031866	Hs.302634	FZD8	10p11.21	296			3		1128	
ENST00000367013	Hs.446946	C1orf133	1q32.2	297			3		1258	
NM_078487	Hs.72901	CDKN2B	9p21.3	298			3		616	NvsT
DB381305	Hs.629427	DB381305	18q21.33	299					617	NvsT
NM_152852	Hs.523702	MS4A6A	11q12.1	300					1466	NvsT
NM_001080425	Hs.184736	BFXL1	Xq22.1	301			3		1198	NvsT
NM_021939	Hs.463035	FKBP10	17q21.2	302	1		3		1625	NvsT
NM_002961	Hs.654444	S100A4	1q21.3	303	1		3		1267	
NM_001423	Hs.693682	EMP1	12p13.1	304			3		1151	NvsT
NM_000700	Hs.494173	ANXA1	9q21.13	305	1				1266	
NM_001008540	Hs.593413	CXCR4	2q21.3	306	1				1492	
NM_002922	Hs.75256	RGS1	1q31.2	307	1				1493	
NM_005627	Hs.510078	SGK	6q23.2	308					1233	NvsT
NM_006732	Hs.590958	FOSB	19q13.32	309					1113	
NM_002928	Hs.413297	RGS16	1q25.3	310	1				1309	NvsT
THC2516687	Hs.8867	THC2516687	1p22.3	311	1				1310	
NM_000399	Hs.1395	EGR2	10q21.2	312	1		3		1311	
NM_001901	Hs.591346	CTGF	6q23.2	313	1		3		1312	
NM_001554	Hs.8867	CYR61	1p22.3	314	1		3		1313	
NM_001781	Hs.208854	CD69	12p13.31	315					1119	
NM_004417	Hs.171695	DUSP1	5q35.1	316	1				1120	
NM_002923	Hs.78944	RGS2	1q31.2	317	1				1121	
NM_004951	Hs.784	EBI2	13q32.3	318					1125	
NM_005261	Hs.654463	GEM	8q22.1	319	1		3		1123	
ENST00000339446	Hs.530443	LOC387763	11p11.2	320	1		3		1124	
NM_006988	Hs.643357	ADAMTS1	21q21.3	321					1122	NvsT
NM_173198	Hs.279522	NR4A3	9q31.1	322					1126	
NM_006108	Hs.654637	SPON1	11p15.2	323	1		3		1220	NvsT
NM_001080393	Hs.431092	GLT8D4	3p13	324	1		3		1268	
AW961597	Hs.675553	AW961597	3p13	325	1		3		1269	
THC2515368		THC2515368	3p13	326			3		1270	
NM_021020	Hs.521432	LZTS1	8p21.3	327	1		3		1635	NvsT
NM_014571	Hs.472566	HEYL	1p34.2	328			3		1636	NvsT

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
AK023854	Hs.526594	OBSL1	2q35	329	1		3		1302	
NM_024563	Hs.13528	C5orf23	5p13.3	330			3		1303	
NM_030762	Hs.177841	BHLHB3	12p12.1	331	1		3		1253	
NM_181847	Hs.121520	AMIGO2	12q13.11	332			3		1254	
NM_000693	Hs.459538	ALDH1A3	15q26.3	333			3		1247	
NM_000014	Hs.212838	A2M	12p13.31	334	1		3		1212	NvsT
AI_833309	Hs.622596	AI_833309	3q26.2	335			3		1210	NvsT
NM_001003793	Hs.221436	RBMS3	3p24.1	336	1		3		1213	NvsT
NM_003277	Hs.505337	CLDN5	22q11.21	337	1				1199	NvsT
NM_004684	Hs.62886	SPARCL1	4q22.1	338	1		3		1200	NvsT
NM_021219	Hs.517227	JAM2	21q21.3	339	1				1201	NvsT
NM_021914	Hs.180141	CFL2	14q13.2	340	1		3		1176	NvsT
NM_015687	Hs.526972	FILIP1	6q14.1	341	1		3		1177	NvsT
NM_032784	Hs.135254	RSPO3	6q22.33	342	1		3		1156	
NM_024769	Hs.591949	ASAM	11q24.1	343	1		3		1157	
AK074473	Hs.559353	C20orf82	20p12.1	344	1		3		1320	
NM_004791	Hs.584836	ITGBL1	13q33.1	345	1		3		1321	
NM_000095	Hs.1584	COMP	19p13.11	346	1		3		1322	
NM_005940	Hs.143751	MMP11	22q11.23	347	1		3		1628	NvsT
NM_003326	Hs.181097	TNFSF4	1q25.1	348	1		3		1385	
NM_000493	Hs.520339	COL10A1	6q22.1	349	1		3		1383	NvsT
BF337308	Hs.619396	BF337308		350	1		3		1386	
NM_007021	Hs.93675	C10orf10	10q11.21	351	1		3		1314	
NM_058229	Hs.403933	FBXO32	8q24.13	352	1		3		1172	NvsT
AK023391	Hs.592775	AK023391	8q24.13	353	1		3		1173	NvsT
NM_022343	Hs.493819	C9orf19	9p13.3	354	1				1235	NvsT
A_23_P421323		A_23_P421323		355	1		3		1174	NvsT
NM_052966	Hs.518662	FAM129A	1q25.3	356	1		3		1175	NvsT
NM_005909	Hs.584777	MAP1B	5q13.2	357	1		3		1276	
AK055112	Hs.637017	AK055112	5q13.2	358	1		3		1277	
NM_015833	Hs.474018	ADARB1	21q22.3	359	1		3		1272	
NM_015103	Hs.301685	PLXND1	3q21.3	360	1				1327	
NM_003062	Hs.604116	SLIT3	5q35.1	361	1		3		1295	
NM_000609	Hs.522891	CXCL12	10q11.21	362	1		3		1275	
NM_012261	Hs.22920	C20orf103	20p12.2	363	1		3		1278	
NM_173653	Hs.302257	SLC9A9	3q24	364	1		3		1202	NvsT
NM_024692	Hs.122927	CLIP4	2p23.2	365	1		3		1207	NvsT
NM_199072	Hs.427236	MDFIC	7q31.2	366	1		3		1351	
NM_002113	Hs.363396	CFHR1	1q31.3	367	1		3		1203	
NM_021023	Hs.575869	CFHR3	1q31.3	368	1		3		1204	NvsT
NM_001014975	Hs.363396	CFH	1q31.3	369	1		3		1205	NvsT

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_003881	Hs.592145	WISP2	20q13.12	370			3		1289	
NM_000618	Hs.160562	IGF1	12q23.2	371			3		1161	
W60781	Hs.391561	W60781	8q21.13	372	1		3		1163	
ENST00000372045	Hs.496587	ENST00000372045	Xq22.3	373	1		3		1162	NvsT
A_32_P210193		A_32_P210193	12p12.3	374	1		3		1164	
NM_003613	Hs.442180	CILP	15q22.31	375	1		3		1297	
NM_000900	Hs.365706	MGP	12p12.3	376	1		3		1298	
NM_000961	Hs.302085	PTGIS	20q13.13	377	1		3		1299	
AK055387	Hs.21639	SPEG	2q35	378	1		3		1178	NvsT
NM_002667	Hs.170839	PLN	6q22.31	379	1		3		1186	NvsT
AK129844	Hs.170839	PLN	6q22.31	380	1		3		1187	NvsT
NM_145056	Hs.515490	DACT3	19q13.32	381	1		3		1185	
AI042308	Hs.666842	AI042308	11q23.3	382	1		3		1183	
NM_213674	Hs.300772	TPM2	9p13.3	383	1		3		1179	
NM_003734	Hs.198241	AOC3	17q21.31	384	1		3		1184	NvsT
NM_181526	Hs.504687	MYL9	20q11.23	385	1		3		1180	
NM_001613	Hs.500483	ACTA2	10q23.31	386	1		3		1181	
NM_001001522	Hs.632099	TAGLN	11q23.3	387	1		3		1182	
NM_004787	Hs.29802	SLIT2	4p15.31	388	1		3		1292	
AW268902	Hs.29802	AW268902	4p15.31	389	1		3		1293	
NM_198148	Hs.656887	CPXM2	10q26.13	390	1		3		1290	
NM_002048	Hs.65029	GAS1	9q21.33	391	1		3		1282	
NM_199168	Hs.522891	CXCL12	10q11.21	392	1		3		1291	NvsT
NM_003480	Hs.512842	MFAP5	12p13.31	393	1		3		1283	
NM_000104	Hs.154654	CYP1B1	2p22.2	394	1		3		1284	
NM_003013	Hs.481022	SFRP2	4q31.3	395	1		3		1285	
NM_003239	Hs.592317	TGFB3	14q24.3	396	1		3		1287	
BE835321	Hs.592317	BE835321	14q24.3	397	1		3		1288	
NM_199511	Hs.477128	CCDC80	3q13.2	398	1		3		1279	
NM_006207	Hs.458573	PDGFRL	8p22	399	1		3		1376	
NM_003485	Hs.8882	GPR68	14q32.12	400	1		3		1340	
NM_001039580	Hs.61271	MAP9	4q32.1	401	1		3		1260	
NM_002023	Hs.519168	FMOD	1q32.1	402	1				1317	
NM_206943	Hs.654497	LTBP1	2p22.3	403	1		3		1338	
AL049443	Hs.660870	AL049443	2p16.3	404	1		3		1170	
NM_000064	Hs.529053	C3	19p13.3	405	1		3		1319	
A_24_P561165		A_24_P561165	11q12.1	406	1		3		1391	
NM_005099	Hs.211604	ADAMTS4	1q23.3	407	1		3		1324	
NM_212482	Hs.203717	FN1	2q35	408	1		3		1325	
NM_170744	Hs.585457	UNC5B	10q22.1	409	1		3		1339	
NM_001001396	Hs.343522	ATP2B4	1q32.1	410	1		3		1273	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
AK024680	Hs.660596	AK024680	2q33.3	411	1		3		1274	
NM_004370	Hs.101302	COL12A1	6q13	412	1		3		1397	NvsT
NM_138440	Hs.372579	VASN	16p13.3	413	1		3		1332	
NM_012445	Hs.302963	SPON2	4p16.3	414	1		3		1335	NvsT
NM_007361	Hs.369840	NID2	14q22.1	415	1		3		1333	NvsT
AB058761	Hs.54925	ZNF469	16q24.2	416	1		3		1337	
NM_001849	Hs.420269	COL6A2	21q22.3	417	1		3		1328	
NM_001848	Hs.474053	COL6A1	21q22.3	418	1		3		1329	
Z74615	Hs.172928	COL1A1	17q21.33	419	1		3		1334	NvsT
NM_002609	Hs.509067	PDGFRB	5q33.1	420	1		3		1336	NvsT
NM_001846	Hs.508716	COL4A2	13q34	421	1				1330	
AW138903	Hs.8546	AW138903	19p13.12	422	1		3		1326	NvsT
BX427588	Hs.669724	BX427588	7q21.3	423	1		3		1392	
NM_002889	Hs.647064	RARRES2	7q36.1	424	1		3		1318	
NM_013372	Hs.40098	GREM1	15q13.3	425	1		3		1381	
BC039414	Hs.655414	BC039414	4q32.1	426	1		3		1382	
NM_002404	Hs.296049	MFAP4	17p11.2	427	1		3		1211	NvsT
NM_005269	Hs.632702	GLI1	12q13.3	428	1		3		1358	
NM_004349	Hs.368431	RUNX1T1	8q21.3	429	1		3		1359	
ENST00000377047	Hs.444329	GPC6	13q32.1	430	1		3		1356	
NM_006486	Hs.24601	FBLN1	22q13.31	431	1		3		1357	
NM_001734	Hs.458355	C1S	12p13.31	432	1		3		1354	
NM_004598	Hs.654695	SPOCK1	5q31.2	433	1		3		1281	
NM_181435	Hs.171929	C1QTNF3	5p13.3	434	1		3		1377	
BC020879	Hs.664877	MGC24103	9p22.3	435	1		3		1379	
NR_002766	Hs.525589	MEG3	14q32.2	436	1		3		1378	
NM_014476	Hs.85862	PDLIM3	4q35.1	437	1		3		1206	
NM_005086	Hs.183428	SSPN	12p12.1	438	1		3		1352	
NM_004105	Hs.76224	EFEMP1	2p16.1	439	1		3		1353	
NM_153026	Hs.524348	PRICKLE1	12q12	440	1		3		1296	
NM_006873	Hs.44385	STON1	2p16.3	441	1		3		1361	
AK095791	Hs.411391	AK095791	11q24.1	442	1		3		1380	
NM_001856	Hs.368921	COL16A1	1p35.2	443	1		3		1348	
NM_016651	Hs.48950	DACT1	14q23.1	444	1		3		1349	
NM_002192	Hs.583348	INHBA	7p14.1	445	1		3		1632	NvsT
NM_014112	Hs.657018	TRPS1	8q23.3	446	1		3		1350	
NM_032812	Hs.658134	PLXDC2	10p12.31	447	1		3		1375	
NM_130386	Hs.464422	COLEC12	18p11.32	448	1		3		1294	
NM_053044	Hs.479119	HTRA3	4p16.1	449	1		3		1395	
NM_001004019	Hs.198862	FBLN2	3p25.1	450	1		3		1389	
NM_024893	Hs.124638	C20orf39	20p11.21	451	1		3		1413	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_000362	Hs.644633	TIMP3	22q12.3	452	1		3		1409	
NM_002615	Hs.532768	SERPINF1	17p13.3	453	1		3		1411	
NM_005576	Hs.65436	LOXL1	15q24.1	454	1		3		1422	
NM_144601	Hs.298198	CMTM3	16q22.1	455	1		3		1423	
NM_003380	Hs.642813	VIM	10p12.33	456	1		3		1368	
THC2574606		THC2574606	10p12.33	457	1		3		1369	
NM_020856	Hs.278436	TSHZ3	19q12	458	1		3		1367	
ENST00000375377	Hs.533953	KIAA1462	10p11.23	459	1		3		1365	
NM_012232	Hs.437191	PTRF	17q21.31	460	1		3		1366	
NM_007085	Hs.269512	FSTL1	3q13.33	461	1		3		1420	
NM_033138	Hs.490203	CALD1	7q33	462	1		3		1362	
NM_000393	Hs.445827	COL5A2	2q32.2	463	1		3		1398	NvsT
NM_000138	Hs.591133	FBN1	15q21.1	464	1		3		1421	
NM_000089	Hs.489142	COL1A2	7q21.3	465	1		3		1399	
NM_000090	Hs.443625	COL3A1	2q32.2	466	1		3		1400	
NM_003118	Hs.111779	SPARC	5q33.1	467	1		3		1401	NvsT
NM_006169	Hs.503911	NNMT	11q23.2	468	1		3		1373	
NM_005110	Hs.30332	GFPT2	5q35.3	469	1		3		1374	
NM_016205	Hs.570855	PDGFC	4q32.1	470	1		3		1410	
NM_002775	Hs.501280	HTRA1	10q26.13	471	1		3		1414	
NM_023110	Hs.264887	FGFR1	8p12	472	1		3		1363	
NM_000168	Hs.21509	GLI3	7p14.1	473	1		3		1364	
NM_001129	Hs.439463	AEBP1	7p13	474	1		3		1390	
NM_003014	Hs.658169	SFRP4	7p14.1	475	1		3		1394	NvsT
NM_005202	Hs.353001	COL8A2	1p34.3	476	1		3		1412	
NM_001850	Hs.654548	COL8A1	3q12.1	477	1		3		1402	NvsT
AL359062	Hs.134830	AL359062	3q12.1	478	1		3		1403	NvsT
NM_015170	Hs.409602	SULF1	8q13.3	479	1		3		1406	NvsT
L12350	Hs.371147	THBS2	6q27	480	1		3		1407	NvsT
NM_032532	Hs.520525	FNDC1	6q25.3	481	1		3		1417	
NM_001797	Hs.116471	CDH11	16q21	482	1		3		1418	NvsT
NM_032208	Hs.165859	ANTXR1	2p14	483	1		3		1419	
NM_005504	Hs.438993	BCAT1	12p12.1	484	1		3		1346	
NM_006475	Hs.136348	POSTN	13q13.3	485	1		3		1345	
NM_006868	Hs.99528	RAB31	18p11.22	486	1		3		1415	
NM_138455	Hs.405614	CTHRC1	8q22.3	487	1		3		1404	NvsT
NM_004460	Hs.654370	FAP	2q24.2	488	1		3		1405	NvsT
NM_017680	Hs.435655	ASPN	9q22.31	489	1		3		1408	
NM_000396	Hs.632466	CTSK	1q21.2	490	1		3		1393	
NM_001920	Hs.156316	DCN	12q21.33	491	1		3		1355	
NM_002345	Hs.406475	LUM	12q21.33	492	1		3		1416	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_152330	Hs.434914	FRMD6	14q22.1	493	1		3		1360	
NM_003596	Hs.421194	TPST1	7q11.21	494	1		3		1372	
NM_016206	Hs.435013	VGLL3	3p12.1	495	1		3		1370	
NM_030786	Hs.655938	SYNC1	1p35.1	496	1		3		1371	
BM999343	Hs.99528	BM999343	18p11.22	497	1		3		1388	
NM_004530	Hs.513617	MMP2	16q12.2	498	1		3		1387	
NM_003247	Hs.371147	THBS2	6q27	499	1		3		1396	
NM_001030059	Hs.40479	PPAPDC1A	10q26.12	500	1		3		1384	
NM_004403	Hs.520708	DFNA5	7p15.3	501	1		3		1341	
NM_018004	Hs.658956	TMEM45A	3q12.2	502	1				1343	
NM_003254	Hs.522632	TIMP1	Xp11.23	503	1		3		1633	NvsT
NM_014583	Hs.475353	LMCD1	3p26.1	504	1		3		1316	
NM_032849	Hs.646647	C13orf33	13q12.3	505	1		3		1286	
AY098593	Hs.439341	KGFLP1	9p11.2	506	1		3		1263	
NM_004995	Hs.2399	MMP14	14q11.2	507	1		3		1304	NvsT
NM_178234	Hs.591845	TUSC3	8p22	508					1261	
NM_006832	Hs.509343	PLEKHC1	14q22.2	509	1		3		1264	
NM_000165	Hs.74471	GJA1	6q22.31	510			3		1265	
AF200348	Hs.332197	PXDN	2p25.3	511	1				1331	
NM_016307	Hs.660115	PRRX2	9q34.11	512	1		3		1305	
AW276332	Hs.406475	AW276332	12q21.33	513	1		3		1306	
NM_002998	Hs.1501	SDC2	8q22.1	514	1		3		1307	
AK022110	Hs.289044	AK022110	5q11.2	515	1		3		1308	
NM_031442	Hs.8769	TMEM47	Xp21.1	516	1		3		1256	
NM_018357	Hs.416755	LARP6	15q23	517			3		1250	
NM_005725	Hs.310458	TSPAN2	1p13.2	518	1		3		1257	
NM_001033564	Hs.591340	LOC619208	6q21	519	1		3		1347	
NM_005711	Hs.482730	EDIL3	5q14.3	520					1148	NvsT
NM_004348	Hs.535845	RUNX2	6p12.3	521	1		3		1271	
NM_013409	Hs.9914	EST	5q11.2	522			3		1262	
NM_016613	Hs.567498	C4orf18	4q32.1	523			3		1255	
AK025997	Hs.655087	MCTP1	5q15	524					1096	
NM_000565	Hs.591492	IL6R	1q21.3	525					1237	NvsT
AB002384	Hs.559459	C6orf32	6p22.2	526					1054	
NM_021615	Hs.655622	CHST6	16q23.1	527	1			4	983	
NM_015184	Hs.202010	PLCL2	3p24.3	528					1153	NvsT
NM_001874	Hs.654387	CPM	12q15	529	1				1234	NvsT
NM_005949	Hs.513626	MT1F	16q13	530					1223	NvsT
NM_176870	Hs.647370	MT1M	16q13	531	1				1226	NvsT
NM_005953	Hs.647371	MT2A	16q13	532	1				1224	NvsT
ENST00000245185	Hs.647371	MT2A	16q13	533	1				1225	NvsT

