

UNIVERSIDAD DE ALCALÁ

FACULTAD DE MEDICINA

DPTO. DE MEDICINA



**ANÁLISIS DE LA EXPRESIÓN GÉNICA
GLOBAL EN ESTADIOS INICIALES DEL
CARCINOMA NO MICROCÍTICO DE PULMÓN:
IDENTIFICACIÓN DE UN PREDICTOR DE
RECIDIVA ASOCIADO CON LAS CÉLULAS B DEL
SISTEMA INMUNE**

TESIS DOCTORAL

SUSANA HERNÁNDEZ PRIETO

Licenciada en Biología, especialidad Biología Celular y Molecular

ALCALÁ DE HENARES, 2014

HOSPITAL CLÍNICO SAN CARLOS

SERVICIO DE ANATOMÍA PATOLÓGICA

SERVICIO DE CIRUGÍA TORÁCICA

SERVICIO DE ONCOLOGÍA MÉDICA



DIRECTOR:

Dr. Julián SANZ ORTEGA

CODIRECTORES:

Dr. Florentino HERNANDO TRANCHO

Dr. José Antonio LÓPEZ GARCÍA-ASENJO

Dra. Beatriz PÉREZ-VILLAMIL SALGADO

El presente estudio ha sido realizado en el Laboratorio de Patología Molecular y Dianas Terapéuticas (Servicio de Anatomía Patológica), Servicio de Cirugía Torácica y Laboratorio de Oncología Molecular (Servicio de Oncología Médica) del Hospital Clínico San Carlos de Madrid y se presenta en el Departamento de Medicina de la Facultad de Medicina de la Universidad de Alcalá por Susana Hernández Prieto para optar al grado de Doctor.

D. Julián Sanz Ortega, Doctor en Medicina y Cirugía, responsable del Laboratorio de Patología Molecular y Biobanco del Servicio de Anatomía Patológica del Hospital Clínico San Carlos de Madrid y Profesor Titular de la Universidad Complutense de Madrid

CERTIFICA:

Que Dña. Susana Hernández Prieto, Licenciada en Biología por la Universidad de Alcalá, especialidad Biología Celular y Molecular ha desarrollado bajo mi dirección el trabajo de investigación titulado **“ANÁLISIS DE LA EXPRESIÓN GÉNICA GLOBAL EN ESTADIOS INICIALES DEL CARCINOMA NO MICROCÍTICO DE PULMÓN: IDENTIFICACIÓN DE UN PREDICTOR DE RECIDIVA ASOCIADO CON LAS CÉLULAS B DEL SISTEMA INMUNE”** al que, tras su revisión, doy mi aprobación para ser presentado y defendido como Tesis Doctoral.

Y para que así conste, firmo el presente certificado en Madrid, a 28 de Marzo de 2014.



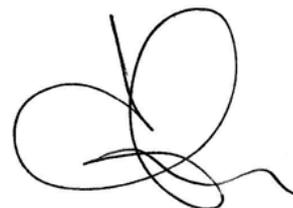
Fdo. Dr. Julián Sanz Ortega

D. Florentino Hernando Trancho, Doctor en Medicina y Cirugía,
Jefe de Servicio de Cirugía Torácica del Hospital Clínico San Carlos de Madrid
y profesor asociado de la Universidad Complutense de Madrid

CERTIFICA:

Que Dña. Susana Hernández Prieto, Licenciada en Biología por la Universidad de Alcalá, especialidad Biología Celular y Molecular ha desarrollado bajo mi codirección el trabajo de investigación titulado **“ANÁLISIS DE LA EXPRESIÓN GÉNICA GLOBAL EN ESTADIOS INICIALES DEL CARCINOMA NO MICROCÍTICO DE PULMÓN: IDENTIFICACIÓN DE UN PREDICTOR DE RECIDIVA ASOCIADO CON LAS CÉLULAS B DEL SISTEMA INMUNE”** al que, tras su revisión, doy mi aprobación para ser presentado y defendido como Tesis Doctoral.

Y para que así conste, firmo el presente certificado en Madrid,
a 28 de Marzo de 2014.



Fdo. Dr. Florentino Hernando Trancho



Universidad
Europea de Madrid

LAUREATE INTERNATIONAL UNIVERSITIES

D. José Antonio López García-Asenjo, Doctor en Medicina y Cirugía,
Director de Anatomía Patológica Labco España y Portugal y profesor asociado de la
Universidad Europea de Madrid

CERTIFICA:

Que Dña. Susana Hernández Prieto, Licenciada en Biología por la Universidad de Alcalá, especialidad Biología Celular y Molecular ha desarrollado bajo mi codirección el trabajo de investigación titulado **“ANÁLISIS DE LA EXPRESIÓN GÉNICA GLOBAL EN ESTADIOS INICIALES DEL CARCINOMA NO MICROCÍTICO DE PULMÓN: IDENTIFICACIÓN DE UN PREDICTOR DE RECIDIVA ASOCIADO CON LAS CÉLULAS B DEL SISTEMA INMUNE”** al que, tras su revisión, doy mi aprobación para ser presentado y defendido como Tesis Doctoral.

Y para que así conste, firmo el presente certificado en Madrid,
a 28 de Marzo de 2014.

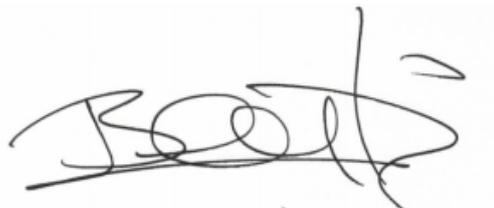
Fdo. Dr. José Antonio López García-Asenjo

Dña. Beatriz Pérez-Villamil Salgado, Doctora en Ciencias Biológicas,
Investigadora del Instituto de Investigación Sanitaria (IdISSC) del Hospital Clínico
San Carlos en el Servicio de Oncología Médica,

CERTIFICA:

Que Dña. Susana Hernández Prieto, Licenciada en Biología por la Universidad de Alcalá, especialidad Biología Celular y Molecular ha desarrollado bajo mi codirección el trabajo de investigación titulado **“ANÁLISIS DE LA EXPRESIÓN GÉNICA GLOBAL EN ESTADIOS INICIALES DEL CARCINOMA NO MICROCÍTICO DE PULMÓN: IDENTIFICACIÓN DE UN PREDICTOR DE RECIDIVA ASOCIADO CON LAS CÉLULAS B DEL SISTEMA INMUNE”** al que, tras su revisión, doy mi aprobación para ser presentado y defendido como Tesis Doctoral.

Y para que así conste, firmo el presente certificado en Madrid,
a 28 de Marzo de 2014.



Fdo. Dra. Beatriz Pérez-Villamil Salgado



**Melchor Álvarez de Mon Soto, Catedrático de Medicina y Director del
Departamento de Medicina y Especialidades Médicas**

INFORMA QUE:

En su opinión, el trabajo de investigación presentado por D. **Susana Hernández Prieto** titulado **“Análisis de la expresión génica global en estadios iniciales del carcinoma no microcítico de pulmón: identificación de un predictor de recidiva asociado con las células B del sistema inmune”**, realizado bajo la dirección del Dr. D. Julián Sanz Ortega y la codirección de los Dres. D. Florentino Hernando Trancho, D. José Antonio López García-Asenjo y D^a. Beatriz Pérez-Villamil Salgado, reúne los requisitos científicos, metodológicos, formales y de originalidad suficientes para ser defendido como Tesis Doctoral ante el Tribunal que legalmente corresponda.

Y para que conste donde proceda, a los efectos oportunos, se firma la presente en Alcalá de Henares a cuatro de abril de dos mil catorce.




*A Félix y Sergio
A mis padres y hermanas*

Han sido varios años los que he dedicado a elaborar este proyecto de tesis doctoral y durante los cuales he conocido a muchas personas que han puesto su granito de arena. Por ello, a todas ellas, gracias. De manera especial quiero agradecer:

Al Profesor Julián Sanz Esponera, gran anatomopatólogo y persona excepcional que siempre apoyó, desde mi incorporación al Servicio de Anatomía Patológica del Hospital Clínico San Carlos, mi actividad tanto asistencial como de investigación: patología molecular, banco de tumores, proyectos de investigación, publicación de resultados, desarrollo de cursos de formación...y por supuesto, mi proyecto de tesis. Por todo ello, muchas gracias jefe.

Al Dr. Julián Sanz Ortega que me dio la oportunidad en el año 2.000, de incorporarme al proyecto del banco de tumores del Hospital Clínico San Carlos y así entrar a formar parte de su laboratorio de Patología Molecular donde comencé a desarrollar mi actividad profesional. Por su gran interés en la investigación y por su confianza en mí para desarrollar este proyecto, que sé que considera, el mejor de los que ha dirigido. También agradecerle de corazón todo el apoyo mostrado en un momento muy duro para mí, cuando nuestra relación laboral finalizó. A pesar de ello, estoy convencida que la amistad y el aprecio que nos une, nunca desaparecerán.

A mis codirectores de tesis, cada uno de ellos aportando su gran conocimiento de cada una de las áreas en las que trabajan. Grandes profesionales y sobre todo grandes personas que siempre me han apoyado en todos los años en los que hemos compartido actividad. Al Dr. Florentino Hernando Tranco que defendió de manera excepcional la utilidad y la implantación del banco de tumores en el hospital y que me transmitió sus inquietudes respecto a los pacientes con cáncer de pulmón. Por apoyar este proyecto de investigación desde que comenzó y transmitir a todo su equipo la necesidad de desarrollar, además de su actividad asistencial, una investigación de calidad. Por sus buenos consejos en la revisión de esta tesis y por su apoyo incondicional hacia mí en todo momento. Al Dr. José Antonio López-García Asenjo por toda su ayuda en la revisión histológica de las muestras del proyecto, transmitiéndome su gran conocimiento de la histopatología del pulmón. Por colaborar en el buen funcionamiento del banco de tumores y en el desarrollo de la investigación en el servicio y por apoyarme y ayudarme durante todos los años en los que hemos compartido actividad profesional. A la Dra. Beatriz Pérez-Villamil Salgado por aplicar en este proyecto, su gran conocimiento de la metodología de microarrays, por toda

la ayuda en el diseño y desarrollo metodológico del mismo, por seguir adelante a pesar de las dificultades y por ayudarme y apoyarme en el desarrollo de mi faceta investigadora.

A mis compañeras del laboratorio de Patología Molecular y Banco de Tumores, Mamen, Mila y Elena, que han hecho posible, mediante su gran ayuda, que pudiera dedicarme más intensamente al desarrollo de mi tesis doctoral. Mamen respecto a mi actividad asistencial diaria, Mila mano a mano en el laboratorio en esas largas jornadas de hibridaciones y de análisis de resultados y Elena en todo, persona que le pidas lo que le pidas siempre está dispuesta a ayudar. Gracias a las tres porque siempre me habéis animado a continuar con la tesis y sobre todo gracias por todos estos años de trabajo (y cafés) que hemos compartido y en los que puedo asegurar que hemos formado un gran equipo.

A todo el personal del Servicio de Anatomía Patológica del Hospital Clínico San Carlos, tanto ya jubilado como en activo, que siempre me han ayudado y animado durante todos estos años. Sois tantos que no puedo nombraros uno a uno pero sabéis que a todos os llevaré siempre en el corazón. Patólogos, personal técnico, supervisores, personal de secretaría, residentes, auxiliares, personal de limpieza...un gran servicio al que sólo puedo decir gracias por todo.

Al personal del laboratorio de Oncología Molecular con la Dra. Trinidad Caldés Llopis, directora, y del servicio de Oncología Médica y Unidad de Ensayos Clínicos, con el Profesor Eduardo Díaz-Rubio, Jefe de Servicio, donde he conocido y he trabajado con personas excepcionales, que me han transmitido sus conocimientos, sus consejos, su apoyo y su interés durante todo el desarrollo de este proyecto. En especial quiero agradecer al Dr. Javier Puente Vázquez y al Dr. José Luis González-Larriba por toda la ayuda prestada en el desarrollo del mismo.

A Alex, del que diga lo que diga siempre será poco. Gracias por tu interés, ayuda en el laboratorio, tus grandes conocimientos de bioinformática (que tantas veces has intentado transmitirme y que para mí eran tan difíciles de entender), por no cansarte de siempre seguir adelante, de buscar información, de aprender todos los días algo nuevo y sobre todo por animarme y por todos tus consejos siempre tan acertados. Valencia puede estar segura de tener un gran profesional trabajando allí y yo, aunque sea un poquito lejos, un gran amigo.

Al personal del Servicio de Cirugía Torácica del Hospital Clínico San Carlos, donde he conocido y he trabajado con grandes cirujanos muy implicados en su actividad asistencial pero sin dejar de lado su inquietud científica. En especial quiero agradecer a la Dra. Ana Gómez Martínez por toda la ayuda prestada durante tantos años y por su gran apoyo, como amiga, en algún que otro duro momento.

Al Dr. José Luís Subiza Lestache, al que tengo que agradecer su colaboración en la etapa final de este proyecto, ayudándome a entender el difícil mundo de la inmunología.

A mi nueva familia profesional del Servicio de Anatomía Patológica y Laboratorio de Dianas Terapéuticas (LDT) del Hospital HM Universitario Sanchinarro, con el Dr. Fernando López-Ríos, Jefe de Servicio y Director, que en este último año me han ofrecido su ayuda desinteresada y me han animado a perder ese miedo al folio en blanco para así poder finalizar esta tesis doctoral. Gracias por vuestra confianza en mí y por permitirme formar parte de un equipo de profesionales y sobre todo de personas espectacular.

A Félix, gracias por estar siempre a mi lado y por animarme con la tesis durante todos estos años, mediante tu estrategia de la psicología inversa. Por creer en mí, lo que me ha ayudado mucho a desarrollarme profesionalmente (aunque sea algo lejos de casa...) e incluso llevar a cabo y finalizar este proyecto. Gracias también por tu gran apoyo que me ha ayudado a superar los duros momentos. A Sergio, gracias por entender y ayudar a mamá cuando tenía que hacer en el ordenador lo que dices que son “todos esos deberes que te mandan en tu cole...”. A los dos, gracias por formar parte de mi vida.

A mis padres, Ignacio e Isabel, por seguir tan unidos, a pesar de las dificultades y siempre al lado de vuestras hijas. Por darme la mejor educación y la libertad para elegir la carrera que quería estudiar. A mis dos hermanas, Silvia y Cristina, por interesarse en mis avances y en los resultados de esta investigación y por animarme a finalizarla. Cristina, a ti que solamente te queda un año de carrera y que desde que comenzaste en la universidad ya te has implicado en el mundo de la investigación, sólo decirte que ánimo, porque merece la pena.

LISTADO DE ABREVIATURAS

A

- AC:** Adenocarcinoma.
- AJCC:** del inglés, American Joint Committee on Cancer.
- ALPI:** del inglés, Adjuvant Lung Project Italy.
- ANITA:** del inglés, Adjuvant Novelbine International Trial Association.
- ATS:** del inglés, American Thoracic Society.

B

- BAC:** del inglés, Bronchioloalveolar Carcinoma.
- BLT:** del inglés, Big Lung Trial.

C

- °C:** grados Celsius.
- CDK:** del inglés, Cyclin-dependent kinases.
- cdDNA:** DNA copia.
- CGA:** Campo de Gran Aumento.
- CI:** Intervalo de Confianza.
- CMP:** Carcinoma Microcítico de Pulmón o Carcinoma de Pulmón de Célula Pequeña.
- CNMP:** Carcinoma No Microcítico de Pulmón o Carcinoma de Pulmón No de Célula Pequeña.
- CTP:** Citidina Trifosfato.

D

- DAVID:** del inglés, Database for Annotation, Visualization and Integrated Discovery.
- DNA:** del inglés, Desoxyribonucleic Acid.
- DNAsa:** Enzima Endonucleasa.
- dNTPs:** Desoxirribonucleótidos Trifosfato.
- DWD:** del inglés, Distance Weighted Discrimination.

E

- ECOG:** del inglés, Eastern Cooperative Oncology Group.
- EDTA:** Ácido Etilendiaminotetraacético.
- EEUU:** Estados Unidos.
- EGFR:** Receptor del Factor de Crecimiento Epidérmico (del inglés, Epidermal Growth Factor Receptor).
- ERCC1:** del inglés, Excision Repair Cross-Complementation group 1.
- ERS:** del inglés, European Respiratory Society.

F

- FAM-6:** fluorocromo 6-carboxifluoresceína.
- Fc:** fragmento constante.

G

- GO:** del inglés, Gene Ontology.
- GSEA:** del inglés, Gene Set Enrichment Analysis.

H

- HCSC:** Hospital Clínico San Carlos.
- HEX-6:** fluorocromo 6-hexaclorofluoresceína
- HZ:** Hazard Ratio

I

- IALT:** del inglés, International Adjuvant Lung Cancer Trial.
- IASLC:** del inglés, International Association for the study of the lung cancer.
- Ig:** inmunoglobulinas.
- IL:** interleucinas.
- ILE:** Intervalo Libre de Enfermedad.

K

- K-M:** curvas Kaplan-Meier.
- KNN:** del inglés, K-Nearest Neighbour.

L

- LACE:** del inglés, The Lung Adjuvant Cisplatin Evaluation.
- LCC:** carcinoma de células grandes (del inglés, Large Cell Carcinoma).
- LOH:** pérdida de heterocigosidad (del inglés, Loss Of Heterocigosity).

M

- MAQC:** del inglés, Microarray Quality Control.
- MIAME:** del inglés, Minimum Information About Microarray Experiment.
- mM:** minimolar.
- MMLV-RT:** Transcriptasa Reversa del Virus de la Leucemia Murina, del inglés Murine Leukemia Virus Reverse Transcriptase.
- mRNA:** RNA mensajero.

N

- NaCl:** cloruro sódico.
- NCCN:** del inglés, National Comprehensive Cancer Network
- NCICBR:** del inglés, National Cancer Institute of Canada Clinical Trials Group
- NER:** del inglés, Nucleotide Excision Repair.
- NSCLC:** Non-Small Cell Lung Cancer.
- NK:** Natural Killer.

O

- OMS:** Organización Mundial de la Salud.

P

- PAM:** del inglés, Prediction Analysis of Microarrays.
- pb:** pares de bases.
- PCA:** Análisis de Componentes Principales (del inglés, Principal Component Analysis).
- PCR:** Reacción en Cadena de la Polimerasa (del inglés, Polymerase Chain Reaction).
- PCR-RFLP:** reacción en cadena de la polimerasa-polimorfismos de longitud de fragmentos de restricción (del inglés, Polymerase Chain Reaction- Restriction Fragment Length Polymorphism).
- PMN:** Polimorfonucleares.
- pmol/μg:** picomoles/microgramo.
- pmol/μl:** picomoles/microlitro.
- PNT:** Procedimientos Normalizados de Trabajo.
- PT:** Pulmón Tumoral.

Q

- QC-Report:** del inglés, Quality-Control Report.

R

- RB:** Retinoblastoma.
- RE:** receptores de estrógeno.
- RIN:** del inglés, RNA Integrity Number.
- RNA:** del inglés, Ribonucleic Acid.
- RNAsa:** Enzima Ribonucleasa.
- RP:** receptores de progesterona.
- rpm:** revoluciones por minuto.
- RT-qPCR:** Retrotranscripción seguida de PCR Cuantitativa (del inglés, Reverse Transcription Quantitative Polymerase Chain Reaction).

S

- SAM:** del inglés, Significant analysis of microarrays.
- SCC:** Carcinoma Epidermoide (del inglés, Squamous Cell Carcinoma).
- SCLC:** del inglés, Small Cell Lung Cáncer.
- SDS:** Dodecilsulfato Sódico.
- SG:** Supervivencia Global.
- SOM:** del inglés, Self Organizing Maps.
- SVM:** del inglés, Support Vector Machines.

T

- TET-6:** Fluorocromo 6-tetraclorofluoresceína.
- TILs:** Linfocitos Intratumorales, del inglés Tumor Infiltrating Lymphocytes.
- TK:** Tirosin-Kinasa (del inglés, Tyrosine-Kinase).
- TKIs:** Inhibidores de Tirosin-Kinasas (del inglés, Tyrosine-Kinase Inhibitors).

U

- U/μl:** unidades microlitro.

V

- VEGF:** Factor de Crecimiento del Endotelio Vascular (del inglés, Vascular Endothelial Growth Factor).
- VEGFR:** Receptor del Factor de Crecimiento del Endotelio Vascular (del inglés, Vascular Endothelial Growth Factor Receptor).
- VPN:** Valor predictivo negativo
- VPP:** Valor predictivo positivo

W

- WT:** Wild Type.

μ

- μM:** micromolar.
- μm:** micra.

ÍNDICE**1. INTRODUCCIÓN**

1.1. Generalidades del cáncer de pulmón	3
1.1.1. Incidencia y mortalidad del cáncer de pulmón	3
1.1.2. Etiología del cáncer de pulmón	4
1.1.3. Clasificación histológica del cáncer de pulmón	5
1.2. Carcinoma no microcítico de pulmón (CNMP)	7
1.2.1. Subtipos histológicos más frecuentes del CNMP	7
1.2.2. Estadificación del CNMP	9
1.2.3. Factores pronósticos del CNMP	10
1.2.3.1. Factores pronósticos dependientes del tumor	10
1.2.3.2. Factores pronósticos dependientes del paciente	11
1.2.4. Tratamiento y tasas de recidiva en el CNMP	12
1.3. Biología molecular del cáncer de pulmón	15
1.3.1. Alteraciones en proto-oncogenes	16
1.3.1.1. <i>RAS</i>	17
1.3.1.2. <i>EGFR</i>	17
1.3.1.3. <i>ALK</i>	18
1.3.1.4. <i>B-RAF</i>	18
1.3.1.5. <i>MET</i>	19
1.3.1.6. <i>PIK3CA</i>	19
1.3.1.7. <i>HER2</i>	19
1.3.1.8. <i>ROS1</i>	19
1.3.1.9. <i>FGFR1</i>	20
1.3.1.10. <i>DDR2</i>	20
1.3.2. Inactivación de genes supresores de tumores	20
1.3.2.1. <i>TP53</i>	20
1.3.2.2. <i>Vía p16INK4a-ciclinaD1-CDK4-RB</i>	20
1.3.3. Activación de la invasión y metástasis	21
1.3.4. Inducción de la angiogénesis	21

1.3.5. Resistencia a la muerte celular programada o apoptosis	21
1.3.6. Alteraciones en genes implicados en las vías de reparación del ADN.....	21
1.4. Sistema inmune y cáncer.....	22
1.5. Microarrays.....	24
1.5.1. Microarrays de expresión	25
1.5.1.1. Marcaje e hibridación de las muestras.....	26
1.5.1.2. Análisis de datos	28
1.5.1.2.1. Pre-procesado de datos	28
1.5.1.2.2. Análisis de datos	29
1.5.2. Fiabilidad y reproducibilidad de la técnica de microarrays de expresión	30
1.5.3. Aplicaciones de los microarrays de expresión en cáncer	31
1.5.4. Aplicaciones de los microarrays de expresión en el CNMP	32
<u>2. JUSTIFICACIÓN E HIPÓTESIS</u>	
2.1. Justificación.....	37
2.2. Hipótesis	37
<u>3. OBJETIVOS DEL ESTUDIO</u>	
3.1. Objetivos del estudio	41
<u>4. MATERIAL Y MÉTODOS</u>	
4.1. Diseño experimental	45
4.1.1. Preselección de los pacientes diagnosticados de CNMP	45
4.1.2. Selección de la plataforma de microarrays.....	46
4.1.3. Selección de la técnica de hibridación de microarrays.....	47
4.2. Muestras tumorales y no tumorales.....	47
4.2.1. Obtención de las muestras de los pacientes.....	47
4.2.2. Criterios de calidad de las muestras de los pacientes	48

4.2.2.1. Muestras tumorales de CNMP: revisión histológica	48
4.2.2.2. Muestra de referencia o “pool”: revisión histológica	51
4.3. Extracción de ácidos nucleicos	51
4.3.1. Extracción de RNA: Método del trizol.....	51
4.3.2. Extracción de DNA	52
4.3.3. Cuantificación de los ácidos nucleicos	53
4.3.4. Control de calidad del RNA	53
4.4. Marcaje e hibridación de las muestras en los microarrays de expresión.....	54
4.4.1. Marcaje y amplificación de la muestra.....	56
4.4.2. Control del marcaje y amplificación de la muestra	57
4.4.3. Hibridación y lavado del microarray	57
4.5. Escaneado de los microarrays	58
4.6. Cuantificación de la imagen	59
4.7. Análisis de datos	60
4.7.1. Control de calidad.....	60
4.7.1.1. Quality Control-Report (QC-Report)	60
4.7.1.2. Análisis de componentes principales o PCA.....	61
4.7.2. Filtrado de datos	62
4.7.3. Métodos no supervisados para el descubrimiento de clases.....	64
4.8. Estudio de mutaciones.....	64
4.8.1. Estudio de mutaciones en el gen <i>EGFR</i>	64
4.8.1.1. Mutaciones en el exón 19 de <i>EGFR</i>	65
4.8.1.2. Mutaciones en el exón 21 de <i>EGFR</i>	66
4.8.2. Estudio de mutaciones en el gen <i>K-RAS</i>	67
4.8.2.1. Mutaciones en el codón 12 de <i>K-RAS</i>	68
4.8.2.2. Mutaciones en el codón 13 de <i>K-RAS</i>	69
4.8.3. Estudio de mutaciones en el gen <i>B-RAF</i>	70
4.9. Estudios de asociación.....	72

4.10. Estudio de las vías moleculares alteradas entre los grupos moleculares obtenidos.....	73
--	-----------

4.11. Validación externa de la clasificación molecular y obtención del predictor ...	73
---	-----------

4.12. Estudio de enriquecimiento de términos “Gene Ontology” en el predictor pronóstico	75
--	-----------

5. RESULTADOS

5.1. Pre-selección de pacientes	79
--	-----------

5.2. Técnica de microarrays y controles de calidad.....	79
--	-----------

5.2.1. Análisis histológico de las muestras tumorales	79
---	----

5.2.2. Control de calidad del RNA extraído	79
--	----

5.2.3. Gráficos MA-Plots.....	80
-------------------------------	----

5.2.4. Estudio de comportamientos anómalos de las muestras mediante PCA	81
---	----

5.3. Características de los 84 pacientes incluidos en el estudio	82
---	-----------

5.3.1. Estudios de asociación entre las variables clínico-histopatológicas y mutacionales con el ILE	84
--	----

5.4. Análisis de microarrays.....	85
--	-----------

5.4.1. Caracterización del patrón de expresión de muestras tumorales y muestras de parénquima pulmonar no tumoral.....	86
--	----

5.4.2. Métodos no supervisados para la clasificación molecular de los estadios iniciales del CNMP	87
---	----

5.4.2.1. Clasificación molecular del CNMP utilizando el listado de 3.232 genes ..	87
---	----

5.4.2.2. Clasificación molecular del CNMP, no relacionada con la histología, utilizando el listado de 2.160 genes.....	90
--	----

5.4.2.2.1. Estudios de asociación de la clasificación molecular en tres clusters no relacionados con la histología y las variables clínico-histopatológicas y mutacionales	94
--	----

5.4.2.2.2. Estudio de las vías moleculares alteradas entre los tres clusters obtenidos.....	96
---	----

5.5. Validación de la nueva clasificación molecular en tres clusters no relacionados con la histología utilizando una matriz de datos externa.....	98
5.5.1. Características clínico-histopatológicas de las muestras de la matriz de validación	98
5.5.2. Combinación de la matriz de desarrollo y la matriz de validación utilizando la técnica de Distance Weighted Discrimination DWD	99
5.5.3. Obtención y validación de un predictor génico para la clasificación molecular en tres clusters no relacionados con la histología.....	101
5.6. Simplificación de la clasificación molecular no relacionada con la histología y obtención de un predictor de recidiva	102
5.6.1. Obtención y validación de un predictor de recidiva	102
5.6.2. Validación del predictor de recidiva de 50 genes para estadios I y II de manera independiente.....	105
5.6.3. Descripción de las funciones de los 50 genes del predictor de recidiva.....	106
5.6.4. Estudio de enriquecimiento de términos de Gene Ontology (GO) en el predictor de 50 genes	109
<u>6. DISCUSIÓN</u>	
6.1. Importancia pronóstica de los factores clínico-histopatológicos y mutacionales convencionales en los estadios iniciales del CNMP	113
6.2. Clasificación molecular del CNMP	114
6.2.1. Firmas génicas en el CNMP, aspectos relacionados con el diseño del estudio... 115	
6.2.2. Interpretación de la nueva clasificación molecular	116
6.2.3. Predictor de recidiva: papel del sistema inmune en los estadios iniciales del CNMP.....	119
6.2.4. Comparación con otros predictores génicos en el CNMP.....	121
6.3. Posibles implicaciones clínicas del predictor de 50 genes	122
<u>7. CONCLUSIONES</u>	125

8. BIBLIOGRAFÍA 131

9. ANEXOS 147

1. INTRODUCCIÓN

1.1. GENERALIDADES DEL CÁNCER DE PULMÓN

1.1.1. Incidencia y mortalidad del cáncer de pulmón

A nivel mundial, el cáncer de pulmón es el tumor maligno más importante en cuanto a incidencia siendo diagnosticados alrededor de 1,6 millones de casos anuales, lo que representa aproximadamente el 14% de todos los nuevos casos de cáncer. Respecto a la mortalidad, es la primera causa de muerte por cáncer en varones y la segunda causa en mujeres con una tasa anual de 1,1 millones de personas y con una tasa de supervivencia global a cinco años de sólo el 15% (Siegel et al., 2013).

En España, las tasas de incidencia y mortalidad muestran una tendencia similar a la descrita anteriormente. En varones, el cáncer de pulmón es el tercer tumor maligno más frecuente. Las mayores tasas de incidencia se mostraron durante la década de los años 90 con un máximo de 85 casos por cada 100.000 varones entre 1997-99. Respecto a la mortalidad, en varones es la primera causa de muerte por cáncer por encima del cáncer de colon y recto y el de próstata. En mujeres, la incidencia del cáncer de pulmón está aumentando de manera progresiva a lo largo de los años. Respecto a la mortalidad, es la tercera causa de muerte por cáncer después del cáncer de mama y el cáncer de colon y recto (figura 1; Sánchez et al., 2010).

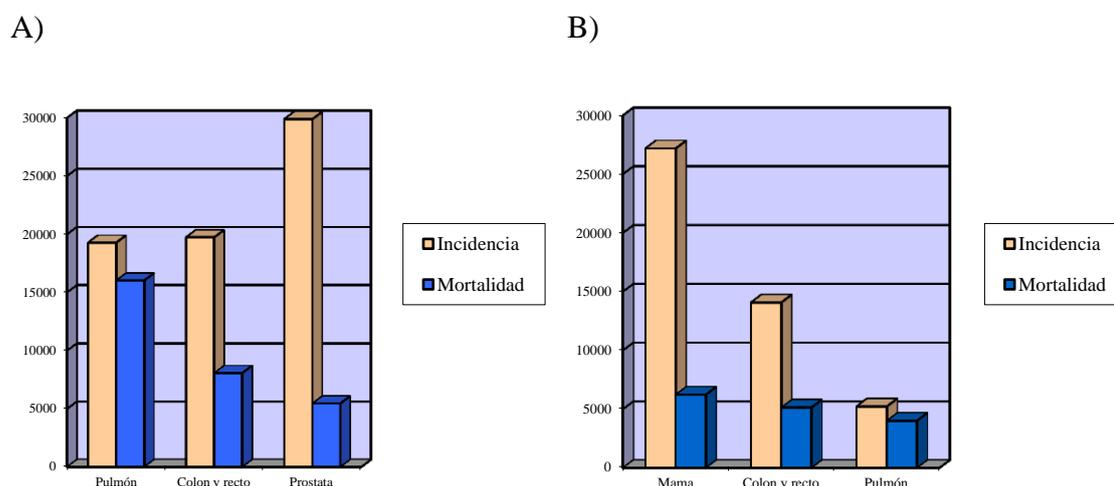


Figura 1. Incidencia y mortalidad en España de los principales tumores malignos separada por sexos.
A) Varones B) Mujeres. (Figura adaptada a partir de Sánchez et al., 2010)

En la población masculina, durante los últimos diez años, se han observado signos de estabilización en la incidencia del cáncer de pulmón y la mortalidad ha disminuido en un 10% cuando se comparan los datos actuales respecto a 2007. Por el contrario, en la

población femenina, estas tasas han ido aumentando. En el año 2007 la tasa de incidencia en varones respecto de mujeres fue 7 veces superior, pero en la actualidad ha disminuido hasta alcanzar una proporción de 4 a 1 y respecto a la tasa de mortalidad, en las mujeres está experimentando un incremento significativo con un crecimiento anual del 7% (Sculier., 2013; Malvezzi et al., 2013). Estos datos reflejan que en España se está reproduciendo el fenómeno experimentado en otros países occidentales, como Estados Unidos (EEUU), donde el cáncer de pulmón ha pasado a ocupar el primer puesto en mortalidad entre los tumores malignos en mujeres, por delante incluso del cáncer de mama, que tradicionalmente ha sido el tumor más letal entre la población femenina (Siegel et al., 2013). Estos cambios reflejan la creciente incorporación de la mujer al consumo de tabaco.

1.1.2. Etiología del cáncer de pulmón

El cáncer de pulmón era una enfermedad rara en ambos sexos a principios del siglo XX, hecho que indica que un elevado porcentaje de casos con esta enfermedad están originados por causas ambientales. Durante los últimos años, se han realizado grandes progresos en la caracterización de las causas ambientales que afectan en la aparición de este tipo de cáncer. Existen diferentes agentes etiológicos causantes del cáncer de pulmón (figura 2) pero los estudios epidemiológicos y experimentales señalan a un agente predominante, el tabaco, que causa aproximadamente entre el 80-90% de los casos de cáncer de pulmón en aquellos países donde el hábito de fumar es común (Alberg et al., 2013).



Figura 2. Agentes etiológicos del cáncer de pulmón. La figura muestra los diferentes agentes etiológicos causantes del cáncer de pulmón.

El humo del tabaco contiene más de 1.200 sustancias tóxicas causantes de cáncer que se dividen en agentes iniciadores (hidrocarburos aromáticos policíclicos del tipo benzopireno), agentes promotores (derivados del fenol), elementos radioactivos (polonio 210, carbono-14 y potasio-40) y otros contaminantes como el arsénico, níquel, mohos y aditivos. Al ser el tabaco el principal factor de riesgo, la incidencia y mortalidad del cáncer de pulmón son reflejo de los cambios sociales respecto al comportamiento de este hábito (dosis, años y tipo de tabaco). En comparación con los fumadores persistentes, el riesgo de padecer cáncer de pulmón se reduce tras dejar de fumar (De la Cruz et al., 2011; Sankpal et al., 2012). En España en el año 2006 el tabaquismo ocasionó 58.573 muertes que constituyen el 16,2% de todos los fallecimientos. Según la causa de muerte, el 29,4% de estas defunciones se debió al cáncer de pulmón (Hernández-García et al., 2010).

Además del tabaco también existen otros carcinógenos relacionados con el cáncer de pulmón, como los presentes en exposiciones ocupacionales (asbestos, níquel, cromo y arsénico) y exposición a radiaciones, incluyendo el gas radón en minas y exposición a la contaminación del aire. Algunos de estos carcinógenos interaccionan de manera sinérgica con el tabaco y aumentan el riesgo de desarrollo de cáncer de pulmón, como por ejemplo, el efecto sinérgico del tabaco y la exposición a asbestos o radón; la sinergia entre ambos induce un riesgo de 50 a 90 veces mayor de desarrollar cáncer de pulmón del que tiene una persona que está expuesta a este carcinógeno pero que no fuma (Saracci, 1987).

En los últimos años se están publicando evidencias científicas de que factores propios del estilo de vida de las personas, como el tipo de dieta o la falta de ejercicio, también podrían influir en el riesgo de desarrollo de cáncer de pulmón (Kushi et al., 2012).

1.1.3. Clasificación histológica del cáncer de pulmón

La clasificación histológica de los tumores de pulmón vigente en la actualidad es la propuesta por la Organización Mundial de la Salud (OMS) en el año 2004 donde se define el término “cáncer de pulmón” como aquel que engloba múltiples neoplasias de tipo epitelial, mesenquimal o linfoproliferativo entre otros (Travis et al., 2004).

Aproximadamente, el 95% de todas las variantes descritas en la práctica clínica corresponden a tumores epiteliales malignos o carcinomas, lo que hace que de rutina clínica, el término carcinoma de pulmón o carcinoma broncogénico se haga sinónimo al de cáncer de pulmón.

Desde el punto de vista histológico, los tumores epiteliales malignos o carcinomas de pulmón se clasifican en dos grandes grupos (tabla 1):

▪ ***Carcinoma Microcítico de Pulmón (CMP) o Carcinoma de Pulmón de Célula Pequeña (SCLC del inglés, Small Cell Lung Cancer)***. Representan el 20% de los casos diagnosticados y su evolución clínica es rápida. Su tratamiento se suele basar en quimioterapia o radioterapia.

▪ ***Carcinoma No Microcítico de Pulmón (CNMP) o Carcinoma de Pulmón No de Célula Pequeña (NSCLC del inglés, Non Small Cell Lung Cancer)***. Representan el 80% restante y su evolución clínica es más lenta, por ello el tratamiento más común es la resección quirúrgica con linfadenectomía, siempre que la estadificación del tumor lo permita.

CLASIFICACIÓN HISTOLÓGICA DE LOS TUMORES EPITELIALES MALIGNOS DE PULMÓN
<p>1. CARCINOMA DE PULMÓN DE CÉLULA PEQUEÑA - Carcinoma de células pequeñas combinado</p>
<p>2. CARCINOMA DE PULMÓN NO DE CÉLULA PEQUEÑA</p> <p>1. Adenocarcinoma (AC) - Adenocarcinoma mixto - Adenocarcinoma acinar - Adenocarcinoma papilar - Carcinoma bronquioloalveolar (no mucinoso, mucinoso o mixto) - Adenocarcinoma sólido con producción de mucina - Adenocarcinoma fetal - Carcinoma coloide - Cistoadenocarcinoma mucinoso - Adenocarcinoma de células “en anillo de sello” - Adenocarcinoma de células claras</p>
<p>2. Carcinoma epidermoide o escamoso (SCC) - Papilar - De células claras - De células pequeñas - Basaloide</p>
<p>3. Carcinoma de células grandes (LCC) - Carcinoma de células grandes neuroendocrino - Carcinoma basaloide - Carcinoma de tipo linfoepitelioma - Carcinoma de células claras - Carcinoma de células grandes con fenotipo rabdoide</p>
<p>4. Carcinoma adenoescamoso</p>
<p>5. Carcinoma sarcomatoide - Carcinoma pleomórfico - Carcinoma de células fusiformes - Carcinoma de células gigantes - Carcinosarcoma - Blastoma pulmonar</p>
<p>6. Tumor carcinoide - Carcinoide típico - Carcinoide atípico</p>
<p>7. Tumores de glándula salivar - Carcinoide mucoepidermoide - Carcinoma adenoide quístico - Carcinoma epi-mioepitelial</p>

Tabla 1. Clasificación histológica de los tumores epiteliales malignos de pulmón
(adaptada de OMS, 2004).

1.2. CARCINOMA NO MICROCÍTICO DE PULMÓN (CNMP)

1.2.1. Subtipos histológicos más frecuentes del CNMP

Dentro del CNMP, los subtipos histológicos más frecuentes son el adenocarcinoma (*AC del inglés, adenocarcinoma*), el carcinoma epidermoide o escamoso (*SCC del inglés, squamous cell carcinoma*) y el carcinoma de células grande (*LCC del inglés, large cell carcinoma*).

- ***El adenocarcinoma (AC)*** representa alrededor del 40% de los CNMP y se define como el tumor epitelial maligno con diferenciación glandular o producción de mucina que muestra patrones de crecimiento acinares, papilares, bronquioloalveolares, sólidos o una mezcla de ellos. Es el subtipo histológico más común, siendo más frecuente en mujeres y, aunque la mayoría de los pacientes en los que se desarrolla son fumadores, es el subtipo que más frecuentemente aparece en individuos que nunca han fumado. Los ACs pueden ser tumores únicos o múltiples y aparecen dentro de un amplio rango de tamaños. Respecto a la localización en el pulmón, suelen ser tumores periféricos de bordes lobulados o estrellados y suelen metastatizar preferentemente vía linfática y hemática. Aproximadamente una quinta parte de los ACs presentan metástasis a distancia en el momento del diagnóstico, siendo el cerebro, hueso, glándulas adrenales e hígado los órganos en los que más comúnmente se detectan estas metástasis. Este subtipo histológico es el que presenta menor tasa de recidiva locoregional después de la resección respecto al resto de subtipos histológicos de CNMP.

- ***El carcinoma epidermoide o escamoso de pulmón (SCC)*** representa entre un 25-30% de los CNMP y se define como el tumor epitelial maligno que muestra queratinización y/o puentes intercelulares y que se origina en el epitelio bronquial. Es un tumor más frecuente en varones y en fumadores. Macroscópicamente son tumores de color blanco o gris, suelen ser de gran tamaño, de localización central en el pulmón y cavitados. Es un tumor localmente agresivo como muestra su alta tasa de recidiva locoregional posterior a la resección, que es más frecuente que en el resto de subtipos. Por el contrario, las metástasis a distancia son mucho menos frecuentes en los SCC que en los AC o que en otros subtipos histológicos.