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_005951	Hs.438462	MT1H	16q13	534	1				1227	NvsT
NM_175617	Hs.534330	MT1E	16q13	535	1				1228	NvsT
NM_005952	Hs.374950	MT1X	16q13	536	1				1229	NvsT
X97261	Hs.647358	MT1L	16q13	537	1				1230	NvsT
NM_005950	Hs.433391	MT1G	16q13	538	1				1231	NvsT
NM_005947	Hs.656629	MT1B	16q13	539	1				1232	NvsT
NM_000450	Hs.89546	SELE	1q24.2	540					1051	
NM_000552	Hs.440848	VWF	12p13.31	541	1				1052	
NM_005282	Hs.17170	GPR4	19q13.32	542	1				1323	NvsT
NM_018371	Hs.655166	ChGn	8p21.3	543	1				1063	
NM_170736	Hs.411299	KCNJ15	21q22.13	544	1				1064	
AA593970	Hs.608251	AA593970	2p14	545				4	1106	NvsT
NM_001124	Hs.441047	ADM	11p15.4	546					1060	
NM_139314	Hs.9613	ANGPTL4	19p13.2	547	1				1061	
A_23_P170719		A_23_P170719	19p13.2	548					1062	
NM_001109	Hs.501574	ADAM8	10q26.3	549	1				1066	
NM_138931	Hs.478588	BCL6	3q27.3	550	1				1067	
A_23_P123234		A_23_P123234	7q22.1	551	1				1315	
BC060766	Hs.655169	SLC2A14	12p13.31	552	1				1069	
NM_000201	Hs.643447	ICAM1	19p13.2	553	1				1456	
NM_006291	Hs.525607	TNFAIP2	14q32.32	554	1				1457	
NM_002133	Hs.517581	HMOX1	22q12.3	555	1				1053	
NM_145699	Hs.348983	APOBEC3A	22q13.1	556					1070	
NM_173842	Hs.81134	IL1RN	2q13	557	1			4	1081	
NM_006018	Hs.458425	GPR109B	12q24.31	558				4	1083	
NM_021935	Hs.528665	PROK2	3p13	559				4	1084	
NM_000576	Hs.126256	IL1B	2q13	560				4	1082	
NM_000584	Hs.624	IL8	4q13.3	561				4	1077	NvsT
NM_007115	Hs.437322	TNFAIP6	2q23.3	562	1				1344	
NM_018643	Hs.283022	TREM1	6p21.1	563	1				1078	NvsT
NM_002984	Hs.75703	CCL4	17q12	564	1				1085	
D00044	Hs.514107	CCL3	17q12	565	1				1086	
NM_003955	Hs.527973	SOCS3	17q25.3	566	1				1068	
NM_020530	Hs.248156	OSM	22q12.2	567	1			4	1079	
NM_015714	Hs.432132	G0S2	1q32.2	568	1				1080	
NM_000636	Hs.487046	SOD2	6q25.3	569					1071	
NM_002965	Hs.112405	S100A9	1q21.3	570					1073	
NM_002432	Hs.153837	MNDA	1q23.1	571	1				1091	
NM_016184	Hs.504657	CLEC4A	12p13.31	572					1090	
NM_014358	Hs.236516	CLEC4E	12p13.31	573	1				1089	
NM_002664	Hs.468840	PLEK	2p14	574	1				1092	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_005565	Hs.304475	LCP2	5q35.1	575	1				1093	
NM_000433	Hs.587558	NCF2	1q25.3	576	1				1094	
NM_006864	Hs.631592	LILRB3	19q13.42	577	1				1095	
NM_020980	Hs.104624	AQP9	15q22.1	578	1				1072	NvsT
NM_002727	Hs.1908	SRGN	10q21.3	579	1				1087	
NM_004049	Hs.227817	BCL2A1	15q25.1	580					1074	
NM_002029	Hs.753	FPR1	19q13.33	581	1				1075	
NM_002964	Hs.416073	S100A8	1q21.3	582					1076	
NM_002852	Hs.591286	PTX3	3q25.32	583					1216	
NM_001001547	Hs.120949	CD36	7q21.11	584					1242	NvsT
NM_052960	Hs.422688	RBP7	1p36.22	585	1		3		1218	
NM_005623	Hs.271387	CCL8	17q12	586	1				1221	NvsT
NM_181755	Hs.195040	HSD11B1	1q32.2	587	1				1342	
NM_002982	Hs.303649	CCL2	17q12	588	1				1217	
NM_182566	Hs.122561	VMO1	17p13.2	589	1			4	1065	
NM_004994	Hs.297413	MMP9	20q13.12	590	1		3		1056	
NM_006770	Hs.67726	MARCO	2q14.2	591	1				1055	
NM_000582	Hs.313	SPP1	4q22.1	592	1		3		1057	NvsT
NM_024508	Hs.136912	ZBED2	3q13.13	593					1431	
L34088	Hs.387679	HLA-DQA1		594					1433	
NM_020056	Hs.591798	HLA-DQA2	6p21.32	595					1434	
X12544	Hs.654405	X12544	6p21.32	596					1435	
NM_002996	Hs.531668	CX3CL1	16q13	597	1				1454	
NM_030643	Hs.115099	APOL4	22q12.3	598					1455	
NM_001165	Hs.127799	BIRC3	11q22.2	599					428	
NM_003004	Hs.558009	SECTM1	17q25.3	600					429	NvsT
NM_016135	Hs.272398	ETV7	6p21.31	601					1444	
NM_005409	Hs.632592	CXCL11	4q21.1	602					1452	
NM_006398	Hs.44532	UBD	6p22.1	603					1453	NvsT
NM_001010919	Hs.381220	FAM26F	6q22.1	604		2			1445	
NM_007315	Hs.642990	STAT1	2q32.2	605					1446	
NM_002053	Hs.62661	GBP1	1p22.2	606	1				1447	
NM_002416	Hs.77367	CXCL9	4q21.1	607					1448	
NM_001565	Hs.632586	CXCL10	4q21.1	608	1				1449	
NM_052941	Hs.409925	GBP4	1p22.2	609					1450	
NM_002164	Hs.840	INDO	8p11.21	610					1451	
NM_173574	Hs.353208	ZNF683	1p36.11	611	1				1440	
NM_001768	Hs.85258	CD8A	2p11.2	612					1443	
NM_033423	Hs.348264	GZMH	14q12	613	1				1441	
NM_002985	Hs.514821	CCL5	17q12	614					1442	
NM_002286	Hs.409523	LAG3	12p13.31	615					1437	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_006144	Hs.90708	GZMA	5q11.2	616					1438	
NM_006433	Hs.105806	GPLY	2p11.2	617	1				1439	
NM_002995	Hs.546295	XCL1	1q24.2	618					430	
NM_006399	Hs.509964	BATF	14q24.3	619					1436	
THC2657355		THC2657355	7q22.3	620					432	NvsT
NM_002838	Hs.654514	PTPRC	1q31.3	621	1				433	
NM_001778	Hs.243564	CD48	1q23.3	622					434	NvsT
NM_002258	Hs.169824	KLRB1	12p13.31	623					435	NvsT
NM_001803	Hs.276770	CD52	1p36.11	624					436	
NM_006564	Hs.34526	CXCR6	3p21.31	625					437	
NM_000732	Hs.504048	CD3D	11q23.3	626					438	
NM_198196	Hs.142023	CD96	3q13.2	627					439	
NM_198517	Hs.534648	TBC1D10C	11q13.1	628		2			440	
M20432	Hs.409934	HLA-DQB1	6p21.32	629					1463	
NM_002122	Hs.387679	HLA-DQA1	6p21.32	630					1464	
NM_198232	Hs.78224	RNASE1	14q11.2	631					1219	NvsT
NR_003491	Hs.517502	MIAT	22q12.1	632	1				1462	
NM_000417	Hs.231367	IL2RA	10p15.1	633					1461	
NM_006840	Hs.655338	LILRB5	19q13.42	634					1236	NvsT
M27126	Hs.654405	M27126		635					1482	
NM_004355	Hs.436568	CD74	5q33.1	636	1				1483	
NM_002118	Hs.654428	HLA-DMB	6p21.32	637	1				1484	
NM_002121	Hs.485130	HLA-DPB1	6p21.32	638	1				1485	
NR_001435		HLA-DPB2	6p21.32	639	1				1486	
BE926212		BE926212	4p13	640	1				1487	
NM_021983	Hs.654405	HLA-DRB4		641	1				1488	
BC106057	Hs.654405	HLA-DRB3	6p21.32	642					1489	
NM_002125	Hs.534322	HLA-DRB5	6p21.32	643	1				1490	
NM_002123	Hs.409934	HLA-DQB1	6p21.32	644	1				1491	
NM_000952	Hs.433540	PTAFR	1p35.3	645	1				1458	
NM_052942	Hs.513726	GBP5	1p22.2	646	1	2			1459	
NM_001953	Hs.592212	ECGF1	22q13.33	647	1	2			1460	
NM_033050	Hs.279575	SUCNR1	3q25.1	648	1		3		1472	
NM_002988	Hs.143961	CCL18	17q12	649	1	2			1471	
ENST00000368054	Hs.610260	CD84	1q23.2	650	1				1476	
NM_000211	Hs.375957	ITGB2	21q22.3	651	1				1477	
NM_001005340	Hs.190495	GPNMB	7p15.3	652	1		3		1478	
NM_001645	Hs.110675	APOC1	19q13.32	653	1				1479	
NM_000041	Hs.654439	APOE	19q13.32	654	1		3		1480	
THC2582296	Hs.654439	THC2582296	19q13.32	655	1				1481	
NM_001017986	Hs.534956	FCGR1B	1p11.2	656	1				1473	NvsT

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_138715	Hs.147635	MSR1	8p22	657	1		3		1467	
BC031655	Hs.10649	C1orf38	1p35.3	658	1				1088	
NM_007268	Hs.8904	VSIG4	Xq12	659	1				1468	
NM_004244	Hs.504641	CD163	12p13.31	660	1				1469	
NM_024021	Hs.325960	MS4A4A	11q12.2	661	1				1470	
NM_000569	Hs.372679	FCGR3A	1q23.3	662	1				1474	
NM_004172	Hs.481918	SLC1A3	5p13.2	663	1		3		1475	
BC022548	Hs.125293	RASGEF1A	10q11.21	664					1495	
NM_005532	Hs.532634	IFI27	14q32.13	665					1497	
NM_004335	Hs.118110	BST2	19p13.11	666					1498	
NM_017631	Hs.591710	FLJ20035	4q32.3	667					1496	NvsT
NM_006417	Hs.82316	IFI44	1p31.1	668	1				1499	
NM_001547	Hs.437609	IFIT2	10q23.31	669					1501	
NM_001548	Hs.20315	IFIT1	10q23.31	670					1502	
NM_080657	Hs.17518	RSAD2	2p25.2	671					1503	
NM_006820	Hs.389724	IFI44L	1p31.1	672					1500	
NM_002462	Hs.517307	MX1	21q22.3	673					1504	
NM_005101	Hs.458485	ISG15	1p36.33	674					1505	
NM_004796	Hs.368307	NRXN3	14q31.1	675					1641	
NM_017855	Hs.143811	ODAM	4q13.3	676					1642	
AB007954	Hs.604754	KIAA0485	1p31.3	677					1688	
NM_017415	Hs.655084	KLHL3	5q31.2	678					1608	
BX537788	Hs.688969	BX537788	10p14	679					1689	
NM_032679	Hs.148322	ZNF577	19q13.33	680					225	
AK125482	Hs.591496	MEX3A	1q22	681					274	NvsT
XM_044166	Hs.591496	MEX3A	1q22	682					275	NvsT
NM_005723	Hs.591706	TSPAN5	4q23	683					1610	NvsT
AK055659	Hs.591706	TSPAN5	4q23	684					1611	NvsT
BC010538	Hs.657197	C18orf18	18p11.31	685					1644	
DQ179139	Hs.558645	C21orf99	9q12	686					1609	
NM_003970	Hs.443683	MYOM2	8p23.3	687					195	
NR_003288	Hs.674810	LOC729603	6q25.3	688					197	
AK128234	Hs.416925	CEACAM19	19q13.31	689					198	
BC030123	Hs.163155	LOC441461	9q31.1	690	1				199	
BC062758	Hs.571424	BC062758	8q21.11	691					200	
BC034319	Hs.363526	BC034319	8p12	692					1722	
NM_014289	Hs.496593	CAPN6	Xq22.3	693					1537	
NM_005654	Hs.519445	NR2F1	5q15	694					1721	
NM_004058	Hs.584744	CAPS	19p13.3	695					201	
AB051549	Hs.525247	ZFH2	14q11.2	696					898	
NM_030770	Hs.46720	TMPRSS5	11q23.2	697					1562	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_152529	Hs.516604	GPR155	2q31.1	698					1533	
NM_001033045	Hs.516604	GPR155	2q31.1	699					1534	
NM_030817	Hs.23388	APOLD1	12p13.1	700					1607	
NM_032866	Hs.148989	CGNL1	15q21.3	701					1251	
NM_021069	Hs.655143	SORBS2	4q35.1	702					1539	
AK094603		AK094603	4q35.1	703					1540	
NM_001853	Hs.126248	COL9A3	20q13.33	704					1567	
NM_014454	Hs.591336	SESN1	6q21	705					1564	
THC2651501		THC2651501	2q14.2	706					1565	
NM_002193	Hs.1735	INHBB	2q14.2	707					1566	
NM_003483	Hs.505924	HMGA2	12q14.3	708					1555	
THC2655811		THC2655811	12q14.3	709					1556	
NM_000131	Hs.36989	F7	13q34	710					1543	
A_24_P863363		A_24_P863363	14q23.1	711					1544	
NM_207381	Hs.306343	TNFAIP8L3	15q21.2	712					1545	
NM_001007139	Hs.523414	IGF2	11p15.5	713					1546	
NM_001971	Hs.348395	ELA1	12q13.13	714					1547	
THC2647658		THC2647658	5q35.3	715					1548	
NM_000612	Hs.523414	IGF2	11p15.5	716					1549	
THC2644672		THC2644672	16p13.3	717					1550	
AK074614		AK074614	11p15.5	718					1552	
AL117401	Hs.534980	DKFZP434P211	22q11.23	719					1551	
THC2532455		THC2532455	2q37.3	720					1553	
BC104421	Hs.645610	BC104421	11p15.5	721					1554	
NM_000371	Hs.427202	TTR	18q12.1	722					415	NvsT
NM_001063	Hs.518267	TF	3q22.1	723					808	
AK055418	Hs.592518	AK055418	10q21.1	724					1568	
NM_006516	Hs.473721	SLC2A1	1p34.2	725					1638	
NM_001202	Hs.68879	BMP4	14q22.2	726					1583	
NM_004626	Hs.108219	WNT11	11q13.5	727					390	
NM_002899	Hs.529571	RBP1	3q23	728					1541	
NR_002196	Hs.533566	H19	11p15.5	729					1542	
NM_015432	Hs.188781	PLEKHG4	16q22.1	730					1614	NvsT
A_24_P916853		A_24_P916853	8q24.21	731					1676	NvsT
AK097639	Hs.355950	CBWD3	9p11.2	732					1664	
NM_198853	Hs.645328	TRIM74	7q11.23	733					1665	
A_24_P852099		A_24_P852099		734					1666	
CR627362	Hs.654650	LOC595101	16p11.2	735					1667	
THC2551927		THC2551927	3q28	736					1668	
AK021738	Hs.632227	TMC6	17q25.3	737					1670	
AK057576	Hs.675850	AK057576	12q24.33	738					1669	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
THC2529512		THC2529512	19q13.11	739					1671	
L40520	Hs.654587	L40520	5q13.2	740					1672	
AL512711	Hs.26593	HDAC10	22q13.33	741					1673	
THC2781073		THC2781073	16q22.1	742					1674	
AF289610	Hs.675915	AF289610	2p11.1	743					1702	
THC2510656		THC2510656	1q21.1	744					1703	
NR_003370	Hs.348979	LOC730092	16p12.1	745					1704	
NR_001442	Hs.72222	FER1L4	20q11.22	746					1705	
NM_006887	Hs.503093	ZFP36L2	2p21	747					1134	
NM_021990	Hs.22785	GABRE	Xq28	748					1706	
AK124788		GGTL3	20q11.22	749					1707	
ENST00000378052	Hs.490512	ZNF783	7q36.1	750					1694	
BC041959	Hs.655573	BC041959	7q11.21	751					1695	
BC010934	Hs.693802	ZBTB20	3q13.31	752					1696	
AL136621	Hs.644041	ZMYM2	13q12.11	753			3		1697	
ENST00000379606	Hs.507783	ENST00000379606	13q13.3	754					1698	
THC2607363		THC2607363	5p13.3	755					1699	
THC2669975		THC2669975	7q31.1	756					1700	
THC2764893		THC2764893	3q26.31	757				4	1701	
AK024934	Hs.532921	KIAA1641	2q11.2	758					1690	
AK057596	Hs.651675	LOC150759	2q11.2	759					1691	
BC015370		BC015370	19q13.41	760					1692	
THC2746807		THC2746807	11p15.4	761					1693	
NR_002802	Hs.648467	TncRNA	11q13.1	762					1677	
THC2664215		THC2664215	11p15.5	763	1				1678	
AK025142		DST	6p12.1	764					1679	
AA837799	Hs.677704	AA837799	22q12.3	765	1		3		1680	
BC051317		RBBP6	16p12.1	766					1681	
BX090412	Hs.445847	BX090412	15q25.1	767					1682	
AL050097	Hs.667735	DKFZP586B0319	3q29	768					1683	
THC2611661		THC2611661	11q21	769					1684	
DB111455	Hs.660706	DB111455	11q21	770					1685	
THC2697162		THC2697162	3q13.31	771					1686	
AK129547	Hs.640055	AK129547	6p22.3	772					1687	
NM_014226	Hs.104119	RAGE	14q32.31	773					1506	
ENST00000360524	Hs.531547	LOC731479	17p11.2	774					1507	
ENST00000339425		ENST00000339425	21q11.2	775					1508	NvsT
BC016022	Hs.679496	BC016022		776					1509	
NM_014141	Hs.655684	CNTNAP2	7q36.1	777					868	
NM_007069	Hs.502775	HRASLS3	11q12.3	778					869	
NM_000153	Hs.513439	GALC	14q31.3	779					870	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_006569	Hs.159525	CGREF1	2p23.3	780					250	NvsT
NM_000849	Hs.2006	GSTM3	1p13.3	781					3	
NM_014505	Hs.525529	KCNMB4	12q15	782					4	
NM_013332	Hs.433213	HIG2	7q32.1	783					247	NvsT
NM_016354	Hs.235782	SLCO4A1	20q13.33	784					246	NvsT
BX640887	Hs.400876	LOC388796	20q11.23	785					66	
FNST00000360896	Hs.400876	FNST00000360896	20q11.23	786					67	NvsT
NM_004490	Hs.411881	GRB14	2q24.3	787					1662	
NM_152559	Hs.647042	WBSCR27	7q11.23	788					1663	
NM_001628	Hs.521212	AKR1B1	7q33	789					682	NvsT
NM_001150	Hs.1239	ANPEP	15q26.1	790					848	NvsT
NM_133637	Hs.191705	DQX1	2p13.1	791					940	
NR_002729	Hs.349110	MSTP9	1p36.13	792					849	
NM_004170	Hs.444915	SLC1A1	9p24.2	793					855	
NM_001935	Hs.368912	DPP4	2q24.2	794					856	
NM_002259	Hs.512576	KLRC1	12p13.2	795					857	
NM_002260	Hs.591157	KLRC2	12p13.2	796					858	
NM_030806	Hs.497159	C1orf21	1q25.3	797					859	NvsT
NM_024420	Hs.497200	PLA2G4A	1q31.1	798					860	
NM_002395	Hs.21160	ME1	6q14.2	799					920	
XM_940314	Hs.144372	NLF2	15q22.2	800					893	
NM_004350	Hs.170019	RUNX3	1p36.11	801					1429	
AF289562	Hs.592200	AF289562	22q13.1	802					863	
AK125361	Hs.276808	AK125361	22q13.1	803					864	
NM_004062	Hs.513660	CDH16	16q22.1	804					865	
NM_001080395	Hs.514575	AATK	17q25.3	805					866	
NM_001040105	Hs.271819	MUC17	7q22.1	806					867	
NM_006681	Hs.418367	NMU	4q12	807					1599	
AK126431	Hs.656660	LOC654433	2q13	808					862	
NM_017577	Hs.24583	GRAMD1C	3q13.31	809					853	NvsT
NM_004293	Hs.494163	GDA	9q21.13	810					854	
NM_020826	Hs.436643	SYT13	11p11.2	811					850	
NM_002847	Hs.490789	PTPRN2	7q36.3	812					852	
NM_001003954	Hs.181107	ANXA13	8q24.13	813					416	
NM_004617	Hs.133527	TM4SF4	3q25.1	814					861	
NM_000531	Hs.117050	OTC	Xp11.4	815					819	
NM_176794	Hs.421848	MRPL43	10q24.31	816					1569	
NM_004852	Hs.194725	ONECUT2	18q21.31	817					1571	NvsT
K03200	Hs.662242	K03200	3q29	818					1659	
AK027091	Hs.464404	AK027091		819					1660	
BC071773	Hs.626592	BC071773		820					1661	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_024786	Hs.659832	ZDHC11	5p15.33	821					904	
ENST00000377492		ENST00000377492	9q12	822					905	
NM_001089	Hs.26630	ABCA3	16p13.3	823					975	
NM_153350	Hs.513244	FBXL16	16p13.3	824					976	
AY030238	Hs.552649	LQK1	1q32.3	825					396	
NM_006896	Hs.660918	HOXA7	7p15.2	826					398	
NM_003317	Hs.94367	TITF1	14q13.3	827					399	
A_23_P215382		A_23_P215382		828		2			400	
NM_152739	Hs.659350	HOXA9	7p15.2	829		2			401	
NM_000522	Hs.592172	HOXA13	7p15.2	830					402	
A_32_P234853		A_32_P234853	7p15.2	831					403	
NR_002795	Hs.587427	HOXA11S	7p15.2	832					404	
AK093987	Hs.634354	AK093987	7p15.2	833					405	
NM_006547	Hs.648088	IGF2BP3	7p15.3	834					1432	
NM_032892	Hs.578544	FRMD5	15q15.3	835					1588	
NM_032876	Hs.655832	JUB	14q11.2	836	1		3		1627	NvsT
NM_032034	Hs.105607	SLC4A11	20p13	837				4	1640	
NM_002639	Hs.55279	SERPINB5	18q21.33	838				4	1594	
NM_012101	Hs.504115	TRIM29	11q23.3	839					1596	
NM_032405	Hs.208600	TMPRSS3	21q22.3	840					1597	
NM_144505	Hs.104570	KLK8	19q13.33	841					1600	
NM_144947	Hs.57771	KLK11	19q13.33	842				4	1601	
NM_002776	Hs.275464	KLK10	19q13.33	843					1602	NvsT
AK095258	Hs.16622	ZNF185	Xq28	844					63	
NM_016584	Hs.98309	IL23A	12q13.2	845					64	
NM_005139	Hs.480042	ANXA3	4q21.21	846					1595	NvsT
NM_001197	Hs.475055	BIK	22q13.2	847					894	
NM_005555	Hs.654613	KRT6B	12q13.13	848					1590	
NM_005554	Hs.433845	KRT6A	12q13.13	849					1591	
NM_005980	Hs.2962	S100P	4p16.1	850				4	122	
NM_003064	Hs.517070	SLPI	20q13.12	851					123	
BC036550	Hs.371980	LOC650392	16p13.3	852					988	
NM_017899	Hs.525709	TESC	12q24.22	853					1579	NvsT
NM_007231	Hs.522109	SLC6A14	Xq23	854					125	
NR_003038	Hs.292457	SNHG5	6q14.3	855					1572	
THC2539634		THC2539634	6q14.3	856					1573	
NM_021021	Hs.655236	SNTB1	8q24.12	857					1582	NvsT
NM_001003845	Hs.368802	SP5	2q31.1	858					1584	NvsT
NM_012342	Hs.533336	BAMBI	10p11.23	859					1585	
NM_003726	Hs.316931	SKAP1	17q21.32	860					1580	
NM_002546	Hs.81791	TNFRSF11B	8q24.12	861					1576	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
AK023647	Hs.43047	AK023647	21q22.3	862					1577	
NM_152773	Hs.135997	MGC33212	3q29	863					15	
THC2697511		THC2697511	Xq25	864					420	
NM_003918	Hs.567381	GYG2	Xp22.33	865					407	
NM_004660	Hs.99120	DDX3Y	Yq11.21	866					408	
NM_001039567	Hs.367761	RPS4Y2	Yq11.223	867					409	
NM_001008	Hs.282376	RPS4Y1	Yp11.31	868					410	
XR_018247	Hs.647382	LOC441774	16q23.1	869					411	
NM_198391	Hs.41296	FLRT3	20p12.1	870					397	
NM_145307	Hs.58559	PLEKHK1	10q21.2	871					72	NvsT
NM_001666	Hs.3109	ARHGAP4	Xq28	872					1716	
NM_032865	Hs.438292	TNS4	17q21.2	873					77	NvsT
NM_003302	Hs.534360	TRIP6	7q22.1	874					1717	
FNST00000380456	Hs.375142	RGL3	19p13.2	875					1655	
BC049823	Hs.380933	RPL22L1	3q26.2	876					985	
NM_000413	Hs.654385	HSD17B1	17q21.31	877					986	
XM_096733	Hs.150840	C14orf72	14q32.31	878					987	
NM_022908	Hs.84753	NT5DC2	3p21.1	879					887	
NM_006163	Hs.75643	NFE2	12q13.13	880					1049	
NM_005217	Hs.654448	DEFA3	8p23.1	881					1050	
NM_001030060	Hs.567973	SAMD5	6q24.3	882					895	
NM_016569	Hs.129895	TBX3	12q24.21	883					1570	
NM_014899	Hs.445030	RHOBTB3	5q15	884					901	
NM_001013680	Hs.441039	LOC401233	6p25.2	885					1720	
NM_001001188	Hs.369759	TRPM2	21q22.3	886					874	
NM_024017	Hs.463350	HOXB9	17q21.32	887					391	
THC2783834		THC2783834	17q21.32	888	1				392	
NM_024016	Hs.514292	HOXB8	17q21.32	889					393	
NM_018952	Hs.98428	HOXB6	17q21.32	890					394	
NM_002146	Hs.654560	HOXB3	17q21.32	891					395	
NM_000148	Hs.69747	FUT1	19q13.33	892					1578	NvsT
NM_152573	Hs.657750	RASEF	9q21.32	893					413	
NM_024087	Hs.19404	ASB9	Xp22.2	894					180	NvsT
NM_001012761	Hs.526902	RGMB	5q21.1	895					1535	
NM_022138	Hs.487200	SMOC2	6q27	896					1536	
NM_004864	Hs.616962	GDF15	19p13.11	897					189	NvsT
NM_203347	Hs.413902	UNQ2541	9q34.3	898					1657	
NM_144691	Hs.653110	CAPN12	19q13.2	899					193	
NM_152599	Hs.213603	FLJ35773	17p13.1	900					949	NvsT
NM_000676	Hs.167046	ADORA2B	17p12	901					182	
NM_017650	Hs.21816	PPP1R9A	7q21.3	902					173	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_017412	Hs.40735	FZD3	8p21.1	903					176	
NM_145168	Hs.87779	HSPC105	16q23.3	904					172	
THC2657737		THC2657737	12p13.2	905					181	
NM_152496	Hs.534562	MANEAL	1p34.3	906					68	
NM_005472	Hs.523899	KCNE3	11q13.4	907					167	
BC041772	Hs.567664	SPNS2	17p13.2	908					168	NvsT
ENST00000261275	Hs.383564	KIAA0574	15q13.1	909					794	NvsT
NM_153426	Hs.643588	PITX2	4q25	910					165	
AF026246		AF026246	8p23.1	911					166	
NM_014861	Hs.6168	ATP2C2	16q24.1	912					791	
NM_018667	Hs.368421	SMPD3	16q22.1	913					792	NvsT
NM_144603	Hs.191762	NOXO1	16p13.3	914					188	
NM_004443	Hs.2913	EPHB3	3q27.1	915					190	
A_32_P177595		A_32_P177595	5q23.3	916					191	
BE714307	Hs.162585	BE714307	5q23.3	917					192	
NM_001004331	Hs.412138	MGC88374	14q23.3	918					1643	
AY358366	Hs.642705	KIAA1324	1p13.3	919			3		945	
A_32_P224040		A_32_P224040	14q32.2	920			3		946	
NM_004496	Hs.163484	FOXA1	14q21.1	921					947	
NM_001049	Hs.248160	SSTR1	14q21.1	922					948	
NM_000597	Hs.438102	IGFBP2	2q35	923					923	
NM_000282	Hs.80741	PCCA	13q32.3	924					924	
NM_032342	Hs.655738	C9orf125	9q31.1	925					941	NvsT
ENST00000374851	Hs.388742	C9orf125	9q31.1	926					942	
NM_002250	Hs.10082	KCNN4	19q13.31	927					194	
NM_001677	Hs.291196	ATP1B1	1q24.2	928				4	943	
ENST00000382790	Hs.204044	ENST00000382790	18q21.33	929					944	
NM_013314	Hs.665244	BLNK	10q23.33	930					555	
AK126354	Hs.228320	C10orf81	10q25.3	931			3		70	NvsT
NM_024889	Hs.228320	C10orf81	10q25.3	932			3		71	NvsT
NM_032387	Hs.105448	WNK4	17q21.31	933					1652	
NM_015225	Hs.262857	KIAA0367	9q21.13	934					1653	
NM_021784	Hs.155651	FOXA2	20p11.21	935					204	
NM_020340	Hs.656215	KIAA1244	6q23.3	936					205	
ENST00000330640	Hs.194408	ENST00000330640	6q23.3	937				4	206	
BG216262	Hs.637431	BG216262	3q21.2	938					207	
AK074886	Hs.434403	LOC389634	12p13.31	939					208	
AB007953		KIAA0484	8p23.1	940					209	
NM_000689	Hs.76392	ALDH1A1	9q21.13	941					933	
NM_000300	Hs.466804	PLA2G2A	1p36.13	942					800	
A_23_P62857		A_23_P62857		943					801	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
A_23_P410456		A_23_P410456	11p15.4	944					876	
NM_203451	Hs.422375	LOC400120	13q13.3	945					877	
CD607715	Hs.255230	CD607715	7q11.21	946					878	
ENST00000325900	Hs.374414	ENST00000325900	Xq26.3	947					880	
AF516696	Hs.656687	CACNA2D3	3p14.3	948					879	
AK125510	Hs.650801	C1orf104	1q22	949					881	
NM_025193	Hs.460618	HSD3B7	16p11.2	950					882	
NM_014337	Hs.438587	PPIL2	22q11.21	951					883	
AK056520	Hs.694296	FLJ31958	10q11.23	952					884	
THC2780599		THC2780599	17q25.1	953					885	
NM_206836	Hs.15250	PECI	6p25.2	954					723	NvsT
A_23_P3083		A_23_P3083	14q22.1	955					724	NvsT
NM_002563	Hs.654526	P2RY1	3q25.2	956					725	NvsT
BC107798	Hs.631558	TNNT1	19q13.42	957					888	
NM_003020	Hs.156540	SCG5	15q13.3	958					900	
THC2585656	Hs.496631	THC2585656	Xq23	959					1656	
NM_024829	Hs.131933	FLJ22662	12p13.1	960					1654	
NM_206927	Hs.369520	SYTL2	11q14.1	961					781	NvsT
NM_182495	Hs.446760	FAM55B	11q23.2	962					782	NvsT
NM_152315	Hs.668531	FAM55A	11q23.2	963					783	NvsT
NM_017678	Hs.179100	FAM55D	11q23.2	964	1				784	NvsT
AK094950	Hs.375762	AK094950	15q21.1	965					909	NvsT
NM_032413	Hs.112242	C15orf48	15q21.1	966					706	NvsT
NM_021199	Hs.511251	SORDL	15q21.1	967					707	NvsT
NM_018690	Hs.200333	APOB48R	16p11.2	968					780	NvsT
NM_030926	Hs.111577	ITM2C	2q37.1	969					705	NvsT
NM_018700	Hs.519514	TRIM36	5q22.3	970					793	NvsT
NM_017636	Hs.467101	TRPM4	19q13.33	971					778	NvsT
NM_013261	Hs.527078	PPARGC1A	4p15.2	972					668	NvsT
ENST00000251847	Hs.101774	ENST00000251847	20p12.1	973					669	NvsT
NM_002456	Hs.89603	MUC1	1q22	974				4	915	
NM_018406	Hs.369646	MUC4	3q29	975					734	NvsT
NM_025080	Hs.535326	ASRGL1	11q12.3	976			3		919	
NM_000240	Hs.183109	MAOA	Xp11.3	977					651	NvsT
NM_003645	Hs.11729	SLC27A2	15q21.2	978			3		580	
NM_000901	Hs.163924	NR3C2	4q31.23	979			3		704	NvsT
NM_015393	Hs.105460	DKFZP564O0823	4q13.3	980	1				786	NvsT
NM_138409	Hs.370055	C6orf117	6q14.3	981	1		3		785	
NM_018423	Hs.24979	STYK1	12p13.2	982			3		760	NvsT
NM_017821	Hs.524626	RHBDL2	1p34.3	983					761	NvsT
NM_138805	Hs.61265	FAM3D	3p14.2	984					787	NvsT