• **El carcinoma de células grandes (LCC)** representa entre el 10-15% de los CNMP. Puede aparecer en cualquier parte del pulmón y tiende a crecer y propagarse rápidamente, lo que hace muy difícil tratarlo.

• **El adenoescamoso** es el subtipo histológico de menor incidencia representando un 2% de todos los CNMP.

A principios del año 2011 fue publicada una nueva clasificación para el subtipo histológico más común del CNMP, el adenocarcinoma (tabla 2). Esta clasificación ha tenido en cuenta todos los avances científicos obtenidos para la comprensión de este subtipo histológico (Travis et al., 2011) y es la primera que se basa en el conocimiento de una plataforma multidisciplinar integrada por representantes de la International Adjuvant Lung Cancer Trial (IASLC), de la American Thoracic Society (ATS) y de la European Respiratory Society (ERS). Su finalidad ha sido proporcionar una terminología y criterios diagnósticos uniformes, especialmente para el carcinoma bronquioloalveolar (BAC), describir la mejor aproximación para la utilización de las muestras de pequeño tamaño obtenidas de tumores no resecables y definir la estrategia para el manejo de tejidos destinados para estudios moleculares e inmunohistoquímicos.

CLASIFICACIÓN HISTOLÓGICA DE LOS ADENOCARCINOMAS
<p>1. LESIONES PREINVASIVAS</p> <ul style="list-style-type: none"> - Hiperplasia adenomatosa atípica. - Adenocarcinoma in situ (≤ 3cm, antiguo bronquioloalveolar) <ul style="list-style-type: none"> - No mucinoso - Mucinoso - Mixto mucinoso / no mucinoso
<p>2. ADENOCARCINOMA MINIMAMENTE INVASIVO (tumor con predominio lepidico de tamaño ≤ 3 cm que invade ≤ 5 mm)</p> <ul style="list-style-type: none"> - No mucinoso - Mucinoso - Mixto mucinoso / no mucinoso
<p>3. ADENOCARCINOMA INVASIVO</p> <ul style="list-style-type: none"> - De patrón predominantemente lepidico (antiguo adenocarcinoma bronquioloalveolar, que invade >5 mm) - De patrón predominantemente acinar - De patrón predominantemente papilar - De patrón predominantemente micropapilar - De patrón predominantemente sólido con formación de mucina
<p>4. VARIANTES DE ADENOCARCINOMA INVASIVO</p> <ul style="list-style-type: none"> - Adenocarcinoma invasivo mucinoso - Coloide - Fetal (alto y bajo grado) - Entérico

Tabla 2. Clasificación histológica de los adenocarcinomas de pulmón
(adaptada de Travis et al., 2011).

1.2.2. Estadificación del CNMP

En los pacientes con CNMP es necesario realizar un estudio que permita conocer el grado de extensión de la enfermedad para así poder determinar su pronóstico y guiar su tratamiento. El sistema de clasificación más utilizado en la actualidad es el *sistema TNM del American Joint Committee on Cancer (AJCC, 7ª edición)* que describe la extensión anatómica de la enfermedad neoplásica basándose en tres parámetros (tabla 3):

- **T:** Tumor primario – extensión.
- **N:** Ausencia o presencia de metástasis en ganglios linfáticos regionales.
- **M:** Ausencia o presencia de metástasis a distancia.

SISTEMA INTERNACIONAL TNM PARA LA ESTADIFICACIÓN DEL CNMP
T (Tumor Primario - Extensión)
<p>TX: Tumor primario que no puede ser evaluado, o tumor probado por la existencia de células tumorales malignas en esputo o lavados bronquiales pero no visualizado por métodos de imagen o broncoscopia.</p> <p>T0: Sin evidencia de tumor primario.</p> <p>Tis: Carcinoma in situ.</p> <p>T1: Tumor ≤ 3 cm en su mayor diámetro, rodeado por pulmón o pleura visceral, sin evidencia broncoscópica de invasión más proximal del bronquio lobar.</p> <p style="padding-left: 20px;">T1a: Tumor ≤ 2 cm en su mayor diámetro</p> <p style="padding-left: 20px;">T1b: Tumor >2cm pero ≤ 3 cm en su mayor diámetro.</p> <p>T2: Tumor >3cm pero ≤ 7 cm en su mayor diámetro o tumor con cualquiera de las siguientes características: afecta al bronquio principal, distante 2 cm o más de la carina principal; invade la pleura visceral; asociado con atelectasia o neumonitis obstructiva que se extiende hasta la región hiliar pero no afecta al pulmón entero.</p> <p style="padding-left: 20px;">T2a: Tumor >3cm pero ≤ 5 cm en su mayor diámetro.</p> <p style="padding-left: 20px;">T2b: Tumor >5cm pero ≤ 7 cm en su mayor diámetro.</p> <p>T3: Tumor >7cm o de cualquier tamaño que invada directamente cualquiera de las siguientes estructuras: pared torácica (incluyendo los tumores del sulcus superior), diafragma, nervio frénico, pleura mediastínica, pericardio parietal; o un tumor a menos de 2 cm de la carina principal pero sin invadirla; o asociado a atelectasia o neumonitis obstructiva del pulmón entero o existencia de nódulo(s) tumoral(es) separado(s) del tumor primario, en su mismo lóbulo.</p> <p>T4: Tumor de cualquier tamaño que invade cualquiera de las siguientes estructuras: mediastino, corazón, grandes vasos, tráquea, nervio recurrente laríngeo, esófago, cuerpo vertebral, carina; o existencia de nódulo(s) tumoral(es) separado(s) del tumor primario, en un lóbulo diferente del pulmón homolateral.</p>
SISTEMA INTERNACIONAL TNM PARA LA ESTADIFICACIÓN DEL CNMP
N (Ausencia o presencia de metástasis en ganglios linfáticos regionales)
<p>NX: Los ganglios linfáticos regionales no pueden ser evaluados.</p> <p>N0: No existen metástasis ganglionares linfáticas regionales.</p> <p>N1: Metástasis en ganglios linfáticos peribronquiales homolaterales y/o hiliares homolaterales e intrapulmonares, incluyendo la afectación por extensión directa.</p> <p>N2: Metástasis en ganglios linfáticos mediastínicos homolaterales y/o subcarinales.</p> <p>N3: Metástasis ganglionares linfáticas mediastínicas contralaterales, hiliares contralaterales, escalénicas homolaterales o contralaterales, o supraclaviculares.</p>
SISTEMA INTERNACIONAL TNM PARA LA ESTADIFICACIÓN DEL CNMP
M (Ausencia o presencia de metástasis a distancia)
<p>MX: Las metástasis a distancia no pueden ser evaluadas.</p> <p>M0: No existen metástasis a distancia.</p> <p>M1: Existen metástasis a distancia.</p> <p style="padding-left: 20px;">M1a: Existencia de nódulo(s) tumoral(es) separado(s) del tumor primario, en un lóbulo del pulmón contralateral; tumor con nódulos pleurales o derrame pleural (o pericárdico) maligno.</p> <p style="padding-left: 20px;">M1b: Existen metástasis a distancia.</p>

Tabla 3. Sistema TNM para la estadificación del CNMP. Se muestran, en tablas separadas, los tres parámetros (T, N y M) en los que se basa este sistema.

En función de estos tres parámetros, este sistema internacional clasifica los tumores de CNMP en *Estadio Tumoral I* y *II* (enfermedad localizada), *Estadio Tumoral III* (enfermedad localmente avanzada) y *Estadio Tumoral IV* (enfermedad metastásica) (tabla 4).

ESTADIO TUMORAL		T	N	M	ENFERMEDAD
ESTADIO I	Estadio IA	T1 a,b	N0	M0	ENFERMEDAD LOCALIZADA
	Estadio IB	T2 a	N0	M0	
ESTADIO II	Estadio IIA	T1 a,b	N1	M0	
		T2a	N1	M0	
		T2b	N0	M0	
	Estadio IIB	T2b	N1	M0	
		T3	N0	M0	
ESTADIO III	Estadio IIIA	T1, T2	N2	M0	
		T3	N1, N2	M0	
		T4	N0, N1	M0	
	Estadio IIIB	T4	N2	M0	
		Cualquier T	N3	M0	
ESTADIO IV	Estadio IV	Cualquier T	Cualquier N	M1a, b	ENFERMEDAD METASTÁSICA

Tabla 4. Estadios de la enfermedad del CNMP. La tabla muestra la clasificación en estadios del CNMP en función del TNM y su correlación con el tipo de enfermedad.

1.2.3. Factores pronósticos del CNMP

En las enfermedades malignas, los factores pronósticos que indican el manejo del paciente se engloban en dos grandes grupos que incluyen factores pronósticos dependientes del tumor y factores pronósticos dependientes del paciente.

1.2.3.1. Factores pronósticos dependientes del tumor

El *estadio tumoral* es, en la actualidad, el principal factor dependiente del tumor que determina el pronóstico y guía el tratamiento de los pacientes afectados de CNMP (Sánchez de Cos et al., 2008; Chansky K et al., 2009). En términos generales el pronóstico del paciente empeora a medida que aumenta el estadio tumoral. Los tumores pequeños y bien localizados son los que tienen mayor posibilidad de supervivencia a largo plazo mientras que la afectación ganglionar peribronquial y mediastínica la disminuye. Un claro ejemplo es que en los estadios III se observan diferencias significativas de la supervivencia según la afectación ganglionar sea N0-1, N2 o N3. La invasión de los vasos linfáticos y

sanguíneos del parénquima pulmonar comporta también peor pronóstico, incluso en los pacientes con enfermedad localizada. En los estadios IV, la localización y el número de metástasis tienen una gran importancia pronóstica. Los pacientes con metástasis óseas responden peor a la quimioterapia y viven menos y también son indicadores de mal pronóstico la afectación hepática y cutánea y la existencia de más de una localización metastásica.

El segundo factor pronóstico más importante dependiente del tumor que se maneja en la práctica clínica diaria es la *histología*. Se han publicado series quirúrgicas que muestran una peor supervivencia del subtipo histológico de células grandes y el adenocarcinoma respecto al subtipo epidermoide, debido principalmente a una mayor tendencia a la recaída metastásica de estos dos subtipos histológicos. Sin embargo, la mayoría de los estudios no encuentran diferencia pronóstica debida al subtipo histológico del tumor.

1.2.3.2. Factores pronósticos dependientes del paciente

El factor pronóstico más importante relacionado con el enfermo es el *estado general* o “*performance status*”. Existen varias escalas para cuantificar la capacidad funcional de los pacientes siendo las más utilizadas las del Eastern Cooperative Oncology Group (ECOG) y la de Karnofsky (tabla 5). La información obtenida de ellas se correlaciona con la supervivencia y predice la respuesta al tratamiento.

ECOG	KARNOFSKY	DEFINICIÓN
0	100	Asintomático
1	80-90	Con síntomas. Ambulatorio
2	60-70	Con síntomas. En cama menos del 50% del día.
3	40-50	Con síntomas. En cama más del 50% del día.
4	20-30	Postrado en cama

Tabla 5. Estado clínico del paciente. Correlación entre los distintos niveles de las escalas de ECOG y Karnofsky.

Otros factores pronósticos dependientes del paciente incluyen *datos analíticos*, *edad* y *sexo*. Marcadores tumorales séricos como al antígeno carcinoembrionario, el antígeno de células escamosas o la enolasa neuronal específica parecen influir en el pronóstico de estos pacientes al igual que ocurre en otros tipos de tumores (Díez et al., 1996).

En general factores de buen pronóstico incluyen el diagnóstico de la enfermedad en estadios iniciales, un buen “performance status” (ECOG 0, 1 o 2) y la ausencia de pérdida significativa de peso (no superior al 5%) (Finkelstein et al., 1986). Algunas revisiones, como la basada en el California Cancer Registry, encuentran que la edad avanzada al diagnóstico y el género masculino influyen negativamente en la supervivencia por recidiva en pacientes resecaos (Ou et al., 2007).

1.2.4. Tratamiento y tasas de recidiva en el CNMP

Hasta la década de los años 60, la resección quirúrgica y la radioterapia fueron las opciones de tratamiento más extendidas en el CNMP. Posteriormente, la quimioterapia basada en agentes citotóxicos, tanto adyuvante como neoadyuvante, se ha sumado al arsenal terapéutico y en la actualidad, se ha consolidado como la principal opción de tratamiento para determinados estadios tanto de tumores primarios como metastáticos. El cisplatino y su análogo el carboplatino son los agentes citotóxicos empleados con mayor frecuencia combinándolos con gemcitabina, vinorelbina, paclitaxel o docetaxel para mejorar su eficacia (Arriagada et al., 2004; Winton et al., 2005; Douillard et al., 2006).

En el CNMP, al igual que ocurre en otros tipos de tumores, la decisión de la pauta de tratamiento de la enfermedad depende de su estadificación, evaluado mediante el sistema de clasificación TNM. Las guías internacionales del National Comprehensive Cancer Network (NCCN; www.nccn.org) indican el tratamiento para cada uno de los estadios de esta enfermedad, los cuales se encuentran en continua evolución a medida que se va teniendo mayor información de este tipo tumoral.

En general, *la resección quirúrgica con linfadenectomía y posterior seguimiento en consulta* es el tratamiento de elección en pacientes operables en estadio IA, IB (T2a, N0) y IIA (T2b, N0) (NCCN Guidelines versión 2. 2014). Hasta hace algunos años también se incluía en este mismo grupo de tratamiento a los pacientes en estadio IIA (T1ab-T2a, N1) y IIB (T3, N0; T2b, N1) aunque en la actualidad, éstos son susceptible de varias opciones de tratamiento por lo que está recomendada una evaluación multidisciplinar de cada caso en la que, en muchas ocasiones, se acaba priorizando la indicación de quimioterapia. Sin embargo, aunque el sistema TNM es la herramienta más poderosa en la toma de decisiones en el paciente con CNMP, no es capaz de explicar toda la heterogeneidad biológica de los estadios iniciales como demuestran las elevadas tasas de recidiva a 5 años para los estadios IA, IB, IIA y IIB que se encuentra entre el 16-19%, 23-27%, 37-37.5% y

39-40% respectivamente (Edge et al., 2010; Saynak et al 2010). Tumores clasificados mediante el TNM como estadios iniciales, muestran una elevada tasa de recidiva posterior a la resección quirúrgica y en consecuencia, una elevada tasa de mortalidad (Fibbla et al., 2013). Para evitar las limitaciones de los indicadores clásicos en predecir el pronóstico de estos estadios, los esquemas de tratamiento se encuentran en continua discusión a la vez que se desarrollan nuevos ensayos clínicos y publican multitud de artículos científicos que tratan de conocer el beneficio que la quimioterapia adyuvante pudiera aportar (Booth et al., 2006; Alam et al., 2006; Pisters et al., 2005). En 1995, el Grupo Colaborativo para el estudio del CNMP (Non-Small Cell Lung Cancer collaborative Group, 1995) publicó el meta-análisis que incluía datos de 14 estudios en los que se comparaba el tratamiento basado sólo en resección quirúrgica frente a la resección quirúrgica seguida por quimioterapia adyuvante. Los resultados de 8 de los 14 estudios que utilizaban combinaciones basadas en cisplatino, mostraron un beneficio absoluto para la supervivencia del 5% a los 5 años aunque esta diferencia no fue estadísticamente significativa. Estos resultados impulsaron la necesidad de llevar a cabo nuevos ensayos clínicos fase III para evaluar la eficacia y seguridad de la quimioterapia basada en platino en el CNMP reseccionado frente a la resección quirúrgica. Los principales estudios realizados fueron International Adjuvant Lung Cancer Trial-IALT (Arraigada et al, 2004), Adjuvant Lung Project Italy-ALPI (Scagliotti et al., 2003), the Big Lung Trial-BLT (Waller et al., 2004), the National Cancer Institute of Canada Clinical Trials Group-NCICBR10 (Winton et al, 2005), the Adjuvant Navelbine International Trial Association-ANITA (Douillard et al 2006) y el Cancer and Leukemia Group B study-CALGB9633 (Strauss et al, 2008). Los resultados mostrados fueron contradictorios respecto al beneficio del tratamiento adyuvante, con lo que para tratar de resolver estas discordancias, The Lung Adjuvant Cisplatin Evaluation (LACE) publicó los datos de 4.854 pacientes incluidos en cinco de estos ensayos aleatorizados (ALPI, ANITA, BLT, IALT y NCICBR10) (Pignon et al., 2008). Los resultados del meta-análisis mostraron que la quimioterapia adyuvante se asocia a una mejora en la supervivencia global con una reducción relativa del riesgo de muerte del 11%, demostrando a su vez que el grado de beneficio varía dependiendo del estadio tumoral. En los estadios II el beneficio para la supervivencia global fue del 5,3% a los 5 años mientras que un pequeño beneficio sin significado estadístico se observó en el estadio IB y un efecto perjudicial en el estadio IA. Como resumen, respecto a los estadios II, la

quimioterapia adyuvante ha demostrado beneficio para la supervivencia global de los pacientes, sin embargo, sigue existiendo un elevado porcentaje de ellos que, a pesar de no recaer tras la resección quirúrgica, reciben este tratamiento y por tanto son pacientes tratados en exceso que desarrollan problemas de salud derivados de la propia quimioterapia (Pignon et al., 2006). Respecto al subgrupo de pacientes estadio IB, la mayoría de los estudios o no han hallado beneficio estadísticamente significativos de la quimioterapia adyuvante (Winton et al, 2005; Douillard et al., 2006) o el beneficio ha sido mínimo (Pignon et al., 2008). Aun así, algunos de ellos recomiendan su uso en pacientes que cumplan determinados factores de riesgo como pobre grado de diferenciación, invasión vascular, resección en cuña y márgenes mínimos (Pignon et al., 2006; Strauss et al., 2008). Por último, respecto a los estadios IA, aunque hay muy pocos datos disponibles para este subgrupo de pacientes debido a que la mayoría de los ensayos clínicos los han excluido, se ha concluido que la quimioterapia adyuvante postoperatoria no ofrece beneficio, con lo que no está indicada (Pignon et al., 2008). De manera general el uso de la quimioterapia en pacientes con enfermedad en estadio I es menos claro por lo que, estos estadios del CNMP no reciben quimioterapia a pesar de presentar elevadas tasas de recidiva a 5 años.

Los estadios IIIA (N2T3) y IIIB son no quirúrgicos, por lo que el *tratamiento multimodal basado en radioterapia y/o quimioterapia* es el indicado en estos pacientes con enfermedad localmente avanzada. Los estadios IIIA (excepto N2T3) pueden ser tratados quirúrgicamente cuando, tras quimioterapia neoadyuvante, se cumplen los siguiente criterios: conseguir una infraestadificación del tumor tras la quimioterapia, que el enfermo no precise neumonectomía, que no haya alteración de la difusión y que el tumor sea resecable. En los estadios IV, estadios metastásicos, el *tratamiento suele ser paliativo*, bien con quimioterapia o con soporte sintomático (Smith et al., 2012).

Las diferentes modalidades descritas anteriormente se encuadran en el esquema clásico de tratamiento del CNMP, según el cual, todos los pacientes reciben el mismo procedimiento basándose fundamentalmente en dos parámetros: el diagnóstico anatomopatológico y el estadio tumoral. En la actualidad, se ha desarrollado una nueva estrategia de tratamiento del CNMP que es la *terapia dirigida* que ha surgido como consecuencia de los resultados obtenidos en la investigación molecular del cáncer de pulmón. Los trabajos científicos publicados han demostrado la heterogeneidad de los tumores y la variabilidad de los mismos en cuanto a su evolución, aun cuando presentan

características anatomopatológicas y clínicas comunes (Farhat et al., 2013). La terapia dirigida es un tipo de tratamiento para el que se utilizan medicamentos u otras sustancias que identifican y atacan a las células cancerosas de manera específica sin dañar las células normales. En la actualidad existen tratamientos de terapia dirigida en el CNMP basados en la utilización de anticuerpos monoclonales e inhibidores de la tirosina-kinasa que están mostrando resultados muy prometedores en la medicina oncológica personalizada.

1.3. BIOLOGÍA MOLECULAR DEL CÁNCER DE PULMÓN

La elevada frecuencia y mortalidad del cáncer de pulmón ha convertido el estudio de esta enfermedad, en el objetivo principal de multitud de proyectos de investigación que tratan de comprender los mecanismos moleculares que contribuyen a su aparición y evolución. El proceso de carcinogénesis se inicia y progresa debido a la acumulación de alteraciones en genes con función en procesos celulares tan importantes como son la división celular, muerte celular programada o apoptosis, detección y reparación del daño del DNA, regulación de expresión génica, adhesión celular y transducción de señales. Respecto a los cambios genéticos que muestran las células tumorales a nivel de su DNA podemos incluir amplificaciones o deleciones, mutaciones puntuales, cambios epigenéticos, traslocaciones, pérdida de heterocigosidad (Loss of heterocigosity [LOH]), alteraciones en la regulación transcripcional y en los niveles de expresión génica, etc.

La alteración de estos procesos celulares vitales, confiere a la célula tumoral un conjunto de características propias, descritas en el año 2.000 por Hanahan y Weinberg (Hanahan et al., 2000) que incluyen las descritas a continuación (figura 3):

- *Autosuficiencia respecto señales de crecimiento.*
- *Evasión al control de factores supresores de crecimiento.*
- *Activación de la capacidad de invasión y metástasis.*
- *Potencial replicativo ilimitado.*
- *Capacidad sostenida de inducción de la angiogénesis.*
- *Resistencia a la muerte celular programada o apoptosis.*

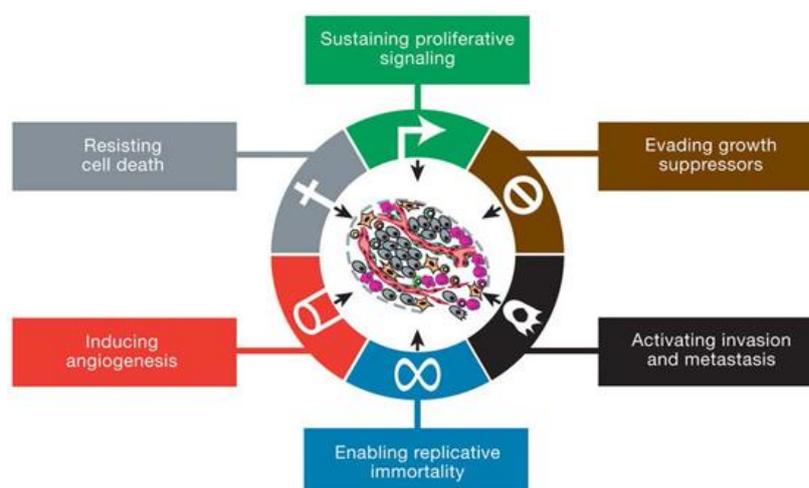


Figura 3. Características de las células tumorales. La figura muestra las seis características de las células tumorales propuestas por Hanahan y Weinberg en el año 2000 (*Figura tomada de Hanahan et al., 2000*).

Los avances en investigación de la última década han añadido, a las seis características anteriores, cuatro nuevas características emergentes que incluyen *la reprogramación del metabolismo energético, la evasión tumoral a la destrucción inmune, la inestabilidad del DNA y la inflamación* (Hanahan et al., 2011).

Al igual que en otros tumores, el proceso de carcinogénesis en el pulmón está relacionado con la activación de proto-oncogenes, que resulta en un aumento de señales de mitosis que promueven crecimiento celular, así como con la inactivación de genes supresores de tumores. También están alterados genes implicados en invasión y metástasis, angiogénesis, resistencia a la apoptosis y vías de reparación de DNA.

1.3.1. Alteraciones en proto-oncogenes

La activación de proto-oncogenes se produce generalmente por amplificación génica, sobreexpresión, mutación puntual o reordenamientos de DNA. En la figura 4 se muestran los más comúnmente alterados, representando su frecuencia en función del subtipo histológico tumoral estudiado. En los adenocarcinomas, las alteraciones moleculares que promueven crecimiento tumoral y conforman dianas terapéuticas están muy bien definidas utilizándose en la rutina oncológica diaria. En el caso de los carcinomas epidermoides, durante los últimos años también se están descubriendo varias alteraciones moleculares que podrían convertirse en potenciales dianas de tratamiento.

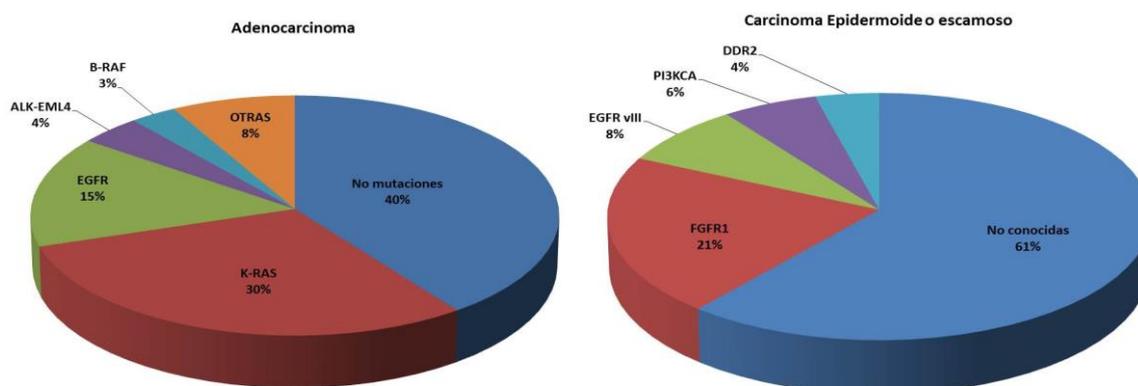


Figura 4. Distribución de mutaciones en proto-oncogenes en el CNMP. Los gráficos representan el porcentaje de mutaciones en función del subtipo histológico estudiado **A)** Adenocarcinoma. En otras se incluyen mutaciones en *N-RAS*, *MET*, *PIK3CA*, *HER2* y *ROS1*. **B)** Carcinoma Epidermoide o escamoso.

1.3.1.1. RAS

La familia *RAS* está constituida por tres miembros, *K-RAS*, *H-RAS* y *N-RAS*, cuya función incluye la transducción de señales hacia el núcleo celular. Pueden estar activados mediante mutaciones en el cáncer de pulmón, especialmente *K-RAS*, lo que se traduce en un aumento de la proliferación celular. Las mutaciones en *K-RAS* aparecen entre el 25-40% de los casos (Ding et al., 2008), afectan especialmente a los adenocarcinomas y se descubrieron mucho antes que las mutaciones de *EGFR*. Suelen afectar a varones, fumadores y no asiáticos (Slebos et al., 1992) y se asocian a una peor supervivencia, tanto en enfermedad localizada como avanzada y a quimio-resistencia al cisplatino. Las mutaciones de *EGFR* y *K-RAS* son mutuamente excluyentes, lo que sugiere que la mutación de uno u otro gen puede ser suficiente para estimular la carcinogénesis. La presencia de mutaciones en los otros dos miembros de esta familia, *H-RAS* y *N-RAS*, son muy raras (Rodenhuis et al., 1992).

1.3.1.2. EGFR

La familia *ERBB*, que incluye 4 receptores con actividad tirosin-kinasa (TK) nombrados como *EGFR*, *ErbB-2 (HER2)*, *ErbB-3* y *ErbB-4*, constituye una vía fundamental en el cáncer de pulmón. El miembro más importante de esta familia es el *EGFR*, el cual regula la proliferación epitelial. Se ha observado sobre-expresión de *EGFR* entre el 50-90% de los CPNM lo que ha hecho que se lleven a cabo multitud de proyectos de investigación para el desarrollo de inhibidores específicos frente a esta diana. Además, en los últimos años, se ha visto renovado el interés por esta vía ya que se han detectado mutaciones somáticas de *EGFR* en pacientes con CPNM que se traducen en cambios

conformacionales del receptor que llevan a un incremento de la sensibilidad a inhibidores TK (TKIs) como son el erlotinib y gefitinib. Se han descrito entre el 10-15% de los pacientes caucásicos y entre el 30-40% de los pacientes del Este asiático (Lynch et al., 2004; Paez et al., 2004; Pao et al., 2004) y son más frecuentes en el subtipo histológico de adenocarcinoma, mujeres, no fumadores y asiáticos del Este. La resistencia a la terapia con TKIs se ha asociado a inserciones o a la presencia de la mutación secundaria T790M en el exón 20 de *EGFR* (Gazdar AF., 2009), a mutaciones en *K-RAS* (Pao et al., 2005) o amplificaciones de *MET* (Engelman et al., 2007). Mientras que las mutaciones de *EGFR* son bastante raras en el carcinoma epidermoide, mutaciones de la variante III que implican a su dominio extracelular (8% de los tumores), ganancias del número de copias o sobre-expresión proteica están más comúnmente asociados a este subtipo histológico (Heist et al., 2012).

1.3.1.3. ALK

El receptor *ALK* se encuentra activado, en un elevado porcentaje de los casos de CNMP, debido a su fusión con *EML4* mediante traslocaciones cromosómicas (Soda et al., 2007). Estos pacientes responden al tratamiento con crizotinib con una reducción muy significativa del tamaño tumoral y con un aumento de la esperanza y calidad de vida. Los reordenamientos de *ALK* se han identificado aproximadamente en el 4% de los CNMP, más comúnmente en el subtipo adenocarcinoma, en pacientes jóvenes y que nunca han fumado (Koivunen et al., 2008).

1.3.1.4. B-RAF

Mientras que las mutaciones activadoras en *B-RAF* son muy comunes en el melanoma, en el CNMP únicamente ocurren en el 3% de los casos (Davies et al., 2002) asociadas al subtipo histológico de adenocarcinoma. Los genes *B-RAF* y *K-RAS* forman parte de la vía de señalización mediada por *EGFR* por lo que no es sorprendente que las mutaciones en estos genes sean mutuamente excluyentes. En la actualidad, las mutaciones en *B-RAF* representan una importante diana terapéutica en el melanoma pero hasta el momento, sólo hay datos limitados sobre la respuesta clínica a estos tratamientos en el paciente con CNMP (Falchook et al., 2012).

1.3.1.5. MET

En el CNMP el gen *MET* se encuentra amplificado entre el 1-7% de los pacientes más frecuentemente en el carcinoma epidermoide y de manera mutuamente excluyente con mutaciones en *K-RAS* (Cappuzzo et al., 2009). Esta alteración se ha constituido en la actualidad como nueva diana para el desarrollo de agentes terapéuticos dirigidos. Tal y como ya se ha comentado anteriormente, el aumento en el número de copias de *MET*, está relacionada con la adquisición de resistencia secundaria a inhibidores de *EGFR* en aproximadamente el 20% de los pacientes (Engelman et al., 2007).

1.3.1.6. PI3KCA

La vía de señalización *PI3K/AKT/mTOR* juega un papel muy importante en la transducción de señales que regulan la proliferación, la supervivencia, la diferenciación y la motilidad celular. Mutaciones activadoras de *PI3KCA*, identificadas entre el 1-3% de los CNMP, desencadenan una activación constitutiva de dicha vía independiente de ligando (Yamamoto et al., 2008) y suelen ocurrir asociadas a mutaciones en *EGFR* o *K-RAS* lo que sugiere que no representa una verdadera mutación “driver”. Tanto las mutaciones de *PI3KCA*, que representan alrededor de un 6% de los casos, como las amplificaciones en este gen están especialmente asociadas al carcinoma epidermoide (Ji et al., 2011).

1.3.1.7. HER2

La activación del gen *HER2* ocurre en una pequeña proporción del cáncer de pulmón mediante sobre-expresión, en el 20% de los casos, amplificación génica, en el 2% de los casos y mutaciones activadoras, entre el 1.6-4% de los CNMP (Tomizawa et al., 2011). Suele ocurrir más frecuentemente en los adenocarcinomas que no presentan mutaciones en *EGFR* o *K-RAS*, asociadas al sexo femenino y en pacientes no fumadores.

1.3.1.8. ROS1

Los reordenamientos en *ROS1* aparecen entre el 1.2-2.6% de los adenocarcinomas estudiados (Bergethon et al., 2012; Takeuchi et al., 2012), parecen ser más comunes en pacientes jóvenes que nunca han fumado y de etnia Asiática y son sensibles a los TKI entre los que se incluye crizotinib (inhibidor *ALK/MET*).

1.3.1.9. *FGFR1*

Las amplificaciones génicas más frecuentemente detectadas en el carcinoma epidermoide incluyen a *SOX*, *PDGFRA* y *FGFR1* (Tran et al., 2013). Alrededor de un 21% de ellos muestran amplificaciones en *FGFR1* las cuales se ha demostrado que son sensibles a inhibición por pequeñas moléculas (Dutt et al., 2011).

1.3.1.10. *DDR2*

Recientemente han sido publicados los resultados de mutaciones en el gen *DDR2* que representan alrededor de un 4% de los carcinomas epidermoides de pulmón estudiados y las cuales, han demostrado in vitro su utilidad como diana terapéutica por su sensibilidad a dasatinib (Hammerman et al., 2011).

1.3.2. Inactivación de genes supresores de tumores

Los genes supresores de tumores son aquellos que inhiben el proceso de carcinogénesis y en los que, la pérdida de su función, se produce por la inactivación de sus dos alelos, uno de ellos por LOH y el otro mediante mutaciones puntuales, epigenéticas o transcripcionales. Genes supresores de tumores clave en el CNMP son *TP53* y la vía *p16INK4a-ciclinaD1-CDK4-RB*.

1.3.2.1. *TP53*

TP53 es un gen que codifica una proteína imprescindible para mantener la integridad genómica al prevenir la acumulación de daño genético en las células hijas obtenidas en la mitosis. Este gen se encuentra mutado en más del 50% de los CNMP normalmente asociado a mutaciones de *EGFR* y *K-RAS*, es independiente del tipo histológico tumoral, está relacionada con el consumo de tabaco y alcohol y confiere mal pronóstico (Steels et al., 2001; Mitsudomi et al., 2000).

1.3.2.2. Vía *p16INK4a-ciclinaD1-CDK4-RB*

El gen *CDKN2A* codifica la proteína p16 que pertenece a la familia de las quinasas dependientes de ciclinas (CDKs) y que junto con el gen *RB* forman parte de la vía *p16INK4a-ciclinaD1-CDK4-RB* cuya función es el control de la transición G1-S del ciclo celular. La inactivación de *CDKN2A* se produce en un 40-60% de los CNMP mientras que la pérdida de función del gen *RB*, aun siendo descrita en un 10-15% de los CNMP (Wikenheiser-Brokamp et al.; 2000) parece una alteración más frecuente de los CMP

(90%). El gen *RB* también se encuentra inactivado por metilación de su promotor, por lo que tiene un gran interés como posible marcador biológico (Palmisano et al., 2000).

1.3.3. Activación de la invasión y metástasis

Las integrinas son una superfamilia de glicoproteínas que participan mayoritariamente en la unión de las células con la matriz extracelular. Varios estudios describen, que la pérdida de adhesión celular puede tener un importante papel en la migración de las células tumorales y por tanto en la aparición de metástasis (Boelens et al., 2007; Zhu et al. 2007).

1.3.4. Inducción de la angiogénesis

Los factores -A, -C y -D de crecimiento del endotelio vascular (*VEGF-A*, *VEGF-C* y *VEGF-D*) y los receptores -1 y -2 de factores de crecimiento del endotelio vascular (*VEGFR-1* y *VEGFR-2*) tienen un papel crucial en la angiogénesis. En el CNMP, la mayoría de los estudios muestran que la sobre-expresión de estos marcadores está asociada con un peor pronóstico (Donnem et al., 2007; Herbst et al., 2005).

1.3.5. Resistencia a la muerte celular programada o apoptosis

La resistencia a la apoptosis por las células tumorales conlleva una acumulación de las mismas y por tanto crecimiento tumoral. *BCL-2* y *BAX* son genes que codifican para moléculas antiapoptóticas y apoptóticas respectivamente que se encuentran en equilibrio en el citoplasma celular. Varios estudios han demostrado sobre-expresión de *BCL-2* en el CNMP (Kren et al., 2004) aunque todavía en la actualidad, el valor pronóstico de este marcador para esta patología es contradictorio.

1.3.6. Alteraciones en genes implicados en las vías de reparación del ADN

Se ha descrito un gran número de genes que codifican para proteínas implicadas en la regulación del ciclo celular y sobre todo en la reparación del ADN aberrante y en impedir su replicación. El gen *ERCC1* (excision repair cross-complementation group 1) es una nucleasa que actúa durante el proceso de reparación de los nucleótidos (NER). Muchas células tumorales resisten los efectos de la quimioterapia mediante un incremento de la capacidad de reparación del ADN mediante la activación de la vía NER debido a unos altos niveles de *ERCC1*. Varios artículos científicos señalan al *ERCC1* como marcador para predecir resistencia a cisplatino, carboplatino y oxaliplatino en el CNMP

(Metzger et al., 1998; Shirota et al., 2001; Rosell et al., 2004). El gen *RRM1* codifica para una subunidad reguladora de la ribonucleótido reductasa (enzima que cataliza la reducción de ribonucleósido difosfato a los correspondientes deoxirribonucleótidos). Se ha demostrado que líneas celulares que expresan diferentes niveles de *RRM1*, presentan distinta sensibilidad a cisplatino de tal modo que a mayor expresión de *RRM1*, más resistencia a cisplatino en comparación con las líneas de baja expresión (Gautman et al., 2003). Varios grupos de investigación han señalado también al *RRM1* como un determinante celular de eficacia de la gemcitabina (Bepler et al., 2006). *BRCA1* tiene también un papel esencial en la vía de reparación del ADN. En líneas celulares de cáncer de ovario, la sobreexpresión de *BRCA1* se correlaciona con un incremento a la resistencia a cisplatino (Taron et al., 2004). Por otro lado, en líneas celulares de cáncer de mama, una baja expresión de mRNA medida por retrotranscripción seguida de PCR cuantitativa (RT-qPCR) se relaciona con una alta sensibilidad a cisplatino y etopóxido pero con resistencia a agentes antimicrotúbulo como el paclitaxel y vincristina (Quinn et al., 2003).

1.4. SISTEMA INMUNE Y CÁNCER

Como ya se ha comentado anteriormente una de las características más recientemente incorporada a la lista de características alteradas en las células tumorales es la evasión tumoral a la destrucción inmune mediada por linfocitos T y B, macrófagos y células natural killer (NK) (figura 5), (Hanahan et al., 2011).

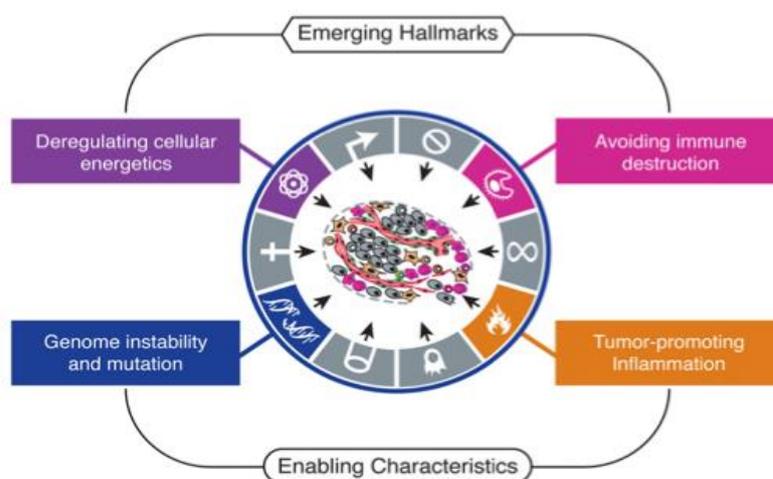


Figura 5. Características emergentes de las células tumorales. La figura muestra las dos características emergentes de las células tumorales propuestas por Hanahan y Weinberg en el año 2011. (Figura tomada de Hanahan et al., 2011).

Ya en el año 1957 fue postulada la teoría de la “inmunovigilancia” por Burnet y Thomas la cual propone que las células y tejidos del organismo están constantemente monitorizados por el sistema inmune el cual es responsable de reconocer y eliminar la gran mayoría de las células tumorales incipientes evitando así el desarrollo de tumores (Burnet, 1957; Thomas, 1959). La “inmunovigilancia” es un fenómeno general que parece afectar no sólo a diferentes órganos en los que se puede originar el cáncer (mama, colon, pulmón, cabeza y cuello, riñón, vejiga, ovarios, próstata...), sino también a los diferentes tipos de células cancerosas (adenocarcinoma, carcinoma de células escamosas, carcinoma de células grandes, melanoma...) (Butterfield et al., 2008; Tahara et al., 2009; Butterfield et al., 2011). En los últimos años se ha postulado que la “inmunovigilancia” funciona como una etapa de un proceso más general conocido como “inmunoección del cáncer” mediante el cual una persona está protegida por su sistema inmune no sólo contra el crecimiento del cáncer si no también frente al desarrollo de la inmunogenicidad del mismo.

La “inmunoección del cáncer” consta de tres etapas principales que son la eliminación, el equilibrio y el escape (figura 6) (Dunn et al., 2002).

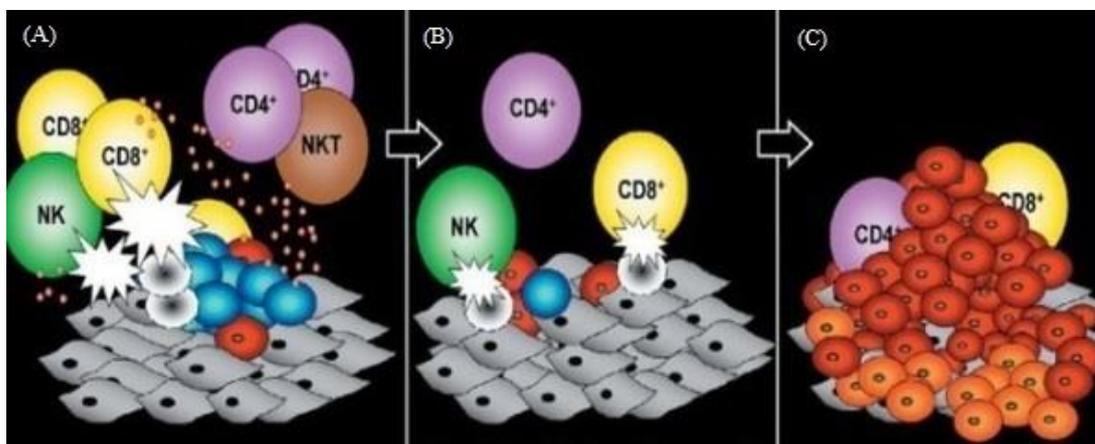


Figura 6. Fases del proceso de “inmunoección del cáncer”. La figura muestra las fases de **A)** Eliminación, **B)** Equilibrio y **C)** Escape del proceso de inmunoección del cáncer. **En azul se muestran las células tumorales en desarrollo, en rojo las variantes tumorales y en gris las células no transformadas.** (Figura tomada de Dunn et al., 2002).

En este contexto más amplio, la “inmunovigilancia” ocurriría durante la etapa de eliminación en la cual el sistema inmunológico puede reconocer y destruir células tumorales recientemente originadas. En este punto el cáncer puede ser eliminado pero también es posible que el cáncer pueda ser editado, es decir, el sistema inmune elimina todas las células con mutaciones fácilmente reconocibles mientras que las variantes

tumorales restantes entran en un periodo de letargo en el que se mantienen bajo control durante la etapa conocida como equilibrio. En la última etapa de escape es posible que el equilibrio se rompa y el tumor, a través de mecanismos de resistencia, pueda evadir la respuesta inmunitaria (Finn et al., 2008; Bindea et al., 2010; Mlecnik et al., 2011).

Los resultados obtenidos en amplias series demuestran que el contexto inmune de los tumores es un factor pronóstico esencial para el intervalo libre de enfermedad (ILE) y para la supervivencia global (SG) de los pacientes. Como consecuencia, multitud de artículos destacan la importancia de incluir biomarcadores inmunológicos que nos ayuden a predecir el pronóstico y la respuesta a terapia de los pacientes con cáncer. Utilizar esta clasificación inmune o “inmuno-score” podría suponer un valor pronóstico que incluso podría ser superior al obtenido con el sistema de clasificación convencional basado en el TNM (Galon et al., 2012. Editorial). En la actualidad se postula el introducir este parámetro como parte de la rutina diagnóstica, pronóstica y terapéutica de los tumores (Pages et al., 2010; Ascierto et al., 2011; Galon et al., 2012. Review).

1.5. MICROARRAYS

La técnica de microarrays surge como un método para estudiar la expresión de miles de genes de manera simultánea. El desarrollo de esta nueva metodología se ha visto impulsado de manera significativa gracias a dos hechos clave de la historia de la biología molecular como ha sido conocer la secuencia del genoma humano (Venter et al., 2001; Lander et al., 2001) y la rápida evolución de las técnicas de análisis molecular. En la actualidad, no sólo es posible el estudio global del RNA mensajero (mRNA) que codifica para proteínas, “*Gene Expression Profiling Analysis*”, sino también el análisis de mutaciones y polimorfismos del genoma, “*Genome Wide Association Studies*”, análisis de alteraciones en el número de copias de regiones de DNA, “*Comparative Genome Hybridization*” e incluso el análisis de parte del RNA no codificante, “*miRNAs Expression Analysis*” (figura 7). A las aplicaciones descritas anteriormente, se han sumado en los últimos años otras nuevas como son el estudio masivo de los procesos de metilación de las regiones promotoras de los genes o el estudio de los factores transcripcionales mediante técnicas de inmunoprecipitación de cromatina.

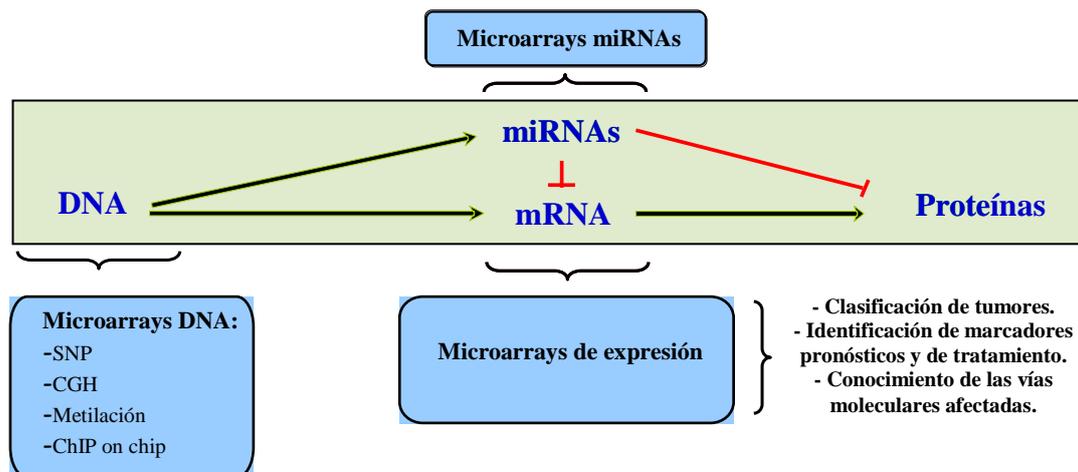


Figura 7. Aplicaciones de los microarrays. El esquema detalla los 3 grandes tipos de microarrays además de las aplicaciones de los microarrays de expresión.

1.5.1. Microarrays de expresión

Los microarrays de expresión (figura 8) se definen como un conjunto de miles de sondas de DNA copia (cDNA) u oligonucleótidos de DNA dispuestas de manera ordenada sobre una superficie sólida que permite analizar simultáneamente la expresión del genoma completo de un organismo. Cada una de las sondas representa un gen determinado al poseer una secuencia complementaria al mRNA transcrito por dicho gen. La utilización de esta tecnología posibilita la medición de los niveles de expresión de todos los genes que conforman el genoma al mismo tiempo y en un único experimento.

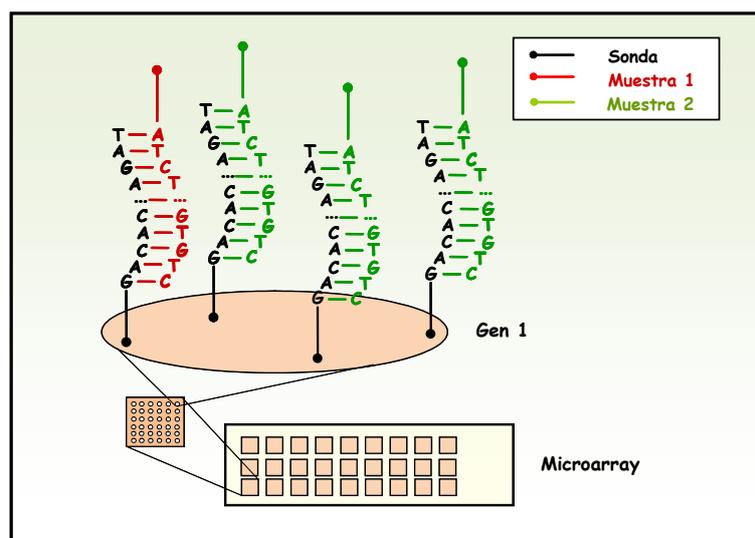


Figura 8. Esquema de un microarray de Agilent Technologies. En el dibujo se observa la disposición ordenada de los spots (compuestos por varias copias de una sonda específica para un único gen) sobre un soporte sólido. El ejemplo representa una hibridación doble (muestra 1 y muestra 2) en la que el mRNA para el gen 1 de cada una de las muestras competirá por la unión a las sondas. La cantidad de mRNA que se une a las sondas es proporcional a la cantidad de mRNA transcrito para cada gen.

En la actualidad se utilizan los microarrays de oligonucleótidos elaborados comercialmente por distintas empresas sustituyendo los que se elaboraban en los laboratorios utilizando como sondas cDNAs de diferentes tamaños. En este tipo de microarrays, los oligonucleótidos que se utilizan son todos 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.

Respecto a la longitud de los oligonucleótidos de los microarrays podemos diferenciar dos subtipos:

- *Microarrays de Oligonucleótidos cortos*: de unas 25 pb aproximadamente, como los utilizados por la plataforma de Affymetrix. El principal problema de este tipo de oligonucleótidos es que debido a su pequeño tamaño tienen menor fuerza de unión al mRNA. Para solucionar este problema, la compañía Affymetrix utiliza la técnica de “mismatch-perfect match” y varios oligonucleótidos diferentes por cada gen.
- *Microarrays de 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 Biotools.

Respecto a la síntesis de los oligonucleótidos para la elaboración de microarrays existen dos posibilidades:

- *Síntesis in situ*: el oligonucleótido se sintetiza sobre el propio microarray.
- *Síntesis previa*: se sintetiza primeramente el oligonucleótido de manera independiente y posteriormente se anclará a la matriz.

1.5.1.1. Marcaje e hibridación de las muestras

En el protocolo de microarrays, el primer paso consiste en “etiquetar” todos los mRNAs 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 (Quackenbush et al., 2006). El método más utilizado es el de Eberwine que consta en primer lugar de una retrotranscripción con un oligo-dT, que tiene la

secuencia de la T7-RNA-polimerasa y posteriormente una amplificación con dicha enzima en la que se incorpora el compuesto fluorescente sintetizándose RNA antisense marcado. Gracias a la introducción de este pequeño paso de amplificación, es posible partir de una menor cantidad de mRNA inicial (0.1-1 μ g) lo que supone una gran ventaja cuando se trabaja con muestras humanas ante la usual escasez de material tumoral congelado disponible.

Una vez realizado el marcaje se lleva a cabo la hibridación al microarray. En función de cómo se quiera estudiar la expresión diferencial entre las muestras, se utilizará o bien la técnica de hibridación simple o bien la técnica de hibridación doble o competitiva:

- *Técnica de Hibridación simple:* cada una de las muestras de estudio se hibridan individualmente a un microarray, marcando todas ellas con el mismo fluorocromo, generalmente Cianina-3 (Cy-3). Esto hace que el error de incorporación o la velocidad de degradación del fluorocromo afecten a todas las muestras por igual. Este diseño permite, respecto a la hibridación doble, una mayor flexibilidad a la hora de realizar las comparaciones. En este caso las fluorescencias pertenecientes a cada gen se transforman en logaritmos en base 10.

- *Técnica de Hibridación doble o competitiva:* se marcan las dos muestras a comparar, muestra de estudio y muestra de referencia con diferentes fluorocromos, generalmente Cianina-5 (Cy-5) y Cianina-3 (Cy-3) respectivamente y se hibridan juntas a un único microarray. La ventaja de este tipo de hibridación es que al usar el mismo microarray, el ruido de fondo del cristal y de los “spots” de DNA son los mismos para las dos muestras. Además, al utilizar la misma muestra de referencia en todos los microarrays, el proceso de normalización y la comparación de las muestras a estudio entre sí resulta más sencillo. Las principales desventajas son que exige corregir los datos de expresión debido a las diferencias de incorporación de cada fluorocromo y que es necesario contar con muestra de referencia suficiente para cualquier tipo de comparación que se quiera llevar a cabo. En este caso los datos de expresión se representan en forma de ratios de los valores de fluorescencia cada gen en el tumor con respecto a su referencia y transformados en logaritmos en base 2.

1.5.1.2. Análisis de datos

La unión de cada mRNA a su sonda dispuesta en un “spot”, produce una intensidad de fluorescencia que es directamente proporcional al nivel de expresión de ese gen. 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. La principal razón por la que este análisis presenta una gran complejidad es debido al gran número de variables (genes) respecto al bajo tamaño muestral utilizado en este tipo de estudios (Ej.: del orden de varios miles de genes cuando se estudia el genoma humano completo y alrededor de 15-100 muestras). Los datos obtenidos se organizan en matrices en donde cada fila representa un gen y cada columna una muestra, pudiendo ser definidos, tanto muestras como genes, en forma de vectores (figura 9). 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.

$$\begin{matrix}
 & m_1 & m_2 & m_3 & m_4 & \dots \\
 g_1 & \left(\begin{array}{ccccc} 12 & -4 & 6 & 0 & \dots \\ 5 & 3 & -1 & 10 & \dots \\ 11 & 3 & 0 & -2 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{array} \right) \\
 g_2 & & & & & \\
 g_3 & & & & & \\
 \dots & & & & &
 \end{matrix}$$

Figura 9. Representación de la matriz de datos de expresión. Cada columna “m” representa una muestra y cada fila “g” representa un gen.

El proceso de análisis de los datos consta de dos fases principales, el pre-procesado de datos y el análisis de datos. A continuación se detallan las principales características de cada una de estas dos fases.

1.5.1.2.1. Pre-procesado de datos: incluye las etapas de filtrado y normalización de los datos (Quackenbush et al 2006; Fan et al., 2006).

- Filtrado de datos: elimina de la matriz de datos aquellas 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: corrige errores o sesgos introducidos durante los procesos de marcaje e hibridación de las muestras que pueden ser originados por la distinta incorporación de los fluorocromos, por diferencias en la cantidad de mRNA 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 de 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.).

1.5.1.2.2. Análisis de datos: los datos están listos para su estudio una vez se han filtrado y normalizado. Para su análisis posterior pueden usarse métodos supervisados o métodos no supervisados dependiendo de cuál sea el objetivo del trabajo (Quackenbush et al., 2001; Quackenbush et al., 2006).

- **Métodos no supervisados:** son técnicas de agrupación que para generar resultados se basan exclusivamente en los datos de expresión obtenidos sin hacer uso de ningún conocimiento previo relativo a la patología que se está estudiando. 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 de interés, basándose solamente en las similitudes entre los perfiles de expresión de las diferentes muestras incluidas en el estudio. 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:** se utilizan parámetros externos a la matriz de expresión (Ej. estadificación del tumor, estatus del receptor de estrógeno, recidiva o no de la enfermedad, etc.) para generar los supuestos que se van a analizar. Cuando el objetivo final es caracterizar las diferencias entre grupos de muestras se utilizan los procesos de comparación de clases (t-test, ANOVA, Significant analysis of microarrays-SAM, 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 neighbour (KNN),

Prediction Analysis Microarrays (PAM), 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 pertenezca.

Los tres supuestos anteriores (descubrimiento, comparación y predicción de clases) exigen, para confirmar la universalidad de sus resultados, la verificación de los hallazgos obtenidos en matrices externas de datos independientes, también conocidas como *matrices de validación*, a la utilizada para obtener los resultados, conocidas como *matrices de entrenamiento* (Simon et al., 2005).

1.5.2. Fiabilidad y reproducibilidad de la técnica de microarrays de expresión

La posibilidad de ser capaces, utilizando la metodología de microarrays, de caracterizar todas las enfermedades humanas en un corto período de tiempo, llevó a la publicación de varios artículos científicos poco rigurosos que utilizaban esta metodología. Como consecuencia y para evitar la continua publicación de este tipo de trabajos, aparecieron varias revisiones en las que se señalaban las numerosas carencias de este tipo de estudios (Simon et al., 2003; Ioannidis et al., 2005) e incluso una comparativa de diversos estudios de microarrays en cáncer en la 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 (Michiels et al., 2005).

A partir del año 2005, como consecuencia de la evolución de la técnica y la progresiva estandarización de los protocolos que incluye, fueron surgiendo una serie de trabajos llevados a cabo por grupos cooperativos dentro del proyecto Microarray Quality Control (MAQC), que demostraron su reproducibilidad tanto entre laboratorios como entre plataformas, siempre y cuando se estandarizasen los protocolos utilizados y se establecieran controles de calidad (Larkin et al., 2005; Irizarry et al., 2005; Bammler et al., 2005; Shi et al., 2006). Además, varios grupos consiguieron demostrar la validez funcional de los predictores moleculares obtenidos en sus estudios utilizando la tecnología de microarrays. Un claro ejemplo es el cáncer de mama, en el que los distintos predictores pronóstico obtenidos reconocían los mismos grupos moleculares de pacientes (Fan et al., 2006). Además, parte de los procesos de estandarización de esta técnica se centraron en determinar la información mínima que debía ser hecha pública por los grupos de investigación en este tipo de trabajos. Así, a partir de 2001 se comenzaron a elaborar las

denominadas reglas MIAME o Minimum Information About Microarray Experiments (Brazma et al., 2001) en donde se especifica la información mínima que se ha de proporcionar de cada análisis de microarrays para facilitar el intercambio de información entre grupos de investigación y la interpretación por otros grupos de los trabajos publicados (dato bruto, dato procesado, identificación y características de las muestras, anotación de los genes, etc...).

1.5.3. Aplicaciones de los microarrays de expresión en cáncer

La oncología es una de las áreas en las que los microarrays han despertado mayor interés, quizás debido a la posibilidad de esta técnica de dar a conocer la gran variabilidad molecular inherente a esta enfermedad. Hasta el momento, el gran número de trabajos científicos publicados en el campo de la oncología que han utilizado esta metodología han permitido:

- *Establecer la individualidad de cada tumor.* Se ha demostrado que molecularmente el tumor primario es más parecido a su metástasis que a otro tumor primario, aunque éste último pertenezca al mismo estadio (Perou et al., 2000).
- *Descubrir nuevas vías moleculares alteradas en el desarrollo de tumores.* Se han descrito tanto vías de carcinogénesis específicas de cada órgano (Ramaswamy et al., 2001) como vías comunes en tumores desarrollados en diferentes órganos (Chung et al., 2002).
- *Desarrollar nuevos sistemas de clasificación y descubrir nuevos subtipos tumorales.* Se han publicado varios artículos científicos que describen diferentes patrones de expresión génica que caracterizan diferentes subtipos moleculares en tumores tan frecuentes como el cáncer de mama (Perou et al., 2000; Sorlie et al., 2001), el cáncer de pulmón (Garber et al., 2001), linfomas (Rosenwald et al., 2002), etc.
- *Desarrollar nuevos predictores con valor pronóstico.* Existen claros ejemplos de que el conocimiento obtenido mediante esta tecnología puede y debe ser trasladado a la clínica. Basándose en los perfiles de expresión de algunos tumores como linfomas, cáncer de colon, CNMP, etc... se han generado varias firmas génicas con importante valor pronóstico (Dave et al., 2004; Kerr et al., 2009; Kratz et al., 2012). Incluso para el cáncer de mama ya están disponibles en el mercado test como el MamaPrint de Agendia que incluye 70 genes, se desarrolla sobre tejido congelado, se basa en la tecnología de microarrays y está aprobado por la Food and Drug Administration (FDA) (Van't Veer et al., 2002), el OncotypeDX de Genomic Health, score de recurrencia para seleccionar pacientes que

puedan obtener el máximo beneficio del tratamiento con tamoxifeno y que incluye 21 genes, se desarrolla sobre tejido fijado en formol e incluido en parafina y que utiliza la tecnología de la RT-PCR (Paik et al., 2004) y el ensayo de recurrencia del cáncer de mama de AviraDx que incluye 8 genes, se desarrolla sobre tejido fijado en formol e incluido en parafina y que utiliza la tecnología de RT-PCR (Ma et al., 2006).

- *Obtener perfiles de expresión que predicen respuesta al tratamiento.* Varios autores han desarrollado y publicado trabajos en los que se han obtenido el perfil de expresión diferencial entre pacientes respondedores y no respondedores frente al mismo tratamiento (Kihara et al., 2001; Hess et al., 2006).
- *Descubrir vías moleculares de resistencia a un tratamiento.* En el cáncer de mama se ha descrito que la alteración en la vía de PI3K es un factor de suma importancia en determinar la resistencia de estas pacientes al trastuzumab (Berns et al., 2007).

1.5.4. Aplicaciones de los microarrays de expresión en el CNMP

Al igual que en otros tumores, el primer objetivo con el que comenzaron a desarrollarse los estudios de expresión génica global en el CNMP fue la clasificación molecular de esta patología (Bhattacharjee et al., 2001; Garber et al., 2001). Con posterioridad, fueron publicados varios estudios que han demostrado una clara asociación entre los perfiles de expresión génica obtenidos en el CNMP y la supervivencia de los pacientes (Wigle et al., 2002; Lu et al., 2006; Kratz et al., 2009; Boutros et al., 2009; Roepman et al., 2009; Zhu et al., 2010) hecho que ha desencadenado que muchos grupos de investigación desarrollen estudios en los que su objetivo principal sea la identificación de nuevos marcadores pronósticos utilizando esta tecnología. Dentro de esta línea de investigación, podemos destacar varios trabajos científicos que han mostrado resultados muy prometedores centrados en los estadios iniciales de CNMP. En el año 2006, Lu y colaboradores publicaron un trabajo de investigación que identificaba un predictor de 64 genes capaz de diferenciar tumores de alto y bajo riesgo, pero únicamente centrado en estadios I de esta enfermedad (Lu et al., 2006). En el año 2009, el grupo de Roepman publicó un predictor de 72 genes relacionado con respuesta inmune, que funciona bien tanto en estadios I como II de CNMP (Roepman et al., 2009). Más recientemente, Zhu y colaboradores publicaron un predictor de supervivencia para estadios iniciales de CNMP que incluye 15 genes (Zhu et al., 2010).

Como muestran los trabajos anteriores, los resultados obtenidos han sido dispares en cuanto al número e identidad de los genes a incluir en el predictor, quizás debido en gran parte al uso de diferentes criterios en la selección de los pacientes a incluir en el estudio respecto a variables tan importantes como la obtención de muestras biológicas, la elección de los estadios tumorales, la exclusión o no de subtipos histológicos de gran importancia en el CNMP, así como a la falta, en algunos casos, de validación independiente. Por tanto es necesario estandarizar protocolos e identificar los problemas metodológicos que deben ser subsanados con el fin de generar información reproducible y aplicable a la práctica clínica. En esta línea se han publicado dos trabajos de investigación clave para, que en la actualidad, puedan desarrollarse de manera correcta estudios de microarrays en CNMP. El primero de ellos fue publicado en el año 2009 por el grupo de Kratz y colaboradores (Kratz et al., 2009) y el segundo en el año 2010 por Jyothi Subramanian y Richard Simon (Subramanian et al., 2010). Estos trabajos incluyen una evaluación crítica del diseño experimental, del análisis estadístico, de la utilidad de los resultados obtenidos en las decisiones clínicas, etc... de todos los estudios desarrollados entre los años 2002 a 2009 que incluyen firmas pronósticas obtenidas a través de perfiles de expresión génica. Ambos trabajos destacan que para que una nueva firma génica sea aplicable en la práctica clínica diaria debe de ser capaz de proveer información pronóstica relevante y adicional a los métodos standard aplicados en la actualidad al CNMP. Así, una nueva firma génica se consideraría clínicamente útil en los estadios I y II de la enfermedad si es capaz de cumplir los requisitos detallados a continuación:

- Ser más efectiva en identificar pacientes de alto riesgo clasificados como estadio I completamente resecaados que pudieran beneficiarse de quimioterapia adyuvante.
- Ser más efectiva en identificar pacientes con tumores clasificados como estadio II que tengan bajo riesgo de recurrencia y en los que la quimioterapia fuera innecesaria.

Como conclusión destacar que la metodología de microarrays es una herramienta muy potente para la identificación de firmas génicas con valor pronóstico pero es necesaria la estandarización de su metodología para poder obtener información que pueda ser trasladada a la rutina clínica diaria. Una de las aplicaciones directas en el CNMP sería la obtención de un método robusto que sea capaz de estratificar pacientes en grupos de buen y mal pronóstico y que genere información adicional al método convencional de clasificación pronóstica del CNMP, que en la actualidad es la clasificación TNM.

2. JUSTIFICACIÓN e HIPÓTESIS

2.1. JUSTIFICACIÓN

El cáncer de pulmón es el tumor maligno más importante a nivel mundial en cuanto a incidencia. Respecto a mortalidad, es la primera casusa de muerte por cáncer en varones y la segunda causa en mujeres con una tasa de supervivencia global a cinco años de sólo el 15%.

En el CNMP, la clasificación por estadios obtenida a partir del TNM es el principal indicador pronóstico que condiciona las indicaciones terapéuticas de los pacientes. En la actualidad, la resección quirúrgica con linfadenectomía es el tratamiento de elección en los estadios IA, IB y IIA, cuando la enfermedad está localizada y se cumplen los criterios de resecabilidad.

El empleo de quimioterapia neoadyuvante y/o adyuvante en los estadios iniciales del CNMP está en continúa discusión ya que, aunque existen suficientes pruebas sobre su eficacia en el estadio IIB, su uso estadios IA, IB y IIA es menos claro por lo que, de manera general, éstos no reciben quimioterapia a pesar de presentar elevadas tasas de recidiva a 5 años que llegan a alcanzar hasta un 30% de los pacientes. Es decir, sujetos clasificados en un mismo estadio, evolucionan en la enfermedad de manera muy diferente.

Por tanto, es necesario identificar marcadores adicionales que puedan ser utilizados clínicamente y sean más efectivos o complementen a los factores de riesgo estándar para explicar toda la heterogeneidad biológica del tumor.

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

2.2. HIPÓTESIS

El análisis del perfil de expresión génica global en pacientes con estadios iniciales del CNMP sometidos a resección quirúrgica con linfadenectomía, permite obtener una clasificación molecular capaz de identificar el riesgo de recidiva para estos pacientes.

3. OBJETIVOS DEL ESTUDIO

3.1. OBJETIVOS DEL ESTUDIO

Teniendo en cuenta estos antecedentes respecto a los estadios iniciales del CNMP, nos planteamos la realización del presente estudio con los objetivos que se detallan a continuación:

1.- Identificar los factores clínico-histopatológicas y mutacionales convencionales de mayor relevancia pronóstica, estudiando su asociación con el intervalo libre de enfermedad.

2.- Clasificación molecular de los pacientes mediante análisis del perfil de expresión génica global.

2.1. Establecer la correlación entre la clasificación molecular y las variables clínico-histopatológicas y mutacionales convencionales.

2.2. Estudio de asociación entre la clasificación molecular y el intervalo libre de enfermedad.

3.- Obtención de un predictor génico basado en la clasificación molecular obtenida y análisis de su aplicabilidad como predictor pronóstico.

4. MATERIAL Y MÉTODOS

4.1. DISEÑO EXPERIMENTAL

4.1.1. Preselección de los pacientes diagnosticados de CNMP

Se preseleccionaron todos los pacientes operados en el Hospital Clínico San Carlos (HCSC) en el periodo 2001-2008 y diagnosticados de CNMP en estadios iniciales de la enfermedad (I y II) que cumplían los requisitos que se detallan a continuación:

- Pacientes no diagnosticados de ningún otro tumor previo al CNMP.
- Pacientes operados por el mismo equipo de Cirugía Torácica del HCSC. Todos ellos fueron sometidos a una de las cuatro técnicas quirúrgicas que se detallan a continuación: neumonectomía (extirpación de todo el pulmón), lobectomía (se extirpa un lóbulo del pulmón), bilobectomía (extirpación de dos lóbulos del pulmón) o segmentectomía (extirpación de parte de un lóbulo). También es importante destacar que, independientemente de la técnica quirúrgica de elección, a todos los pacientes se les sometió a linfadenectomía.
- Tumores completamente resecaados o tumores R0 (Rami-Porta, 2005).
- Pacientes no tratados con quimioterapia / radioterapia previa a la cirugía (neoadyuvante).
- Pacientes no tratados con quimioterapia / radioterapia posterior a la cirugía (adyuvante), al menos hasta el momento de presentar recidiva.
- Seguimiento mínimo del paciente durante 3 años en consulta de Cirugía Torácica.

De la historia clínica se obtuvieron los datos relativos a la edad, sexo, hábito de fumar, localización pulmonar, subtipo histológico, afectación ganglionar, estadio tumoral, recidiva y tipo de recidiva (tabla 6). Respecto al hábito de fumar, los pacientes se clasificaron en la categoría de no fumador si no habían fumado en ningún momento de su vida, exfumador cuando habían abandonado el hábito tabáquico como mínimo 1 año antes

del diagnóstico de la enfermedad y fumador activo cuando el paciente fumaba en el momento del diagnóstico.

Del informe anatomopatológico se obtuvieron los datos relativos al grado de diferenciación del tumor y a la presencia o no de queratinización (tabla 6). El grado de diferenciación del tumor se clasificó en 3 categorías definidas como pobre, moderado o bien siempre basándose en los criterios histopatológicos convencionales. Por último, la queratinización se consideró presente cuando se observaban láminas de queratina en las preparaciones histológicas de los tumores.

PARÁMETRO	CATEGORÍA	PARÁMETRO	CATEGORÍA
Edad	Años	Afectación ganglionar	Ausente (N0) Presente (N1)
Sexo	Varón / Mujer	Estadio tumoral	I / II
Hábito de fumar	No fumador Exfumador Fumador activo	Recidiva	Si / No
Localización	Pulmón derecho Pulmón izquierdo	Tipo de recidiva	Locoregional A distancia
Subtipo histológico	Adenocarcinoma (AC) Epidermoide o escamoso (SCC) Célula grande (LCC) Adenoescamoso	Grado diferenciación	Pobre Moderado Bien
		Queratinización	Ausente Presente

Tabla 6. Parámetros clínico-histopatológicos. La tabla muestra los parámetros clínicos de los pacientes incluidos en el estudio y los parámetros histopatológicos de las muestras tumorales de dichos pacientes que fueron recopilados para ser incluidos en el estudio.

4.1.2. Selección de la plataforma de microarrays

Se utilizó la tecnología de microarrays de Agilent Technologies (Palo Alto, California, EEUU) en la que sus matrices de expresión están formadas por oligonucleótidos largos (60pb) de secuencia específica 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).

4.1.3. Selección de la técnica de hibridación de microarrays

Como modelo de hibridación se utilizó la técnica de hibridación doble o competitiva en la que cada una de las muestras a estudio es hibridada frente a la misma muestra de referencia o “pool”, consiguiéndose así facilitar y homogeneizar el proceso de normalización. En nuestro caso, la muestra de referencia consistió en una mezcla o “pool” de RNA de parénquima pulmonar no tumoral obtenido de los propios pacientes incluidos en el estudio.

4.2. MUESTRAS TUMORALES Y NO TUMORALES

4.2.1. Obtención de las muestras de los pacientes

Una vez obtenido el listado de los pacientes preseleccionados es necesario confirmar que de cada uno de ellos se dispone, en el Banco de Tumores del HCSC, muestra congelada tanto tumoral como no tumoral. El Banco de Tumores del HCSC se rige bajo las normas establecidas por la Ley 14/2007, de 3 de julio, de Investigación biomédica y por el Real Decreto 1716/2011, de 18 de noviembre, por el que se establecen los requisitos básicos de autorización y funcionamiento de los biobancos con fines de investigación biomédica y del tratamiento de las muestras biológicas de origen humano, y se regula el funcionamiento y organización del Registro Nacional de Biobancos para investigación biomédica. Uno de los principales requisitos que cabe destacar es la obligación de información por parte del facultativo correspondiente al paciente y la consecuente firma del Consentimiento Informado por éste demostrando que está de acuerdo en que el Banco de Tumores almacene en sus instalaciones el material tumoral y no tumoral excedente del diagnóstico para destinarlo a investigación biomédica.

Desde su creación en el año 2000, el Banco de Tumores del HCSC, utiliza protocolos estandarizados para la selección, congelación y el mantenimiento de las muestras. Es el patólogo el que sobre la pieza pulmonar en fresco selecciona las muestras a congelar identificando macroscópicamente las zonas afectas, de donde obtiene las muestras tumorales y de regiones alejadas mínimo entre 7-10 cm, las muestras de parénquima pulmonar no tumoral. Posteriormente se depositan en criomoldes perfectamente identificados y se embeben en medio criosolidificable OCT compound Tissue-Tek (Sakura. Tokio, Japón) para su congelación en isopentano a -80°C. Por último, los bloques

de tejido congelado se almacenan de manera organizada en el arcón congelador de -80°C hasta el momento de su uso.

4.2.2. Criterios de calidad de las muestras de los pacientes

La inclusión final de los pacientes en el estudio depende de que sus correspondientes muestras congeladas cumplan los siguientes criterios de calidad respecto a la representación mínima de células para el estudio y a la integridad del RNA extraído.

- *Representación mínima de células para el estudio.* Las muestras congeladas y almacenadas en el Banco de Tumores del HCSC se someten a revisión histológica por parte de dos patólogos de manera independiente para evaluar:

- En el caso de las muestras tumorales, si poseen una representación $\geq 70\%$ de células tumorales.

- En el caso de las muestras de parénquima pulmonar no tumoral, si poseen una representación del 100% de células normales.

- *Calidad del RNA.* El RNA extraído de las muestras congeladas, tanto tumorales como no tumorales, debía superar el criterio mínimo de calidad establecido e indicado mediante RIN (RNA Integrity Number) teniendo que ser siempre ≥ 7.5 .

4.2.2.1. Muestras tumorales de CNMP: revisión histológica.

Es fundamental realizar el análisis histológico de las muestras tumorales congeladas que se van a utilizar, ya que es imprescindible que en ellas exista un porcentaje representativo de las células que queremos estudiar, en nuestro caso células tumorales de CNMP, y por el contrario, debemos evitar una elevada contaminación con parénquima pulmonar no tumoral.

Se consideró que para el estudio de microarrays únicamente se incluyeran aquellas muestras congeladas que superaran el 70% de células tumorales.

El protocolo para la revisión histológica de las muestras congeladas (figura 10) se basa en realizar en cada una de ellas un corte de $4\ \mu\text{m}$ en criostato Leica Biosystems CM1850 (Wetzlar, Alemania) que se tiñe con Hematoxilina-Eosina (HE) y que se utiliza para la obtención del porcentaje de células tumorales. Si se supera el límite mínimo establecido, se recogen alrededor de 20 cortes de $25\ \mu\text{m}$ en eppendorff de 2 ml para la extracción de ácidos nucleicos. Los eppendorff utilizados son estériles y libres de

DNA/RNA. Por último, se realiza un nuevo corte de 4 μm para una segunda tinción de HE en la que se valora de nuevo el porcentaje de células tumorales para así confirmar que se seguía cumpliendo con el criterio mínimo de representación tumoral exigido.

Los cortes iniciales y finales teñidos con HE de cada una de las muestras tumorales de CNMP fueron evaluados al microscopio por dos patólogos de manera independiente para evaluar si el porcentaje de células tumorales superaba el mínimo establecido en cada uno de ellos para poder incluir la muestra tumoral en el estudio.

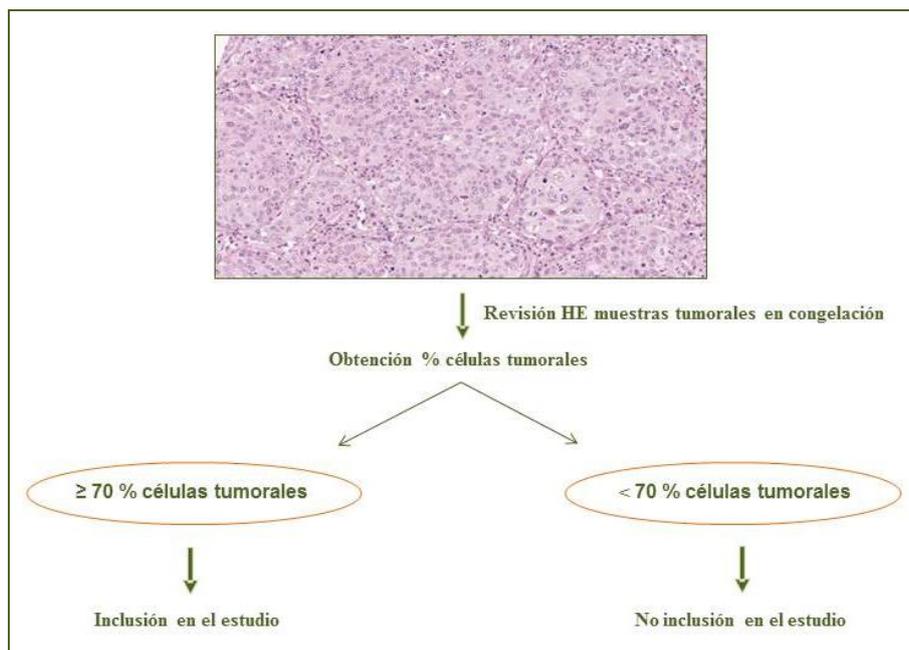


Figura 10. Estudio histológico para la selección de muestras de pacientes con CNMP. El criterio establecido fue poseer como mínimo un 70% de células tumorales para que la muestra de ese paciente pudiera ser incluida en el estudio.

Por otra parte, se utilizaron las secciones de tejido teñidas con HE y destinados al diagnóstico anatomopatológico de los pacientes finalmente incluidos en el estudio, para llevar a cabo el análisis de los parámetros histopatológicos de los tumores que no podían ser obtenidos del informe anatomopatológico como los descritos en la tabla 6 del presente trabajo. En la rutina de trabajo del Servicio de Anatomía Patológica, las biopsias diagnósticas se fijan en formol tamponado durante 24 horas, se tallan para ser incluidas en cassettes y se introducen en el procesador de tejidos el cual, de manera automática, va realizando diferentes pases para deshidratarlas y depositarlas en parafina. En este punto ya es posible elaborar los bloques de parafina a partir de los cuales se realizan cortes en micrótopo que se tiñen con HE y que pueden ser revisados en el microscopio de luz

óptica. Todo este proceso mantiene más intacta la histología de los tumores que el proceso de congelación al que son sometidas las muestras de Biobanco.

Ya que la evaluación de los parámetros histopatológicos que se detallan a continuación requiere una muy buena preservación de la histología de los tumores, se optó porque fueran evaluados en las secciones de tejido teñidas con HE y destinadas al diagnóstico anatomopatológico de los pacientes finalmente incluidos en el estudio.

La presencia de linfocitos intratumorales (TILs), la inflamación aguda (caracterizada por la presencia de leucocitos polimorfonucleares, PMN, entre los que se incluyen los neutrófilos, eosinófilos y basófilos), la inflamación crónica (caracterizada por la presencia de células mononucleares entre las que se incluyen los macrófagos, linfocitos y células plasmáticas), la necrosis y la cantidad de estroma tumoral fueron valorados por dos patólogos de manera independiente.

Los TILs se consideraron positivos o presentes cuando aparecían más de 2 linfocitos por campo de gran aumento (CGA) en el componente epitelial del tumor. La inflamación aguda se consideró positiva o presente cuando aparecían más de 2 PMN por CGA tanto en el componente epitelial como en el estroma. Respecto a la inflamación crónica se consideró moderada/fuerte cuando, en el estroma tumoral, se identificaron células linfoplasmocitarias que representaban al menos el 10% de la población celular total. La necrosis se consideró presente cuando más del 10% del tumor mostraba áreas necróticas. La cantidad de estroma tumoral se consideró baja cuando representaba menos del 20% de todo el componente tumoral y media / alta cuando representaba más del 20% (tabla 7).

PARÁMETRO	CATEGORÍA
TILs	Ausente / Presente
Inflamación aguda	Ausente / Presente
Inflamación crónica	Ausente / Débil Moderada / Fuerte
Necrosis	Ausente / Presente
Estroma tumoral	Bajo Medio / Alto

Tabla 7. Parámetros histopatológicos de las muestras del estudio.

4.2.2.2. Muestra de referencia o “pool”: revisión histológica

En los estudios de microarrays con hibridación doble o competitiva, todas las muestras a estudio se hibridan frente a una única muestra común denominada “muestra referencia” o “pool” que en nuestro caso consistió en una mezcla de RNAs de parénquima pulmonar no tumoral obtenido de 42 pacientes seleccionados al azar de los 84 finalmente incluidos en el estudio. Mediante la revisión histológica se confirmó que todas las muestras utilizadas para constituir la muestra de referencia o “pool” únicamente incluyeran parénquima pulmonar sin representación de células tumorales (100% células normales). Una vez extraído y cuantificado el RNA se constituyó el “pool” utilizando la misma concentración de RNA de cada una de las 42 muestras en un volumen final suficiente para poder llevar a cabo todas las hibridaciones del estudio.

4.3. EXTRACCIÓN DE ÁCIDOS NUCLEICOS

4.3.1. Extracción de RNA: Método del trizol

El Trizol® (Invitrogen, California. EEUU) es una solución monofásica de fenol e isotiocianato de guanidina capaz de mantener la integridad del RNA mientras lisa las células y disuelve los componentes celulares.

Se siguió el protocolo de extracción indicado por el fabricante con pequeñas modificaciones que se detallan a continuación:

- Homogenización mecánica en frío de los cortes de tejido congelado usando Trizol® (1ml) y un homogenizador Ultra Turrax T8 (S8N-5G) (Rose Scientific Ltd, Canadá).
- Separación del homogeneizado en tres fases añadiendo cloroformo (200µl x ml de Trizol®) y llevando a cabo una centrifugación a 13.000 rpm durante 10 minutos. Las tres fases que se obtienen son:
 - Fase acuosa (RNA).
 - Fase orgánica (restos celulares y DNA).
 - Fase fenólica (restos de Trizol® y cloroformo).
- Precipitación en frío (4°C) del RNA de la fase acuosa mediante isopropanol (600µl x ml de Trizol®) durante toda la noche. Centrifugación a 13.000 rpm durante 30 minutos para

recuperar el pellet de RNA y lavado con etanol al 75 % para eliminar impurezas y restos de sales.