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_003561	Hs.567366	PLA2G10	16p13.12	985					730	NvsT
NM_003196	Hs.446354	TCEA3	1p36.12	986					779	NvsT
NM_006633	Hs.291030	IQGAP2	5q13.3	987			3		768	NvsT
NM_153229	Hs.224630	TMEM92	17q21.33	988			3		134	
NM_016357	Hs.525419	LIMA1	12q13.13	989					750	NvsT
NM_001015880	Hs.524491	PAPSS2	10q23.2	990					758	NvsT
NM_001218	Hs.210995	CA12	15q22.2	991			3		752	NvsT
NM_001200	Hs.73853	BMP2	20p12.3	992					773	NvsT
NM_025047	Hs.287702	ARL14	3q26.1	993					774	NvsT
NM_002245	Hs.208544	KCNK1	1q42.2	994					908	
NM_001993	Hs.62192	F3	1p21.3	995					928	
A_23_P103951		A_23_P103951	1q32.2	996				4	929	
NM_000574	Hs.527653	CD55	1q32.2	997				4	930	
NM_000716	Hs.99886	C4BPB	1q32.2	998			3		126	
NM_016619	Hs.546392	PLAC8	4q21.22	999			3		767	NvsT
NM_025214	Hs.120790	CCDC68	18q21.2	1000					766	NvsT
NM_153840	Hs.256897	GPR110	6p12.3	1001					124	
ENST00000267857	Hs.194710	GCNT3	15q22.2	1002					764	NvsT
NM_004751	Hs.194710	GCNT3	15q22.2	1003					765	NvsT
NM_017723	Hs.657689	C9orf167	9q34.3	1004					906	
ENST00000366874		ENST00000366874	1q41	1005					907	
NM_139248	Hs.68864	LIPH	3q27.2	1006					769	NvsT
BM768581	Hs.351126	BM768581	3q27.2	1007					770	NvsT
NM_020672	Hs.288998	S100A14	1q21.3	1008					762	
NM_015873	Hs.103665	VILL	3p22.2	1009					763	NvsT
NM_181644	Hs.567714	MFSD4	1q32.1	1010					775	NvsT
NM_001285	Hs.194659	CLCA1	1p22.3	1011					776	NvsT
THC2654007		THC2654007	2p14	1012					777	NvsT
NM_152338	Hs.632195	ZG16		1013					628	NvsT
NM_017625	Hs.50813	ITLN1	1q23.3	1014					736	NvsT
NM_138938	Hs.567312	REG3A	2p12	1015					128	
NM_001008387	Hs.447084	REG3G	2p12	1016					129	
NM_006507	Hs.4158	REG1B	2p12	1017					130	
NM_002909	Hs.49407	REG1A	2p12	1018			3		131	
NM_001002236	Hs.525557	SERPINA1	14q32.13	1019			3	4	956	
NM_007329	Hs.279611	DMBT1	10q26.13	1020			3		133	
NM_000439	Hs.78977	PCSK1	5q15	1021				4	989	
NM_021800	Hs.260720	DNAJC12	10q21.3	1022					990	
NM_001085	Hs.534293	SERPINA3	14q32.13	1023					991	
BX104982	Hs.192124	BX104982	3p22.1	1024					636	NvsT
NM_002407	Hs.97644	SCGB2A1	11q12.3	1025					756	NvsT

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_005559	Hs.270364	LAMA1	18p11.31	1026					757	NvsT
BC087840	Hs.57971	HES5	1p36.32	1027					788	
AK126207	Hs.299329	B3GNT7	2q37.1	1028					726	NvsT
A_32_P169353		A_32_P169353	2q37.1	1029					727	NvsT
NM_032717	Hs.99196	MAG1	4q21.23	1030					751	NvsT
NM_005209	Hs.415790	CRYBA2	2q35	1031					728	NvsT
NM_024533	Hs.156784	CHST5	16q23.1	1032					729	NvsT
NM_181726	Hs.693753	ANKRD37	4q35.1	1033					1102	
NM_022073	Hs.135507	EGLN3	14q13.1	1034					911	
AK123483	Hs.446388	AK123483	11p15.5	1035					912	
NM_024590	Hs.22895	ARSJ	4q26	1036					984	
NM_004633	Hs.25333	IL1R2	2q11.2	1037				4	743	NvsT
NM_198569	Hs.318894	GPR126	6q24.1	1038				4	932	
NM_000436	Hs.278277	OXCT1	5p13.1	1039					918	
A_23_P120644		A_23_P120644	20q11.23	1040					746	NvsT
NM_014056	Hs.7917	HIGD1A	3p22.1	1041					747	NvsT
NM_004235	Hs.376206	KLF4	9q31.2	1042					749	NvsT
NM_033342	Hs.487412	TRIM7	5q35.3	1043				4	917	
NM_138340	Hs.397978	ABHD3	18q11.2	1044					748	NvsT
NM_012465	Hs.154296	TLL2	10q24.1	1045					982	
THC2660636		THC2660636	17p13.1	1046					742	NvsT
NM_007312	Hs.75619	HYAL1	3p21.31	1047					931	
NM_004970	Hs.839	IGFALS	16p13.3	1048				4	953	
NM_138969	Hs.170673	RDHE2	8q12.1	1049				4	954	
NM_005423	Hs.2979	TFF2	21q22.3	1050				4	955	
NM_002257	Hs.123107	KLK1	19q13.33	1051					735	NvsT
NM_174953	Hs.513870	ATP2A3	17p13.2	1052					732	NvsT
NM_005476	Hs.5920	GNE	9p13.3	1053				4	722	
NM_024101	Hs.102406	MLPH	2q37.3	1054				4	968	
BC022217	Hs.132340	C6orf85	6p25.2	1055					733	NvsT
AK091508		AK091508	9q12	1056				4	958	
NM_139242	Hs.531615	MTFMT	15q22.31	1057				4	959	
THC2647276		THC2647276	8q24.13	1058				4	960	NvsT
NM_006278	Hs.591947	ST3GAL4	11q24.2	1059				4	822	
NM_174921	Hs.576320	C4orf34	4p14	1060				4	744	NvsT
A_24_P798431		A_24_P798431	4p14	1061					745	NvsT
NM_003225	Hs.162807	TFF1	21q22.3	1062				4	965	
BC010526	Hs.383197	LOC401127	4p14	1063				4	966	
THC2736540		THC2736540	21q22.3	1064					967	
NM_006408	Hs.530009	AGR2	7p21.1	1065			3	4	969	
NM_176813	Hs.100686	AGR3	7p21.1	1066			3	4	970	NvsT

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_018414	Hs.105352	ST6GALNAC1	17q25.1	1067			3	4	739	NvsT
NM_017417	Hs.511985	GALNT8	12p13.32	1068					957	
NM_182532	Hs.663950	TMEM61	1p32.3	1069					740	NvsT
NM_003657	Hs.400556	BCAS1	20q13.2	1070					737	NvsT
NM_005187	Hs.513811	CBFA2T3	16q24.3	1071				4	961	NvsT
NM_003890	Hs.111732	FCGBP	19q13.2	1072			3	4	738	NvsT
NM_032044	Hs.660883	REG4	1p12	1073				4	971	
ENST00000358378	Hs.534332	ENST00000358378	11p15.5	1074				4	972	
NR_001558		C20orf56	20p11.21	1075				4	974	
NM_014471	Hs.555934	SPINK4	9p13.3	1076			3	4	973	
NM_198151	Hs.443169	LOC253012	7q21.3	1077			3	4	741	NvsT
NM_019079	Hs.562195	LTD1	1p31.3	1078					921	
NM_024565	Hs.14070	CCNJL	5q33.3	1079					962	NvsT
NM_032579	Hs.307047	RETNLB	3q13.13	1080					950	
NM_080734	Hs.2719	WFDC2	20q13.12	1081					951	
NM_080736	Hs.2719	WFDC2	20q13.12	1082					952	
NM_033317	Hs.417795	DMKN	19q13.12	1083					897	
NM_024989	Hs.229988	PGAP1	2q33.1	1084					896	
NM_153225	Hs.439040	RPESP	8q21.11	1085					922	
NM_032229	Hs.525105	SLITRK6	13q31.1	1086					963	
NM_004056	Hs.654388	CA8	8q12.1	1087				4	964	
NM_006533	Hs.646364	MIA	19q13.2	1088					1026	
NM_000239	Hs.524579	LYZ	12q15	1089					925	
THC2682885		THC2682885	17q23.1	1090					926	
NM_138799	Hs.467634	MBOAT2	2p25.1	1091					927	
NM_003071	Hs.3068	HLTF	3q24	1092					892	
NM_006072	Hs.131342	CCL26	7q11.23	1093					1529	
NM_002888	Hs.131269	RARRES1	3q25.33	1094				4	934	
NM_020169	Hs.478067	LXN	3q25.32	1095					935	NvsT
NM_014751	Hs.336994	MTSS1	8q24.13	1096					1143	
NM_002863	Hs.282417	PYGL	14q22.1	1097					1142	
NM_014747	Hs.654808	RIMS3	1p34.2	1098					936	
NM_002353	Hs.23582	TACSTD2	1p32.1	1099					1598	
NM_001311	Hs.70327	CRIP1	14q32.33	1100				4	937	
A_23_P21882		A_23_P21882		1101				4	938	
A_24_P110601		A_24_P110601	22q11.23	1102				4	939	
NM_018013	Hs.445244	SOBP	6q21	1103					1144	
NM_080685	Hs.436142	PTPN13	4q21.3	1104					1021	
NM_000853	Hs.268573	GSTT1	22q11.23	1105					406	
NM_144676	Hs.658619	TMED6	16q22.1	1106					570	
AF275804	Hs.237396	AF275804	16q12.2	1107					1575	NvsT

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
BC042557	Hs.31409	LOC120376	11q23.1	1108					902	
NM_030929	Hs.632079	KAZALD1	10q24.31	1109					903	
NM_003695	Hs.415762	LY6D	8q24.3	1110					899	
NM_001505	Hs.20961	GPR30	7p22.3	1111					569	NvsT
BE798911	Hs.662555	BE798911	2q31.1	1112					872	
NM_030583	Hs.189445	MATN2	8q22.2	1113					754	NvsT
BM802662	Hs.586812	BM802662	13q13.3	1114					1538	
NM_000683	Hs.123022	ADRA2C	4p16.2	1115					1606	
NM_173660	Hs.122110	DOK7	4p16.2	1116					1574	NvsT
NM_004004	Hs.591234	GJB2	13q12.11	1117					1024	
XM_001131389	Hs.646725	LOC730999	10q22.1	1118					1025	
NM_018192	Hs.374191	LEPREL1	3q28	1119					585	NvsT
NM_000222	Hs.479754	KIT	4q12	1120					586	NvsT
NM_005130	Hs.1690	FGFRP1	4p15.32	1121					587	NvsT
NM_001007097	Hs.494312	NTRK2	9q21.33	1122					1023	
NM_012385	Hs.513463	NUPR1	16p11.2	1123					1494	
NM_000163	Hs.125180	GHR	5p12	1124					806	
NM_020436	Hs.517113	SALL4	20q13.2	1125	1				1675	
BC041856	Hs.536395	DUXAP10	14q11.1	1126					1710	
BG118529	Hs.536633	BG118529	14q11.1	1127					1711	
NM_005769	Hs.251383	CHST4	16q22.3	1128					62	
NM_153488	Hs.670781	MAGEA2B	Xq28	1129					1712	
NM_005367	Hs.169246	MAGEA12	Xq28	1130					1713	
NM_006546	Hs.144936	IGF2BP1	17q21.32	1131					1714	
NM_153479	Hs.423718	CSAG1	Xq28	1132					1715	
NM_015973	Hs.278959	GAL	11q13.2	1133					11	
NM_004086	Hs.21016	COCH	14q12	1134					12	
NM_000363	Hs.644596	TNNI3	19q13.42	1135					889	
DQ249310	Hs.644234	UCA1	19p13.12	1136					1589	
NM_002704	Hs.2164	PPBP	4q13.3	1137					1586	
NM_002619	Hs.81564	PF4	4q13.3	1138					1587	
NM_004131	Hs.1051	GZMB	14q12	1139					112	NvsT
NM_001276	Hs.382202	CHI3L1	1q32.1	1140					111	NvsT
NM_152737	Hs.111164	RNF182	6p23	1141					114	
NM_012168	Hs.132753	FBXO2	1p36.22	1142					115	
NM_198951	Hs.517033	TGM2	20q11.23	1143					113	NvsT
NM_012254	Hs.292177	SLC27A5	19q13.43	1144					107	
NM_147175	Hs.385956	HS6ST2	Xq26.2	1145					108	
NM_031901	Hs.405880	MRPS21	1q21.2	1146					386	
NM_004272	Hs.591761	HOMER1	5q14.1	1147					109	NvsT
NM_006404	Hs.647450	PROCR	20q11.22	1148					110	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_024693	Hs.22242	ECHDC3	10p14	1149					1	
NM_022350	Hs.591249	LRAP	5q15	1150					9	
BP871540	Hs.515176	BP871540	17p11.2	1151					890	
AL833005	Hs.675929	AL833005		1152					891	
NM_021158	Hs.516826	TRIB3	20p13	1153					286	NvsT
NM_058179	Hs.494261	PSAT1	9q21.2	1154					73	NvsT
NM_006623	Hs.487296	PHGDH	1p12	1155					74	
ENST00000357303	Hs.24951	ENST00000357303	2q37.3	1156					54	
BM129308	Hs.653792	BM129308	8p21.2	1157					55	
NM_173567	Hs.201555	ABHD7	1p22.1	1158					56	NvsT
NM_032181	Hs.302346	TMEM166	2p12	1159					57	
NM_145051	Hs.211374	RNF183	9q32	1160					58	
BX107298	Hs.502418	BX107298	11p11.2	1161					196	
NM_001147	Hs.583870	ANGPT2	8p23.1	1162					53	NvsT
ENST00000309178		ENST00000309178	5q34	1163					153	
NM_002300	Hs.446149	LDHB	12p12.1	1164					14	
NM_005329	Hs.592069	HAS3	16q22.1	1165					51	
NM_033300	Hs.576154	LRP8	1p32.3	1166					76	NvsT
NM_144492	Hs.660278	CLDN14	21q22.13	1167					52	
NM_174936	Hs.18844	PCSK9	1p32.3	1168					47	NvsT
NM_001216	Hs.63287	CA9	9p13.3	1169					48	
NM_015715	Hs.149623	PLA2G3	22q12.2	1170					49	
BC014113	Hs.663029	LOC254848	4p16.1	1171					50	
NM_001236	Hs.154510	CBR3	21q22.12	1172					916	
NM_152450	Hs.531168	FAM81A	15q22.2	1173					152	
NM_018092	Hs.444046	NETO2	16q12.1	1174			3		155	
NM_139160	Hs.280990	DEPDC7	11p13	1175					156	
AB051443	Hs.505597	KIAA1656	22q12.2	1176					164	
NM_000662	Hs.591847	NAT1	8p22	1177					572	NvsT
NM_152562	Hs.33366	CDCA2	8p21.2	1178					79	NvsT
NM_001017420	Hs.99480	ESCO2	8p21.1	1179					80	
NM_018492	Hs.104741	PBK	8p21.1	1180					81	
XR_018387	Hs.646624	LOC647298	13q31.1	1181				4	82	NvsT
BC043212	Hs.120196	LOC402125	3p24.3	1182					154	
NM_145060	Hs.134726	C18orf24	18q21.1	1183					100	
ENST00000375256	Hs.494557	ZNF367	9q22.32	1184					102	NvsT
NM_024094	Hs.315167	DCC1	8q24.12	1185					84	NvsT
NM_018685	Hs.62180	ANLN	7p14.2	1186					93	NvsT
NM_001274	Hs.24529	CHEK1	11q24.2	1187					85	NvsT
NM_003318	Hs.169840	TTK	6q14.1	1188					86	NvsT
NM_080668	Hs.434886	CDCA5	11q13.1	1189					98	NvsT