- El RNA se resuspende en agua libre de RNAsas (Qiagen. Hilden, Alemania).

Parte del volumen recuperado con el RNA se sometió posteriormente a purificación con columnas RNeasy Micro Kit (Qiagen) y tratamiento con DNAsa para eliminar cualquier resto de contaminación con DNA genómico. El protocolo que se siguió es el indicado por el fabricante:

- Unión del RNA a la matriz de la columna. Se incorporan 45µg de RNA + 350 µl Buffer RLT-β-mercaptoetanol + 250 µl de etanol absoluto al interior de la columna. Centrifugación a 10.000 rpm durante 15 segundos.
- Incubación del material retenido en la columna con DNAsa (30 unidades) + Buffer RDD (70 µl) durante 15 minutos. Lavado con buffer RW1 (350 µl) y centrifugación a 10.000 rpm durante 15 segundos.
- Lavados con RPE (500 µl) y etanol al 80% (500 µl). Centrifugación a 10.000 rpm durante 15 segundos y 2 minutos respectivamente. Cambiar el tubo de recolección.
- Añadir 25 µl de agua libre de RNAsas, centrifugación a 14.000 rpm durante 1 minuto. Obtención del RNA eluido.

4.3.2. Extracción de DNA

Se utilizó como material de partida para la extracción del DNA las fases orgánica y fenólica procedentes de la purificación con Trizol®. El método de extracción que se utilizó fue el siguiente:

- Adición de etanol absoluto (600 µl por ml de Trizol®), agitación mediante vortex y centrifugación (8.000 rpm durante 10 minutos). Eliminación del sobrenadante obtenido.
- Lavados sucesivos con citrato de sodio con etanol al 10% (1ml) para eliminar los restos fenólicos y posteriormente con una solución de NaCl 0,075 M y EDTA 0,024 M 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 Proteinasa K (0,6 mg/ml) mediante incubación a 55 °C durante 48 horas.
- Precipitación con NaCl (2,1 M) de los restos celulares digeridos y obtención del DNA en disolución mediante centrifugación a 14.000 rpm durante 15 minutos.
- Precipitación del DNA con isopropanol (volumen / volumen) durante 30 minutos y centrifugación (14.000 rpm durante 30 minutos). Lavado con etanol al 75% y resuspensión del pellet en agua destilada.

4.3.3. Cuantificación de los ácidos nucleicos

Se evaluó la cantidad de DNA y RNA obtenido y su pureza mediante el uso del espectrofotómetro Nanodrop ND-1000 (Thermo Fisher Scientific Inc. Wilmington, EEUU) en 1µl de muestra. La concentración de DNA y RNA se obtuvo mediante la medición de la absorbancia y la aplicación de la siguiente 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 (1 µl en nuestro caso).

Además, se tuvieron en cuenta los siguientes ratios a la hora de evaluar la pureza de la muestra:

- Ratio 260/280 para descartar la contaminación con proteínas.
- Ratio 260/230 para descartar la contaminación con restos de sales.

4.3.4. Control de calidad del RNA

Los niveles de expresión del RNA mensajero se ven afectados directamente por la degradación de la molécula de RNA. La calidad o nivel de integridad del RNA se valoró utilizando Bioanalyzer 2100 (Agilent Technologies) (figura 11) en combinación con el software “RNA Integrity Number (RIN)” (Agilent Technologies) el cual integra, mediante algoritmos matemáticos, diversos parámetros del perfil del RNA y proporciona un valor (en una escala del 0-10, en donde el 0 indica degradación completa) que refleja la calidad

de la muestra (Imbeaud et al 2005; Strand et al., 2007). Sólo muestras tumorales y no tumorales congeladas con un RIN ≥ 7.5 fueron seleccionadas para el estudio.

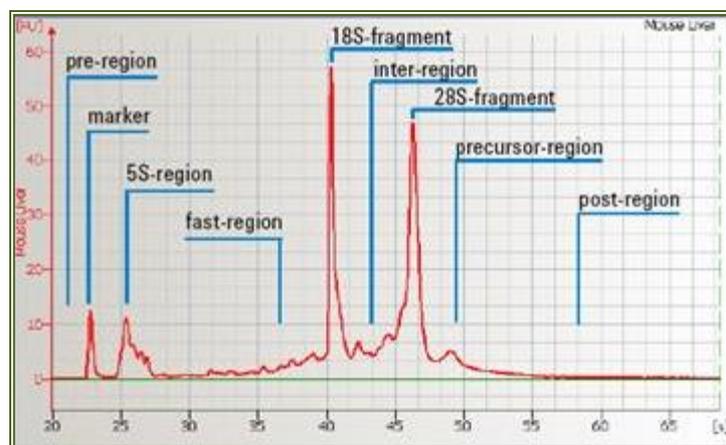
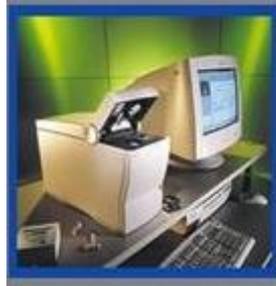


Figura 11. Fotografía del Bioanalyzer 2100 junto con el electroferograma resultante. En dicho electroferograma se indican las regiones que son utilizadas para calcular el valor del RIN en cada una de las muestras.

4.4. MARCAJE E HIBRIDACIÓN DE LAS MUESTRAS EN LOS MICROARRAYS DE EXPRESIÓN

Se utilizaron microarrays del tipo “Two-color Human Whole Genome Oligo Microarrays” con formato 4 X 44 K (producto number G4112F. Agilent Technologies).

<http://www.genomics.agilent.com/CollectionSubpage.aspx?PageType=Product&SubPageType=ProductData&PageID=1516>.

Las características de este tipo de microarrays se presentan en la tabla 8 destacando las que se detallan a continuación:

- Los puntos a estudio representan el conjunto de mRNAs (RNAs mensajero) que se transcribe en las células.

- Los controles negativos son oligonucleótidos que al contacto con el líquido de hibridación van a formar un “hairpin” consigo mismo evitando que se pueda unir ningún transcrito marcado y su fluorescencia se considera inespecífica de la técnica.
- Los controles positivos son oligonucleótidos de producción (utilizado como control de la síntesis del microarray por la compañía que lo fabrica) y de marcaje de los extremos del microarray (utilizado para identificar la posición exacta del array en el cristal).

	Whole Human Oligo Microarray
<i>Formato del array</i>	4 x 44 K
<i>Dimensiones del array</i>	25 mm x 75 mm
<i>Total puntos array</i>	45.015
<i>Tamaño del punto</i>	65 micrones Ø
<i>Longitud de la sonda</i>	60 pb
<i>Total de puntos a estudio</i>	43.376
<i>Total de sondas a estudio</i>	41.000
<i>Controles Positivos</i>	1486
<i>Controles Negativos</i>	153

Tabla 8. Características del formato de microarrays utilizado en el estudio. La tabla muestra las características del formato de microarray de 4 x 44 K que fue utilizado en el estudio.

Como control interno de la linealidad, sensibilidad y precisión en los pasos de retro-transcripción, amplificación y marcaje de las muestras, se utilizó el kit “Two-Color RNA Spike-In” (Agilent Technologies).

<http://www.genomics.agilent.com/CollectionSubpage.aspx?PageType=Product&SubPageType=ProductData&PageID=1428>.

Este kit incluye dos “Spike-In”: el “Spike-A” que se mezcla con las muestras que en el experimento se marcan con Cianina 3 (CTP-Cy3) y que en nuestro caso es la muestra de referencia o “pool” y el “Spike-B” que se mezcla con las muestras que en el experimento se marcan con Cianina 5 (CTP-Cy5) y que en nuestro caso son los CPNM. Cada Spike-In contiene 10 transcritos sintetizados in vitro y derivados del Adenovirus E1A de los que conocemos su concentración y por tanto la fluorescencia que van a emitir.

A continuación se detallan los pasos seguidos para el proceso de marcaje e hibridación de las muestras en los microarrays.

4.4.1. Marcaje y amplificación de la muestra

Se siguió el protocolo del kit “Two-Color Microarray-Based Gene Expression Analysis (Quick Amp Labeling) versión 5.7, marzo 2008” (Agilent Technologies) cuyo procedimiento se describe en la figura 12.

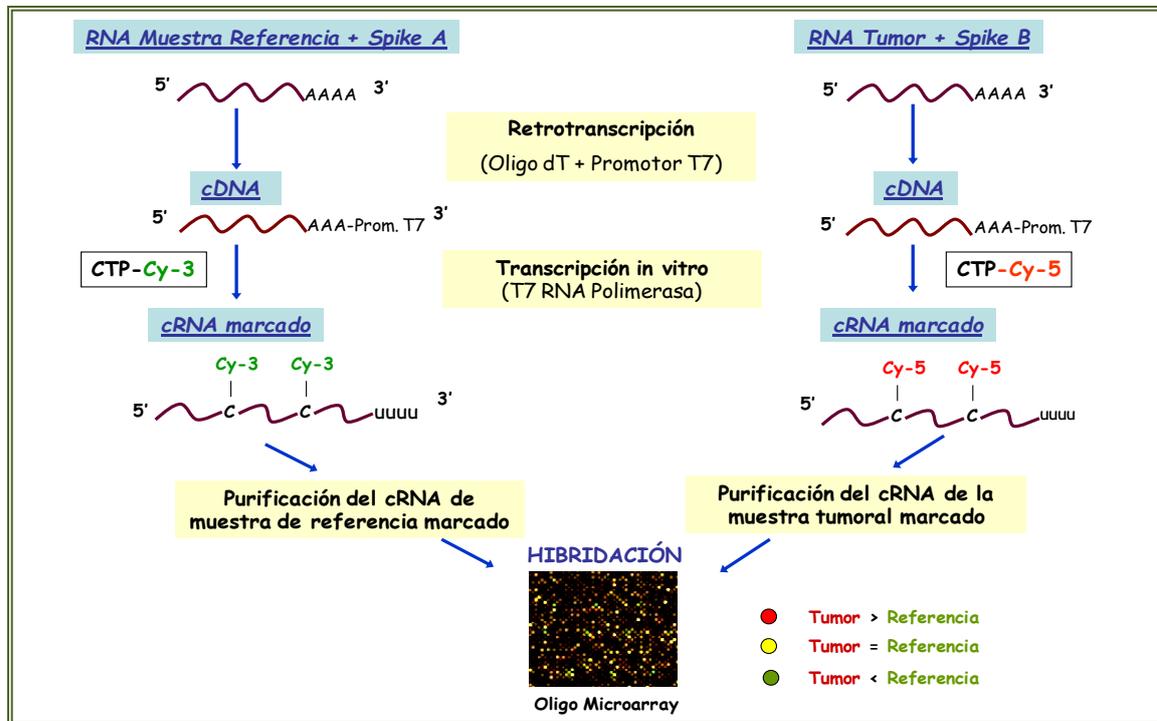


Figura 12. Esquema detallado de la técnica de microarrays. La figura representa los pasos de amplificación, marcaje e hibridación competitiva de un tumor y del pool de referencia sobre un cristal de microarray.

- 500 ng de RNA total se retrotranscribieron usando como cebador un oligo poli dT con la secuencia del promotor de la T7 RNA Polimerasa en su extremo 5' y como enzima retrotranscriptasa la MMLV-RT. En este paso se incluyen los controles internos Spike A y Spike B para el RNA de la muestra de referencia y para el RNA de cada uno de los tumores de CNMP respectivamente.
- Utilizando como molde el cDNA de doble cadena obtenido anteriormente, se sintetizó cRNA utilizando la enzima T7-RNA Polimerasa. En este paso se consigue de manera simultánea la amplificación y la incorporación del marcaje fluorescente. En el caso de la muestra de referencia o “pool” de tejido normal se utilizó CTP-Cy3 (color verde) y en caso del tejido tumoral se utilizó como marcaje CTP-Cy5 (color rojo).

- El cRNA marcado obtenido se purificó mediante columnas RNeasy Mini Kit (Qiagen) siguiendo el protocolo establecido. En este paso se consiguió eliminar restos de reactivos y el excedente de cianina no incorporada.

4.4.2. Control del marcaje y amplificación de la muestra

Como control del proceso de amplificación y marcaje de las muestras, se mide por duplicado en el Nanodrop ND-1000 (Nanodrop Technologies. Delaware, EEUU) la concentración del cRNA obtenido y la tasa de incorporación de cianina. Las muestras con una cantidad de cRNA total inferior a 825 ng o una incorporación de cianina menor de 8 pmol/μg deben ser descartadas.

4.4.3. Hibridación y lavado del microarray

Los procesos de hibridación y lavado del microarray constan de los pasos que se detallan a continuación:

- El cRNA es fragmentado en secuencias más pequeñas e hibridado al microarray. El proceso de hibridación se lleva a cabo en un horno de hibridación Agilent G2545A Hybridization Oven (Agilent Technologies) a 65 °C durante un periodo de 17 horas. En la tabla 9 se muestran los detalles referentes a los reactivos utilizados.

Reactivos proceso hibridación	Whole Human Oligo Microarray 4 x 44 k
<i>Cantidad inicial cRNA-Cy5</i>	825 ng
<i>Cantidad inicial cRNA-Cy3</i>	825 ng
<i>10X Blocking Agent</i>	11 μl
<i>Agua destilada</i>	hasta 52.8 μl
<i>25X Fragmentation Buffer</i>	2.2 μl
<i>Buffer de Hibridación</i>	2X GE Hybridization Buffer HI-RPM (55 μl)
<i>Volumen final de hibridación</i>	100 μl
<i>Temperatura de hibridación</i>	65 °C
<i>Tiempo de hibridación</i>	17 horas
<i>Velocidad del rotor del horno</i>	10 rpm

Tabla 9. Reactivos utilizados en el proceso de hibridación de las muestras a los microarrays.

- Como paso final se llevó a cabo el lavado de los microarrays para eliminar todo el excedente no unido a los oligonucleótidos. El proceso de lavado consta de dos pasos cuyas características se detallan en la tabla 10.

	Buffer lavado	Temperatura	Tiempo
<i>Desmontaje de cristales</i>	GE Wash Buffer 1	Ambiente	
<i>1^{er} lavado</i>	GE Wash Buffer 1	Ambiente	1' en agitación
<i>2^o lavado</i>	GE Wash Buffer 2	37 °C	1' en agitación

Tabla 10. Buffer, temperatura y tiempos del proceso de lavado.

4.5. ESCANEADO DE LOS MICROARRAYS

El escaneado de los cristales se realizó utilizando el escáner confocal “Agilent Microarray Scanner” (Agilent Technologies) modelo G2565BA con autofocus y el programa “Agilent Scan Control” versión 7 (SC v.7) (Agilent Technologies) (tabla 11).

Con este equipo es posible medir al mismo tiempo la intensidad de señal emitida por los dos fluorocromos incorporados (Cy3 y Cy5) cuando son excitados mediante láser SHG-TAG (532nm) y Helio-Neon (633nm) respectivamente.

Parámetros	Protocolo de escaneado SC v.7
<i>Área de escaneado</i>	61 mm x 21 mm
<i>“Extended Dinamic Range”</i>	SI
<i>PMT Canal Rojo</i>	100% - 10 %
<i>PMT Canal Verde</i>	100 % - 10 %
<i>Resolución del escáner</i>	5 µm
<i>Modo de escaneado</i>	2 veces

Tabla 11. Parámetros del proceso de escaneado del microarray.

Durante el proceso, se crea una imagen de alta definición en formato TIFF que refleja los niveles de expresión para cada uno de los puntos del microarray (figura 13).

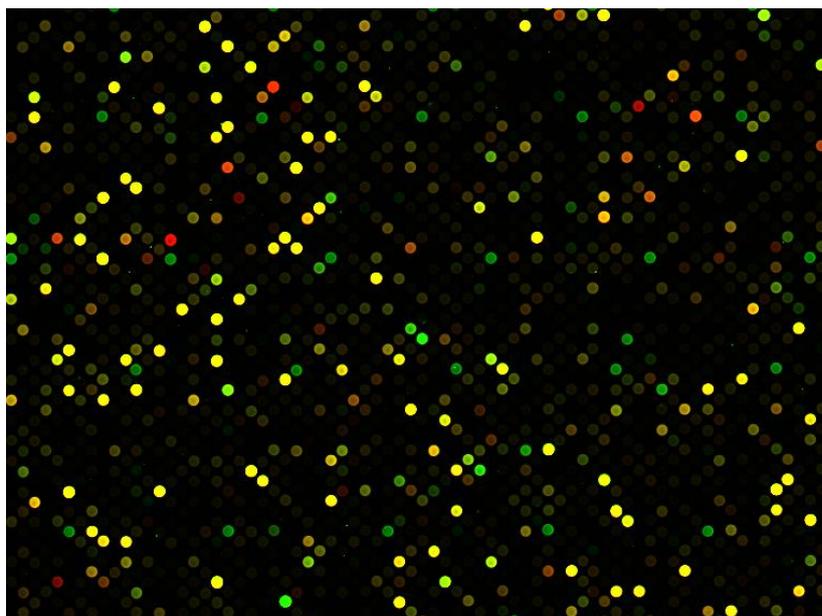


Figura 13. Imagen obtenida del escaneo de los microarrays.

4.6. CUANTIFICACIÓN DE LA IMAGEN

La extracción de los datos contenidos en las imágenes escaneadas en formato TIFF se llevó a cabo utilizando el programa “Feature Extraction” versión 10.7.3 (Agilent Technologies), siguiendo los pasos que se describen a continuación:

- Mediante una plantilla de localización de genes, “grid”, diseñada por Agilent Technologies (014850_D_20070820) se lleva a cabo la identificación y localización exacta de los mismos.
- Utilizando el algoritmo “Cookie Cutter” se realiza la extracción de los valores de intensidad de fluorescencia y ruido de fondo para cada uno de los puntos del array.
- Identificación de “puntos problemáticos” o “de baja calidad” (Ej. aquellos que no poseen una medida de la intensidad uniforme, que están saturados o cuya relación señal/ruido era muy baja, etc...).
- Cálculo de la fluorescencia procesada del punto. Se llevó a cabo siguiendo estos tres pasos: se elimina el efecto del ruido de fondo, se realiza una normalización de los valores de expresión (LOWESS) y por último y sólo en el caso de sondas con señales muy bajas, se establece un valor de fluorescencia mínimo. En el resto de los casos, el valor final de fluorescencia será el obtenido tras el proceso de normalización.

Con los datos obtenidos del programa “Feature Extraction” versión 10.7.3” (Agilent Technologies) se llevó a cabo el posterior análisis de los resultados del estudio.

4.7. ANÁLISIS DE DATOS

4.7.1. Control de calidad

4.7.1.1. Quality Control-Report (QC-Report)

El programa “Feature Extraction” (Agilent Technologies) genera en paralelo al proceso de extracción de datos, un archivo denominado “Quality-Control Report” (QC-Report) que recoge los parámetros que se detallan a continuación (tabla 12) y mediante los cuales nos mostrará una evaluación inicial de la calidad de cada uno de los experimentos, englobándolos como bueno o evaluable. En el caso de que el experimento no muestre los valores esperados, será necesario descartarlo ya que habrá sufrido algún fallo durante el proceso de hibridación y/o escaneado.

Parámetros	Bueno	Evaluable
<i>IsGoodGrid</i>	>1	<1
<i>AnyColorPrcntFeatNonUniform</i>	<1	>1
<i>gNegCtrlAveBGSubSignal</i>	-20 to 10	<-20 or >10
<i>gNegCtrlSDevBGSubSignal</i>	<15	>15
<i>rNegCtrlAveBGSubSignal</i>	-20 to 4	<-20 or >4
<i>rNegCtrlSDevBGSubSignal</i>	<6	>6
<i>gNonCntrlMedCVBkSubSignal</i>	0 to 18	<0 or >18
<i>rNonCntrlMedCVBkSubSignal</i>	0 to 18	<0 or >18
<i>gE1aMedCVBkSubSignal</i>	0 to 18	<0 or >18
<i>rE1aMedCVBkSubSignal</i>	0 to 18	<0 or >18
<i>absE1aObsVsExpCorr</i>	>0.86	<0.86
<i>absE1aObsVsExpSlope</i>	>0.85	<0.85

Tabla 12. Lista de parámetros evaluados en el QC-Report. La tabla muestra los doce parámetros evaluados en el control de calidad del QC-Report y el rango de resultados de cada uno de ellos para clasificarlo como bueno o evaluable. En caso de no mostrar los resultados esperados, será necesario descartar ese experimento.

El QC-Report también recoge los resultados obtenidos en el control interno para la linealidad, sensibilidad y precisión de los pasos de retro-transcripción, amplificación y marcaje, los “Spike-In”. El QC-Report muestra los valores esperados y observados para los 10 transcritos derivados del Adenovirus E1A que se incluyen en cada experimento (tabla 13).

Nombre transcrito	Valor esperado
<i>(+)E1A_r60_n9</i>	-1.00
<i>(+)E1A_r60_a107</i>	-0.48
<i>(+)E1A_r60_a135</i>	-0.48
<i>(+)E1A_r60_n11</i>	-0.48
<i>(+)E1A_r60_1</i>	0.00
<i>(+)E1A_r60_a20</i>	0.00
<i>(+)E1A_r60_3</i>	0.48
<i>(+)E1A_r60_a104</i>	0.48
<i>(+)E1A_r60_a97</i>	0.48
<i>(+)E1A_r60_a22</i>	1.00

Tabla 13. Transcritos que conforman los “Spike-In”. La tabla muestra los diez transcritos derivados del Adenovirus E1A que se incluyen en cada experimento y los valores esperados para cada uno de ellos.

4.7.1.2. Análisis de componentes principales o PCA

Una vez realizada la normalización y superado el QC-Report de cada microarray individualmente, se realizó un Análisis de Componentes Principales o PCA que nos permitiese detectar un comportamiento anómalo de alguna de las muestras, tanto tumorales como normales, debido a sesgos que hubiesen pasado inadvertidos durante los controles previos. Se llevó a cabo un estudio de PCA en el que se analizaron de manera independiente los siguientes parámetros:

- *Semana en la que se realizó la hibridación de las muestras tumorales.* Las hibridaciones de las 84 muestras tumorales se realizaron en 7 semanas diferentes. Para llevar a cabo el PCA, cada una de las muestras se identificó en función de la semana en la que se llevó a cabo su hibridación en los microarrays.

- *Tasa de incorporación de Cy5 en las muestras tumorales.* La tasa de incorporación de Cy5 debe alcanzar como mínimo, para superar el paso de control de marcaje y amplificación de la muestra, 8 pmol/μg. Para llevar a cabo el PCA, las muestras tumorales fueron clasificadas como grupo 1 si la tasa de incorporación de Cy5 se encontraba entre 8-10 pmol/μg y como grupo 2 si la tasa de incorporación de Cy5 había sido superior a 11 pmol/μg.
- *Tasa de incorporación de Cy3 en las muestras de referencia o “pool”.* Al igual que en el caso de las muestras tumorales, para llevar a cabo el PCA, las muestras incluidas en el “pool” fueron clasificadas como grupo 1 si la tasa de incorporación de Cy3 se encontraba entre 8-10 pmol/μg y como grupo 2 si la tasa de incorporación de Cy3 había sido superior a 11 pmol/μg.
- *Número RIN.* En el estudio, el nivel de integridad o número RIN que nos indica la calidad del RNA de las muestras debía alcanzar como mínimo 7.5. Para llevar a cabo el PCA, las muestras fueron clasificadas como grupo 1 si el RIN se encontraba entre 7.5-8.5 y como grupo 2 si el RIN se encontraba entre <8.5-10.

Para este análisis se utilizó el programa “GeneSpring” versión 11.5 (GS11.5) (Agilent Technologies) y el listado de sondas que pasaron el filtrado de “sondas problemáticas” que referimos a continuación.

4.7.2. Filtrado de datos

Se utilizó el programa “GeneSpring” versión 11.5 (GS11.5) (Agilent Technologies) para el proceso de filtrado de sondas el cual engloba los siguientes pasos (figura 14):

- *Exclusión de sondas “problemáticas”.* Se eliminan las sondas que en más de un 10 % de las muestras presentan baja fluorescencia o problemas en el proceso de hibridación (listado resultante que incluye 24.617 sondas).
- *“Gene Level Experiment”.* Proceso mediante el cual se realiza un promedio de todas las sondas que poseen un mismo identificador (Entrez ID). Por tanto, a partir de este paso del filtrado se deja de trabajar con sondas y se comienza a trabajar con genes individuales (listado resultante que incluye 17.881 genes).

- *Filtrado por expresión.* Se seleccionaron los genes con una variación de expresión de al menos 3 veces con respecto a la media de ese gen en el 10% de las muestras. Se obtuvo un listado de 3.232 genes.

En etapas posteriores del estudio y en base a los resultados que se obtuvieron, se llevó a cabo un nuevo filtrado:

- *Filtrado histológico.* Sobre el listado de 3.232 genes se llevó a cabo un segundo filtrado mediante la identificación y exclusión de los genes diferencialmente expresados entre los subtipos histológicos de AC y SCC utilizando un T-Test con corrección para comparaciones múltiples de Benjamini y Hochberg (B&H), $p < 0.01$ y una diferencia de expresión de más de 1.5 veces. Los genes que cumplieron estos criterios fueron excluidos del listado inicial de 3.232 genes resultando un listado final de 2.160 genes con el que se realizó la clasificación molecular del CNMP.

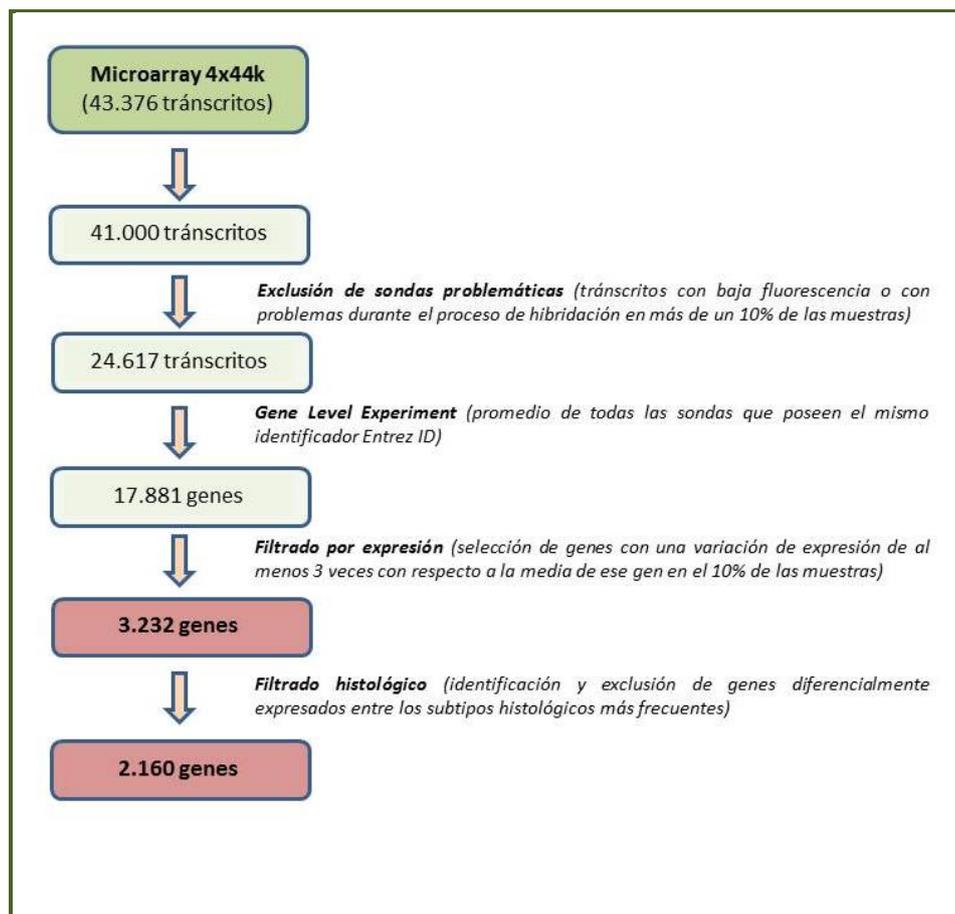


Figura 14. Esquema representativo proceso de filtrado de genes. La figura representa el proceso de filtrado de genes para conseguir la matriz final con la que llevar a cabo la clasificación molecular. Previo al proceso de filtrado de genes se eliminaron los transcritos que codificaban para los controles internos.

4.7.3. Métodos no supervisados para el descubrimiento de clases

Con la finalidad de clasificar las muestras en grupos molecularmente homogéneos, se llevó a cabo un procedimiento de dos pasos en el que primeramente se utilizó la técnica de “clustering jerárquico” posteriormente mejorada mediante la técnica de “k-means”. Con este procedimiento conseguimos disminuir la heterogeneidad intra-cluster y aumentar la variabilidad inter-cluster. Para su desarrollo se utilizó el programa “GeneSpring” versión 11.5 (GS11.5) (Agilent Technologies). En ambas técnicas se aplicó la correlación centrada de Pearson, el “average linkage” y se realizó un centrado de genes.

En el “clustering jerárquico”, el número de grupos moleculares obtenidos y seleccionados mediante criterio visual en el árbol obtenido, fue validado mediante el programa “BRB Array Tools” (versión 4.2.0 Beta 2) <http://linus.nci.nih.gov/BRB-ArrayTools.html> el cual utiliza algoritmos matemáticos que tienen en cuenta dos parámetros (McShane et al., 2002):

- R-Index: reproducibilidad del número y tamaño de los grupos encontrados.
- D-Index: reproducibilidad en la composición de 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 (100.000 perturbaciones). El D-Index incluye las omisiones y adiciones de muestras entre los grupos.

El método de “k-means” se desarrolló utilizando como base el número de grupos moleculares generados a partir del “clustering jerárquico”.

4.8. ESTUDIO DE MUTACIONES

4.8.1. Estudio de mutaciones en el gen EGFR

El análisis de las mutaciones de los exones 19 y 21 del EGFR se realizó basándonos en la técnica descrita por Qiulu P. et al (2005) con pequeñas modificaciones.

De manera general, se amplifican mediante dos PCR independientes los exones 19 y 21 del gen. Las mutaciones descritas en la literatura para el exón 19 son de tipo delección (la más frecuente implica la pérdida de 15 pb, aunque también se han descrito pérdidas de 12 pb, 18 pb y 21 pb) por lo que el diseño del estudio se basó en el análisis del tamaño de los productos generados en la PCR. En el caso del exón 21, la mutación más frecuente

(L858R) consiste en un cambio en el nucleótido 2573 que conlleva un cambio de timina (T) por guanina (G) con el consiguiente cambio de aminoácido, por lo que el diseño del estudio se basa en el análisis del tamaño de los productos generados en la PCR tras ser sometidos a restricción enzimática (PCR-RFLP -polimorfismo de longitud de fragmentos de restricción-) utilizando la enzima de restricción Sau96I (New England Biolabs. Ipswich. USA).

4.8.1.1. Mutaciones en el exón 19 de EGFR

Utilizando los primers EGFR-Ex19-FWD1 y EGFR-Ex19-REV1 se amplifica por PCR un fragmento de 206 pb que incluye todo el exón 19 (tabla 14).

EGFR-Ex19-FWD1	GCACCATCTCACAATTGCCAGTTA (24 bases)
EGFR-Ex19-REV1	(FAM-6)-AAAAGGTGGGCCTGAGGTTCA (21 bases)

Tabla 14. Secuencia de los primers para PCR del exón 19 de EGFR (5'→ 3'). El primer reverse se encuentra marcado con el fluorocromo FAM en su extremo 5'.

El primer reverse se encuentra marcado con el fluorocromo FAM-6 en su extremo 5'. La mezcla de reacción de PCR contiene: Biotaq DNA Polimerasa 5 U/μl. (Bioline. London. England), buffer 10xNH₄ (Bioline), MgCl₂ 50 mM (Bioline), dNTPs 200 μM cada uno (Applied Biosystems. Branchburg. New Jersey. EEUU), y primers 10 pmol/μl cada uno (Sigma-Aldrich. St. Louis. Missouri. EEUU). La PCR se llevó a cabo con los siguientes tiempos y temperaturas: 94°C durante 5 minutos, 35 ciclos de 94°C durante 30 segundos, 65°C durante 30 segundos y 72°C durante 40 segundos, y una extensión final a 72°C durante 10 minutos. El correcto funcionamiento de la PCR se comprueba en un gel de acrilamida al 6 %. Del volumen restante se realiza una dilución 1:20 en formamida y se añade 0.8 μl de marcador de peso molecular GeneScan 500(-250) LIZ (Applied Biosystems). Las muestras se desnaturalizan a 95°C durante 15 minutos y se mantienen a 4 °C durante 5 minutos. Se someten a una electroforesis capilar con polímero POP7 usando el analizador genético ABI 3100 (Applied Biosystems) y el programa de análisis de fragmentos GeneMapper versión 4.0 (Applied Biosystems).

En el análisis de los resultados observamos que las muestras con secuencia normal para el exón 19 de EGFR (wild type [WT]) muestran únicamente un pico que corresponde a 206 pb. Por el contrario, si la muestra tumoral analizada presenta una delección, además

del pico anterior obtendremos un pico adicional, que en el caso de la pérdida más frecuente (15 pb) correspondería al tamaño de 191 pb (figura 15).

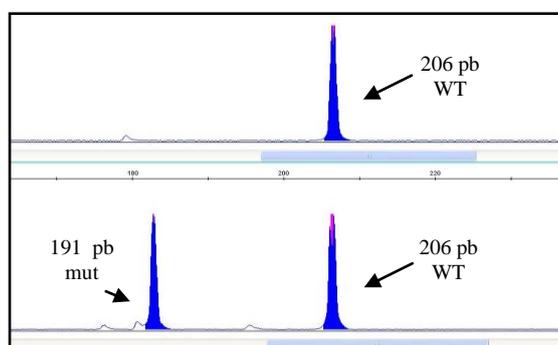


Figura 15. Imagen del programa GeneMapper. En la parte superior, muestra de CPNM que no presenta delección en el exón 19 de EGFR y en la parte inferior, muestra de CPNM que presenta delección de 15 pb en el exón 19 de EGFR.

4.8.1.2. Mutaciones en el exón 21 de EGFR

Se utilizan los primers EGFR-Ex21-FWD1 y EGFR-Ex21-REV1 que generan un producto de PCR de 222 pb. El primer reverse se encuentra marcado con el fluorocromo HEX-6 en su extremo 5' (tabla 15).

EGFR-Ex21-FWD1	CCTCACAGCAGGGTCTTCTCTGT (23 bases)
EGFR-Ex21-REV1	(HEX-6)-TCAGGAAAATGCTGGCTGACCTA (23 bases)

Tabla 15. Secuencia de los primers para PCR del exón 21 de EGFR (5'→ 3'). El primer reverse se encuentra marcado con el fluorocromo HEX en su extremo 5'.

Tanto la mezcla de reacción como las condiciones de la PCR son las mismas que las descritas para el exón 19 de EGFR. El correcto funcionamiento se comprueba en un gel de acrilamida al 6 %. Del volumen restante de PCR se utilizan 5 µl para llevar a cabo la restricción enzimática añadiéndoles 1,5 µl de Sau96I (New England Biolabs), 1,5 µl de buffer 4 (New England Biolabs) y 2 µl de agua. La mezcla de reacción se incuba durante 3 horas a 37°C en termobloque. Del producto de PCR-RFLP se realiza una dilución 1:3 en formamida y se añade 0.8 µl de marcador de peso molecular GeneScan 500(-250) LIZ (Applied Biosystems). Las muestras se desnaturalizan a 95°C durante 15 minutos y se mantienen a 4 °C durante 5 minutos. Se someten a una electroforesis capilar con polímero POP7 usando el analizador genético ABI 3100 (Applied Biosystems) y los resultados se obtuvieron utilizando el programa de análisis de fragmentos, GeneMapper versión 4.0 (Applied Biosystems).

En la secuencia WT del exón 21, existe un sitio de restricción para el enzima Sau96I Applied Biosystems. En los pacientes con la mutación L858R, como consecuencia del cambio de timina por guanina, aparece una segunda diana de restricción para este enzima. Así, al someter el producto de PCR de 222 pb del exón 21 a digestión con Sau96I (New England Biolabs), podemos diferenciar los pacientes con CPNM que poseen la mutación puntual ya que poseen dos dianas de restricción para el enzima.

Cuando la secuencia del exón 21 es normal, en el analizador de fragmentos aparece un único pico de 170 pb, consecuencia del corte de Sau96I (New England Biolabs) en la única diana de restricción existente. En aquellos pacientes con mutación L858R, además del pico anterior, aparecerá un segundo pico de 82 pb como consecuencia del corte de Sau96I (New England Biolabs) en la segunda diana de restricción generada por la existencia de la mutación puntual (figura 16).

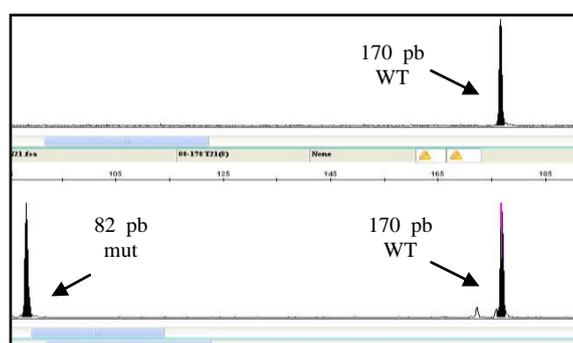


Figura 16. Imagen del programa GeneMapper. En la parte superior, muestra de CPNM que no presenta mutación puntual L858R en el exón 21 de EGFR y en la parte inferior, muestra de CPNM que presenta la mutación puntual L858R en el exón 21 de EGFR.

4.8.2. Estudio de mutaciones en el gen K-RAS

El análisis de las mutaciones del codón 12 y 13 del gen K-RAS se realizó siguiendo la técnica descrita por Hatzaki A et al (2001) con pequeñas modificaciones. De manera general el estudio se basa en el análisis del tamaño de los productos generados en la PCR tras ser sometidos a restricción enzimática (PCR-RFLP) utilizando la enzima de restricción BstN1 y Hae III (New England Biolabs) para el codón 12 y 13 de K-RAS respectivamente. Se utilizaron los primers mutados diseñados por este grupo los cuales poseen una base alterada ("mismatch") con respecto a la secuencia original, creando así un sitio de restricción artificial que permite diferenciar las secuencias mutadas de las no mutadas. Uno de los primers de cada pareja fue marcado con fluorocromo para llevar a cabo la posterior electroforesis capilar en el secuenciador ABI Prism 3100 con polímero (POP7). Para la

obtención de resultados se utilizó el programa de análisis “Gene Mapper version 4.0” (Applied Biosystems).

4.8.2.1. Mutaciones en el codón 12 de K-RAS

Utilizando los primers K-RAS-12-FWD1 y K-RAS-12-REV1 se amplifica por PCR un fragmento de 109 pb que incluye el codón 12 de K-RAS. El primer forward incluye una sustitución de una G por una C para crear un sitio de restricción para el enzima BstN1 y el primer reverse se encuentra marcado con el fluorocromo FAM-6 en su extremo 5’ (tabla 16).

K-RAS-12-FWD1	ACTGAATATAAACTTGTGGTAGTTGGAC <u>C</u> CT (30pb)
K-RAS-12-REV1	(FAM-6)-CCTCTATTGTTGGATCATATTCGTC (25pb)

Tabla 16. Secuencia de los primers para PCR del codón 12 de K-Ras (5’→ 3’). En el primer KRAS-12-FWD1 se sustituyó una G por una C y el primer KRAS-12-REV1 se encuentra marcado con el fluorocromo FAM en su extremo 5’.

La mezcla de reacción de PCR contiene: BioTaq DNA Polimerasa 5 U/μl (Bioline), buffer reacción de PCR 10xNH₄ (Bioline), MgCl₂ 50 mM (Bioline), dNTPs 200 μM cada uno (Applied Biosystems) y primers 10 pmol/μl cada uno (Sigma-Aldrich). La PCR se llevó a cabo con los siguientes tiempos y temperaturas: 94°C durante 5 minutos, 35 ciclos de 94°C durante 45 segundos, 58°C durante 45 segundos y 72°C durante 45 segundos, y una extensión final a 72°C durante 10 minutos. El correcto funcionamiento de la PCR se comprueba en un gel de acrilamida al 6 %.

Del volumen restante de PCR se utilizan 5 μl para llevar a cabo la restricción enzimática en la que se utiliza 1,5 μl de BstN1 (New England Biolabs), 1,5 μl de buffer 2 (New England Biolabs) y 2 μl de agua. La mezcla de reacción se incuba durante 3 horas a 37°C en termobloque. Del producto de PCR-RFLP se realiza una dilución 1:3 en formamida y se añade 0.8 μl de marcador de peso molecular GeneScan 500(-250) LIZ (Applied Biosystems).

Las muestras se desnaturalizan a 95°C durante 15 minutos y se mantienen a 4 °C durante 5 minutos. Se someten a una electroforesis capilar con polímero POP7 usando el analizador genético ABI 3100 y el programa de análisis de fragmentos GeneMapper versión 4.0 (Applied Biosystems).