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_001254	Hs.405958	CDC6	17q21.2	1190					94	NvsT
NM_031966	Hs.23960	CCNB1	5q13.2	1191					95	NvsT
NM_022346	Hs.567567	NCAPG	4p15.32	1192					96	NvsT
NM_001012271	Hs.514527	BIRC5	17q25.3	1193					99	NvsT
NM_001809	Hs.1594	CENPA	2p23.3	1194					97	NvsT
NM_018131	Hs.14559	CEP55	10q23.33	1195					91	NvsT
NM_001786	Hs.334562	CDC2	10q21.2	1196					92	NvsT
NM_002497	Hs.153704	NEK2	1q32.3	1197					89	NvsT
NM_145697	Hs.651950	NUF2	1q23.3	1198					90	NvsT
NM_032117	Hs.294088	MND1	4q31.3	1199					87	NvsT
NM_002358	Hs.591697	MAD2L1	4q27	1200					88	NvsT
NM_001067	Hs.156346	TOP2A	17q21.2	1201					83	NvsT
A_24_P221285		A_24_P221285	Xp11.3	1202			3		101	NvsT
A_32_P152696		A_32_P152696	9p21.3	1203					103	NvsT
BC011243	Hs.642598	ZNF587	19q13.43	1204					105	NvsT
NM_181803	Hs.93002	UBE2C	20q13.12	1205					106	NvsT
NM_005573	Hs.89497	LMNB1	5q23.2	1206					873	
ENST00000335534	Hs.135094	LOC146909	17q21.31	1207					104	NvsT
A_32_P124773		A_32_P124773	3q25.31	1208					16	
BC082758	Hs.644701	BC082758	2q35	1209					17	
NM_030641	Hs.257352	APOL6	22q12.3	1210					18	
THC2598362		THC2598362	16p11.2	1211					19	
NM_000023	Hs.463412	SGCA	17q21.33	1212					45	
NM_015668	Hs.120021	RGS22	8q22.2	1213					46	
NM_015229	Hs.22616	KIAA0664	17p13.3	1214					22	
BC031278	Hs.375797	BC031278	1p21.1	1215					23	
THC2739443	Hs.603732	THC2739443	2q32.1	1216					24	
NM_080621	Hs.27189	SAMD10	20q13.33	1217					25	
NM_006439	Hs.584852	MAB21L2	4q31.3	1218					26	
AB016898	Hs.667982	LOC653483	6q27	1219					27	
NM_019060	Hs.110196	CRCT1	1q21.3	1220					28	
THC2655508		THC2655508	3p21.31	1221					29	
NM_024076	Hs.221873	KCTD15	19q13.11	1222					30	
AJ276555	Hs.152129	AJ276555	12q24.11	1223					31	
A_24_P289854		A_24_P289854	11q22.3	1224					32	
AF264625	Hs.675021	AF264625	10q24.32	1225					33	
A_24_P934306		A_24_P934306	1q32.1	1226					34	
NM_052999	Hs.15159	CMTM1	16q21	1227					35	
AA417913	Hs.87889	AA417913	14q32.13	1228					36	
ENST00000331096		ENST00000331096	12q24.11	1229					37	
NM_005309	Hs.103502	GPT	8q24.3	1230					38	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_014507	Hs.349111	MCAT	22q13.2	1231					39	
NM_145659	Hs.528111	IL27	16p11.2	1232					40	
NM_173502	Hs.256632	PRSS36	16p11.2	1233					41	
A_24_P752362		A_24_P752362	8p23.1	1234					21	
ENST00000307662	Hs.654723	SYNPO	5q33.1	1235					42	
ENST00000344415	Hs.677674	LOC400581	17p11.2	1236					43	
XR_018043	Hs.646473	LOC442308	7p11.2	1237					44	
NM_001004313	Hs.462230	LOC388335	17p13.1	1238					612	NvsT
NM_000204	Hs.312485	CFI	4q25	1239					117	NvsT
ENST00000301171	Hs.250615	ENST00000301171	19q13.2	1240					566	NvsT
ENST00000342584	Hs.481464	ENST00000342584	5p15.33	1241				4	567	NvsT
NM_152997	Hs.320147	C4orf7	4q13.3	1242					442	
NM_001192	Hs.2556	TNFRSF17	16p13.13	1243					655	NvsT
BC107852	Hs.449585	BC107852	22q11.22	1244					444	NvsT
CD709370	Hs.559369	CD709370	22q11.22	1245					445	NvsT
ENST00000390256	Hs.570141	ENST00000390256	2p11.2	1246					443	NvsT
A_24_P930327		A_24_P930327	19q13.2	1247		2			441	
ENST00000390559	Hs.510635	ENST00000390559	14q32.33	1248		2			446	NvsT
ENST00000283694	Hs.458446	ENST00000283694	15q11.2	1249					448	NvsT
AK026408	Hs.449601	AK026408	22q11.22	1250		2			449	
ENST00000390625	Hs.631480	ENST00000390625	14q32.33	1251		2			450	NvsT
NM_001995	Hs.406678	ACSL1	4q35.1	1252		2			453	NvsT
BQ880155	Hs.604483	BQ880155	10q11.23	1253					454	NvsT
A_23_P72330		A_23_P72330		1254		2			455	NvsT
BC063599	Hs.449621	IGKV2-24	2p11.2	1255					456	NvsT
A_24_P384604		A_24_P384604	2q11.2	1256					457	NvsT
ENST00000390260	Hs.449621	ENST00000390260	2p11.2	1257					458	NvsT
BC034142	Hs.449621	IGKV1-5	2p11.2	1258					459	NvsT
ENST00000390267	Hs.449609	ENST00000390267	2p11.2	1259		2			460	NvsT
ENST00000327926	Hs.525894	ENST00000327926	22q11.1	1260		2			461	NvsT
A_24_P203886		A_24_P203886	2q11.2	1261					462	NvsT
ENST00000312946		ENST00000312946	2p11.2	1262					463	NvsT
THC2497143	Hs.449585	THC2497143	22q11.22	1263					467	NvsT
ENST00000390312	Hs.449585	ENST00000390312	22q11.22	1264					468	NvsT
ENST00000390294	Hs.655198	ENST00000390294	22q11.22	1265					526	NvsT
BC012876	Hs.449585	BC012876	22q11.22	1266					469	NvsT
BC023973	Hs.535668	IGLV6-57	22q11.22	1267					470	NvsT
A_23_P159163		A_23_P159163		1268		2			474	NvsT
A_23_P84791		A_23_P84791		1269		2			475	NvsT
AY998685	Hs.652878	AY998685		1270		2			476	
ENST00000390243	Hs.652878	ENST00000390243	2p11.2	1271		2			477	NvsT

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_021181	Hs.517265	SLAMF7	1q23.3	1272		2			486	NvsT
AF063695	Hs.449585	AF063695		1273		2			471	
A_23_P72252		A_23_P72252		1274		2			472	
X57802	Hs.449585	X57802	22q11.22	1275		2			473	
NM_198440	Hs.593679	DERL3	22q11.23	1276					484	
X57818	Hs.449585	X57818	22q11.22	1277		2			466	
AF035035	Hs.574701	IGKV1D-8	2p11.2	1278		2			488	NvsT
NM_006235	Hs.654525	POU2AF1	11q23.1	1279		2			487	NvsT
AF267875	Hs.535668	AF267875	22q11.22	1280		2			485	
X01147	Hs.654512	X01147	22q11.22	1281		2			489	NvsT
A_23_P61068		A_23_P61068		1282		2			490	NvsT
AJ399872	Hs.449621	AJ399872		1283		2			491	NvsT
ENST00000390595	Hs.510635	ENST00000390595	14q32.33	1284		2			479	NvsT
AJ252276	Hs.510635	AJ252276	2q11.1	1285		2			478	
A_23_P113056		A_23_P113056	14q32.33	1286		2			480	
BC070333	Hs.634941	IGHV1-69	14q32.33	1287		2			481	
ENST00000390610	Hs.644810	ENST00000390610	14q32.33	1288		2			482	
ENST00000390622	Hs.644810	ENST00000390622	14q32.33	1289		2			483	
NM_002281	Hs.658118	KRT81	12q13.13	1290		2			530	NvsT
A_24_P204574		A_24_P204574	22q11.22	1291		2			517	NvsT
A_23_P112957		A_23_P112957	22q11.22	1292		2			527	NvsT
AF076205	Hs.449574	AF076205	22q11.22	1293		2			528	NvsT
BC009479		TTLL3	3p25.3	1294		2			518	NvsT
NM_016459	Hs.409563	MGC29506	5q31.2	1295		2			519	NvsT
BI521983	Hs.449574	ARPC1B	22q11.22	1296		2			520	NvsT
ENST00000328419		ENST00000328419	18p11.31	1297		2			521	NvsT
A_23_P435390		A_23_P435390		1298		2			522	NvsT
AY172962	Hs.449585	AY172962	22q11.22	1299		2			523	NvsT
NM_001013618	Hs.567636	CTA-246H3.1	22q11.23	1300		2			524	NvsT
NM_020070	Hs.348935	IGLL1	22q11.23	1301		2			525	NvsT
BC024289	Hs.510635	IFI6	14q32.33	1302		2			451	
ENST00000390543	Hs.510635	ENST00000390543	14q32.33	1303		2			452	
ENST00000390630	Hs.643624	ENST00000390630	14q32.33	1304		2			492	
A_23_P158868		A_23_P158868		1305		2			493	
AJ245002	Hs.631647	AJ245002	15q11.2	1306		2			494	
AF067420	Hs.648398	IGHA1	14q32.33	1307		2			495	
AY003763	Hs.510635	AY003763		1308		2			496	NvsT
ENST00000390636	Hs.547404	ENST00000390636	14q32.33	1309		2			501	
A_24_P323298		A_24_P323298	14q32.33	1310		2			502	
AF103312	Hs.547404	AF103312	14q32.33	1311		2			503	
A_24_P110487		A_24_P110487	16p11.2	1312		2			504	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
AB063751	Hs.533963	AB063751	14q32.33	1313		2			511	
BX640624	Hs.648398	IGHA1	14q32.33	1314		2			512	NvsT
AF471475		AF471475	14q32.33	1315		2			505	
ENST00000390615	Hs.633485	ENST00000390615	14q32.33	1316		2			506	
Y11328	Hs.650507	Y11328	16p11.2	1317		2			507	
AF471454	Hs.633485	AF471454	14q32.33	1318		2			508	
ENST00000360102		ENST00000360102	16p11.2	1319		2			509	
A 24 P384119		A 24 P384119	16p11.2	1320		2			510	NvsT
ENST00000390632	Hs.644810	ENST00000390632	14q32.33	1321		2			513	
AJ519285	Hs.551925	AJ519285	14q32.33	1322		2			514	
EF178107	Hs.644810	EF178107	14q32.33	1323		2			515	
NP109842		NP109842	16p11.2	1324		2			497	
ENST00000390624	Hs.644810	ENST00000390624	14q32.33	1325		2			498	
A 24 P161853		A 24 P161853	16p11.2	1326		2			499	NvsT
ENST00000390609	Hs.510635	ENST00000390609	14q32.33	1327		2			500	NvsT
AK128476	Hs.648398	IGHA1	14q32.33	1328		2			531	NvsT
NM_012390	Hs.651168	SMR3A	4q13.3	1329		2			535	NvsT
THC2699069		THC2699069	17q21.31	1330		2			536	NvsT
BC073764	Hs.449621	IGKC	2p11.2	1331		2			532	NvsT
BC032451	Hs.449621	BC032451	2p11.2	1332		2			537	NvsT
AY062331	Hs.449621	AY062331	2p11.2	1333		2			538	NvsT
RC008026	Hs.676124	MGC16025	2q37.3	1334		2			533	NvsT
BX110985	Hs.143004	BX110985	3p25.3	1335		2			534	NvsT
ENST00000390263	Hs.693601	ENST00000390263	2p11.2	1336		2			539	NvsT
ENST00000390253	Hs.649954	ENST00000390253	2p11.2	1337		2			540	NvsT
ENST00000390261	Hs.693601	ENST00000390261	2p11.2	1338		2			541	NvsT
ENST00000310579		ENST00000310579	9q12	1339		2			542	NvsT
ENST00000331696		ENST00000331696	2p11.1	1340		2			543	NvsT
BC067092	Hs.449621	IGKC	2p11.2	1341		2			544	NvsT
A 24 P490109		A 24 P490109	9q12	1342		2			545	NvsT
BC030813	Hs.449621	BC030813	2p11.2	1343		2			546	NvsT
ENST00000390281	Hs.449621	ENST00000390281	2p11.2	1344		2			547	NvsT
BC095489	Hs.449621	IGKC	2p11.2	1345		2			548	NvsT
ENST00000390275	Hs.390427	ENST00000390275	2p11.2	1346		2			549	NvsT
A 23 P44053		A 23 P44053	2p11.2	1347		2			550	NvsT
ENST00000390258	Hs.693601	ENST00000390258	2p11.2	1348		2			551	NvsT
ENST00000390271		ENST00000390271	2p11.2	1349		2			516	NvsT
A 24 P413286		A 24 P413286		1350		2			529	NvsT
ENST00000360329	Hs.649259	ENST00000360329	14q32.33	1351		2			464	NvsT
NM_152513	Hs.116419	RP5-821D11.2	22q13.2	1352		2			465	NvsT
BC022362	Hs.449621	BC022362	2p11.2	1353					447	NvsT