El hecho de presentar la mutación heterocigota para el codón 12 de K-RAS implica que, en la hebra de DNA mutada, desaparece la diana de restricción para el enzima BstN1 (New England Biolabs) por lo que en el analizador de fragmentos observamos dos picos uno de 109 pb y otro de 80 pb que corresponden a la hebra de DNA con secuencia mutada y a la hebra de DNA con secuencia WT respectivamente. Por el contrario, en las muestras con secuencia WT para el codón 12 la diana de restricción para el enzima se mantiene con lo que es capaz de actuar sobre las dos hebras de DNA y en consecuencia aparece un único pico de 80 pb (figura 17).

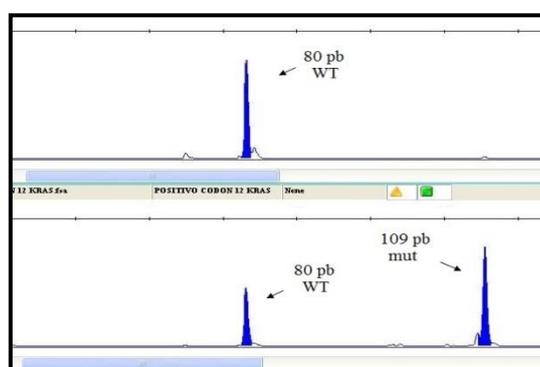


Figura 17. Imagen del programa GeneMapper. En la parte superior, muestra de CPNM que no presenta mutación en el codón 12 de K-RAS y en la parte inferior, muestra de CPNM que presenta mutación en el codón 12 de K-RAS.

4.8.2.2. Mutaciones en el codón 13 de K-RAS

Se utilizan los primers K-RAS-13-FWD1 y K-RAS-13-REV1 para amplificar por PCR un fragmento de 106 pb que incluye el codón 13 de K-RAS. El primer forward incluye una sustitución de una G por C para eliminar un sitio de restricción para el enzima HaeIII y se encuentra marcado con el fluorocromo TET-6 en su extremo 5'. El primer reverse incluye una sustitución de una C por una G para crear un sitio de restricción para el enzima HaeIII (New England Biolabs) (tabla 17).

K-RAS-13-FWD1	(TET-6)-TAA <u>C</u> GCCTGCTGAAAATGACTG (22pb)
K-RAS-13-REV1	GTATCGTCAAGGCACTCTTGCCTA <u>G</u> G (26pb)

Tabla 17. Secuencia de los primers para PCR del codón 13 de K-RAS (5'→3'). En el primer K-RAS-13-FWD1 se sustituyó una G por una C y se encuentra marcado en su extremo 5' con fluorescencia y en el primer K-RAS-13-REV1 se sustituyó una C por una G.

Tanto la mezcla de reacción como las condiciones de la PCR son las mismas que las descritas para el codón 12 de K-RAS. El correcto funcionamiento se comprueba en un gel

de acrilamida al 6 %. Del volumen restante de PCR se utilizan 5 µl para llevar a cabo la restricción enzimática añadiéndoles 1,5 µl de HaeIII (New England Biolabs), 1,5 µl de buffer 4 (New England Biolabs) y 2 µl de agua. La mezcla de reacción se incuba durante 3 horas a 37°C en termobloque. Del producto de PCR-RFLP se realiza una dilución 1:3 en formamida y se añade 0.8 µl de marcador de peso molecular GeneScan 500(-250) LIZ (Applied Biosystems).

Las muestras se desnaturalizan a 95°C durante 15 minutos y se mantienen a 4 °C durante 5 minutos. Se someten a una electroforesis capilar con polímero POP7 usando el analizador genético ABI 3100 (Applied Biosystems) y los resultados se obtuvieron utilizando el programa de análisis de fragmentos GeneMapper versión 4.0 (Applied Biosystems). El hecho de presentar la mutación en el codón 13 de K-RAS implica que desaparece un sitio de restricción para el enzima Hae III (New England Biolabs) por lo que en el análisis de los resultados observamos que las muestras con mutación heterocigota para el codón 13 de K-RAS muestran dos picos, uno de 79 pb y otro de 53 pb que corresponden a la hebra de DNA con secuencia mutada y a la hebra de DNA con secuencia WT respectivamente. Por el contrario, en las muestras con secuencia WT para el codón 13 en ambas hebras de DNA, la diana de restricción para el enzima se mantiene con lo que es capaz de actuar y en consecuencia aparece un único pico de 53 pb (figura 18).

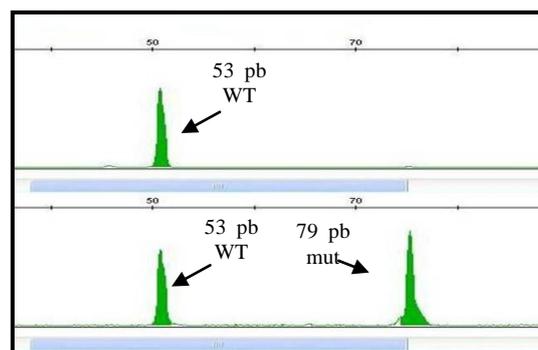


Figura 18. Imagen del programa GeneMapper. En la parte superior, muestra de CPNM que no presenta mutación en el codón 13 de K-RAS y en la parte inferior, muestra de CPNM que presenta mutación en el codón 13 de K-RAS.

4.8.3. Estudio de mutaciones en el gen B-RAF

El análisis de la mutación T1796A (exón 15) se realizó siguiendo la técnica descrita por Cohen y cols. (2003). De manera general el diseño del estudio se basa en el análisis del tamaño de los productos generados en la PCR tras ser sometidos a restricción enzimática

(PCR-RFLP). Se utilizaron los primers B-RAF-Ex15-FWD1 y B-RAF-Ex15-REV en los que el primer forward fue marcado con fluorocromo FAM-6 (Tabla 18).

B-RAF-Ex15-FWD1	(FAM-6)-GCTTGCTCTGATAGGAAAATGAG (23 bases)
B-RAF-Ex15-REV1	GTA ACTCAGCAGCATCTCAGG (21 bases)

Tabla 18. Secuencia de los primers para PCR de B-RAF (5' → 3'). El primer B-RAF-Ex15-FWD1 se encuentra marcado con fluorescencia en su extremo 5

La mezcla de reacción de PCR contiene: BioTaq DNA Polimerasa 5 U/μl (Bioline), buffer reacción de PCR 10xNH₄ (Bioline), MgCl₂ 50 mM (Bioline), dNTPs 200 μM cada uno (Applied Biosystems) y primers 10 pmol/μl cada uno (Sigma-Aldrich).

La PCR se llevó a cabo con los siguientes tiempos y temperaturas: 94°C durante 5 minutos, 35 ciclos de 94°C durante 45 segundos, 58°C durante 45 segundos y 72°C durante 45 segundos, y una extensión final a 72°C durante 10 minutos. El correcto funcionamiento de la PCR se comprueba en un gel de acrilamida al 6 %.

Del volumen restante de PCR se utilizan 5 μl para llevar a cabo la restricción enzimática en la que se utiliza 1,5 μl de TspR1 (New England Biolabs), 1,5 μl de buffer 2 (New England Biolabs) y 2 μl de agua. La mezcla de reacción se incuba durante 3 horas a 65°C en termobloque. Del producto de PCR-RFLP se realiza una dilución 1:3 en formamida y se añade 0.8 μl de marcador de peso molecular GeneScan 500(-250) LIZ (Applied Biosystems). Las muestras se desnaturalizan a 95°C durante 15 minutos y se mantienen a 4 °C durante 5 minutos. Se someten a una electroforesis capilar con polímero POP7 usando el analizador genético ABI 3100 (Applied Biosystems) y el programa de análisis de fragmentos GeneMapper versión 4.0 (Applied Biosystems).

El hecho de presentar la mutación T1796A implica la desaparición de la diana de restricción del enzima TspR1 (New England Biolabs) obteniéndose una banda prominente de 212 pb correspondiente al alelo mutado y las bandas residuales del alelo normal (125 pb y 87 pb). Cuando la secuencia es WT, la diana de restricción del enzima no ha desaparecido, con lo que es capaz de digerir el fragmento amplificado obteniéndose únicamente los dos fragmentos de 125 pb y 87 pb (figura 19).

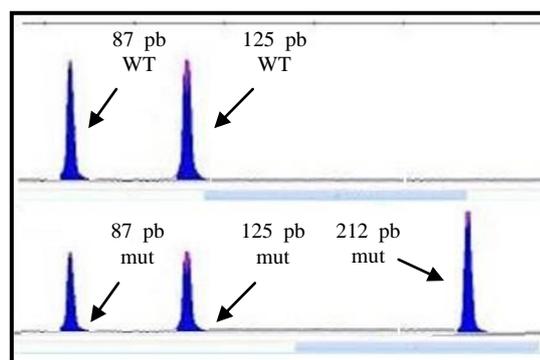


Figura 19. Imagen del programa GeneMapper. En la parte superior, muestra de CPNM que no presenta mutación en el exón 15 de B-RAF y en la parte inferior, muestra de CPNM que presenta mutación en el exón 15 de B-RAF.

4.9. ESTUDIOS DE ASOCIACIÓN

Los estudios de asociación entre los subtipos moleculares y los parámetros clínico-histopatológicos entre sí y de ambos con el ILE se realizaron utilizando el programa estadístico SPSS v13.0 (SPSS Inc., Chicago, EEUU).

Las variables cuantitativas se describen como media \pm desviación standard y las variables cualitativas se describen con su distribución de frecuencias.

La asociación entre los subtipos moleculares y los parámetros clínico-histopatológicos se llevó a cabo utilizando métodos no paramétricos para múltiples categorías (test de Kruskal-Wallis). Las comparaciones por pares cuando fueron necesarias se llevaron a cabo mediante el test U-Mann-Whitney.

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.

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.

El Intervalo Libre de Enfermedad (ILE) se define como el momento desde la cirugía en el que confirma la primera recidiva. Un dato es censurado cuando un paciente muere o sigue vivo sin recidiva en el último contacto. Las curvas Kaplan-Meier (K-M) y el test log-rank se utilizaron para calcular la probabilidad de estar libre de recidiva para cada subgrupo molecular y los otros parámetros evaluados en el estudio. El modelo de riesgos

proporcionales de Cox se utilizó para estimar la distribución de supervivencia y los Hazard Ratios (HR). Por último, un modelo multivariante de Cox se desarrolló incluyendo aquellas variables con $p < 0.1$ en el análisis univariante y variables pronósticas que se consideraron clínicamente relevantes.

4.10. ESTUDIO DE LAS VÍAS MOLECULARES ALTERADAS ENTRE LOS GRUPOS MOLECULARES OBTENIDOS

Se realizó un análisis de las vías moleculares que se encuentran alteradas de manera significativa entre los grupos moleculares obtenidos utilizando la herramienta GSEA (“Gene Set Enrichment Análisis” o análisis de enriquecimiento de conjuntos de genes) disponible en GS11.5 y utilizando los *C2-Functional Gene Sets* (disponibles en la página web de Broad Institute <http://www.broadinstitute.org/gsea/downloads.jsp>) que incluye vías moleculares de las bases de datos Biocarta y Kegg. Sólo se evaluaron las vías moleculares con una representación mínima de más de 15 genes y se utilizaron 100.000 permutaciones para asegurar los resultados. Se utilizó el listado original de genes (17.881 genes), puesto que cuando se analizan vías moleculares conviene incluir todos los genes disponibles que cumplan los controles de calidad, ya que diferencias no significativas de expresión en un grupo de genes pueden sin embargo ser claves, para definir qué caminos de señalización (“pathways”) están alterados entre los grupos. Diferencias con un p-valor ≤ 0.05 fueron consideradas estadísticamente significativas.

4.11. VALIDACIÓN EXTERNA DE LA CLASIFICACIÓN MOLECULAR Y OBTENCIÓN DEL PREDICTOR

Se realizó una validación de la clasificación molecular obtenida con las muestras del HCSC utilizando los resultados procedentes de un estudio de microarrays previamente publicado por el grupo de Roepman et al. (2009).

Los criterios para la selección de esta publicación como matriz externa de validación fueron:

- Estudio desarrollado en pacientes con CNMP.
- Incluye los mismos subtipos histológicos y los mismos estadios iniciales de la enfermedad.

- Utilización de la misma plataforma de microarrays.
- Los datos de expresión génica se encontraban públicamente disponibles.

El proceso de validación de la clasificación molecular obtenida en la matriz de desarrollo mediante el estudio de microarrays de expresión y la generación de un predictor para dicha clasificación molecular constó de las fases que a continuación se detallan:

- **Fusión de las dos matrices de datos de expresión en una única matriz.**

Se utilizó el programa “Distance Weighted Discrimination” [DWD] (<https://genome.unc.edu/pubsup/dwd/>) (Hu et al., 2006) que se basa en técnicas multivariantes para identificar y eliminar los efectos propios de las diferentes plataformas, eliminando los sesgos de la fase experimental existentes en ambas matrices y haciendo posible el análisis conjunto de varias matrices.

- **Obtención de un predictor para la clasificación molecular de las muestras.**

Se utilizó la aplicación “Prediction Analysis of Microarrays” [PAM] (Tibshirani et al, 2002) para la obtención del predictor capaz de clasificar las muestras de la matriz externa de validación o matriz de Roepman en los grupos moleculares generados en la matriz de entrenamiento o matriz del HCSC. La técnica PAM se basa en el método del centroide más cercano o “shrunken centroid”. El punto de partida es el cálculo de un “valor de referencia” o “shrunken centroid” para cada uno de los grupos moleculares generados en la matriz de desarrollo o matriz de entrenamiento ya que fue en ella en la que se definió lo que era el grupo de buen y mal pronóstico. Una vez calculado el “valor de referencia” o “shrunken centroid” para cada grupo, el PAM es capaz de asignar las nuevas muestras de la matriz de validación a cada uno de los grupos previamente definidos al calcular la distancia entre los valores de expresión de los 50 genes de la nueva muestra con respecto a las 50 componentes del “valor de referencia” o “shrunken centroid” de cada grupo. Si la distancia entre la nueva muestra y el “valor 1 de referencia” es menor que la distancia entre la nueva muestra y el “valor 2 de referencia”, se podrá determinar el pronóstico favorable para el nuevo paciente. Por el contrario, si la distancia entre la nueva muestra y el “valor 1 de referencia” es

mayor que la distancia entre la nueva muestra y “valor 2 de referencia”, se podrá determinar el pronóstico desfavorable para el nuevo paciente. Es decir, la nueva muestra se asignará al grupo molecular de cuyos valores de expresión esté más próximo. Además la técnica PAM permite seleccionar automáticamente los genes con mayor poder predictivo.

- **Evaluación del valor pronóstico del predictor obtenido.** Se llevó a cabo comparando el ILE entre los clusters o grupos moleculares obtenidos en la matriz externa de validación o matriz de Roepman mediante las curvas de K-M y el modelo de riesgos proporcionales de Cox.

4.12. ESTUDIO DE ENRIQUECIMIENTO DE TÉRMINOS “GENE ONTOLOGY” EN EL PREDICTOR PRONÓSTICO

Se utilizó la base de datos DAVID (Database for Annotation, Visualization and Integrated Discovery) que incluye un conjunto de herramientas de anotación funcional para entender el significado biológico de largas listas de genes al identificar los términos de Gene Ontology (GO) más relevantes asociados a dicha lista mediante el proceso conocido como enriquecimiento de términos GO (<http://www.geneontology.org/>).

En el sistema de anotación de DAVID, se utiliza el test de Fisher para identificar el enriquecimiento de términos GO, teniendo en cuenta que un p-valor ≤ 0.01 indica una asociación estadísticamente significativa.

5. RESULTADOS

5.1. PRE-SELECCIÓN DE PACIENTES

El estudio comenzó con un total de 105 pacientes que cumplían los criterios de inclusión referentes al diagnóstico, técnica quirúrgica, tratamiento y seguimiento descritos en el apartado 2, selección de pacientes, dentro de la sección material y métodos de esta tesis. De este total, 2 pacientes (PT3 y PT29) tuvieron que ser descartados por no disponer de material congelado en el Banco de Tumores del HCSC y 6 pacientes (PT31, PT35, PT41, PT55, PT60 Y PT63) ya que, aun existiendo material tumoral congelado en el Banco de Tumores, la cantidad remanente tras la utilización de estos bloques en otros proyectos no fue suficiente.

5.2. TÉCNICA DE MICROARRAYS Y CONTROLES DE CALIDAD

En la técnica de microarrays existen etapas no estandarizadas que únicamente dependen del criterio que el investigador quiera aplicar en cada una de ellas. En estas etapas, el diseño experimental debe ser aún más cuidadoso que en el resto para poder obtener unos resultados fiables y libres de sesgos. A continuación se describen estas etapas críticas no estandarizadas y los controles de calidad establecidos en cada una de ellas para el desarrollo de este proyecto.

5.2.1. Análisis histológico de las muestras tumorales

Tras la revisión en el microscopio de las HE obtenidas de las muestras tumorales de los 97 pacientes previamente seleccionados, se descartaron un total de 13 pacientes (PT9, PT22, PT25, PT32, PT42, PT53, PT65, PT67, PT74, PT77, PT81, PT87 y PT91) ya que sus correspondientes muestras congeladas en el Banco de Tumores presentaron un porcentaje de tejido tumoral inferior al mínimo establecido (70%). Tras el análisis histológico se incluyeron un total de 84 muestras que mostraron un porcentaje medio de células tumorales del 80.9%.

5.2.2. Control de calidad del RNA extraído

Es importante destacar que ninguna muestra de las 84 preseleccionadas y finalmente incluidas en el estudio se descartó por no alcanzar el criterio mínimo de calidad exigido a los RNAs extraídos ($RIN \geq 7.5$). Por tanto, la calidad de los RNAs que se incluyeron en el estudio fue muy elevada obteniéndose un RIN medio de 8.6 (en una escala de 0-10) con un mínimo de 7.6 (PT15) y un máximo de 9.7 (PT52) (figura 20).

Estos datos son un reflejo directo de la profesionalidad del personal implicado y del óptimo diseño de los Procedimientos Normalizados de Trabajo (PNT) que se utilizan para la congelación de las muestras que llegan al Servicio de Anatomía Patológica y que son incluidas en el Banco de Tumores del HCSC.

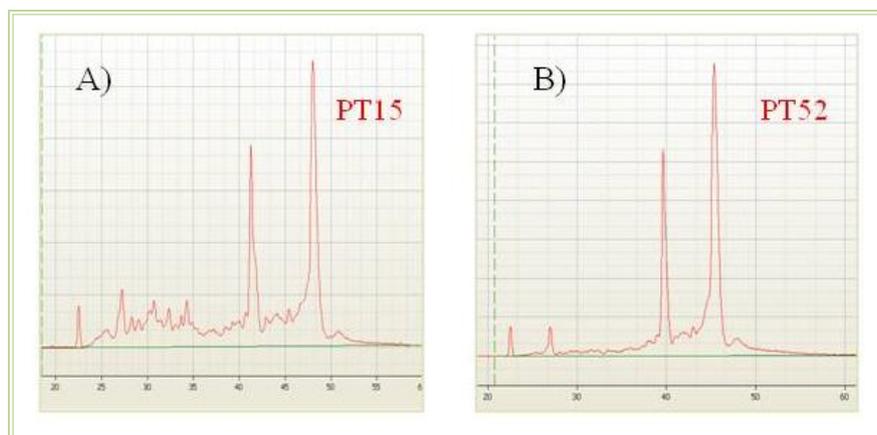


Figura 20. Representación del rango de calidad de los RNAs utilizados en el estudio. A) Muestra PT15 con un valor de RIN cercano al límite mínimo de calidad aceptado (≥ 7.5). **B)** muestra PT52 con un valor de RIN cercano al límite máximo de calidad posible (10).

5.2.3. Gráficos MA-Plots

Durante el estudio se llevaron a cabo un total de 92 hibridaciones mediante doble marcaje (Cy5 y Cy3) que comprendían las 84 muestras tumorales incluidas en el estudio a las que además se sumaron 8 tejidos de parénquima pulmonar no tumoral. El diseño de las hibridaciones fue el que se describe a continuación:

- 84 hibridaciones de cada uno de los 84 tumores frente a la muestra de referencia o “pool” (Tumor-Cy5/Pool-Cy3) (figura 21-A).
- 8 hibridaciones de los 8 tejidos de parénquima pulmonar no tumoral frente a la muestra de referencia o “pool” (Normal-Cy5/Pool-Cy3) (figura 21-B).

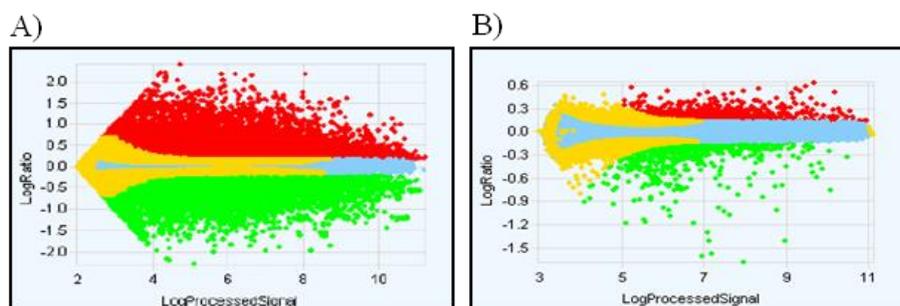


Figura 21. Pie de figura en página siguiente

Figura 21. Representación mediante MA-Plot de los genes diferencialmente expresados en los dos tipos de hibridaciones realizadas. A) Tejido tumoral vs. muestra de referencia o “pool”. B) Tejido parénquima pulmonar no tumoral vs. muestra de referencia o “pool”. El color rojo indica genes sobreexpresados 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 .

5.2.4. Estudio de comportamientos anómalos de las muestras mediante PCA

Tal y como muestra el PCA de la figura 22, ni en las muestras tumorales analizadas, ni en las muestras de parénquima pulmonar no tumoral que formaron parte de la muestra de referencia o “pool” se observaron sesgos significativos debidos a la semana en que se realizó la hibridación, al resultado del RIN del RNA extraído o la tasa de incorporación de Cy5 en las muestras tumorales y de Cy3 para las muestras del “pool”.

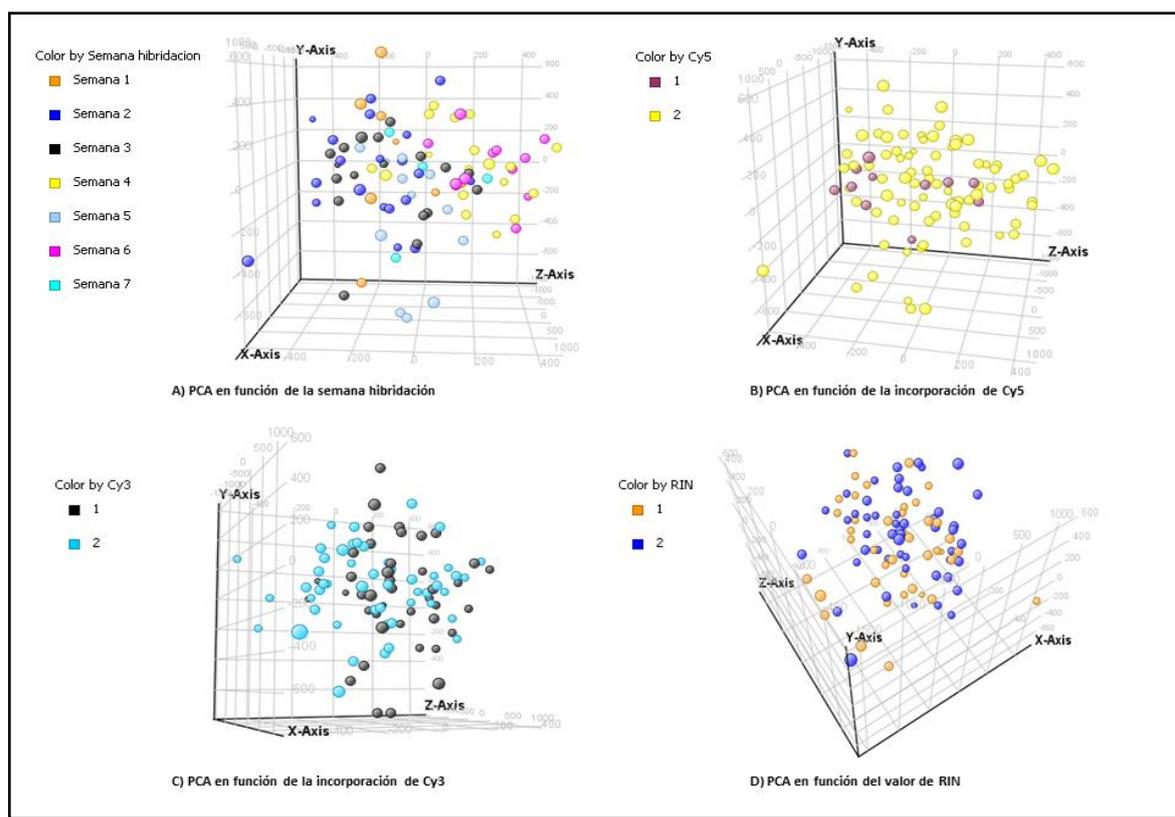


Figura 22. Análisis de componentes principales o PCA. La figura muestra la distribución de las muestras del estudio una vez realizado el PCA en función de **A)** semana de hibridación **B)** incorporación de Cy5 **C)** incorporación de Cy3 **D)** valor de RIN.

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

5.3. CARACTERÍSTICAS DE LOS 84 PACIENTES INCLUIDOS EN EL ESTUDIO

Los datos relativos a las variables clínico-histopatológicas y al estudio mutacional de los 84 pacientes que cumplieron los criterios de inclusión y superaron los controles de calidad para ser finalmente incluidos en el estudio se muestran en la tabla 19.

Variable		Nº Pacientes	Porcentaje
Edad (años)	Media (min-max)	66.5 (36-82)	---
Sexo	Varón	72	86
	Mujer	12	14
Hábito de fumar*	No fumador	5	6
	Ex-fumador	42	51
	Fumador	35	43
Subtipo Histológico	Adenocarcinoma	40	48
	Epidermoide o escamoso	39	46
	Carcinoma células grandes	2	2
	Carcinoma adenoescamoso	3	4
Afectación ganglios linfáticos	N0	81	96
	N1	3	4
Estadio (7ª edición AJCC)	Estadio I	60	71
	Estadio II	24	29
Recidiva	Ausente	55	66
	Presente	29	34
Localización del tumor	Pulmón derecho	49	58
	Pulmón izquierdo	35	42
Grado de diferenciación	Pobre	37	44
	Moderado	25	30
	Bien	22	26
Estroma Tumoral	Bajo	22	26
	Medio	46	55
	Alto	16	19
Necrosis	Baja	57	68
	Alta	27	32
Queratinización	Ausente	76	90
	Presente	24	10
TILs	Ausente	31	37
	Presente	53	63
Inflamación crónica	Baja	67	80
	Moderada-Fuerte	17	20
Inflamación aguda (PMN)	Ausente	48	57
	Presente	36	43

Variable		Nº Pacientes	Porcentaje
Mutaciones <i>K-RAS</i>	Wild Type	73	87
	Mutado	11	13
Mutaciones <i>EGFR</i>	Mutado	83	99
	Mutado	1	1
Mutaciones <i>B-RAF</i>	Wild Type	84	100
	Mutado	0	0

Tabla 19. Descripción de los parámetros clínico-histopatológicos y estudio mutacional de los pacientes incluidos en el estudio. (* El hábito de fumar para 2 pacientes no se conoce).

La edad media de diagnóstico del tumor en los pacientes incluidos en el estudio fue de 66.5 años. En nuestra serie, la relación de pacientes en función del sexo demostró que el CNMP es un tumor más frecuente en el sexo masculino (86% hombres) que en el femenino (14% mujeres). Respecto al hábito de fumar, considerado el principal factor etiológico en el cáncer de pulmón, únicamente el 6% de los pacientes incluidos fueron no fumadores mientras que el 43% de ellos fueron fumadores. En la serie también se incluyeron un 51% de pacientes exfumadores en el momento del diagnóstico, considerando como exfumador al paciente que había dejado el hábito tabáquico al menos un año antes del diagnóstico de la enfermedad.

Tal y como aparece descrito en la literatura, los subtipos histológicos más frecuentes del CNMP representaron el 94% de los tumores. En nuestra serie, el subtipo AC representó el 48% de los tumores analizados y el SCC representó el 46%. El resto de subtipos histológicos del CNMP menos frecuentes, LCC y el carcinoma adenoescamoso, representaron un 4% y un 2% respectivamente. Únicamente el 4% de los pacientes presentaron afectación ganglionar en el momento del diagnóstico.

Respecto al estadio tumoral, un 71% de los pacientes fueron diagnosticados de CNMP en estadio I mientras que el 29% restante fue diagnosticado de CNMP en estadio II. En la serie de estudio, la tasa de recidiva fue del 34% y en cuanto a la localización, los tumores analizados parecieron desarrollarse de manera más frecuente en el pulmón derecho (58%) que en el pulmón izquierdo (42%).

En relación al grado de diferenciación del tumor, un 44% de las muestras presentaron un grado de diferenciación pobre y un 55% de los tumores presentaron una cantidad media de estroma tumoral. La necrosis fue baja en un 68% de los tumores y la queratinización estaba ausente en un 90% de los tumores.

Respecto a las variables relacionadas con la respuesta inmune intratumoral es importante destacar que un 63% de los tumores mostraron linfocitos intratumorales (TILs) y signos de inflamación aguda se observaron en un 43% de los casos. Respecto a los signos de inflamación crónica fueron detectados en la mayoría de los pacientes pero sólo en un 20% de ellos fueron clasificados con signos moderados-fuertes.

En cuanto a las mutaciones estudiadas, ninguno de los tumores presentaron mutaciones en *B-RAF* y únicamente un 1% mostraron mutaciones en *EGFR*. Respecto a *K-RAS*, fue la mutación génica más ampliamente detectada en un 13% de los casos.

5.3.1. Estudios de asociación entre las variables clínico-histopatológicas y mutacionales con el ILE

En este estudio no se encontró asociación estadísticamente significativa entre ninguna de las variables clínico-histopatológicas estudiadas con el ILE. Solamente la presencia de mutaciones de *K-RAS* mostró tendencia respecto al ILE con una $p=0.072$ (tabla 20). No se llevó a cabo ni el estudio de asociación ni de las mutaciones de *EGFR*, por presentar una incidencia muy baja, ni de las mutaciones de *B-RAF*, por estar ausentes, en los CNMP incluidos en el estudio.

Variable	Asociación con ILE	Resultado
Edad	0.676	No significativo
Sexo	0.357	No significativo
Hábito de fumar	0.954	No significativo
Subtipo Histológico	0.733	No significativo
Afectación ganglios linfáticos	0.699	No significativo
Estadio (7ª edición AJCC)	0.989	No significativo
Localización del tumor	0.199	No significativo
Grado de diferenciación	0.784	No significativo
Estroma Tumoral	0.234	No significativo
Necrosis	0.235	No significativo

Variable	Asociación con ILE	Resultado
Queratinización	0.399	No significativo
TILs	0.425	No significativo
Inflamación crónica	0.523	No significativo
Inflamación aguda (PMN)	0.668	No significativo
Mutaciones <i>K-RAS</i>	0.072	No significativo

Tabla 20. Asociación entre las variables clínico-histopatológicas y el ILE. Ninguna de las variables clínico-histopatológicas mostró asociación estadísticamente significativa con el ILE. Únicamente las mutaciones de *K-RAS* mostraron tendencia respecto ILE.

Las curvas de Kaplan-Meier para las dos principales características histopatológicas, el estadio tumoral y la clasificación histopatológica, no mostraron diferencias estadísticamente significativas en nuestra serie (figura 23).

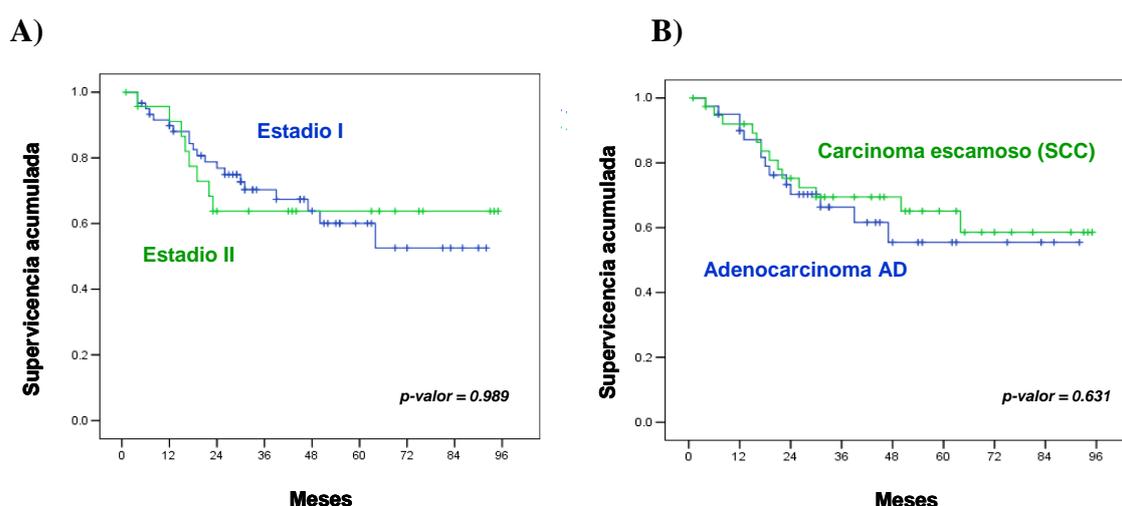


Figura 23. Curvas de Kaplan-Meier. A) Probabilidad de ILE en función del estadio (I o II) del CNMP. B) Probabilidad de ILE en función del subtipo histológico (únicamente representados los dos subtipos histológicos principales del CNMP que en nuestro estudio representan el 94% de las muestras tumorales incluidas).

5.4. ANÁLISIS DE MICROARRAYS

El estudio de los perfiles de expresión génica global se llevó a cabo a partir de 84 muestras tumorales y 8 muestras de parénquima pulmonar no tumoral. Los dos tipos de muestras anteriormente descritas pasaron los diferentes controles de calidad establecidos para el desarrollo del presente proyecto. En la matriz final se incluyeron un total de 17.881 genes según los criterios de filtrado explicados en la sección 3.7.2 (filtrado de datos). Para el propósito de la clasificación y caracterización tumoral sólo fueron utilizados aquellos

genes que mostraban una mayor variación en la expresión a lo largo de las muestras del estudio, utilizando de esta forma, en una primera aproximación, un grupo más reducido de 3.232 genes que finalmente disminuyó hasta los 2.160 genes, listado obtenido al eliminar los genes diferencialmente expresados entre AD y SCC.

5.4.1. Caracterización del patrón de expresión de muestras tumorales y muestras de parénquima pulmonar no tumoral

Aun no siendo el objetivo principal de este estudio, el estudio comparativo entre el tejido tumoral y el tejido de parénquima pulmonar no tumoral nos permitió verificar la capacidad discriminadora de la técnica de microarrays. Mediante técnicas no supervisadas (“clustering jerárquico”) realizamos una agrupación de muestras (84 tumores y las 8 muestras de parénquima pulmonar no tumoral) y de genes. El resultado obtenido, como cabía esperar, separó por un lado los tejidos tumorales y por otro lado las muestras de parénquima pulmonar no tumoral y nos permitió identificar los patrones de expresión diferencial entre ambos tipos de muestras (figura 24). Con este experimento se confirmó la solidez de la técnica de microarrays para detectar alteraciones basadas en perfiles globales de expresión apoyando por tanto su utilización en la creación de una clasificación molecular del CNMP en base a estos perfiles.

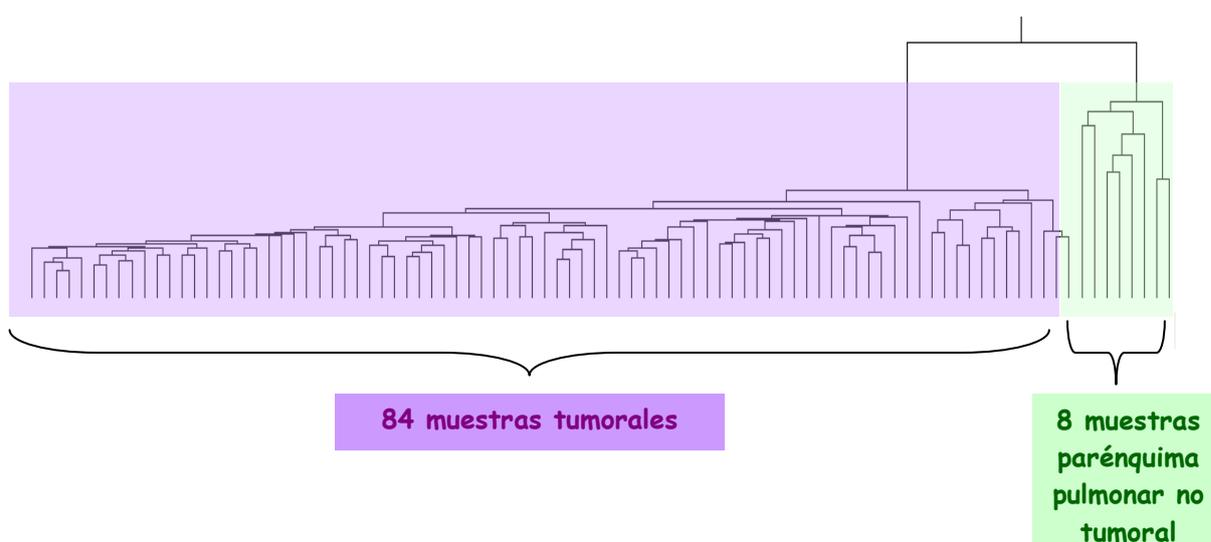


Figura 24. “Clustering” jerárquico. La figura muestra el resultado del “clustering” jerárquico utilizando las 84 muestras tumorales y las 8 muestras de parénquima pulmonar no tumoral. (*Correlación centrada de Pearson, “average linkage” y centrado de genes*).

La validación matemática mediante el D-Index y el R-Index de las agrupaciones obtenidas en el “clustering” jerárquico confirmaron que, molecularmente el tejido normal y el tejido de parénquima pulmonar no tumoral son completamente diferentes entre sí, siendo capaces de asignar correctamente las muestras a su grupo molecular correspondiente en el 100% de los casos, como muestran los valores de robustez (siendo igual a 1) y de omisiones y adiciones (siendo igual a 0) en ambos grupos.

La existencias de grandes diferencias entre tejido tumoral y tejido normal ya habían sido observadas cuando se analizaron los resultados de expresión mediante los gráficos MA-plot.

5.4.2. Métodos no supervisados para la clasificación molecular de los estadios iniciales del CNMP

5.4.2.1. Clasificación molecular del CNMP utilizando el listado de 3.232 genes

La clasificación molecular para distribuir los tumores en grupos molecularmente homogéneos se llevó a cabo utilizando los 84 tumores incluidos en el estudio y el listado de 3.232 genes obtenido del filtrado por expresión (variación de expresión de al menos 3 veces con respecto a la media de ese gen en el 10% de las muestras).

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 pasó 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” posteriormente mejorado mediante la técnica de “K-means”. En ambas técnicas se utilizó como distancia la correlación centrada de Pearson y como método de agrupación de elementos el “average linkage”.

El estudio del dendrograma obtenido tras el análisis mediante “clustering jerárquico” permitió diferenciar dos grupos moleculares o clusters en lo referente a la distribución de las muestras (figura 25):

- El Cluster A (cluster amarillo)
- El Cluster B (cluster naranja)

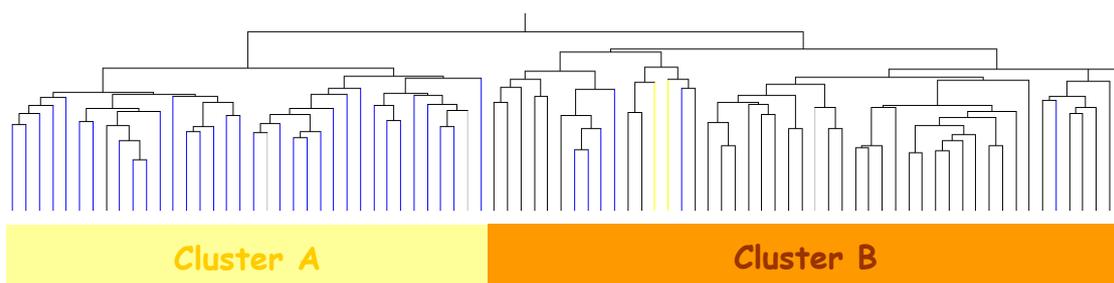


Figura 25. Resultado del “clustering” jerárquico de las 84 muestras tumorales analizadas en función de su perfil molecular con el listado de 3.232 genes. Las muestras se encuentran coloreadas en función del subtipo histológico al que pertenecen. Línea azul: subtipo SCC. Línea negra: subtipo AD. Línea amarilla: carcinoma de célula grande. Línea gris: carcinoma adenoescamoso. La clasificación molecular mostró dos cluster principales fuertemente asociados a la histología de los tumores. El Cluster A incluye un elevado número de CEP mientras que el Cluster B incluye la mayoría de los AD.