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_004944	Hs.476453	DNASE1L3	3p14.3	1354	1		3		657	NvsT
NM_144646	Hs.643431	IGJ	4q13.3	1355					658	NvsT
AW136683	Hs.643431	AW136683	4q13.3	1356					659	NvsT
ENST00000390539	Hs.648398	ENST00000390539	14q32.33	1357					660	NvsT
AF343666	Hs.648398	AF343666	1q23.1	1358					661	NvsT
NM_153259	Hs.591446	MCOLN2	1p22.3	1359					795	NvsT
NM_018284	Hs.656774	GBP3	1p22.2	1360					163	
NM_018974	Hs.567508	UNC93A	6q27	1361					65	
THC2520542		THC2520542	5q23.1	1362					716	NvsT
NM_001080535	Hs.149219	LINCRC	2q11.2	1363					2	
NM_017662	Hs.272225	TRPM6	9q21.13	1364					663	NvsT
THC2709754		THC2709754	13q14.11	1365					378	
NM_001353	Hs.460260	AKR1C1	10p15.1	1366					372	
NM_003739	Hs.78183	AKR1C3	10p15.1	1367					373	
BC040210	Hs.460260	AKR1C1	10p15.1	1368					374	
ENST00000377116	Hs.525093	ENST00000377116	13q31.1	1369					376	
NM_033255	Hs.546467	EPST11	13q14.11	1370					370	
ENST00000313624	Hs.546467	EPST11	13q14.11	1371					371	
NM_020208	Hs.413095	SLC6A20	3p21.31	1372					369	
NM_000253	Hs.195799	MTTP	4q23	1373					298	
NM_080860	Hs.661069	TSGA2	21q22.3	1374					914	
NM_138809	Hs.192586	CMBL	5p15.2	1375					592	NvsT
NM_021147	Hs.3041	CCNU	5q11.2	1376					75	NvsT
NM_018670	Hs.447531	MESP1	15q26.1	1377					143	
AL832534	Hs.684035	AL832534	2p22.1	1378					145	
A_24_P358606		A_24_P358606	5q32	1379					146	
S73202		S73202	9q34.11	1380					147	
THC2786072		THC2786072	7q31.1	1381					148	
XR_019510	Hs.650952	LOC652810	Yq11.21	1382					149	
NM_002288	Hs.43803	LAIR2	19q13.42	1383					116	NvsT
NM_001710	Hs.69771	CFB	6p21.32	1384					118	NvsT
NM_006498	Hs.531776	LGALS2	22q13.1	1385					717	NvsT
NM_031409	Hs.46468	CCR6	6q27	1386					223	
NM_000625	Hs.434386	NOS2A	17q11.1	1387			3		913	
NM_002426	Hs.1695	MMP12	11q22.2	1388					1045	
NM_033292	Hs.2490	CASP1	11q22.3	1389			3		144	
NM_016434	Hs.434878	RTEL1	20q13.33	1390					135	
BC004960	Hs.516108	MGC10955	2p13.1	1391					136	
NM_032945	Hs.434878	TNFRSF6B	20q13.33	1392					137	NvsT
NM_145252	Hs.105887	LOC124220	16p13.3	1393					851	
NM_001080494	Hs.112949	C1orf34	1p32.3	1394	1		3		553	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_148672	Hs.656904	CCL28	5p12	1395	1		3		670	NvsT
NM_002644	Hs.497589	PIGR	1q32.1	1396			3		671	NvsT
THC2591546	Hs.497589	THC2591546	1q32.1	1397			3		672	NvsT
NM_014080	Hs.71377	DUOX2	15q21.1	1398					138	
NM_005564	Hs.204238	LCN2	9q34.11	1399			3		139	
BC029479	Hs.525513	LOC283587	14q31.3	1400			3		140	
NM_024780	Hs.115838	TMC5	16p12.3	1401			3		141	
NM_005764	Hs.431099	PDZK1IP1	1p33	1402			3		142	
NM_022154	Hs.288034	SLC39A8	4q24	1403	1		3		554	
NM_032744	Hs.126409	C6orf105	6p24.1	1404			3		683	NvsT
NM_000715	Hs.1012	C4BPA	1q32.2	1405	1		3		127	
NM_001719	Hs.473163	BMP7	20q13.31	1406					389	
NM_002165	Hs.504609	ID1	20q11.21	1407	1				367	
CN479126	Hs.504609	CN479126	20q11.21	1408					368	
NM_000681	Hs.249159	ADRA2A	10q25.2	1409					591	NvsT
NM_006829	Hs.642660	C10orf116	10q23.2	1410					387	
NM_000494	Hs.117938	COL17A1	10q25.1	1411					388	
NM_001003395	Hs.591347	TPD52L1	6q22.31	1412					170	
NM_144696	Hs.658505	C1orf125	1q25.2	1413			3		573	
NM_183393	Hs.654933	CADPS	3p14.2	1414					171	
BI836406	Hs.130203	BI836406	4q22.1	1415					215	
AK090515	Hs.181297	LOC283663	15q21.3	1416					418	
NM_014859	Hs.499758	KIAA0672	17p12	1417					697	NvsT
NM_001823	Hs.173724	CKB	14q32.33	1418					589	NvsT
A_24_P538328		A_24_P538328	16q11.2	1419					590	
NM_025163	Hs.518403	PIGZ	3q29	1420					588	NvsT
AK023526	Hs.670477	AK023526	4q22.1	1421					1658	
NM_015957	Hs.447794	APIP	11p13	1422					183	
A_24_P170203		A_24_P170203	9q31.2	1423					184	
NM_020997	Hs.656214	LEFTY1	1q42.12	1424					186	
NM_080818	Hs.352218	OXGR1	13q32.1	1425					185	
NM_001482	Hs.75335	GATM	15q21.1	1426					578	
THC2687713		THC2687713	4q33	1427					579	
NM_019000	Hs.481704	FAM134B	5p15.1	1428	1				771	NvsT
THC2675062		THC2675062	2q24.3	1429	1				772	NvsT
NM_194298	Hs.499709	SLC16A9	10q21.1	1430					755	NvsT
NM_006834	Hs.287714	RAB32	6q24.3	1431	1				582	
NM_207578	Hs.487325	PRKACB	1p31.1	1432					595	NvsT
NM_001010971	Hs.591445	SAMD13	1p31.1	1433					596	NvsT
NM_001077	Hs.575083	UGT2B17	4q13.2	1434					597	NvsT
NM_001076	Hs.651166	UGT2B15	4q13.2	1435					598	NvsT

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
A_24_P575267		A_24_P575267	4q13.2	1436					599	NvsT
NM_001073	Hs.339811	UGT2B11	4q13.2	1437					600	NvsT
NM_001075	Hs.201634	UGT2B10	4q13.2	1438					601	NvsT
NM_001074	Hs.654424	UGT2B7	4q13.2	1439	1				602	
NM_053039	Hs.653154	UGT2B28	4q13.2	1440	1				603	
NM_153699	Hs.553652	GSTA5	6p12.1	1441					604	NvsT
RC064586	Hs.410126	LOC145837	15q24.1	1442					605	NvsT
NM_004212	Hs.367833	SLC28A2	15q21.1	1443					799	
NM_153486	Hs.380929	LDHD	16q23.1	1444					690	NvsT
NM_170601	Hs.10056	SIAE	11q24.2	1445					875	
BC063383	Hs.306721	LOC400573	17p13.1	1446			3		711	NvsT
NM_031264	Hs.165619	MUPCDH	11p15.5	1447	1				702	NvsT
NM_016341	Hs.655033	PLCE1	10q23.33	1448					703	NvsT
NM_080489	Hs.657015	SDCRP2	20p13	1449					687	NvsT
A_32_P218707		A_32_P218707	2p11.2	1450					688	NvsT
NM_002842	Hs.179770	PTPRH	19q13.42	1451					689	NvsT
NM_005629	Hs.540696	SLC6A8	Xq28	1452					910	
NM_024307	Hs.289015	GDPD3	16p11.2	1453					677	NvsT
BC103878	Hs.288568	MOGAT2	11q13.5	1454	1		3		678	NvsT
NM_181795	Hs.486354	PKIB	6q22.31	1455	1				684	NvsT
NM_005771	Hs.179608	DHRS9	2q31.1	1456					685	NvsT
AK026328	Hs.232604	FLJ22675	11q13.1	1457	1				686	NvsT
NM_033128	Hs.655515	SCIN	7p21.3	1458					679	NvsT
NM_000860	Hs.655491	HPGD	4q34.1	1459					680	NvsT
NM_019010	Hs.84905	KRT20	17q21.2	1460					676	NvsT
NM_007072	Hs.252351	HHLA2	3q13.13	1461					675	NvsT
AF088076	Hs.445045	AF088076	1q23.2	1462	1		3		665	NvsT
NM_015328	Hs.600789	KIAA0828	7q32.1	1463			3		692	NvsT
NM_020299	Hs.116724	AKR1B10	7q33	1464					693	NvsT
XR_018726	Hs.646726	LOC340888	10q21.3	1465					694	NvsT
NM_197975	Hs.225949	BTNL3	5q35.3	1466	1		3		666	NvsT
NM_024850	Hs.189109	BTNL8	5q35.3	1467	1				667	
NM_032321	Hs.389311	MGC13057	2q32.2	1468					637	NvsT
NM_002153	Hs.162795	HSD17B2	16q23.3	1469					691	NvsT
NM_001041	Hs.429596	SI	3q26.1	1470					759	NvsT
NM_000035	Hs.530274	ALDOB	9q31.1	1471			3		160	
NM_024803	Hs.163079	TUBAL3	10p15.1	1472					681	NvsT
NM_198998	Hs.534650	AQP12A	2q37.3	1473					310	
CR749831	Hs.282795	CR749831	10q11.23	1474	1				349	
NM_175873	Hs.13308	ANKRD43	5q31.1	1475	1				350	
NM_005588	Hs.179704	MEP1A	6p12.3	1476	1				351	NvsT

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_012338	Hs.16529	TSPAN12	7q31.31	1477	1				375	
NM_004063	Hs.591853	CDH17	8q22.1	1478	1				353	
AB018258	Hs.109358	ATP10B	5q34	1479	1				354	
NM_004963	Hs.524278	GUCY2C	12p13.1	1480	1				355	
NM_024791	Hs.374726	PDZD3	11q23.3	1481					719	NvsT
NM_000341	Hs.112916	SLC3A1	2p21	1482	1				359	
NM_005379	Hs.5394	MYO1A	12q13.3	1483					718	NvsT
NM_003869	Hs.282975	CES2	16q22.1	1484					721	NvsT
NM_145202	Hs.15951	PRAP1	10q26.3	1485				4	297	
NM_022097	Hs.178589	LOC63928	16p12.1	1486					720	NvsT
AK026416	Hs.280892	FLJ22763	3q13.13	1487	1			4	664	
NM_018663	Hs.430299	PXMP2	12q24.33	1488	1				695	NvsT
NM_000017	Hs.507076	ACADS	12q24.31	1489					696	NvsT
NM_006149	Hs.5302	JGAL S4	19q13.2	1490			3		713	NvsT
NM_003226	Hs.82961	TFF3	21q22.3	1491					162	
NM_005971	Hs.301350	FXYD3	19q13.12	1492					714	NvsT
NM_016614	Hs.403010	TTRAP	6p22.2	1493					708	NvsT
NM_022901	Hs.128071	LRRC19	9p21.2	1494	1			4	709	NvsT
A_24_P334208		A_24_P334208	2p11.2	1495					710	NvsT
NM_019062	Hs.124835	RNF186	1p36.13	1496	1		3		352	
NM_005814	Hs.651244	GPA33	1q24.1	1497	1				712	NvsT
NM_001644	Hs.560	APOBEC1	12p13.31	1498	1				161	
NM_004616	Hs.170563	TSPAN8	12q21.1	1499					574	
NM_001039792	Hs.208081	UNQ338	9p13.3	1500					575	NvsT
NM_003944	Hs.632460	SELENBP1	1q21.3	1501					698	NvsT
NM_015480	Hs.293917	PVRL3	3q13.13	1502	1				576	
BC017572	Hs.293917	PVRL3	3q13.13	1503	1				577	
NM_002591	Hs.1872	PCK1	20q13.31	1504					715	NvsT
NM_001815	Hs.11	CEACAM3	19q13.2	1505					316	
NM_004363	Hs.466814	CEACAM5	19q13.2	1506					317	
NM_006890	Hs.74466	CEACAM7	19q13.2	1507					699	NvsT
NM_002483	Hs.466814	CEACAM6	19q13.2	1508					314	
BC005008	Hs.466814	CEACAM6	19q13.2	1509					315	
NM_001712	Hs.512682	CEACAM1	19q13.2	1510					700	NvsT
ENST00000343933	Hs.113094	CORO2A	9q22.33	1511	1		3		562	
NM_000196	Hs.1376	HSD11B2	16q22.1	1512					701	NvsT
AL137761	Hs.8379	FAM110C	2p25.3	1513					556	
THC2650074		THC2650074	9q32	1514					169	NvsT
NM_014585	Hs.643005	SLC40A1	2q32.2	1515	1				558	
NM_000765	Hs.111944	CYP3A7	7q22.1	1516					559	
NM_000777	Hs.150276	CYP3A5	7q22.1	1517					560	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_017460	Hs.654391	CYP3A4	7q22.1	1518					561	
NM_001979	Hs.212088	EPHX2	8p21.1	1519					571	NvsT
BC038245	Hs.335413	BC038245	22q13.2	1520					293	NvsT
CR624517	Hs.586321	CR624517	7p21.3	1521					292	
NM_194284	Hs.183617	CLDN23	8p23.1	1522					662	NvsT
NM_005218	Hs.32949	DEFB1	8p23.1	1523					419	NvsT
NM_004861	Hs.17958	GAL3ST1	22q12.2	1524					557	
NM_003500	Hs.444959	ACOX2	3p14.3	1525					158	
THC2706212		THC2706212	4p16.1	1526					159	
NM_000903	Hs.406515	NQO1	16q22.1	1527					157	
AK131023	Hs.662541	AK131023	17q25.1	1528	1				563	NvsT
THC2770152		THC2770152	20q11.23	1529					224	
NM_007197	Hs.31664	FZD10	12q24.33	1530					1007	
NM_002523	Hs.3281	NPTX2	7q22.1	1531					1008	
NM_058186	Hs.473877	FAM3B	21q22.3	1532					291	
NM_004683	Hs.77854	RGN	Xp11.3	1533					1018	
NM_004010	Hs.495912	DMD	Xp21.2	1534					1019	
THC2729899		THC2729899	4q21.22	1535					1020	
NM_002986	Hs.54460	CCL11	17q12	1536					423	
NM_014879	Hs.2465	P2RY14	3q25.1	1537					654	NvsT
NM_014479	Hs.521459	ADAMDEC1	8p21.2	1538					656	NvsT
CR605298	Hs.676511	LOC283904	16q24.1	1539					363	
NM_024768	Hs.134807	CCDC48	3q21.3	1540					364	
XM_001133802	Hs.111902	LOC729464	16q24.1	1541					365	NvsT
NM_033380	Hs.369089	COL4A5	Xq22.3	1542	1				361	
NM_004101	Hs.42502	F2RL2	5q13.3	1543					271	
NM_000029	Hs.19383	AGT	1q42.2	1544					270	
NM_004887	Hs.483444	CXCL14	5q31.1	1545					360	
NM_001463	Hs.128453	FRZB	2q32.1	1546					1646	
NM_021229	Hs.201034	NTN4	12q22	1547					1647	
AK091132	Hs.120633	AK091132	11q21	1548					1648	
NM_005822	Hs.440168	DSCR1L1	6p12.3	1549					1208	NvsT
ENST00000256861		ITIH5	10p14	1550					362	
NM_015429	Hs.477015	ABI3BP	3q12.2	1551					1149	NvsT
NM_002736	Hs.433068	PRKAR2B	7q22.3	1552					1150	NvsT
NM_004657	Hs.26530	SDPR	2q32.3	1553					1209	NvsT
NM_022469	Hs.98206	GREM2	1q43	1554					614	NvsT
NM_002825	Hs.371249	PTN	7q33	1555					1146	NvsT
THC2609493		THC2609493	7q31.1	1556					1147	NvsT
NM_001017915	Hs.601911	INPP5D	2q37.1	1557					239	
NM_032432	Hs.233404	ABLIM2	4p16.1	1558					377	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
NM_015205	Hs.29189	ATP11A	13q34	1559					1621	NvsT
NM_021101	Hs.439060	CLDN1	3q28	1560					1622	NvsT
NM_033260	Hs.591352	FOXQ1	6p25.3	1561					1623	NvsT
ENST00000309874	Hs.599221	ENST00000309874	15q15.1	1562					1563	
BE004814	Hs.647409	BE004814	6p22.3	1563					1615	
NM_002763	Hs.585369	PROX1	1q41	1564					1559	
NM_004655	Hs.156527	AXIN2	17q24.1	1565					280	
AF078165	Hs.156527	AXIN2	17q24.1	1566	1				281	
NM_178493	Hs.106137	NOTUM	17q25.3	1567					1560	
NM_033119	Hs.592059	NKD1	16q12.1	1568					1561	
NM_153000	Hs.293274	APCDD1	18p11.22	1569					1557	
THC2652817		THC2652817	16q12.1	1570					1558	
NM_020416	Hs.479069	PPP2R2C	4p16.1	1571					1581	
NM_018689	Hs.459088	KIAA1199	15q25.1	1572					61	NvsT
NM_006393	Hs.5025	NEBL	10p12.31	1573					60	NvsT
NR_002307	Hs.381329	MSX2P	17q22	1574					59	
BC005081	Hs.663621	BCAN	1q23.1	1575					240	
BU540282	Hs.600805	BU540282	2p16.1	1576					241	
NM_002334	Hs.4930	LRP4	11p11.2	1577					1645	
AF132203	Hs.597496	AF132203	10q24.31	1578					287	NvsT
NM_002083	Hs.2704	GPX2	14q23.3	1579					69	
NM_000740	Hs.7138	CHRM3	1q43	1580					236	
NM_001017928	Hs.220594	CCDC58	3q21.1	1581					78	
BC030122	Hs.559426	BC030122	5q14.3	1582					151	
XM_933693	Hs.181156	LOC440731	1q42.2	1583					150	
D89479	Hs.129742	SULT1B1	4q13.3	1584	1				583	NvsT
NM_014465	Hs.129742	SULT1B1	4q13.3	1585	1				584	NvsT
NM_006418	Hs.559736	OLFM4	13q21.1	1586					174	
BX538051	Hs.149540	BX538051	1q25.2	1587					175	
NM_021010	Hs.655233	DEFA5	8p23.1	1588					177	
NM_001926	Hs.711	DEFA6	8p23.1	1589					178	
NM_006017	Hs.614734	PROM1	4p15.32	1590					132	
NM_024677	Hs.590923	NSUN7	4p14	1591	1				203	
THC2732721		THC2732721	7q35	1592					379	
NM_032501	Hs.529353	ACSS1	20p11.21	1593					380	
AB075864	Hs.370555	KIAA1984	9q34.3	1594					381	
NM_032867	Hs.128196	MICALCL	11p15.3	1595					593	NvsT
NM_006671	Hs.104637	SLC1A7	1p32.3	1596					300	
BC004287	Hs.434957	BC004287	8q21.13	1597					1649	
NM_007210	Hs.505575	GALNT6	12q13.13	1598					1612	
NM_152673	Hs.599259	MUC20	3q29	1599	1				319	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
AK095831	Hs.34447	AK095831	5p15.2	1600					318	
BF803942	Hs.27621	BF803942	5p15.2	1601					320	
NM_000218	Hs.95162	KCNQ1	11p15.4	1602					384	
NM_001185	Hs.546239	AZGP1	7q22.1	1603					290	NvsT
NM_000238	Hs.647099	KCNH2	7q36.1	1604					382	
NM_021258	Hs.110915	IL22RA1	1p36.11	1605					383	
NM_004413	Hs.109	DPEP1	16q24.3	1606					278	NvsT
AL359055	Hs.547764	AL359055	7p15.3	1607					253	NvsT
THC2512199		THC2512199	7p15.3	1608					254	NvsT
BC033117	Hs.224879	LOC645249	7q36.3	1609					251	NvsT
AW979273	Hs.598388	AW979273	7p15.3	1610					252	NvsT
A_32_P25065		A_32_P25065	19p13.11	1611	1			4	385	
NM_000245	Hs.132966	MET	7q31.2	1612					242	NvsT
NM_032951	Hs.647055	MLXIP1	7q11.23	1613					243	
NM_021977	Hs.567337	SLC22A3	6q25.3	1614					244	NvsT
NM_001955	Hs.511899	EDN1	6p24.1	1615					1708	
NM_002655	Hs.14968	PLAG1	8q12.1	1616			3		1709	
AK094963	Hs.657374	FLJ37644	17q24.3	1617	1				302	
NM_001399	Hs.105407	EDA	Xq13.1	1618					301	
BC032913	Hs.97540	LOC389023	2q14.1	1619					565	NvsT
NM_153270	Hs.448572	KLHL34	Xp22.12	1620					304	
NM_005396	Hs.423598	PNLIPRP2	10q25.3	1621					295	
NM_004164	Hs.655516	RBP2	3q23	1622	1				296	
AK130456	Hs.200573	AK130456	6p22.1	1623					303	
NM_022144	Hs.132957	TNMD	Xq22.1	1624					305	
NM_001045	Hs.591192	SLC6A4	17q11.2	1625					306	
THC2669157		THC2669157	10q22.1	1626					307	
NM_006308	Hs.41707	HSPB3	5q11.2	1627					308	
THC2706471		THC2706471	9q13	1628	1				309	
AW901755		AW901755	5q15	1629					231	
NM_025152	Hs.288981	NUBPL	14q12	1630					233	
NM_001005353	Hs.10862	AK3L1	1p31.3	1631				4	232	
NM_080759	Hs.129452	DACH1	13q21.33	1632					294	
BC037328	Hs.49329	BC037328	17q24.3	1633	1				552	
AY358469	Hs.632511	LOC646627	1q44	1634					731	NvsT
NM_019018	Hs.591751	FAM105A	5p15.2	1635					216	
AK125648	Hs.155085	AK125648	5p15.2	1636					217	
NM_000880	Hs.591873	IL7	8q21.12	1637					214	
THC2692456		THC2692456	1p32.1	1638					228	
NM_017781	Hs.272795	CYP2W1	7p22.3	1639					265	
NM_032251	Hs.98564	CCDC88	11q13.1	1640					311	

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
ENST00000305119	Hs.489355	MUC12	7q22.1	1641					1650	NvsT
A_32_P112263		A_32_P112263	6q13	1642					219	
NM_004923	Hs.145932	MTL5	11q13.2	1643					202	
NM_018291	Hs.444301	FLJ10986	1p32.1	1644					218	
NM_152891	Hs.280658	PRSS33	16p13.3	1645					227	
NM_030615	Hs.150013	KIF25	6q27	1646					179	
AK093351	Hs.297967	FLJ36032	1q21.3	1647					312	
NM_152358	Hs.414175	C19orf41	19q13.33	1648					229	
XR_017750		NANOS3	19p13.12	1649					226	
NM_025243	Hs.221597	SLC19A3	2q36.3	1650				4	220	
NM_003636	Hs.440497	KCNAB2	1p36.31	1651					230	
NM_001333	Hs.660866	CTSL2	9q22.33	1652					264	
NM_152709	Hs.37636	STOX1	10q21.3	1653	1				221	
THC2700133		THC2700133	19p12	1654					222	
NM_006644	Hs.36927	HSPH1	13q12.3	1655				4	288	NvsT
ENST00000379426		ENST00000379426	6p24.1	1656					998	
NM_014467	Hs.306339	SRPX2	Xq22.1	1657					1616	NvsT
NM_017549	Hs.563491	EPDR1	7p14.1	1658					257	
NM_015352	Hs.472409	POFUT1	20q11.21	1659				4	261	
NM_080672	Hs.473218	PHACTR3	20q13.32	1660			3		258	
NM_015515	Hs.9029	KRT23	17q21.2	1661					259	
NM_001010853	Hs.652106	ACY1L2	6q15	1662				4	260	
NM_003279	Hs.182421	TNNC2	20q13.12	1663				4	341	
NM_002623	Hs.91161	PFDN4	20q13.2	1664					245	NvsT
NM_002467	Hs.202453	MYC	8q24.21	1665					272	NvsT
NM_144583	Hs.580464	ATP6V1C2	2p25.1	1666					255	
NM_004485	Hs.591531	GNG4	1q42.3	1667					256	
NM_030667	Hs.160871	PTPRO	12p12.3	1668					235	
NM_016653	Hs.444451	ZAK	2q31.1	1669					248	NvsT
NM_004289	Hs.404741	NFE2L3	7p15.2	1670	1				249	NvsT
ENST00000211092	Hs.130714	LOC284422	19p13.3	1671					299	
NM_003260	Hs.332173	TLE2	19p13.3	1672					321	
NM_020686	Hs.336768	ABAT	16p13.2	1673				4	322	
NM_002657	Hs.154104	PLAGL2	20q11.21	1674				4	262	
BC010544	Hs.660769	BC010544	19p13.3	1675				4	263	NvsT
NM_000142	Hs.1420	FGFR3	4p16.3	1676					366	
ENST00000353250		PIWIL2	8p21.3	1677					323	
NM_152321	Hs.162143	ERP27	12p12.3	1678	1				339	
NM_003122	Hs.407856	SPINK1	5q32	1679	1				340	
NM_001008494	Hs.567637	ISX	22q12.3	1680					335	NvsT
NM_001445	Hs.519719	FABP6	5q33.3	1681	1				289	NvsT

Unique Id	UG cluster	Gene symbol	Location	Disposición Heatmap Clasificación tumoral	Cluster 1 vs resto	Cluster 2 vs resto	Cluster 3 vs resto	Cluster 4 vs resto	Disposición Heatmap Norm vs Tum	Dif Exp NvsT
THC2671575		THC2671575		1682	1				581	
NM_003114	Hs.591866	SPAG1	8q22.2	1683					187	
NM_007052	Hs.592227	NOX1	Xq22.1	1684					324	
NM_013955	Hs.592227	NOX1	Xq22.1	1685					325	
NM_013954	Hs.592227	NOX1	Xq22.1	1686					326	
NM_000790	Hs.359698	DDC	7p12.2	1687	1			4	358	
NM_021804	Hs.178098	ACF2	Xp22.2	1688					336	
NM_000047	Hs.386975	ARSE	Xp22.33	1689	1			4	337	
THC2696614		THC2696614	Xp22.33	1690	1				338	
NM_003889	Hs.7303	NR1I2	3q13.33	1691	1				356	
NM_170693	Hs.300863	SGK2	20q13.12	1692	1				357	
NM_173216	Hs.207459	ST6GAL1	3q27.3	1693				4	238	
NM_033427	Hs.592285	CTTNBP2	7q31.2	1694				4	234	
NM_000492	Hs.489786	CFTR	7q31.2	1695				4	237	
NM_017763	Hs.656319	RNF43	17q22	1696	1			4	282	NvsT
NM_005170	Hs.152475	ASCL2	11p15.5	1697				4	283	NvsT
NM_014324	Hs.508343	AMACR	5p13.3	1698	1			4	342	
NM_030949	Hs.486798	PPP1R14C	6q25.1	1699	1			4	348	
NM_025113	Hs.98117	C13orf18	13q14.12	1700				4	346	
NM_014298	Hs.513484	QPRT	16p11.2	1701	1			4	343	
BG114486	Hs.592544	BG114486	16p11.2	1702				4	344	
NM_003270	Hs.43233	TSPAN6	Xq22.1	1703	1			4	345	
NM_006113	Hs.267659	VAV3	1p13.3	1704	1			4	347	
NM_144704	Hs.163543	AIFM3	22q11.21	1705					332	NvsT
NM_015265	Hs.516617	SATB2	2q33.1	1706					333	
THC2678045		THC2678045	6q27	1707	1			4	331	
NM_000845	Hs.449625	GRM8	7q31.33	1708					313	
NM_000933	Hs.472101	PLCB4	20p12.2	1709	1				327	
NM_005021	Hs.486489	ENPP3	6q23.2	1710	1				328	
NM_018296	Hs.125139	LRRC36	16q22.1	1711					276	
NM_003212	Hs.385870	TDGF1	3p21.31	1712					277	
NM_014553	Hs.156471	TFCP2L1	2q14.2	1713	1				334	NvsT
L33930	Hs.644105	CD24	Yq11.222	1714					329	
BX413550	Hs.688325	BX413550	Yq11.222	1715					330	
NM_001657	Hs.270833	AREG	4q13.3	1716					266	
NM_001432	Hs.115263	EREG	4q13.3	1717					267	
NM_018407	Hs.492314	LAPTM4B	8q22.1	1718					285	
NM_003878	Hs.78619	GGH	8q12.3	1719					279	
NM_013296	Hs.584901	GPSM2	1p13.3	1720					284	NvsT
NM_001338	Hs.693697	CXADR	21q21.1	1721	1				268	
AL834140	Hs.677892	AL834140	15q11.2	1722					269	