El número de grupos moleculares obtenidos y seleccionados mediante criterio visual en el árbol se utilizó como base para mejorar esos grupos mediante la técnica de “K-means”. El Cluster A englobó un total de 45 muestras y el Cluster B un total de 39 muestras. El 98% de los SCC fueron clasificados en el Cluster A y el 92% de los AC fueron clasificados en el Cluster B. Respecto a los tres carcinomas adenoescamosos, uno de ellos fueron clasificados en el Cluster A y los dos restantes en el Cluster B y los dos carcinomas de célula grande fueron clasificados en el Cluster A (figura 26).

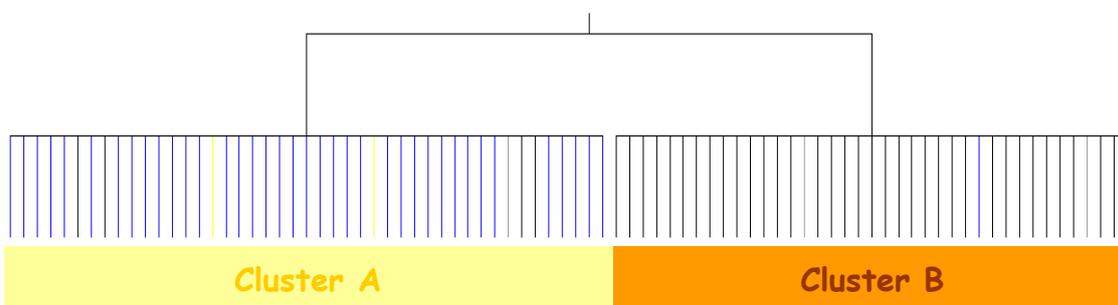


Figura 26. Resultado del “K-Means” obtenido tras el desarrollo de esta técnica sobre los dos grupos moleculares o clusters obtenidos a partir del “clustering” jerárquico de las 84 muestras tumorales analizadas en función de su perfil molecular con el listado de 3.232 genes. Las muestras se encuentran coloreadas en función del subtipo histológico al que pertenecen. Línea azul: subtipo SCC. Línea negra: subtipo AD. Línea amarilla: carcinoma de célula grande. Línea gris: carcinoma adenoescamoso. El Cluster A incluye un elevado número de SCC mientras que el Cluster B incluye la mayoría de los AD.

La curva de Kaplan-Meier no mostró diferencias estadísticamente significativas respecto al ILE entre los clusters A y B obtenidos ($p=0.350$) (figura 27).

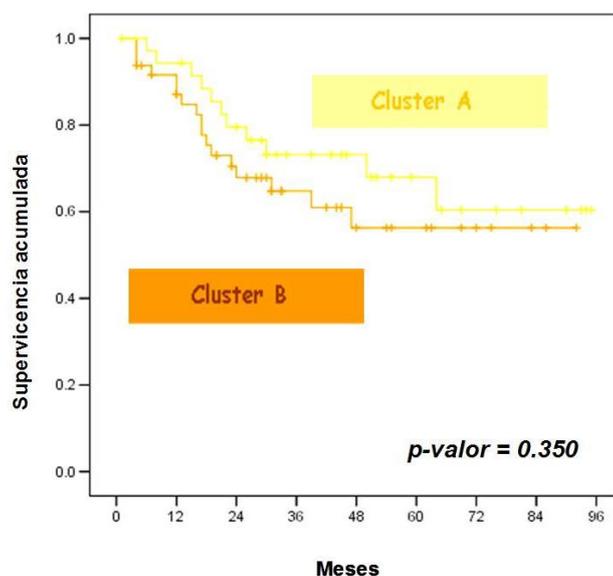


Figura 27. Representación mediante la curva de Kaplan-Meier de los dos clusters obtenidos mediante la técnica de “k-Means”. El gráfico muestra que no existe diferencia estadísticamente significativa respecto al ILE entre los dos clusters moleculares obtenidos con el listado de 3.232 genes.

Como ya se había descrito en el apartado 4.3.1 (estudios de asociación entre las variables clínico-histopatológicas e ILE – página 70), en la serie de 84 pacientes con CNMP en estadios iniciales incluidos en el estudio, ni el estadio tumoral ni la clasificación en subtipos histológicos mostraron asociación estadísticamente significativa con el ILE.

En la clasificación molecular generada para estos 84 CNMP en estadios iniciales utilizando el listado de 3.232 genes se puede observar, de manera muy clara, que dicha clasificación se encuentra condicionada por la histología de los tumores (Cluster A o amarillo incluye el 98% de los SCC y Cluster B o naranja incluye el 92% de los AC). Aunque SCC y ACs son evidentemente tumores distintos a nivel morfológico y molecular, como ya hemos visto anteriormente, tienen un comportamiento clínico similar, lo que podría explicar que no haya diferencias estadísticamente significativas respecto al ILE entre los dos clusters moleculares obtenidos.

Como consecuencia de estos hallazgos se decidió generar una nueva clasificación molecular no relacionada con la histología del CNMP para posteriormente evaluar su implicación biológica y pronóstica. La obtención de esta nueva clasificación molecular se vio apoyada por el criterio oncológico que se utiliza para el tratamiento del paciente con CNMP y según el cual, la clasificación en subtipos histológicos sólo es importante en la enfermedad metastásica (estadio IV) y sólo en relación con el tratamiento de elección para el paciente.

5.4.2.2. Clasificación molecular del CNMP, no relacionada con la histología, utilizando el listado de 2.160 genes

La clasificación molecular no relacionada con los subtipos histológicos se llevó a cabo utilizando los 84 tumores incluidos en el estudio y el listado de 2.160 genes (obtenidos a partir del listado de 3.232 genes y tras la eliminación de aquellos diferencialmente expresados entre los dos principales subtipos histológicos del CNMP).

De nuevo, 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” mejorado posteriormente mediante la técnica “K-means”. En ambas técnicas se utilizó como distancia la correlación centrada de Pearson y como método de agrupación de elementos el “average linkage”.

El estudio del dendrograma obtenido tras el análisis mediante “clustering jerárquico” permitió diferenciar tres grupos moleculares o clusters (Cluster 1, Cluster 2 y Cluster 3) en lo referente a la distribución de las muestras (figuras 28 y 29).

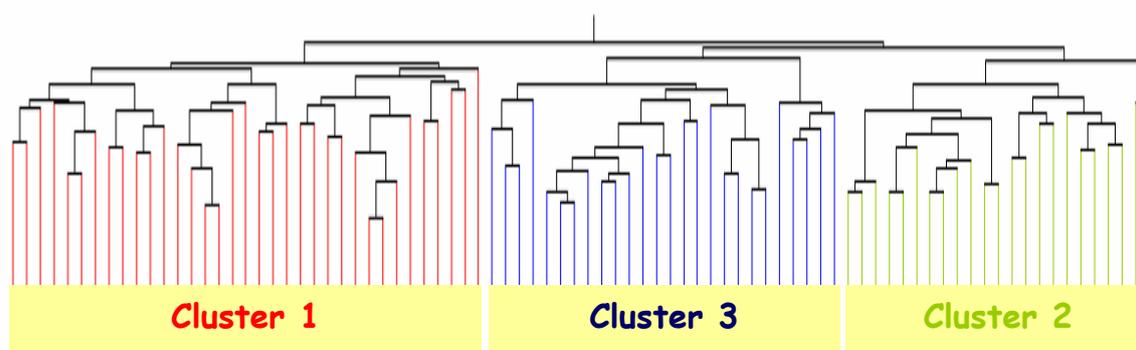


Figura 28. Resultado del “clustering jerárquico” utilizando 2.160 genes en las 84 muestras tumorales. La figura muestra el “clustering jerárquico” resultante para las 84 muestras utilizando el listado de 2.160 genes. Las 84 muestras se distribuyen en 3 clusters (Cluster 1, Cluster 2 y Cluster 3). En 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”.

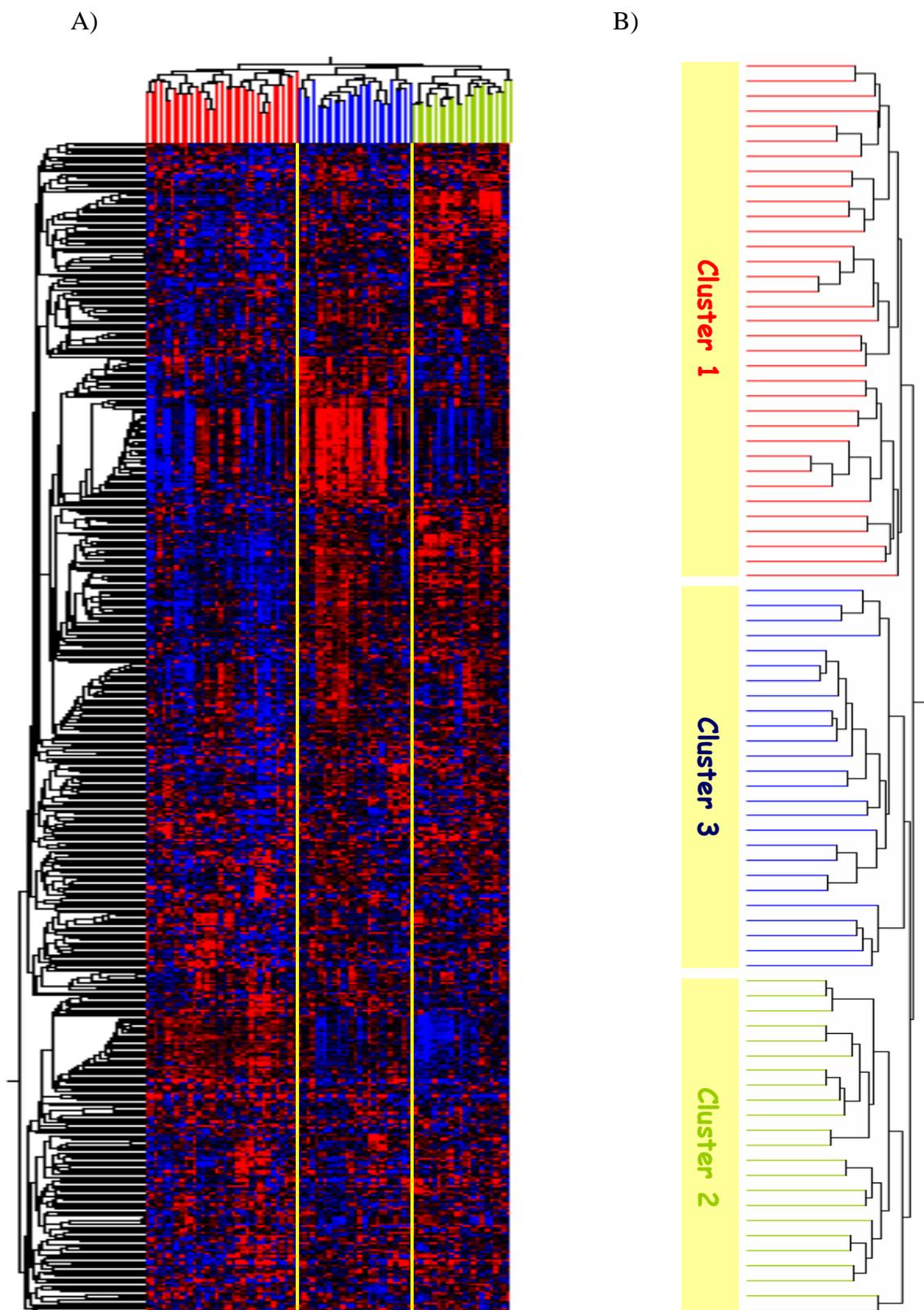


Figura 29. Pie de figura en página siguiente.

Figura 29. Heatmap y “clustering jerárquico” utilizando 2.160 genes en las 84 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)** En horizontal y en la parte superior se muestra el “clustering jerárquico” de las 84 muestras. En vertical se muestra la matriz de expresión o “heatmap”, ordenada tras la realización de un “clustering” jerárquico para los 2.160 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. **B)** Ampliación del “clustering” jerárquico de muestras en el que se describen los 3 grupos moleculares (Cluster 1, Cluster 2 y Cluster 3).

Los resultados obtenidos de la validación matemática, tras la perturbación de la matriz (100.000 veces), reflejaron que las agrupaciones propuestas no eran debidas al azar. Como muestra la tabla 21, los tres grupos moleculares o clusters definidos poseen una reproducibilidad en cuanto a la forma, R-Index o robustez, superior al 85% (en una escala de 0-100%), siendo en el caso del Cluster 2 superior al 90%. En cuanto a la estabilidad en las muestras (D-Index) que conformaron cada uno de los 3 clusters, se observó que en ninguno de los casos, el número promedio de muestras que tras ser asignadas a un determinado cluster cambió de agrupación al alterar la matriz superó las 3.5 muestras por grupo (omisiones). En cuanto al número promedio de muestras que tras ser incluidas en cualquiera de los otros dos cluster fueron incluidas en el grupo de referencia al alterar la matriz (adiciones), fue incluso más bajo ya que no superaron las 1.9 muestras por grupo.

Clusters	Numero de elementos	Robustez	Omisiones	Adiciones	Reproducibilidad Global R-Index = 0.87 D-Index = 5.745
Cluster 1	35	0.857	3.16	1.791	
Cluster 2	23	0.921	0.948	1.872	
Cluster 3	26	0.853	2.243	1.476	

Tabla 21. Resultado de la validación matemática (tras 100.000 perturbaciones) para cada uno de los grupos moleculares obtenidos. A la derecha se muestra la reproducibilidad global de todos los grupos analizados.

El número de clusters o grupos moleculares obtenidos mediante el “clustering jerárquico” fue el que se utilizó posteriormente en la técnica de “K-Means” en la que obtuvimos tres clusters o grupos moleculares mejorados respecto a los obtenidos anteriormente al conseguir disminuir la heterogeneidad intra-cluster y aumentar la variabilidad inter-cluster. El Cluster 1 englobó un total de 30 muestras, el Cluster 2 un total de 25 muestras y el Cluster 3 un total de 29 muestras (figura 30).

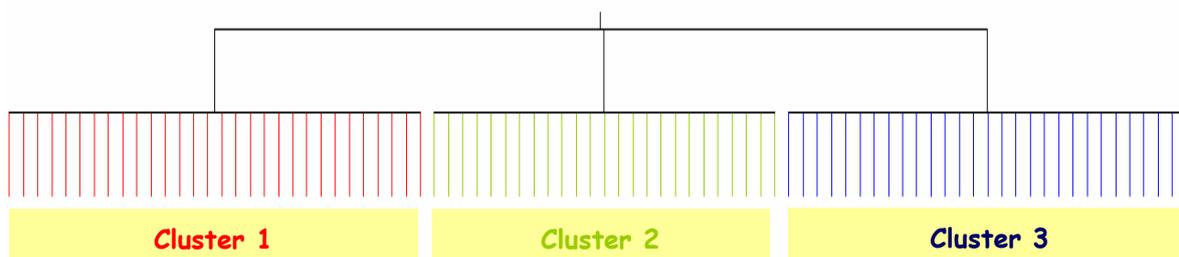


Figura 30. Representación de los tres clusters obtenidos mediante la técnica de “k-Means”. El Cluster 1 incluyó 30 muestras, el Cluster 2 incluyó 25 muestras y el Cluster 3 un total de 29 muestras.

La curva de Kaplan-Meier mostró diferencias estadísticamente significativas respecto al ILE ($p=0.004$) entre los tres clusters obtenidos mediante “clustering jerárquico” mejorado con técnica de “K-Means”, indicando una menor probabilidad de recidiva para el Cluster 3 respecto del Cluster 2 y cluster 1 (figura 31).

Mediante el modelo de riesgos proporcionales de Cox se obtuvieron los Hazard ratios para cada uno de los cluster de mal pronóstico (*Cluster 1 vs Cluster 3*. $HR=6.4$ (IC 95%: 1.8-22.3; $p=0.004$) y *Cluster 2 vs Cluster 3*. $HR=4.9$ (IC 95%: 1.4-17.8; $p=0.014$)) respecto del cluster de buen pronóstico.

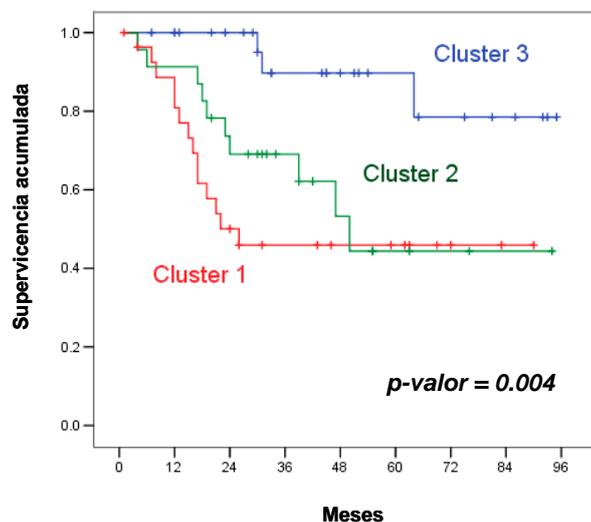


Figura 31. Representación mediante la curva de Kaplan-Meier de los tres clusters obtenidos mediante “clustering jerárquico” mejorado mediante la técnica de “k-Means”. La figura muestra la diferencia estadísticamente significativa respecto al ILE entre los tres clusters obtenidos. El Cluster 3 es el grupo molecular con mejor pronóstico frente al Cluster 1 y Cluster 2.

5.4.2.2.1. Estudios de asociación de la clasificación molecular no relacionados con la histología y las variables clínico-histopatológicas y mutacionales convencionales.

Mediante este estudio se encontró que existe asociación estadísticamente significativa de los tres clusters no relacionados con la histología generados a partir del listado de 2.160 genes y el hábito tabáquico ($p=0.002$), la afectación ganglionar ($p=0.041$) y la inflamación crónica ($p=0.002$) (tabla 22). El tamaño y el estadio tumoral mostraron tendencia con $p=0.066$ y $p=0.083$ respectivamente.

Variable	3 clusters con 2160 genes	Resultado
Edad	0.109	No significativo
Sexo	0.598	No significativo
Hábito de fumar	0.002	Significativo
Subtipo Histológico	0.254	No significativo
Localización del tumor	0.676	No significativo
Grado de diferenciación	0.331	No significativo
Tamaño tumor (T)	0.066	No significativo
Afectación ganglios linfáticos (N)	0.041	Significativo
Estadio tumoral	0.083	No significativo
Mutaciones K-Ras	0.704	No significativo
Necrosis	0.111	No significativo
Estroma Tumoral	0.745	No significativo
Inflamación crónica	0.002	Significativo
TILs	0.384	No significativo
Queratinización	0.328	No significativo
PMN	0.703	No significativo

Tabla 22. Resultado del estudio de asociación entre la clasificación molecular en tres clusters y las variables clínico-histopatológicas y mutacionales incluidas en el estudio. La tabla muestra asociación estadísticamente significativa entre la clasificación en tres clusters no relacionados con la histología y tres de las variables clínico-patológicas incluidas en el estudio (hábito de fumar, afectación ganglionar e inflamación crónica).

Los siguientes gráficos de barras (figura 32) muestran la distribución de las variables con resultado significativo en relación a los tres clusters en los que se llevó a cabo la clasificación molecular de los tumores.

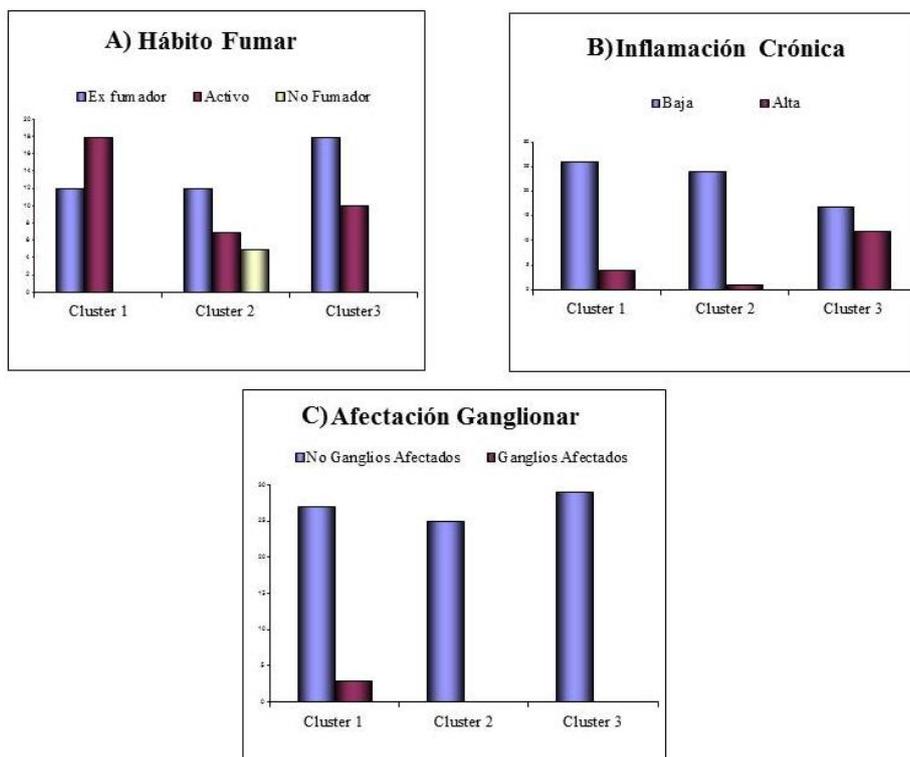


Figura 32. Representación mediante gráficos de barras de las variables clinicopatológicas con resultado significativo en los tres clusters moleculares obtenidos. A) Hábito de fumar: en morado se representan los pacientes exfumadores, en granate los pacientes fumadores activos y en dorado los pacientes no fumadores **B) Inflamación crónica:** en morado se representan los pacientes cuyo tumor fue clasificado con inflamación crónica baja y en granate los pacientes cuyo tumor fue clasificado con inflamación crónica alta. **C) Afectación ganglionar:** en morado se representan los pacientes sin afectación en ganglios linfáticos y en granate los pacientes con afectación en ganglios linfáticos.

El Cluster 1 es que el incluye el mayor número de muestras tumorales provenientes de pacientes que fuman de manera activa en el momento del diagnóstico. Además incluye los tres únicos tumores con afectación ganglionar y un elevado número de tumores caracterizados con inflamación crónica baja o leve.

El Cluster 2 se caracteriza por incluir los cinco únicos pacientes no fumadores de todo el estudio y también un elevado número de tumores caracterizados con inflamación crónica baja o leve.

El Cluster 3 es el que incluye el mayor número de pacientes exfumadores en el momento del diagnóstico y el que incluye el mayor número de tumores caracterizados con una inflamación crónica alta.

5.4.2.2.2. Estudio de las vías moleculares alteradas entre los tres clusters obtenidos

Este análisis reveló un total de 34 vías moleculares diferencialmente expresadas entre los 3 grupos moleculares obtenidos (tabla 23. Pie de tabla en página 89). Entre el Cluster 1 (alto riesgo de recidiva) y el Cluster 3 (bajo riesgo de recidiva) observamos expresión diferencial en 29 vías moleculares en las que, la gran mayoría de ellas, se encuentran relacionadas con **sistema inmune** como las vías de señalización y activación del receptor de células T (M19784 y M6327/M1467/M18215), vías de polarización y desarrollo de respuesta de células T (M4319 y M4047), apoptosis de células T mediada por MEF2D (M5290), vía de señalización del receptor de células B (M9494), vías de respuesta a inflamación y citoquinas y transducción de señales mediada por IL-7 (M6910 y M1296 y citotoxicidad mediada por células NK (M16355). También se encuentran alteradas diferentes **vías de señalización celular** como la vía de la hormona del crecimiento (M9043), motilidad celular mediada por Rac 1 (M8601), vía de los fosfolípidos (M10579), vía de señalización mediada por ATM (M10628), señalización en mastocitos (M1908) o vía de señalización mediada por proteínas G (M9664). Por último, entre estos dos clusters también se encuentran alteradas **vías de control del ciclo celular** (G2/M Checkpoint M8560) y **regulación de la replicación del DNA por kinasas dependientes de ciclinas (CDK)** (M6682).

Las principales diferencias entre el Cluster 2 (alto riesgo de recidiva) y el Cluster 3 se centran en vías moleculares relacionadas con **sistema inmune** (polarización y desarrollo de respuesta de células T (M4319 y M4047), vías de señalización y activación del receptor de células T (M19784 y M6327 / M1467), vía de señalización del receptor de células B (M9494)) y **señalización mediada por IL** (transducción de señales mediada por IL-7 (M1296) y vías de señalización y activación mediada por IL-2 (M7747 y M8615)).

Por último, las vías moleculares alteradas entre los dos cluster de alto riesgo de recidiva se centran en **control del ciclo celular** (G2/M Checkpoint M8560), **regulación de la replicación del DNA por CDK** (M6682) y **vías de señalización** (señalización en mastocitos a través de la fracción Fc Epsilon del receptor I (M1908), vía de señalización mediada por ATM (M10628) y vía de señalización del TNFR2 (M10082)). Además, encontramos diferencias en **vías de susceptibilidad al cáncer dependientes de BRCA1, BRCA2 y ATR** (M9703) y **activación del EGFR** (M7552).

		Cluster 1 vs. Cluster 3	Cluster 1 vs. Cluster 2	Cluster 2 vs. Cluster 3
M7747	IL 2 signaling pathway			
M8615	IL-2 Receptor Beta Chain in T cell Activation			
M4319	IL12 and Stat4 Dependent Signaling Pathway in Th1 Development			
M1296	IL-7 Signal Transduction			
M6327	Activation of Csk by cAMP-dependent Protein Kinase Inhibits Signaling through the T Cell Receptor			
M19784	T Cell Receptor Signaling Pathway			
M1467	The Co-Stimulatory Signal During T-cell Activation			
M4047	Selective expression of chemokine receptors during T-cell polarization			
M9494	B Cell Receptor Signaling Pathway			
M18215	Role of Tob in T-cell activation			
M5290	Role of MEF2D in T-cell Apoptosis			
M16355	Ras-Independent pathway in NK cell-mediated cytotoxicity			
M6910	Cytokines and Inflammatory Response			
M2349	Pertussis toxin-insensitive CCR5 Signaling in Macrophage			
M8601	Rac 1 cell motility signaling pathway			
M9043	Growth Hormone Signaling Pathway			
M11420	CTCF: First Multivalent Nuclear Factor			
M1547	Control of skeletal myogenesis by HDAC and calcium/calmodulin-dependent kinase (CaMK)			
M3430	Effects of calcineurin in Keratinocyte Differentiation			
M11650	Nitric Oxide Signaling Pathway			
M16811	Role of Erk5 in Neuronal Survival			
M10287	fMLP induced chemokine gene expression in HMC-1 cells			
M8731	Aspirin Blocks Signaling Pathway Involved in Platelet Activation			
M2288	NFAT and Hypertrophy of the heart (Transcription in the broken heart)			
M9664	Signaling Pathway from G-Protein Families			
M15181	Regulation of PGC-1a			
M10579	Phospholipids as signalling intermediaries			
M1908	Fc Epsilon Receptor I Signaling in Mast Cells			
M8560	Cell Cycle: G2/M Checkpoint			
M6682	CDK Regulation of DNA Replication			
M10628	ATM Signaling Pathway			
M9703	Role of BRCA1, BRCA2 and ATR in Cancer Susceptibility			
M7552	Role of EGF Receptor Transactivation by GPCRs in Cardiac Hypertrophy			
M10082	TNFR2 Signaling Pathway			

Tabla 23. Listado de las 34 vías moleculares diferencialmente expresadas entre los tres clusters moleculares obtenidos. La tabla muestra la codificación y la descripción de las vías moleculares alteradas obtenidas en el análisis de GSEA. El color verde indica que esa vía se encuentra diferencialmente expresada ($p \leq 0.05$) entre los dos clusters moleculares comparados.

5.5. VALIDACIÓN DE LA NUEVA CLASIFICACIÓN MOLECULAR EN TRES CLUSTERS NO RELACIONADOS CON LA HISTOLOGÍA UTILIZANDO UNA MATRIZ DE DATOS EXTERNA

La validación de una nueva clasificación mediante el uso de datos externos independientes, matriz de validación, 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 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, para este propósito, la matriz de datos publicada por el grupo de Roepman et al. (2009).

5.5.1. Características clínico-histopatológicas de las muestras de la matriz de validación

La serie que conformó la matriz de validación estaba formada por pacientes que cumplían criterios de inclusión muy similares a los establecidos para la matriz de desarrollo y sobre las que se llevaron a cabo estudios con microarrays de doble marcaje y utilizando la misma plataforma que en nuestro proyecto. La serie publicada por este grupo incluía un total de 172 muestras provenientes de pacientes operados de CPNM en estadio I o II, que no fueron sometidos a tratamiento adyuvante a la cirugía, de los que se disponía de un seguimiento en consulta de al menos 2 años y de los que se tenía almacenado material tumoral congelado en el Banco de Tumores.

Respecto a los subtipos histológicos que incluía, hay que destacar que para poder comparar los pacientes de ambas series, únicamente utilizamos en nuestro estudio aquellas muestras cuyo subtipo histológico coincidía con los incluidos en nuestra matriz de desarrollo. Por ello tuvimos que prescindir de un total de 10 muestras y conformar la matriz de validación con 162 muestras que incluían 92 SCC (57%), 58 AC (36%),

10 LCC (6%) y 2 carcinomas adenoescamosos (1%). En cuanto al análisis histológico de las muestras tumorales al microscopio, el grupo de Roepman et al. incluyó en el estudio toda muestra que al menos presentara un 50% de representación tumoral. La tasa de recidiva post-quirúrgica en la matriz de validación fue del 34.5% ya que un total de 56 pacientes desarrollaron recidiva tumoral posterior a la resección quirúrgica con linfadenectomía

5.5.2. Combinación de la matriz de desarrollo y de la matriz de validación utilizando la técnica Distance Weighted Discrimination DWD

Para llevar a cabo la validación era necesario poder aplicar de manera precisa los algoritmos resultantes de un estudio en el otro y viceversa. Esto implicaba eliminar previamente los sesgos de la fase experimental existentes en ambas matrices. Tras combinar los dos conjuntos mediante el método Distance Weighted Discrimination DWD (Hu et al.; 2006) se obtuvo una matriz de expresión con 246 muestras y 17.881 genes.

Se llevó a cabo un análisis no supervisado de todas las muestras tumorales utilizando el listado de 2.160 genes. Lo primero que se observó tras la realización del “clustering jerárquico” fue la validez del método DWD para eliminar los sesgos y conseguir que las muestras provenientes de las dos matrices se entremezclaran de manera notable (figura 33).

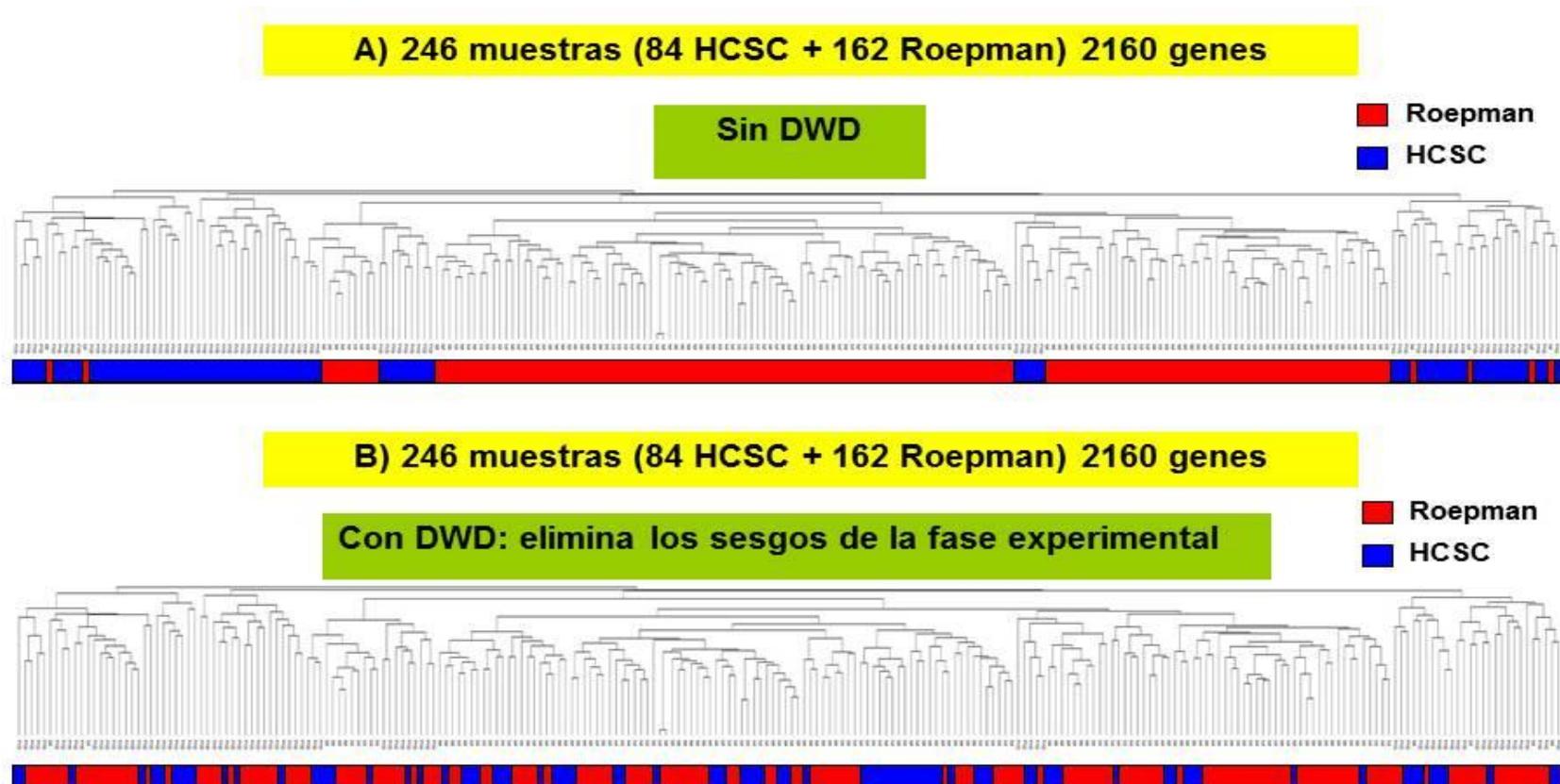


Figura 33. Resultado del “clustering jerárquico” utilizando las muestras (un total de 246) de las dos matrices (HCSC y Roepman) y el listado de 2160 genes. En azul se identifican las muestras pertenecientes a la matriz de desarrollo (HCSC) y en rojo las muestras procedentes de la matriz de validación (Roepman). A) “Clustering” jerárquico sin utilizar el programa DWD. B) “Clustering” jerárquico tras utilizar el método DWD para fusión de matrices de datos.

5.5.3. Obtención y validación de un predictor génico para la clasificación molecular en tres clusters no relacionados con la histología

Para confirmar el valor pronóstico asociado a la clasificación molecular en tres clusters, se desarrolló un predictor génico utilizando como base la matriz del HCSC o matriz de desarrollo y aplicando dicho predictor a la matriz externa o matriz de validación obtenida del estudio de Roepman et al. (2009).

A partir del listado de 2.160 genes y utilizando la técnica Prediction Analysis of Microarrays (PAM) se generó un predictor que incluyó 1.000 genes y que se aplicó en la matriz de validación externa identificando en ella los tres clusters moleculares obtenidos en la matriz de desarrollo, Cluster 1, Cluster 2 y Cluster 3 que incluyeron un total de 71, 40 y 51 muestras respectivamente.

La curva de K-M mostró que la nueva clasificación molecular en tres clusters está asociada de manera estadísticamente significativa con el ILE ($p=0.022$) (figura 34). Mediante el modelo de riesgos proporcionales de Cox se obtuvieron los Hazard Ratios [HR] para cada uno de los cluster de mal pronóstico (*Cluster 1 vs Cluster 3. HR=2.26 (IC 95%: 1.13-4.52; $p=0.012$)* y *Cluster 2 vs Cluster 3. HR=2.51 (IC 95%: 1.14-5.50; $p=0.019$)*) respecto del cluster de buen pronóstico, una vez ajustado por el estadio tumoral.

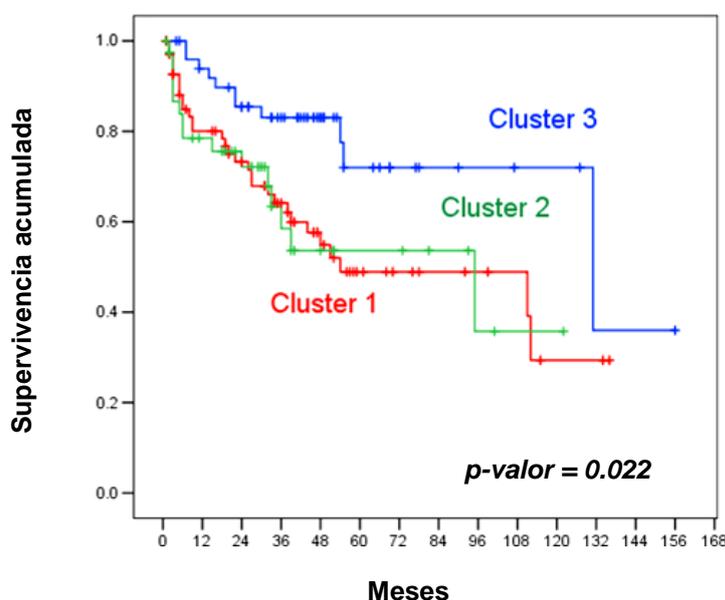


Figura 34. Representación mediante la curva de Kaplan-Meier de los tres clusters obtenidos en la matriz de validación utilizando el predictor de 1.000 genes. La figura representa la supervivencia acumulada de los cluster 1, 2 y 3 obtenidos en los que se dividen los 162 pacientes de la matriz de validación al aplicar el predictor de 1.000 genes. Se obtuvo una $p=0.022$ lo que muestra que esta clasificación molecular está asociada de manera estadísticamente significativa con el ILE.

5.6. SIMPLIFICACION DE LA CLASIFICACIÓN MOLECULAR NO RELACIONADA CON LA HISTOLOGÍA Y OBTENCION DE UN PREDICTOR DE RECIDIVA.

Ya que no existen diferencias estadísticamente significativas respecto al ILE entre los dos clusters de mal pronóstico obtenidos utilizando el listado de 2.160 genes (Cluster 1 y Cluster 2; $p = 0.526$), se simplificó la clasificación molecular de tres a dos clusters, uno de mal pronóstico o alto riesgo de recidiva (Cluster 1 + Cluster 2) y otro de buen pronóstico o bajo riesgo de recidiva (Cluster 3) para así facilitar la identificación de las claves moleculares implicadas en las recidivas de los pacientes en estadios I y II del CNMP (figura 35).

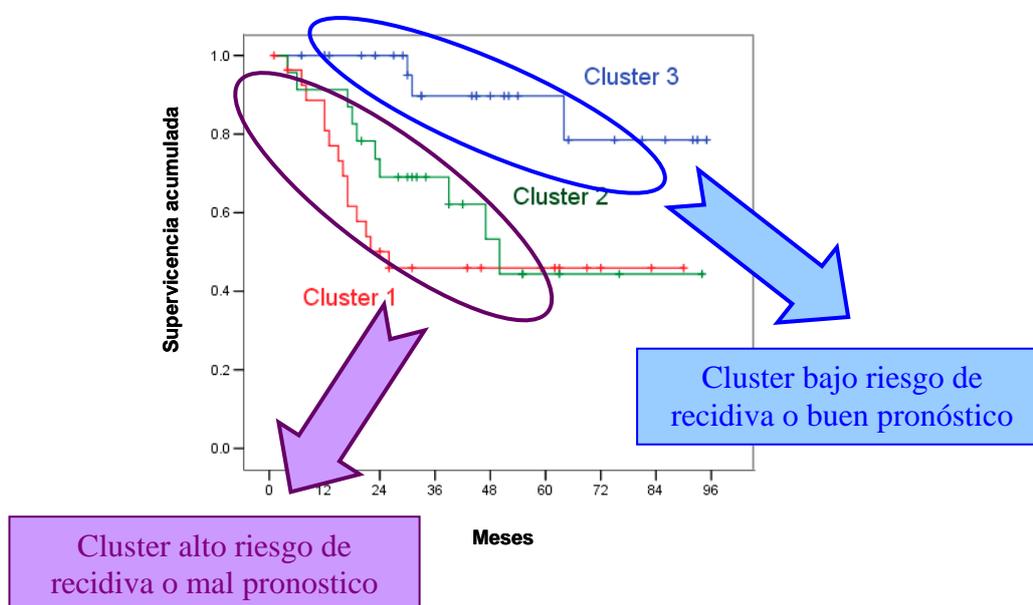


Figura 35. Representación sobre la curva de Kaplan-Meier de la transformación de los tres clusters obtenidos en la clasificación molecular con 2.160 genes a dos clusters. La figura muestra la unificación de los dos clusters de mal pronóstico en un único denominado cluster de alto riesgo de recidiva o mal pronóstico frente al cluster de bajo riesgo de recidiva o buen pronóstico.

5.6.1. Obtención y validación de un predictor de recidiva.

Utilizando la matriz de desarrollo con la clasificación molecular en dos clusters (punto 5.6) y aplicando la técnica PAM, se generó un predictor de 50 genes que se utilizó para clasificar los pacientes de la matriz de validación en uno de los dos grupos moleculares definidos (alto riesgo de recidiva o mal pronóstico y bajo riesgo de recidiva o buen pronóstico).

Una vez asignados los 162 pacientes de la matriz de validación a uno de estos dos grupos moleculares, se llevó a cabo el estudio estadístico para conocer si esta distribución de pacientes se asocia o no de manera estadísticamente significativa al ILE. Las curvas de K-M mostraron que entre los dos cluster existen diferencias estadísticamente significativas respecto al ILE ($p=0.001$) (figura 36). El modelo de riesgos proporcionales de Cox confirmó la habilidad de nuestro predictor de 50 genes para distinguir entre pacientes de alto y bajo riesgo de recidiva (HR=3.44; IC 95%: 1.6-7.3; $p=0.001$).

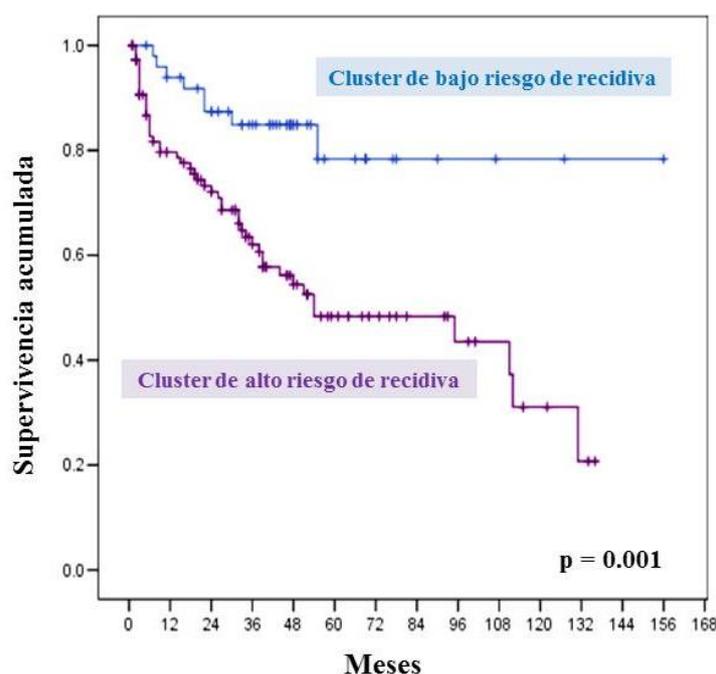


Figura 36. Representación mediante la curva de Kaplan-Meier de los dos clusters obtenidos en la matriz de validación utilizando el predictor de 50 genes. La curva muestra que existen diferencias estadísticamente significativas respecto al ILE entre el cluster de alto riesgo de recidiva o mal pronóstico y el cluster de bajo riesgo de recidiva o buen pronóstico.

Se calcularon los valores de sensibilidad, especificidad, valor predictivo positivo (VPP) y valor predictivo negativo (VPN) para la clasificación molecular con el predictor de 50 genes que se muestran en la tabla 24.

Grupos Moleculares	Sensibilidad	Especificidad	VPP	VPN
Bajo riesgo de recidiva	0,966	0,891	0,824	0,98
Alto riesgo de recidiva	0,891	0,966	0,98	0,824

Tabla 24. Valores de Sensibilidad, Especificidad, VPP y VPN para el predictor de 50 genes

A continuación se muestra la tabla 25 que incluye los 50 genes que forman parte del predictor que clasifica los pacientes con CNMP totalmente resecaos y en estadios iniciales de la enfermedad en dos clusters de alto y bajo riesgo de recidiva. Es importante destacar que estos 50 genes están sobreexpresados en el grupo de bajo riesgo respecto del grupo de alto riesgo de recidiva.

ID Entrez	Símbolo	Descripción
270	<i>AMPD1</i>	Adenosina monofosfato deaminasa
608	<i>TNFRSF17</i>	Miembro 17 de la superfamilia del receptor de necrosis tumoral
930	<i>CD19</i>	CD19
939	<i>CD27</i>	CD27
952	<i>CD38</i>	CD38
973	<i>CD79A</i>	Molécula alfa asociada a inmunoglobulina
974	<i>CD79B</i>	Molécula beta asociada a inmunoglobulina
3002	<i>GZMB</i>	Granzima B
3493	<i>IGHA1</i>	Inmunoglobulina pesada constante alfa 1
3494	<i>IGHA2</i>	Inmunoglobulina pesada constante alfa 2
3500	<i>IGHG1</i>	Inmunoglobulina pesada constante gamma 1
3512	<i>IGJ</i>	Polipéptido J de inmunoglobulina
3535	<i>IGL@</i>	Locus lambda de inmunoglobulina
3543	<i>IGLL1</i>	Polipéptido 1 "like" inmunoglobulina lambda
3662	<i>IRF4</i>	Factor 4 regulador del interferón
3782	<i>KCNN3</i>	Miembro 3 subfamilia N de canales de potasio activados por calcio
3887	<i>KRT81</i>	Keratina 81
4283	<i>CXCL9</i>	Ligando 9 del motivo CXC de quimioquinas
5368	<i>PNOC</i>	Preonocipeptina
5450	<i>POU2AF1</i>	Factor 1 asociado a POU clase 2
8419	<i>BFSP2</i>	Faquinina
9834	<i>KIAA0125</i>	KIAA0125
10563	<i>CXCL13</i>	Ligando 13 del motivo CXC de quimioquinas
11040	<i>PIM2</i>	Oncogén Pim2
26952	<i>SMR3A</i>	Proteína 3A regulada por glándula submaxilar
28904	<i>IGKVID-8</i>	Cadena variable kappa 1D-8 de inmunoglobulina
51237	<i>MZB1</i>	Proteína específica de las células B1 y B de la zona marginal
51303	<i>FKBP11</i>	Proteína 11 de unión a FK506
54900	<i>LAX1</i>	Adaptador 1 transmembrana de linfocitos
57699	<i>CPNE5</i>	Copina V
57823	<i>SLAMF7</i>	Miembro 7 de la familia SLAM
78986	<i>DUSP26</i>	Fosfatasa 26 específica dual
79368	<i>FCRL2</i>	"like" receptor 2 Fc
80307	<i>FER1L4</i>	Pseudogen 4 "like" fer-1
83416	<i>FCRL5</i>	Receptor Fc "like" 5
84824	<i>FCRLA</i>	Receptor Fc "like" A
90925	<i>IGHV5-78</i>	Pseudogen región pesada variable de cadena pesada de inmunoglobulina
91319	<i>DERL3</i>	Miembro 3 de la familia de dominio parecido a Der-1
92154	<i>MTSS1L</i>	"like" supresor metástasis 1
126306	<i>JSRP1</i>	Proteína 1 de retículo sarcoplasmático
140947	<i>C5orf20</i>	Marco de lectura abierta 20 del cromosoma 5
150365	<i>MEI1</i>	Inhibidor de meiosis 1

ID Entrez	Símbolo	Descripción
221188	<i>GPR114</i>	Receptor 114 acoplado a proteína G
401847	<i>LOC401847</i>	Proteína hipotética LOC401847
642424	<i>LOC642424</i>	Región Walker "like" de la cadena kappa V-I de la inmunoglobulina
100132941	<i>LOC100132941</i>	Proteína hipotética LOC100132941
100133862	<i>LOC100133862</i>	Región V35 "like" de la cadena pesada V-I de la inmunoglobulina
100287723	<i>LOC100287723</i>	Región Walker "like" de la cadena kappa V-I de la inmunoglobulina
100290415	<i>IGHV1-24</i>	Inmunoglobulina pesada variable 1-24
100293440	<i>LOC100293440</i>	Similar a la cadena lambda de inmunoglobulina

Tabla 25. Listado de los 50 genes incluidos en el predictor. La tabla muestra tres columnas que corresponden al ID Entrez, al símbolo y a la descripción de los 50 genes incluidos en el predictor que distingue los pacientes con CNMP en estadios iniciales con alto y bajo riesgo de recidiva. En esta tabla los 50 genes se encuentran ordenados de manera ascendente en base a su ID Entrez.

5.6.2. Validación del predictor de recidiva de 50 genes para estadios I y II de manera independiente.

Con la finalidad de comprobar si el predictor de 50 genes funcionaba cuando lo aplicábamos de manera separada sobre muestras en estadios I de CNMP o cuando lo utilizábamos sobre muestras en estadios II de CNMP, los 162 pacientes de la matriz de validación se separaron en dos grupos en función del estadio tumoral del paciente (I o II). Posteriormente se aplicó el predictor de 50 genes obteniendo en cada uno de los subgrupos, un cluster de alto riesgo de recidiva o mal pronóstico y otro cluster de bajo riesgo de recidiva o buen pronóstico entre los que existen diferencias estadísticamente significativas respecto al ILE ($p=0.016$ para los estadios I; $p=0.029$ para los estadios II) (figura 37). El modelo de riesgos proporcionales de Cox confirmó la habilidad de nuestro predictor de 50 genes para distinguir entre pacientes de alto y bajo riesgo de recidiva incluso cuando éstos han sido separados en función del estadio del tumor (HR=3.0; IC 95%: 1.2-8; $p=0.022$ para los estadios I) (HR=3.5; IC 95%: 1.1-12; $p=0.041$ para los estadios II).

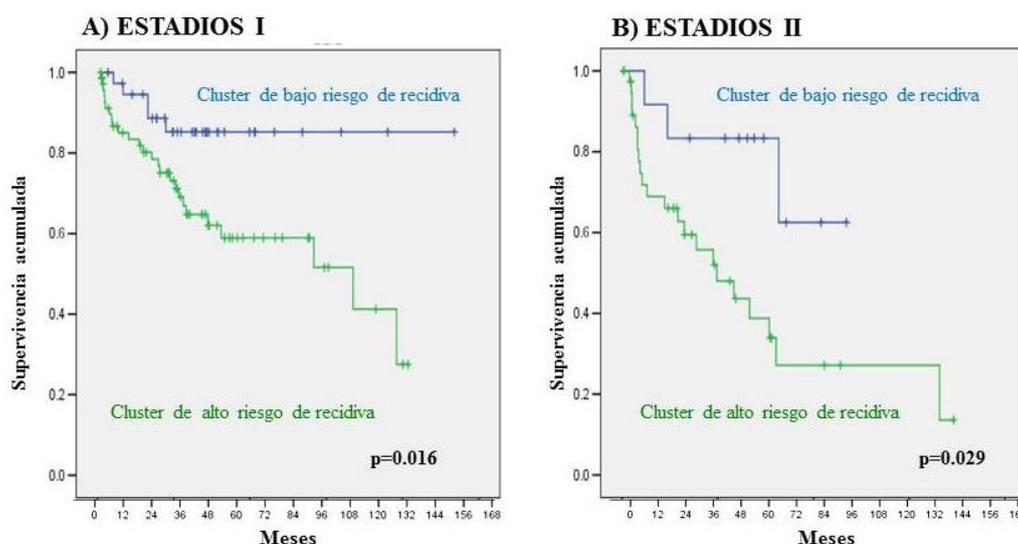


Figura 37. Representación mediante curvas de Kaplan-Meier de los dos clusters obtenidos en la matriz de validación utilizando el predictor de 50 genes de manera separada en los estadios I (A) y en los estadios II (B) del CNMP. Las curvas muestran que existen diferencias estadísticamente significativas respecto al ILE entre el cluster de alto riesgo de recidiva o mal pronóstico y el cluster de bajo riesgo de recidiva o buen pronóstico incluso cuando se separan las muestras de la matriz de validación en función del estadio tumoral.

5.6.3. Descripción de las funciones de los 50 genes del predictor de recidiva

A continuación se muestra una breve descripción de algunas de las funciones de los genes que conforman el predictor de 50 genes y que han sido presentados en la tabla 20:

AMPD1: Cataliza la deaminación de la adenosina monofosfato (AMP) a inosina 5 monofosfato (IMP) en el músculo esquelético y tiene un importante papel en el ciclo de las purinas.

TNFRSF17: Este receptor se expresa en linfocitos B maduros y es importante para el desarrollo de las células B y en la respuesta autoinmune. Tiene como ligando al miembro 13b de la superfamilia del factor de necrosis tumoral y activa el factor nuclear del potenciador del gen polipeptídico de la cadena ligera Kappa en células B (NF-kappaB) y la proteína quinasa activada por mitógeno 8 (MAPK8/JNK). También se une a otros ligandos y envía señales de supervivencia celular y proliferación.

CD19: Molécula que se une al receptor de antígenos de los linfocitos B para disminuir el umbral de estimulación de los linfocitos a través de la estimulación por antígeno.

CD27: Miembro de la superfamilia del receptor del factor de necrosis tumoral. El receptor tiene la función de generar y mantener durante largo tiempo la inmunidad de las células T. El ligando CD70 se une a él y funciona en la activación de las células B y en la síntesis de

inmunoglobulinas. Las proteínas adaptadoras denominadas Factor Asociado a Receptores de Factores de Necrosis Tumoral 2 y 5 (TRAF2 y TRAF5) median en este proceso. La proteína de unión a CD27 (SIVA) es una proteína pro-apoptótica que juega un importante papel en la apoptosis mediada por este receptor.

CD38: Es una ectoenzima multifuncional que se expresa en multitud de células y tejidos especialmente en leucocitos. CD38 también tiene funciones en la adhesión celular, transducción de señales y señalización por calcio.

CD79A y CD79B: codifican para las proteínas Ig-alpha e Ig-beta que son componentes del receptor antigénico de linfocitos B. Las moléculas Ig-alfa e Ig-beta son necesarias para la expresión y función de este receptor.

Granzima B (GZMB): Los linfocitos T citolíticos (CTL) y las células "natural killer" (NK) tienen la habilidad de reconocer, unir y lisar células diana específicas. La GZMB es crucial para la rápida inducción de la apoptosis de las células diana a través de la respuesta inmune generada por los linfocitos T citolíticos.

IGHA1 e IGHA2: Anticuerpo con una importante presencia en las secreciones mucosas y que representa la primera línea de defensa del organismo. Existen dos subclases Inmunoglobulina A1 (IgA1) e Inmunoglobulina (IgA2).

IGHG1: Este gen se encuentra traslocado en la leucemia linfocítica crónica de células B con el gen Ciclina D1 (CCND1) y en subclases de linfomas MALT (Tejido Linfoide Asociado a Mucosa) con los genes "LIM homeobox 4" (LHX4) y "Forkhead box P1" (FOXP1).

IGJ: Su función es unir dos monómeros o bien de Inmunoglobulina M (IgM) o bien de Inmunoglobulina A (IgA). También tiene la función de unir estas inmunoglobulinas al componente secretor.

IGL@: Cada molécula de inmunoglobulina tiene dos cadenas pesadas idénticas y dos cadenas ligeras idénticas. Hay dos clases de cadenas ligeras que son kappa y lambda. Este gen abarca el locus de la cadena ligera lambda que incluye el segmento V (variable), segmento J (unión) y segmento C (constante).

IGLL1: Es un gen de la superfamilia de las inmunoglobulinas que codifica para la cadena ligera sustitutiva del receptor de células preB. Mutaciones en este gen pueden producir deficiencia de células B o agammaglobulinemia.

IRF4: Pertenece a la familia de factores reguladores del interferón. Es específico de linfocitos y regula negativamente los receptores tipo Toll (o TLR), que es una molécula central en la activación de la respuesta inmune innata y adaptativa.

KCNN3: Regula la excitabilidad neuronal.

KRT81: Es un miembro de la familia de keratinas.

CXCL9: Su función no está bien definida pero parece que está implicado en el tráfico de células T.

PNOC: Es un neuropéptido que actúa como ligando endógeno del receptor "Opiate Receptor-Like 1" (ORL1).

POU2AF1: Es un coactivador específico de células B y su ausencia parece estar relacionada con defectos en el desarrollo de células B y con la falta de centros germinales.

BFSP2: también denominada faquinina, es una proteína estructural de filamentos del citoesqueleto. Junto a la filensina forma el BF ("beaded filament").

CXCL13: Promueve la migración de linfocitos B preferentemente frente a linfocitos T y macrófagos mediante estimulación con calcio.

PIM2: Es una serin/treonin/protein kinasa. Previene apoptosis y promueve supervivencia celular.

SMR3A: Es un homólogo funcional del gen *Vcsa1* ("Variable Coding Sequence A1"). Se ha asociado como un marcador de la disfunción eréctil asociada con etiología tanto diabética como no diabética.

MZB1: Está asociada con las cadenas pesadas y ligeras de la inmunoglobulina tipo M (IgM), promoviendo el ensamblaje de la IgM y su secreción.

FKBP11: Pertenece a la familia FKBP las cuales catalizan el plegamiento de los polipéptidos que contienen prolina. Su función es inhibida por FK506 y por rapamicina.

LAX1: Un regulador negativo de la señalización de linfocitos.

CPNE5: Proteína de unión a membrana dependiente de calcio que parece estar implicada en la regulación de fenómenos moleculares en la interfase de la membrana celular y en el citoplasma.

SLAMF7: Está implicada en la activación de células NK y en la regulación de la proliferación de linfocitos B durante la respuesta inmune.

DUSP26: Está asociado con la inactivación de la Proteína Kinasa activada por mitógenos 1 y 3 (MAPK1 y MAPK3), así como con la inhibición de la proliferación de células epiteliales, lo que podría sugerir un papel como gen supresor de tumores.

FCRL2: Forma parte de la superfamilia de receptores de inmunoglobulinas. Puede ser un marcador pronóstico de leucemia linfocítica crónica.

FCRL5: También es parte de la superfamilia de receptores de inmunoglobulinas. Está implicado en el desarrollo de células B y en la linfomagénesis.

FCRLA: Este receptor media la destrucción de los antígenos reconocidos por la Inmunoglobulina G (IgG). Es proteína selectiva de células B y puede estar implicada en su desarrollo.

DERL3: Proteína que se ubica en el retículo endoplasmático con la función de degradar glicoproteínas mal plegadas.

MTSS1L: Puede estar implicada en el empaquetamiento de la actina. Pertenece a la familia MTSS1 (Supresores de Metástasis Tipo 1).

JSRPI: El retículo sarcoplasmático es un compartimento celular que controla la concentración de calcio intracelular y está implicado en las funciones de excitación-contracción de este compartimento celular. En ratones se ha visto que esta proteína interacciona con proteínas claves implicadas en estos procesos de excitación-contracción.

C5orf20: Este gen se expresa en células dendríticas, que son potentes células presentadoras de antígenos implicadas en activar las células T nativas para iniciar la respuesta inmune específica de antígeno.

MEI1: Defectos en su expresión están relacionados con parada en meiosis y se asocia a fenómenos de azoospermia.

GPR114: Proteína G asociada a receptores con un extremo N terminal que contiene regiones ricas en serina/treonina. Se ha descrito su expresión en linfocitos citotóxicos.

IGHV5-78, FER1L4, IGKVID-8, KIAA0125, LOC401847, LOC642424, LOC100132941, LOC100133862, LOC100287723, IGHV1-24 y LOC100293440: a día de hoy, todavía no se conoce la función de estos genes.

5.6.4. Estudio de enriquecimiento de términos de “Gene Ontology” (GO) en el predictor de recidiva de 50 genes.

El análisis de enriquecimiento de términos de Gene Ontology para el predictor de 50 genes reveló una sobre-representación de términos relacionados con *procesos del sistema inmune* (GO:0002376; p=3.47E-08), *respuesta inmune* (GO:0006955; p=1.98E-08) y *respuesta a los estímulos* (GO:0050896; p=0.023).

6. DISCUSIÓN

6.1. IMPORTANCIA PRONÓSTICA DE LOS FACTORES CLINICO-HISTOPATOLÓGICOS Y MUTACIONALES CONVENCIONALES EN LOS ESTADIOS INICIALES DEL CNMP

La edad, el sexo, el subtipo histológico y el estadio tumoral obtenido mediante el sistema TNM, son importantes variables pronósticas en el CNMP. De todas ellas, el sistema de estadificación TNM basado en el tamaño del tumor (T), la afectación ganglionar (N) y la presencia de metástasis a distancia (M) es, en la actualidad, el factor pronóstico más utilizado en el manejo diario de los estadios iniciales del CNMP, clasificando a los pacientes en estadio IA, IB, IIA y IIB. A su vez, el tratamiento de cada paciente vendrá determinado por el estadio en el que se clasifique su tumor en el momento del diagnóstico.

La resección quirúrgica con linfadenectomía es el tratamiento de elección en los estadios iniciales del CNMP. Ya en la introducción de este proyecto de tesis se ha analizado la elevada tasa de recurrencia posterior a la resección quirúrgica en estos estadios, que oscila entre un 16-19% para estadio IA, 23-27% en IB, 37-37.5% en IIA y 39-40% IIB. Aunque recientemente se recomienda el empleo de quimioterapia adyuvante en estadios IB que cumplan determinados factores de riesgo, estadios IIA (T1ab-T2a, N1) y estadios IIB (T3, N0; T2b, N1), los beneficios obtenidos para disminuir la recidiva son limitados alcanzando únicamente un 5% en estadios II y un pequeño beneficio sin significado estadístico en IB.

En nuestro estudio, los estadios I y II no mostraban diferencias pronósticas en su asociación con ILE, es decir, pacientes con enfermedades de similares características según estos criterios, muestran comportamientos diferentes en la evolución de su enfermedad. En consecuencia, es necesario desarrollar herramientas alternativas capaces de predecir con mayor precisión el riesgo de recidiva y con posible proyección sobre la selección de pacientes candidatos a tratamiento adyuvante. Se requiere un método robusto capaz de estratificar pacientes en estadios iniciales de CNMP totalmente resecaos en grupos de buen y mal pronóstico, los primeros candidatos a evitar la toxicidad de la quimioterapia adyuvante y los segundos, por el contrario, candidatos a beneficiarse de ella.

En nuestro estudio, tampoco encontramos asociación estadísticamente significativa del resto de variables clínico-histopatológicas estudiadas con el ILE, tal y como mostraron las curvas de Kaplan-Meier obtenidas. Ni la edad, ni el sexo, ni el subtipo histológico, ni el

hábito tabáquico, ni el grado de diferenciación, ni la localización pulmonar, ni el resto de variables morfológicas, mostraron asociación estadísticamente significativa. Aunque en algunas series previamente publicadas (Jarabo et al., 2008; Bernal et al., 2008; Agarwal et al., 2010), algunas de estas variables se muestran como predictores pronósticos para el CNMP, ésta no es una observación universal tal y como muestran tanto nuestros resultados como los obtenidos en otras series en las que además se destaca las limitaciones de estos factores estándar para predecir el “outcome” de los pacientes e incluso concluyen la necesidad de identificar nuevos biomarcadores que suplan estas necesidades (Lu et al., 2006; Agullo-Ortuño et al., 2010).

Respecto al estudio del estado mutacional de genes relacionados con cáncer, solamente la presencia de mutaciones en *K-RAS* parece ser un indicador de peor pronóstico mostrando, en nuestra serie, tendencia respecto al ILE. Estos resultados concuerdan con los publicados por otros autores (Mascaux et al., 2005; Karachaliou et al., 2013) los cuales describen que pacientes con mutaciones en *K-RAS* presentan peor pronóstico que los pacientes wild type al no responder a quimioterapia adyuvante o a inhibidores de *EGFR*. La baja incidencia de mutaciones en *EGFR* (Lynch et al., 2004; Paez et al., 2004; Pao et al., 2004) o *B-RAF* (Paik et al., 2011) en los estadios iniciales del CNMP comparándolos con los estadios avanzados, sugiere un papel menos importante de estas alteraciones en la carcinogénesis inicial del cáncer de pulmón con la consecuente pérdida de implicación pronóstica.

6.2. CLASIFICACIÓN MOLECULAR DEL CNMP

La utilización de la tecnología de microarrays en los tumores malignos más frecuentes, ha generado la identificación de grupos moleculares con diferencias pronósticas útiles en la práctica clínica diaria, como por ejemplo en el cáncer de mama y el cáncer de colon. En el cáncer de mama se utilizan perfiles de expresión génica para identificar el riesgo de recurrencia de los pacientes en estadios iniciales de la enfermedad, existiendo tres test basados en esta metodología, el MammaPrint (Van't Veer et al., 2002), el Oncotype DX (Paik et al., 2004) y Aviara Dx (Ma et al., 2006). Por su parte, en el cáncer de colon se ha desarrollado una firma que incluye 12 genes y que funciona como predictor independiente de recurrencia en pacientes en estadio II de la enfermedad y que puede ayudar a identificar a aquellos pacientes que no se beneficiarán del tratamiento con quimioterapia (Kerr et al., 2009).

En el CNMP también se han desarrollado varios estudios que utilizan plataformas de análisis masivo para la obtención de perfiles de expresión que puedan ser utilizados como biomarcadores pronósticos que, junto con el TNM o de manera independiente, ayuden en la práctica clínica diaria en la indicación del mejor tratamiento de cada uno de los pacientes de manera individual en la conocida como oncología personalizada.

6.2.1. Firmas génicas en el CNMP, aspectos relacionados con el diseño del estudio

Los estudios de expresión génica han generado resultados muy prometedores respecto del pronóstico de los pacientes con CNMP (Zhu et al., 2010; Wan et al., 2010; Lu et al., 2006; Hou et al., 2010; Roepman et al., 2009) pero obtenido resultados dispares en cuanto a los genes a incluir en el biomarcador. Este hecho posiblemente sea consecuencia de las limitaciones derivadas de la falta de un diseño experimental apropiado, tal y como ha referido recientemente el grupo de Subramanian et al (2010) por la utilización de diferentes criterios de inclusión de pacientes, diferencias en la obtención de muestras, inclusión de distintos estadios tumorales, la exclusión o no de subtipos histológicos de gran importancia, así como la falta, en algunos casos, de validación independiente (Raponi et al., 2006; Chen et al., 2007; Raz et al., 2008; Roepman et al., 2009).

En nuestro estudio, en un intento de superar estas deficiencias, se presenta una descripción muy exhaustiva de los criterios de inclusión de los pacientes, de las características de los mismos, de la manipulación de los tejidos a estudio y de los métodos analíticos y estadísticos empleados. Es importante destacar el análisis histológico al que han sido sometidas las muestras a estudio para comprobar si cumplían con el porcentaje mínimo necesario de tejido tumoral (70%) el cual, se ha llevado a cabo sobre la misma muestra congelada que iba a ser hibridada en los microarrays. Además, previamente a la realización de la clasificación molecular de los tumores, realizamos una caracterización molecular de las 84 muestras tumorales frente a 8 tejidos de parénquima pulmonar no tumoral con la finalidad de verificar la capacidad discriminadora de la técnica de microarrays. Aunque el tamaño muestral del grupo de tejidos de parénquima pulmonar no tumoral era bajo, este inconveniente se vio sin duda compensado por el exhaustivo control histológico realizado y por las grandes diferencias que existen entre estos dos fenotipos. Por último hay que añadir que una gran ventaja de la clasificación molecular que se presenta en este proyecto de tesis es que, al haber sido generada mediante métodos no supervisados, es capaz de crear grupos moleculares más homogéneos. Como consecuencia

nos permite analizar no sólo la asociación pronóstica si no también, conocer las vías moleculares alteradas en el CNMP obteniendo una interpretación biológica más completa del desarrollo tumoral a la vez que somos capaces de identificar potenciales biomarcadores y dianas génicas útiles en el tratamiento personalizado del CNMP.

6.2.2. Interpretación de la nueva clasificación molecular

Inicialmente, nuestra clasificación molecular utilizando el listado de 3.232 genes identificó dos clusters que no muestran diferencias estadísticamente significativas respecto al ILE, con una clara asociación con los subtipos histológicos más representados en el CNMP, resultados concordantes con los obtenidos previamente por otros autores (Tomida et al., 2004). Para evitar este handicap, varios estudios solamente incluyeron en su serie uno de los dos subtipos histológicos, adenocarcinomas (Bhattacharjee et al., 2001; Beer et al., 2002) o carcinomas epidermoides o escamosos (Sun et al., 2004; Inamura et al, 2005). Al utilizar estos criterios de inclusión, fueron capaces de obtener mayor información molecular de cada uno de estos subtipos histológicos de manera independiente, pero generalmente con series limitadas. Incluso en algunos de estos estudios (Raponi et al., 2006), se generaron dos clasificadores pronósticos independientes, uno para adenocarcinomas y otro para carcinomas epidermoides o escamosos que después se combinaron en uno único que pudiera ser utilizado en ambos subtipos histológicos. El principal inconveniente de estos clasificadores, es que únicamente conseguían precisiones predictivas relativamente bajas que se encontraban alrededor del 70% de los casos analizados.

Con la finalidad de superar los inconvenientes anteriores descritos y con la premisa de que todos los estadios iniciales del CNMP van a ser sometidos a tratamiento quirúrgico independientemente del subtipo histológico, realizamos un filtrado y exclusión del listado inicial de 3.232 genes de aquellos que caracterizan las diferencias histológicas entre los dos subtipos principales del CNMP. Este planteamiento, totalmente novedoso y hasta el momento no realizado por ningún otro grupo, nos ofrece la posibilidad de estudiar molecularmente a todos los pacientes sin tener en cuenta la diferenciación histológica en la enfermedad inicial del CNMP, y así obtener mayor información de las claves moleculares implicadas en las recidivas de los pacientes con enfermedad localizada.

Utilizando el listado de 2.160 genes realizamos una segunda clasificación molecular que tenía implicación pronóstica y no estaba relacionada con la histología de los

tumores. La nueva clasificación permite diferenciar, independientemente del TNM, un grupo de pacientes con bajo riesgo de recurrencia (Cluster 3) y dos grupos de pacientes con alto riesgo de recurrencia (Cluster 1 y 2) entre los que no existe diferencia estadísticamente significativa para el riesgo.

En el estudio de asociación de los tres clusters con las variables clínico-histopatológicas incluidas en el estudio, observamos que existe asociación estadísticamente significativa de la nueva clasificación molecular con el hábito tabáquico, la afectación ganglionar y la inflamación crónica (valorada en función de la presencia de células linfoplasmocitarias), variables que por sí mismas no mostraban asociación estadística significativa con el ILE cuando los pacientes no eran clasificados en clusters molecularmente homogéneos. En el Cluster 1 se incluyen los tres únicos pacientes con afectación ganglionar, el Cluster 2 contiene los cinco únicos pacientes no fumadores y el Cluster 3 es el que incluye el mayor número de pacientes con una alta presencia de células linfoplasmocitarias intratumorales. El análisis histopatológico revela que los tumores frecuentemente contienen un infiltrado variable de células relacionadas con el sistema inmune (Finn OJ., 2008). La cuantificación de la infiltración de estos tipos celulares se puede realizar utilizando diferentes métodos como el microscopio de luz óptica o la inmunohistoquímica. En nuestra serie, la cuantificación de las células inmunes en el microscopio no es un indicador pronóstico significativo hasta que los pacientes de CNMP son clasificados en clusters en función de su perfil de expresión.

Varios artículos han descrito recientemente marcadores pronósticos con base inmunológica en el CNMP. Suzuki et al. (2011) publicaron una revisión sobre marcadores pronósticos inmunes tanto en el microambiente tumoral como en sangre periférica. Es importante destacar que dos de las publicaciones incluidas en la revisión, centradas en el microambiente tumoral de estadios iniciales de CNMP (Pelletier et al. (2001); Al-Shibli et al. (2008)), describen la existencia de los linfocitos intratumorales (TILs) con fenotipo CD20+. En ambos estudios, la presencia de estas células linfoplasmocitarias correlaciona con un aumento de la supervivencia de los pacientes, datos concordantes con los obtenidos en nuestro estudio.

Los estudios de microarrays de expresión previos al desarrollado por nuestro grupo y que incluían el análisis de vías moleculares, mostraron que en el CNMP las principales vías alteradas estaban relacionadas con ciclo celular y regulación de la replicación y reparación del DNA (Daraselia et al., 2012). Nuestro estudio aporta que la alteración de las

vías relacionadas con señalización, ciclo celular y regulación de la replicación del DNA, son claves para diferenciar los dos cluster de mal pronóstico (Cluster 2 y Cluster 1) pero que las principales diferencias entre el cluster de buen pronóstico (Cluster 3) respecto a los otros dos, estaban basadas en la alteración de vías moleculares relacionadas con sistema inmune. Algunas de estas vías ya habían sido descritas previamente por otros autores vinculándolas con la respuesta inmune antitumoral en el cáncer. Además, nuestros resultados están en consonancia con los obtenidos previamente por Bryant et al. (2010) en una serie de 432 adenocarcinomas, en los que la alteración combinada de vías moleculares relacionadas con ciclo celular e inhibición de la respuesta inmune se asociaba con mal pronóstico. Por tanto, en los estadios iniciales del CNMP la información pronostica obtenida a partir de la clasificación histológica o de la clasificación molecular relacionada con la histología puede ser marginal comparada con la información proporcionada por las vías moleculares no relacionadas con la histología e identificadas mediante la metodología descrita.

Para la validación de la clasificación molecular obtenida utilizando el listado de 2.160 genes, hemos empleado la única matriz de datos externa en la que se utilizaron criterios de inclusión de pacientes muy similares a los empleados en nuestro estudio y que además utilizó la misma plataforma de microarrays. Esta matriz de validación incluía un total de 162 pacientes de los cuatro subtipos histológicos de CNMP, de los que sus datos de expresión se encuentran disponibles online para su utilización por la comunidad científica. El primer paso fue generar un predictor de 1.000 genes a partir de nuestra nueva clasificación molecular el cual, aplicándolo en la matriz de validación, conseguía clasificar los 162 nuevos pacientes en los tres clusters descritos. De nuevo, la curva de Kaplan-Meier mostró que la nueva clasificación molecular está asociada de manera estadísticamente significativa con el intervalo libre de enfermedad. Por tanto, la validación externa demostró la utilidad pronostica de nuestra clasificación molecular, independientemente de la serie de CNMP que se someta a estudio, cumpliendo así con las recomendaciones de la última revisión de firmas génicas del grupo de Subramanian et al publicada en 2010.

Con la finalidad de obtener una clasificación molecular más fácilmente aplicable, simplificamos a dos clusters de alto y bajo riesgo y desarrollamos y validamos un predictor de recidiva basado en la expresión de tan sólo 50 genes. Este predictor es útil en los diferentes subtipos histológicos de CNMP, lo que hace más fácil su aplicación en otras bases de datos, evitando las limitaciones descritas por el grupo de Hou et al. (2010)

los cuales observaron que los predictores desarrollados utilizando únicamente un subtipo histológico no funcionan bien cuando se aplican a otras series de datos que incluyen subtipos histológicos diferentes.

6.2.3. Predictor de recidiva: papel del sistema inmune en los estadios iniciales del CNMP

Nuestro predictor refleja, por la composición de sus 50 genes, la gran importancia que tiene la respuesta inmune en el pronóstico de los pacientes en estadios iniciales de CNMP, tal y como habíamos observado previamente en los resultados obtenidos del estudio de las vías moleculares alteradas entre los tres clusters generados. En función de su nivel de expresión, todos los genes del predictor se encuentran sobre-expresados en el grupo de bajo riesgo de recidiva respecto del grupo de alto riesgo, con lo que podemos inferir que dicha sobre-expresión nos indica la presencia de una mayor respuesta inmune intratumoral que además se asocia a un mejor pronóstico de estos pacientes.

El estudio funcional de los genes del predictor nos muestra que éste incluye tanto genes que codifican para elementos estructurales (*CD19*, *moléculas de inmunoglobulinas (Ig)* o *el receptor de células B (CD79)*) como para elementos funcionales del linaje de las células B (co-activador transcripcional específico de células B (*POU2AF1*) y activador de células B1 de la zona marginal (*MZB1*)). En el predictor también se incluyen otros genes, que aunque no están restringidos únicamente a las células B, tienen una gran influencia en la homeostasis de este tipo celular como son *TNFRSF17*, *SLAM7F*, *CXCL13* y *IRF4*. La presencia de la *IGHA* entre los 50 genes del predictor parece indicar que las células B/plasmáticas intratumorales han migrado desde el torrente circulatorio ya que en este, la mayoría son fenotipo IGA+ve (Perez-Andres et al., 2010).

Por el contrario, no encontramos ningún gen estructural específico del linaje de células T, células NK o macrófagos pero sí genes cuya función tradicionalmente ha sido asociada a estos tipos celulares como *Granzima B*, *CD38*, *CD27* y receptores para el fragmento constante (Fc) de Ig (*FCRLA*, *FCRL5*, *FCRL2*). Este hecho nos lleva a la conclusión de que, en los estadios iniciales del CNMP, la respuesta inmune intratumoral está mayormente asociada a células B. La presencia en el predictor de genes funcionales tradicionalmente asociados a células T en ausencia de genes estructurales de este tipo celular podría explicarse, tal y como describen varios artículos, por la posibilidad de que

estos genes también puedan ser expresados por células B/plasmáticas. Por ejemplo, *CD38* y *CD27* se encuentran fuertemente sobreexpresados en células plasmáticas promoviendo su proliferación y supervivencia (Pérez-Andrés M. et al., 2010), *CD27* es también un marcador de células B memoria (Tangye SG et al., 2007) y *Pim-2* y *Granzima B* funcionan como mediadores antiapoptóticos para el linaje de células B (Asano et al., 2011; Trinidad et al., 2009). Respecto a éste último, puede ser incluso secretada por células B “citotóxicas” las cuales han sido recientemente descritas como componentes de una nueva vía de diferenciación de este tipo celular (Hagn et al., 2012). Además, respecto al único gen específico de función de células T que se encuentra incluido en el predictor LAX1, recientemente se ha demostrado que su sobreexpresión resulta en un bloqueo de la activación de este tipo celular (Shapiro et al., 2008). Más concretamente, en el CNMP, se ha demostrado que la ausencia de células T (fenotipo efector CD4+), supone un predictor pronóstico favorable para este grupo de pacientes (Zikos et al., 2011).

Aunque los mecanismos inmunes con valor pronóstico se han relacionado principalmente con las células T del sistema inmune, la participación de células B/células plasmáticas también se ha descrito, de acuerdo con nuestros resultados, tanto en estadios iniciales de CNMP (Pelletier et al., 2001; Al-Shibli et al. (2008)) como en otros tumores, por ejemplo, el cáncer de mama (Ascierto et al., 2012; Coronella et al., 2001; Fridman et al., 2005). El grupo de Pelletier, en una serie de 113 pacientes encontró que la presencia de células B intratumorales se asocia a una mejor supervivencia de los pacientes, sugiriendo la importancia de la inmunidad humoral en este tipo de tumores. Al-Shibli et al., en su serie de 371 CNMP encuentra resultados similares a los de Pelletier para el análisis univariante de la presencia de células B intratumorales postulando, al igual que ya habían hecho otros grupos previamente que la inmunidad mediada por anticuerpo puede contribuir a la erradicación del tumor (Gilboa et al., 2004). En el cáncer de mama, Ascierto et al describieron una firma molecular que incluye 5 genes relacionados con células B infiltrantes, capaz de predecir bajo riesgo de recaída en pacientes con cáncer de mama, independientemente del estatus de receptores de estrógeno (RE), receptores de progesterona (RP) o *HER-2/neu*. La relación entre la infiltración por células B/células plasmáticas y la supervivencia también ha sido mostrada en los carcinomas medulares de mama en los que se ha descrito una respuesta inmune humoral desencadenada mediante antígenos tumorales (Coronella et al., 2001; Fridman et al., 2005). Basándonos en los datos

obtenidos en nuestra serie podríamos sugerir que un antígeno similar desencadenaría la respuesta de células B/células plasmáticas en el CNMP abriendo la posibilidad de su utilidad como diana terapéutica.

Recientemente, multitud de artículos sugieren que el desarrollo del cáncer está controlado por el sistema inmune del propio paciente, destacando la importancia de incluir biomarcadores inmunológicos para la predicción del pronóstico y la respuesta a terapia. El grupo de Galon et al (2012) propuso la utilización de un "inmunoscore" o método para clasificar la respuesta inmune intratumoral, que junto con el estadio TNM, sea capaz de conocer de manera más precisa el pronóstico de los pacientes con cáncer. Dicho inmunoscore debería ser incluido como parte de la rutina diagnóstica y pronóstica de los tumores. A la vista de los resultados obtenidos en nuestro proyecto podemos sugerir que nuestra firma de 50 genes podría utilizarse como inmunoscore en el CNMP o como biomarcador predictor de respuesta a inmunoterapia.

6.2.4. Comparación con otros predictores génicos en el CNMP

Previamente a nuestro estudio, varios grupos publicaron predictores génicos asociados con pronóstico en estadios I y II de CNMP. Lu et al. (2006) publicaron un predictor pronóstico para pacientes en estadios I de CNMP que incluía 64 genes, la mayoría de ellos relacionados con adhesión celular, motilidad, proliferación y apoptosis. Utilizaron 5 series externas en las que se validó dicho predictor y mediante las que confirmaron, que existían diferencias entre los pacientes de alto y bajo riesgo, demostrando un mayor poder clasificatorio de este predictor génico respecto de la estadificación tumoral. La principal limitación de este estudio fue que los resultados obtenidos no eran reproducibles cuando se analizaban de manera independiente los estadios IA.

El grupo de Zhu et al. (2010) desarrolló un predictor de 15 genes para la supervivencia de pacientes en estadios iniciales de CNMP, independiente del estadio tumoral. Utilizaron un total de 133 muestras congeladas del ensayo JBR.10 que incluye tanto adenocarcinomas como carcinomas epidermoides o escamosos y utilizaron un límite inferior de representación tumoral muy bajo (20%). El predictor generado fue capaz de crear dos grupos moleculares con distinta supervivencia, siendo así capaces de seleccionar los pacientes que, "a priori", obtendrán un mayor beneficio del uso de quimioterapia adyuvante. Esta clasificación fue validada en 3 de las 4 series externas testadas y la anotación funcional de los 15 genes del predictor indica un primer set de genes reguladores

transcripcionales (*MDM2*, *ZNF236*, *FOSL2*, *HEXIM1*, *MYT1L* y *IKBKAP*), un segundo set de genes que codifican para proteínas transmembrana con función en vías de señalización (*MLANA*, *ATP1B1*, *LICAM*, y *STMN2*) y por último 2 genes involucrados en el metabolismo de purinas y pirimidinas (*ATP1B1* y *UMPS*).

Por último, el grupo de Roepman et al. publicó en 2.009 un predictor de recurrencia que incluye 72 genes, la mayoría relacionados con respuesta inmune, que funciona cuando se analizan de manera independiente los estadios I y II del CNMP. El porcentaje de coincidencia entre los 72 genes de este predictor y los 50 genes incluidos en el nuestro es únicamente de 4 genes. El porcentaje de representación tumoral en nuestras muestras es mayor que en las incluidas en el de Roepman (70% frente a 50%), lo que seguramente evite las posibles interferencias del tejido normal en los genes incluidos en el predictor y respecto a los métodos de análisis, en nuestro estudio fueron no supervisados frente a los métodos supervisados utilizados en este estudio. Por su parte, nuestra matriz de validación está compuesta por muestras externas a nuestro laboratorio, lo que confirma la universalidad de nuestra firma genética, mientras que desafortunadamente, la firma génica del grupo de Roepman no fue validada en estudios independientes.

En resumen, el predictor de 50 genes aportado en nuestro proyecto con respecto al publicado por Roepman et al, aporta las ventajas de ser un predictor más sencillo al disminuir un 30% el número de genes necesarios para identificar pacientes con bajo y alto riesgo de recaída, de ser un predictor validado en un número sensiblemente más elevado de muestras y de haber sido validado en datos completamente externos generados por otro laboratorio, de estar más enfocado al grupo de pacientes sin afectación ganglionar en donde existe una mayor necesidad de este tipo de predictores y de que en un futuro pudiera llegar a tener implicaciones terapéuticas

6.3. POSIBLES IMPLICACIONES CLÍNICAS DEL PREDICTOR DE 50 GENES

El predictor obtenido genera una información pronóstica muy relevante clasificando, con mayor claridad que el TNM, a los pacientes en función del riesgo de recidiva.

Mediante el estudio de estos 50 genes somos capaces de identificar los pacientes en estadio I con pronóstico desfavorable que podrían beneficiarse de un tratamiento adicional y pacientes en estadio II con buen pronóstico que no parecen obtener beneficio de la

quimioterapia e incluso sufren los efectos secundarios tóxicos de este tratamiento. Además, como el predictor de recidiva obtenido incluye genes relacionados principalmente con las células B del sistema inmune, éstos podrían convertirse en potenciales dianas para la manipulación génica con la finalidad de potenciar la respuesta inmune anti-tumoral que reduce significativamente el riesgo de recidiva en los estadios iniciales del CNMP.

En conclusión, los resultados presentados son capaces de mejorar la selección de pacientes susceptibles de tratamiento adicional a la resección quirúrgica con linfadenectomía a la vez que abre la puerta a nuevas posibilidades para la inmunoterapia y el desarrollo de vacunas.

7. CONCLUSIONES

- 1.-** En nuestra serie de 84 pacientes en estadios iniciales de carcinoma no microcítico de pulmón, ninguna de las principales variables clínico-histopatológicas ni mutacionales estudiadas se asocia de manera estadísticamente significativa con el intervalo libre de enfermedad.

- 2.-** La obtención de resultados fiables mediante la metodología de microarrays implica la necesidad tanto de una correcta elección de los criterios para la inclusión de pacientes en el estudio como de la estandarización de procesos clave como la utilización de un criterio histológico para la selección de muestras, control de calidad de los RNA y de los procesos de hibridación y el estudio del comportamiento anómalo de muestras mediante el análisis de componentes principales.

- 3.-** Utilizando la metodología de microarrays de expresión y tras eliminar los genes diferencialmente expresados entre adenocarcinomas y carcinomas de células escamosas, somos capaces de obtener una clasificación molecular en tres clusters que se asocian de manera estadísticamente significativa con el intervalo libre de enfermedad.

- 4.-** En esta clasificación molecular no relacionada con la histología, diferenciamos dos cluster de mal pronóstico o alto riesgo de recidiva y un cluster de buen pronóstico o bajo riesgo de recidiva.

- 5.-** La validación de la clasificación molecular obtenida en una matriz externa indica la idoneidad de la misma para poder ser utilizada de manera universal.

6.- Al no existir diferencias estadísticamente significativas entre los dos clusters de mal pronóstico, éstos pueden ser unificados en uno único y así simplificar la clasificación molecular a un cluster de buen pronóstico o bajo riesgo de recidiva y un cluster de mal pronóstico o alto riesgo de recidiva.

7.- El predictor obtenido incluye únicamente 50 genes y es capaz de clasificar a los pacientes, en función de su riesgo de recidiva, en uno de los dos clusters moleculares generados.

8.- La validación del predictor de 50 genes indica la idoneidad del mismo para poder ser utilizado en cualquier serie. También se ha demostrado su utilidad para clasificar el riesgo de recidiva cuando se aplica en estadios I y II de manera independiente.

9.- Este predictor podría utilizarse para identificar, en estadios I de la enfermedad, los pacientes de mal pronóstico que pudieran beneficiarse de un tratamiento adyuvante y en los estadios II, los pacientes de buen pronóstico que no necesitarían tratamientos complementarios a la resección quirúrgica.

10.- El predictor de 50 genes refleja, por su composición, la importancia de la respuesta inmune asociada a células B en el pronóstico de los pacientes en estadios iniciales del carcinoma no microcítico de pulmón.

11.- Todos los genes del predictor se encuentran sobre-expresados en el cluster de bajo riesgo de recidiva respecto del de alto riesgo, lo que nos indica la presencia de una mayor respuesta inmune intra-tumoral que además se asocia a un mejor pronóstico de estos pacientes.

12.- Los genes del predictor podrían convertirse en dianas para la inmunoterapia a través de las cuales potenciar la respuesta inmune anti-tumoral capaz de reducir significativamente el riesgo de recidiva en los estadios iniciales del CNMP.

8. BIBLIOGRAFÍA

- Agarwala M, Brahmandaya G, Chmielewskib GW et al. “*Age, tumor size, type of surgery, and gender predict survival in early stage (stage I and II) non-small cell lung cancer after surgical resection*”. Lung Cancer 68: 398–402. 2010.
- Agulló-Ortuño MT, López-Ríos F, Paz-Ares L. “*Lung cancer genomic signatures*”. Journal of Thoracic Oncology 5(10): 1673-91. 2010.
- Alam N, Darling G, Shepherd FA et al. “*Postoperative chemotherapy in nonsmall cell lung cancer: a systematic review*”. The Annals of Thoracic Surgery 81(5): 1926-36. 2006.
- Alberg AJ, Brock MV, Ford JG et al. “*Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines*”. Chest 143(5 Suppl): e1S-29S. 2013.
- Al-Shibli KI, Donnem T, Al-Saad S et al. “*Prognostic effect of epithelial and stromal lymphocyte infiltration in non-small cell lung cancer*”. Clinical Cancer Research 14: 5220-7. 2008.
- Arriagada R, Bergman B, Dunant A et al. “*Cisplatin-based adjuvant chemotherapy in patients with completely resected non-small-cell lung cancer*”. New England Journal of Medicine 350(4): 351-60. 2004.
- Asano J, Nakano A, Oda A et al. “*The serine/threonine kinase Pim-2 is a novel anti-apoptotic mediator in myeloma cells*”. Leukemia 25(7): 1182-8. 2011.
- Ascierto ML, De Giorgi V, Liu Q et al. “*An immunologic portrait of cancer*”. Journal of Translational Medicine 9:146. 2011.
- Bammler T, Beyer RP, Bhattacharya S et al. “*Standardizing global gene expression analysis between laboratories and across platforms*”. Nature Methods 2(5): 351-6. 2005.
- Beer D, Kardia S, Huang C et al. “*Gene-expression profiles predict survival of patients with lung adenocarcinoma*”. Nature Medicine 8(8): 816-824. 2002.
- Bepler G, Kusmartseva I, Sharma S et al. “*RRM1-modulated in vitro and in vivo efficacy of gemcitabine and platinum in non-small-cell lung cancer*”. Journal of Clinical Oncology 24: 4731-4737. 2006.
- Bergethon K, Shaw AT, Ou SH et al. “*ROS1 rearrangements define a unique molecular class of lung cancers*”. Journal of Clinical Oncology 30: 863-70. 2012.
- Bernal R, León A, Jaén J et al. “*Recidiva local tras la cirugía del cáncer de pulmón no microcrítico. Factores pronósticos*”. Anales de Medicina Interna. 25 (2): 55-60. 2008.

- Berns K, Horlings HM, Hennessy BT et al. ***“A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer”***. *Cancer Cell* 12: 395-402. 2007.
- Bhattacharjee A, Richards WG, Staunton J et al. ***“Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses”***. *Proceedings of the National Academy of Sciences of USA* 98(24): 13790–13795. 2001.
- Bindea G, Mlecnik B, Fridman WH et al. ***“Natural immunity to cancer in humans”***. *Current Opinion in Immunology* 22: 215-222. 2010.
- Boelens MC, van den Berg A, Vogelzang I et al. ***“Differential expression and distribution of epithelial adhesion molecules in non-small-cell lung cancer and normal bronchus”***. *Journal of Clinical Pathology* 60: 608-614. 2007.
- Booth CM, Shepherd FA. ***“Adjuvant chemotherapy for resected non-small cell lung cancer”***. *Journal of Thoracic Oncology* 1: 180–187. 2006.
- Boutros PC, Lau SK, Pintilie M et al. ***“Prognostic gene signatures for non-small-cell lung cancer”***. *Proceedings of the National Academy of Sciences of USA*. 106(8): 2824–2828. 2009.
- Brazma A, Hingamp P, Quackenbush J et al. ***“Minimum information about a microarray experiment (MIAME)-toward standards for microarray data”***. *Nature Genetics* 29(4): 365-371. 2001.
- Bryant CM, Albertus DL, Kim S et al. ***“Clinically relevant characterization of lung adenocarcinoma subtypes based on cellular pathways: an international validation study”***. *PLoS One* 5(7): 11712. 2010.
- Burnet FM. ***“Cancer a biological approach”***. *British Medical Journal* 1: 841-847. 1957.
- Butterfield LH, Disis ML, Fox BA et al ***“A systematic approach to biomarker discovery; preamble to the iSBTc-FDA taskforce on immunotherapy biomarkers”***. *Journal of Translational Medicine*. 6:81. 2008.
- Cappuzzo F, Marchetti A, Skoka M. ***“Increased MET Gene Copy Number Negatively Affects Survival of Surgically Resected Non–Small-Cell Lung Cancer Patients”*** *Journal of Clinical Oncology* 27:1667-167. 2009.
- Chansky K, Sculier JP, Crowley JJ et al. ***“The International Association for the Study of Lung Cancer Staging Project: prognostic factors and pathologic TNM***

- stage in surgically managed non-small cell lung cancer*". Journal of Thoracic Oncology 4: 792-801. 2009.
- Chen H, Yu S, Chen C et al. "*A Five-Gene Signature and Clinical Outcome in Non-Small-Cell Lung Cancer*". New England Journal of Medicine 356:11-20. 2007.
 - Chung CH, Bernard PS, Perou CM. "*Molecular portraits and the family tree of cancer*". Nature Genetics 32: 533-540. 2002.
 - Cohen Y, Xing M, Mambo E et al. "*BRAF mutation in papillary thyroid carcinoma*". The Journal of the National Cancer Institute 95(8): 625-7. 2003.
 - Coronella J, Spier C, Welch M et al. "*Antigen-Driven Oligoclonal Expansion of Tumor-Infiltrating B Cells in Infiltrating Ductal Carcinoma of the Breast*". The Journal of Immunology 169: 1829-1836. 2002.
 - Daraselia K, Wang Y, Budoff A et al. "*Molecular signature and pathway analysis of human primary squamous and adenocarcinoma lung cancers*". American Journal of Cancer Research 2 (1): 93-103. 2012.
 - Dave SS, Wright G, Tan B et al "*Prediction of survival in follicular lymphoma based on molecular features of tumor-infiltrating immune cells*". New England Journal of Medicine 351: 2159-2169. 2004.
 - Davies H, Bignell GR, Cox C et al. "*Mutations of the BRAF gene in human cancer*". Nature 417(6892): 949-54. 2002.
 - De la Cruz CS, Tanoue LT, Matthay RA. "*Lung cancer: epidemiology, etiology and prevention*". Clinics in Chest Medicine 32(4): 605-44. 2011.
 - Díez M, Torres A, Maestro ML et al. "*Prediction of survival and recurrence by serum and cytosolic levels of CEA, CA125 and SCC antigens in resectable non-small-cell lung cancer*". British Journal of Cancer 73(10): 1248-54. 1996.
 - Ding L, Getz G, Wheeler DA et al. "*Somatic mutations affect key pathways in lung adenocarcinoma*". Nature 455: 1069-75. 2008.
 - Donnem T, Al-Saad S, Al-Shibli K et al. "*Inverse prognostic impact of angiogenic marker expression in tumor cells versus stromal cells in non-small-cell lung cancer*". Clinical Cancer Research 13(22 Pt 1): 6649-6657. 2007.
 - Douillard JY, Rosell R, De Lena M et al "*Adjuvant vinorelbine plus cisplatin versus observation in patients with completely resected stage IB-IIIa non-small-cell lung*

cancer (Adjuvant Navelbine International Trialist Association [ANITA]): a randomised controlled trial". Lancet Oncology (9): 719-27. 2006.

•Dunn GP, Bruce AT, Ikeda H et al. "*Cancer immunoediting: from immunosurveillance to tumor escape*". Nature Immunology 3(11): 991-8. 2002.

•Dutt A, Ramos AH, Hammerman PS et al. "*Inhibitor-sensitive FGFR1 amplification in human non-small cell lung cancer*". PLoS One 6: e20351. 2011.

•Edge SB, Byrd DR, Compton CC et al. "*Lung. AJCC Cancer Staging Manual. 7th edition*". New York. Springer: 253-704. 2010.

•Engelman JA, Zejnullahu K, Mitsudomi T et al. "*MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling*". Science 316(5827): 1039-43. 2007.

•Falchook GS, Long GV, Kurzrock R et al. "*Dabrafenib in patients with melanoma, untreated brain metastases, and other solid tumours: a phase 1 dose-escalation trial*". Lancet 379: 1893-90. 2012.

•Fan J, Ren Y. "*Statistical analysis of DNA microarray data in cancer research*". Clinical Cancer Research. 12(15): 4469-4473. 2006.

•Farhat FS, Houhou W. "*Targeted therapies in non-small cell lung carcinoma: what have we achieved so far?*" Therapeutic Advances in Medical Oncology 5(4): 249-70. 2013.

•Fibla JJ, Cassivi SD, Decker PA et al. "*Validation of the lung cancer staging system revisions using a large prospective clinical trial database (ACOSOG Z0030)*". European Journal of Cardio-Thoracic Surgery 43(5):911-4. 2013.

•Finkelstein D, Ettinger D, Ruckdeschel J et al. "*Long-Term Survivors in Metastatic Non-Small-Cell Lung Cancer: An Eastern Cooperative Oncology Group Study*" Journal of Clinical Oncology 4(5): 702-70. 1986.

•Finn OJ. "*Cancer immunology*". New England Journal of Medicine 358(25): 2704-15. 2008

•Fridman WH, Toth J, McKnight M et al. "*Novel ganglioside antigen identified by B cells in human medullary breast carcinomas: the proof of principle concerning the tumor-infiltrating B lymphocytes*". Journal of Immunology 175(4): 2278-85. 2005.

•Galon J, Pagès F, Marincola FM et al. "*The immune score as a new possible approach for the classification of cancer*". Journal of Translational Medicine 10:1. 2012.

- Galon J. ***“The Immune Score as a New Possible Approach for the Classification of Cancer”*** Journal of Translational Medicine 29: 1093-1102. 2012.
- Garber ME, Troyanskaya OG, Schluens K et al. ***“Diversity of gene expression in adenocarcinoma of the lung”***. Proceedings of the National Academy of Sciences of USA, 98(24): 13784–13789. 2001.
- Gautman A, Li ZR, Bepler G. ***“RRM1-induced metastasis through PTEN-regulated pathways”***. Oncogene 22: 2135-2142. 2003.
- Gazdar AF. ***“Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors”***. Oncogene 28 Suppl 1: S24-31. 2009.
- Gilboa E. ***“The promise of cancer vaccines”***. Nature Reviews Cancer 4:401. 2004.
- Hagn M, Sontheimer K, Dahlke K et al. ***“Human B cells differentiate into granzyme B-secreting cytotoxic B lymphocytes upon incomplete T-cell help”*** Immunology & Cell Biology 90(4): 457-67. 2012.
- Hammerman PS, Sos ML, Ramos AH et al. ***“Mutations in the DDR2 kinase gene identify a novel therapeutic target in squamous cell lung cancer”***. Cancer Discovery 1: 78-8. 2011.
- Hanahan D, Weinberg RA. ***“The hallmarks of cancer”***. Cell 100(1): 57-70. 2000.
- Hanahan D, Weinberg RA. ***“Hallmarks of cancer: the next generation”***. Cell 144(5): 646-74. 2011.
- Hatzaki A, Razi E, Anagnostopoulou K et al. ***“A modified mutagenic PCR-RFLP method for K-ras codon 12 and 13 mutations detection in NSCLC patients”***. Molecular and Cellular Probes 15: 243-7. 2001.
- Heist RS, Sequist LV, Engelman JA. ***“Genetic changes in squamous cell lung cancer: a review”***. Journal of Thoracic Oncology 7: 924-33. 2012.
- Herbst RS, Onn A, Sandler A. ***“Angiogenesis and lung cancer: prognostic and therapeutic implications”***. Journal of Clinical Oncology 23(14): 3243-3256. 2005.
- Hernández-García I, Sáenz-González MC, González-Celador R et al. ***“Mortality attributable to smoking in Spain in 2006”***. Anales Sistema Sanitario Navarro 33(1): 23-33. 2010.

- Hess KR, Anderson K, Symmans WF et al. ***“Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer”***. Journal of Clinical Oncology 24: 4236-44. 2006.
- Hou J, Aerts J, den Hamer B, et al. ***“Gene Expression-Based Classification of Non-Small Cell Lung Carcinomas and Survival Prediction”***. Plos One 5(4): e10312. 2010.
- Hu Z, Fan C, Oh DS et al. ***“The molecular portraits of breast tumors are conserved across microarray platforms”***. BMC Genomics 7: 96. 2006.
- Inamura K, Fujiwara T, Hoshida Y et al. ***“Two subclasses of lung squamous cell carcinoma with different gene expression profiles and prognosis identified by hierarchical clustering and non-negative matrix factorization”***. Oncogene 24(47): 7105-13. 2005.
- Ioannidis JP. ***“Microarrays and molecular research: noise discovery?”*** Lancet 365(9458): 454-455. 2005.
- Irizarry RA, Warren D, Spencer F, et al. ***“Multiple-laboratory comparison of microarray platforms”***. Nature Methods 2(5): 345-50. 2005.
- Jarabo JR, Torres AJ. ***“Recidiva local tras cirugía del cáncer de pulmón”***. Anales de Medicina Interna. 25 (2): 53-54. 2008.
- Ji M, Guan H, Gao C, et al. ***“Highly frequent promoter methylation and PIK3CA amplification in non-small cell lung cancer (NSCLC)”***. BMC Cancer 11:147. 2011.
- Karachaliou N, Mayo C, Costa C et al. ***“KRAS mutations in lung cancer”***. Clinical Lung Cancer 14(3): 205-14. 2013.
- Kerr D, Gray R, Quirke P et al. ***“A quantitative multigene RT-PCR assay for prediction of recurrence in stage II colon cancer: selection of the genes in four large studies and results of the independent, prospectively designed QUASAR validation study”***. Journal of Clinical Oncology 27 (Suppl): 169s. 2009.
- Kihara C, Tsunoda T. ***”Prediction of sensitivity of esophageal tumors to adjuvant chemotherapy by cDNA microarray analysis of gene-expression profiles”***. Cancer Research 61: 6474-9. 2001.
- Koivunen JP1, Mermel C, Zejnullahu K et al. ***“EML4-ALK fusion gene and efficacy of an ALK kinase inhibitor in lung cancer”***. Clinical Cancer Research 14(13): 4275-83. 2008.

-
- Kratz JR, Jablons DM. “*Genomic prognostic models in early-stage lung cancer*”. *Clinical Lung Cancer* 10(3): 151–157. 2009.
 - Kren L, Brazdil J, Hermanova M, et al. “*Prognostic significance of anti-apoptosis proteins survivin and bcl-2 in non-small cell lung carcinomas: a clinicopathologic study of 102 cases*”. *Applied immunohistochemistry & molecular morphology* 12(1): 44-49. 2004.
 - Kushi LH, Doyle C, McCullough M, et al. “*American Cancer Society Guidelines on nutrition and physical activity for cancer prevention: reducing the risk of cancer with healthy food choices and physical activity*”. *Cancer Journal for Clinicians* 62(1): 30-67. 2012.
 - Lander ES, Linton LM, Birren B, et al. “*Initial sequencing and analysis of the human genome*”. *Nature* 409 (6822): 860-921. 2001.
 - Larkin JE, Frank BC, Gavras H, et al. “*Independence and reproducibility across microarray platforms*”. *Nature Methods* 2(5): 337-344. 2005.
 - Lynch TJ, Bell DW, Sordella R et al. “*Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib*”. *New England Journal of Medicine* 350: 2129-2139. 2004.
 - Lu Y, Lemon W, Liu P et al. “*A Gene Expression Signature Predicts Survival of Patients with Stage I Non-Small Cell Lung Cancer*”. *PLoS Medicine* 3(12): e467. 2006.
 - Ma XJ, Hilsenbeck SG. “*The HOXB13: IL17BR expression index is a prognostic factor in early-stage breast cancer*”. *Journal of Clinical Oncology* 24: 4611-9. 2006.
 - Malvezzi M, Bertuccio P, Levi F et al. “*European cancer mortality predictions for the year 2013*” *Annales of Oncology* 24(3):792-800. 2013.
 - Mascaux C, Iannino N, Martin B, et al. “*The role of RAS oncogene in survival of patients with lung cancer: a systematic review of the literature with meta-analysis*”. *British Journal of Cancer* 92: 131–135. 2005.
 - Metzger R, Leichman DG, Danenberg KD et al. “*ERCC1 levels complement thymidilate synthase mRNA levels in predicting response and survival for gastric cancer patients receiving combination cisplatin and fluorouracil chemotherapy*”. *Journal of Clinical Oncology* 16: 309-316. 1998.
 - Michiels S, Koscielny S, Hill C. “*Prediction of cancer outcome with microarrays: a multiple random validation strategy*”. *Lancet* 365 (9458):488-492. 2005.

- Mitsudomi T, Hamajima N, Ogawa M et al. ***“Prognostic alterations in patients with non-small-cell lung cancer: a meta-analysis”***. *Clinical Cancer Research* 6: 4055-4063. 2000.
- Mlecnik B, Bindea G, Pagès F et al. ***“Tumor immunosurveillance in human cancers”***. *Cancer and Metastasis Reviews* 30(1): 5-12. 2011.
- Ou SH, Zell JA, Ziogas A et al. ***“Prognostic factors for survival of stage I nonsmall cell lung cancer patients: A population-based analysis of 19,702 stage I patients in the California Cancer Registry from 1989 to 2003”***. *Cancer* 110(7): 1532-41. 2007.
- Páez JG, Janne PA, Lee JC et al. ***“EGFR mutations in lung cancer. Correlation with clinical response to gefitinib therapy”***. *Science* 304: 1497-1500. 2004.
- Pagès F, Galon J, Dieu-Nosjean MC, et al. ***“Immune infiltration in human tumors: a prognostic factor that should not be ignored”***. *Oncogene* 29(8): 1093-102. 2010.
- Paik S, Shak S, Tang G et al. ***“A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer”***. *New England Journal of Medicine* 351: 2817-26. 2004.
- Palmisano W, Divine K, Saccomanno G, et al. ***“Predicting lung cancer by detecting aberrant promoter methylation in sputum”***. *Cancer Research* 60(21): 5954-5958. 2000.
- Pan Q, Pao W, Ladanyi M. ***“Rapid Polymerase Chain Reaction-Based Detection of Epidermal Growth Factor Receptor Gene Mutations in Lung Adenocarcinomas”***. *Journal of Molecular Diagnostics* 7 (3): 396- 403. 2005.
- Pao W, Miller V, Zakowski M et al. ***“EGF receptor gene mutations are common in lung cancer from “never smokers” and are associated with sensitivity of tumors to gefitinib and erlotinib”***. *Proceedings of the National Academy of Sciences of USA* 101: 13306-13311. 2004.
- Pao W, Wang TY, Riely GJ, et al. ***“KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib”***. *PLoS Medicine* 2(1):e17. 2005.
- Pelletier MP, Edwardes MD. ***“Prognostic markers in resectable non-small cell lung cancer: a multivariate analysis”***. *Journal Canadien de Chirurgie* 44 (3): 180-188. 2001.
- Perez-Andres M, Paiva B, Nieto WG et al. ***“Human Peripheral Blood B-Cell Compartments: A Crossroad in B-Cell Traffic”***. *Cytometry Part B Clinical Cytometry*. 78B (Suppl. 1): S47–S60. 2010.

- Perou CM, Sorlie T, Eisen MB et al. “*Molecular portraits of human breast tumours*”. Nature 406(6797): 747-752. 2000.
- Pignon JP, Burdett S. “*Comment on Survival improvement in resectable non-small cell lung cancer with (neo) adjuvant chemotherapy: results of a meta-analysis of the literature*” Lung Cancer 51(2): 261-2; author reply 263-5. 2006.
- Pignon JP, Tribodet H, Scagliotti GV et al. “*Lung adjuvant cisplatin evaluation: a pooled analysis by the LACE Collaborative Group*”. Journal of Clinical Oncology 26 (21): 3552-9. 2008.
- Pisters KMW, Le Chevalier T. “*Adjuvant chemotherapy in completely resected non-small cell lung cancer*”. Journal of Clinical Oncology 23: 3270–3278. 2005.
- Quackenbush J. “*Computational analysis of microarray data*”. Nature Reviews Genetics 2(6): 418-427. 2001.
- Quackenbush J. “*Microarray analysis and tumor classification*”. New England Journal of Medicine 354(23): 2463-2472. 2006.
- Quinn JE, Kennedy RD, Mullan PB et al. “*BRCA1 functions as a differential modulator of chemotherapy-induced apoptosis*”. Cancer Research 63: 6221-6228. 2003.
- Ramaswamy S, Tamayo P, Rifkin R et al. “*Multiclass cancer diagnosis using tumor gene expression signatures*”. Proceedings of the National Academy of Sciences of USA. 98(26): 15149-15154. 2001.
- Rami-Porta R, Wittekind C, Goldstraw P. “*Complete resection in lung cancer surgery: proposed definition*”. Lung Cancer 49 (1): 25–33. 2005.
- Raponi M, Zhang Y, Yu J et al. “*Expression Signatures for Predicting Prognosis of Squamous Cell and Adenocarcinomas of the Lung*”. Cancer Research 66(15): 7466-7472. 2006.
- Raz D, Ray MR, Kim Y et al. “*A Multigene Assay Is Prognostic of Survival in Patients with Early-Stage Lung Adenocarcinoma*”. Clinical Cancer Research 14: 5565-5570. 2008.
- Rodenhuis S, Slebos RJ. “*Clinical significance of ras oncogene activation in human lung cancer*”. Cancer Research 52: 2665s-9s. 1992.
- Roepman P, Jassem J, Smit EF et al. “*An immune response enriched 72-gene prognostic profile for early-stage non-small-cell lung cancer*”. Clinical Cancer Research 15(1): 284–290. 2009.

- Rosell R, Danenberg KD, Alberola V et al. ***“Ribonucleotide reductase messenger RNA expression and survival in gemcitabine/cisplatin-treated advanced non-small-cell lung cancer patients”***. *Clinical Cancer Research* 10:1318-1325. 2004.
- Rosenwald A, Wright G, Chan WC et al. ***“The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma”***. *New England Journal of Medicine* 346(25): 1937-47. 2002.
- Sánchez M.J, Payer T, De Angelis R et al. ***“Cancer incidence and mortality in Spain: estimates and projections for the period 1981–2012”***. *Annals of Oncology* 21: 30-36. 2010.
- Sánchez de Cos J, Miravet L, Abal J et al. ***“Lung cancer survival in Spain and prognostic factors: a prospective, multiregional study”***. *Lung Cancer* 59: 246-254. 2008.
- Sankpal UT, Pius H, Khan M et al. ***“Environmental factors in causing human cancers: emphasis on tumorigenesis”***. *Tumour Biology* 33(5): 1265-74. 2012.
- Saracci R. ***“The interactions of tobacco smoking and other agents in cancer etiology”***. *Epidemiologic Reviews* 9(1): 175–193. 1987.
- Saynak M, Veeramachaneni NK, Hubbs JL et al. ***“Local failure after complete resection of N0-1 non-small cell lung cancer”***. *Lung Cancer* 71(2): 156-65. 2011.
- Scagliotti GV, Fossati R, Torri V et al. ***“Randomized study of adjuvant chemotherapy for completely resected stage I, II, or IIIA non-small-cell Lung cancer”***. *The Journal of the National Cancer Institute* 95(19): 1453-61. 2003.
- Sculier JP. ***“Nonsmall cell lung cancer”***. *European Respiratory Review* 22(127): 33–3. 2013.
- Shapiro MJ, Nguyen CT, Aghajanian H et al. ***“Negative regulation of TCR signaling by linker for activation of X cells via phosphotyrosine-dependent and -independent mechanisms”***. *Journal of Immunology* 181(10): 7055-61. 2008.
- Shi L, Reid LH, Jones WD et al. ***“The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements”***. *Nature Biotechnology* 24(9): 1151-1161. 2006.
- Shirota Y, Stoehlmacher J, Brabender J et al. ***“ERCC1 and thymidylate synthase mRNA levels predict survival for colorectal cancer patients receiving combination oxaliplatin and fluorouracil chemotherapy”***. *Journal of Clinical Oncology* 19: 4298-4304. 2001.

- Siegel R, Naishadham D, Jemal A. “*Cancer statistics*”. CA Cancer Journal for Clinicians 63(1): 11-30. 2013.
- Simon R, Radmacher MD, Dobbin K et al. “*Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification*”. The Journal of the National Cancer Institute 95(1): 14-18. 2003.
- Simon R. “*Roadmap for developing and validating therapeutically relevant genomic classifiers*”. Journal of Clinical Oncology 23(29): 7332-7341. 2005.
- Slebos RJ, Rodenhuis S. “*The ras gene family in human non-small-cell lung cancer*”. Journal of the National Cancer Institute Monographs 13: 23-29. 1992.
- Smith TJ, Temin S, Alesi ER et al. “*American Society of Clinical Oncology provisional clinical opinion: the integration of palliative care into standard oncology care*”. Journal of Clinical Oncology 30(8): 880-7. 2012.
- Soda M, Choi YL, Enomoto M et al. “*Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer*”. Nature. 448(7153): 561-6. 2007.
- Sorlie T, Perou CM, Tibshirani R et al. “*Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications*”. Proceedings of the National Academy of Sciences of USA 98(19): 10869-10874. 2001.
- Steels E, Paesmans M, Berghmans T et al. “*Role of p53 as a prognostic factor for survival in lung cancer: a systematic review of the literature with a meta-analysis*”. European Respiratory Journal 18: 705-719. 2001.
- Strauss GM, Herndon JE, Maddaus MA et al. “*Adjuvant paclitaxel plus carboplatin compared with observation in stage IB non-small-cell lung cancer: CALGB 9633 with the Cancer and Leukemia Group B, Radiation Therapy Oncology Group, and North Central Cancer Treatment Group Study Groups*”. Journal of Clinical Oncology 26: 5043-5051. 2008.
- Subramanian J, Simon R. “*Gene expression-based prognostic signatures in lung cancer: ready for clinical use?*” Journal of National Cancer Institute 102 (7): 464-74. 2010.
- Sun Z, Yang P, Aubry MC et al. “*Can gene expression profiling predict survival for patients with squamous cell carcinoma of the lung?*” Molecular Cancer 3: 35. 2004.
- Suzuki K, Kachala S, Kadota K et al. “*Prognostic Immune Markers in Non-Small Cell Lung Cancer*”. Clinical Cancer Research 17: 5247-5256. 2011.

-
- Tahara H, Sato M, Thurin M et al. *“Emerging concepts in biomarker discovery. The US-Japan Workshop on Immunological Molecular Markers in Oncology”*. Journal of Translational Medicine 7: 45. 2009.
 - Takeuchi K, Soda M, Togashi Y et al. *“RET, ROS1 and ALK fusions in lung cancer”*. Nature Medicine 18: 378-81. 2012.
 - Tangye SG, Good KL. *“Human IgM+CD27+ B cells: memory B cells or “memory” B cells?”* Journal of Immunology 179 (1): 13-9. 2007.
 - Taron M, Rosell R, Felip E et al. *“BRCA1 mRNA expression levels as an indicator of chemoresistance in lung cancer”*. Human Molecular Genetics 13: 2443-2449. 2004.
 - Thomas L. *“In cellular and humoral aspects of the hypersensitive states”* (Ed. Lawrence, HS) Hoeber-Harper; New York. 529-532. 1959.
 - Tibshirani RJ, Hastie TJ, Narasimhan B et al. *“Diagnosis of multiple cancer types by shrunken centroids of gene expression”* Proceedings of the National Academy of Sciences 99: 6567–657. 2002.
 - Tomida S, Koshikawa K, Yatabe Y et al. *“Gene expression-based, individualized outcome prediction for surgically treated lung cancer patients”*. Oncogene 23(31): 5360-70. 2004.
 - Tomizawa K1, Suda K, Onozato R et al. *“Prognostic and predictive implications of HER2/ERBB2/neu gene mutations in lung cancers”*. Lung Cancer. 74(1): 139-44. 2011.
 - Tran TN, Selinger CI, Kohonen-Corish MR et al. *“Fibroblast Growth Factor Receptor 1 (FGFR1) copy number is an independent prognostic factor in non-small cell lung cancer”*. Lung Cancer 2013 [Epub ahead of print]
 - Travis WD, Brambilla E, Müller-Hermelink HK et al. *“World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the Lung, Pleura, Thymus and Heart”*. Lyon: IARC Press. 2004.
 - Travis WD, Brambilla E, Noguchi M et al. *“International association for the study of lung cancer/american thoracic society/European respiratory society international multidisciplinary classification of lung adenocarcinoma”*. Journal of Thoracic Oncology 6: 244-285. 2011.
 - Trinidad EM, Ballesteros M, Zuloaga J et al. *“An impaired transendothelial migration potential of chronic lymphocytic leukemia (CLL) cells can be linked to ephrin-A4 expression”*. Blood 114(24): 5081-90. 2009.

- Van 't Veer LJ, Dai H, van de Vijver MJ et al. “*Gene expression profiling predicts clinical outcome of breast cancer*”. *Nature* 415: 530-6. 2002.
- Venter JC, Adams MD, Myers EW et al. “*The sequence of the human genome*”. *Science* 291(5507):1304-1351. 2001.
- Waller D, Le Chevalier T. “*Lung adjuvant cisplatin evaluation: a pooled analysis by the LACE Collaborative Group*”. *Journal of Clinical Oncology* 26(21): 3552-9. 2008.
- Wigle DA, Jurisica I, Radulovich N et al. “*Molecular profiling of non-small cell lung cancer and correlation with disease-free survival*”. *Cancer Research* 62(11): 3005–3008. 2002.
- Wikenheiser-Brokamp KA. “*Retinoblastoma regulatory pathway in lung cancer*”. *Current Molecular Medicine* 6: 783-793. 2006.
- Winton TL, Livingston R, Johnson D et al. “*Vinorelbine plus cisplatin vs. observation in resected non-small cell lung cancer*”. *New England Journal of Medicine* 352: 2589–2597. 2005.
- Yamamoto H, Shigematsu H, Nomura M et al. “*PIK3CA mutations and copy number gains in human lung cancers*”. *Cancer Research* 68: 6913-2. 2008.
- Zhu CQ, Popova SN, Brown ER et al. “*Integrin alpha 11 regulates IGF2 expression in fibroblasts to enhance tumorigenicity in human non-small-cell lung cancer cells*”. *Proceedings of the National Academy of Sciences of USA* 104: 11754-11759. 2007.
- Zikos TA1, Donnenberg AD, Landreneau RJ et al. “*Lung T-cell subset composition at the time of surgical resection is a prognostic indicator in non-small cell lung cancer*”. *Cancer Immunology and Immunotherapy* 60(6): 819-27. 2011.

9. ANEXOS

Los datos obtenidos en este proyecto han generado la solicitud de dos patentes nacionales y una patente internacional cuyos datos se muestran a continuación:

1.- Patentes Nacionales:

1.1.- “Método de pronóstico del carcinoma no microcítico de pulmón de estadio I o II”. Inventores: Sanz Ortega, Julián; Ferrer Aldea, Milagros; Hernández Prieto, Susana; Romera López, Alejandro; Pérez-Villamil Salgado, Beatriz; Hernando Trancho, Florentino; Gómez Martínez, Ana María; Jarabo Sarceda, Jose Ramón; Torres García, Antonio José, López García-Asenjo, Jose Antonio; González Larriba, Jose Luis; Puente Vázquez, Javier; Díaz-Rubio García, Eduardo. P201132151. 30 Diciembre 2011. Oficina Española de Patentes y Marcas. Madrid.

1.2. “Método de clasificación del carcinoma no microcítico de pulmón basado en la identificación de una respuesta inmune intratumoral”. Inventores: Sanz Ortega, Julián; Ferrer Aldea, Milagros; Hernández Prieto, Susana; Romera López, Alejandro; Pérez-Villamil Salgado, Beatriz; Hernando Trancho, Florentino; Gómez Martínez, Ana María; Jarabo Sarceda, Jose Ramón; Torres García, Antonio José, López García-Asenjo, Jose Antonio; González Larriba, Jose Luis; Puente Vázquez, Javier; Díaz-Rubio García, Eduardo; Subiza Garrido-Lestache, Jose Luis. P201230250. 17 Febrero 2012. Oficina Española de Patentes y Marcas. Madrid.

2.- Patentes Internacionales:

2.1. “Método de clasificación del carcinoma no microcítico de pulmón basado en la identificación de una respuesta inmune intratumoral”. Inventores: Sanz Ortega, Julián; Ferrer Aldea, Milagros; Hernández Prieto, Susana; Romera López, Alejandro; Pérez-Villamil Salgado, Beatriz; Hernando Trancho, Florentino; Gómez Martínez, Ana María; Jarabo Sarceda, Jose Ramón; Torres García, Antonio José, López García-Asenjo, Jose Antonio; González Larriba, Jose Luis; Puente Vázquez, Javier; Díaz-Rubio García, Eduardo; Subiza Garrido-Lestache, Jose Luis. PPCT/ES2012/070919. 28 enero 2013. Oficina Española de Patentes y Marcas. Madrid.



Justificante de presentación electrónica de solicitud de patente

Este documento es un justificante de que se ha recibido una solicitud española de patente por vía electrónica, utilizando la conexión segura de la O.E.P.M. Asimismo, se le ha asignado de forma automática un número de solicitud y una fecha de recepción, conforme al artículo 14.3 del Reglamento para la ejecución de la Ley 11/1986, de 20 de marzo, de Patentes. La fecha de presentación de la solicitud de acuerdo con el art. 22 de la Ley de Patentes, le será comunicada posteriormente.

Número de solicitud:	P201132151	
Fecha de recepción:	30 diciembre 2011, 11:40 (CET)	
Oficina receptora:	OEPM Madrid	
Su referencia:	ES2604.1	
Solicitante:	FUNDACIÓN PARA LA INVESTIGACIÓN BIOMÉDICA DEL HOSPITAL CLÍNICO SAN CARLOS	
Número de solicitantes:	1	
País:	ES	
Título:	Método de pronóstico del carcinoma no microcítico de pulmón de estadio I o II.	
Documentos enviados:	Descripción-1.pdf (56 p.) Reivindicaciones-1.pdf (4 p.) Dibujos-1.pdf (5 p.) Resumen-1.pdf (1 p.) OLF-ARCHIVE.zip FEERCPT-1.pdf (1 p.) SEQLPDF.pdf (12 p.) SEQLTXT.txt	package-data.xml es-request.xml application-body.xml es-fee-sheet.xml feesheet.pdf request.pdf
Enviados por:	CN=ENTIDAD PONS CONSULTORES DE PROPIEDAD INDUSTRIAL SA - CIF A28750891 - NOMBRE PONS ARIÑO ANGEL - NIF 50534279J,OU=703015345,OU=fnmt clase 2 ca,O=FNMT,C=es	
Fecha y hora de recepción:	30 diciembre 2011, 11:40 (CET)	
Codificación del envío:	0F:40:28:38:7F:C7:23:E7:E5:2A:DE:76:47:F8:AB:F8:94:FF:37:35	



Justificante de presentación electrónica de solicitud de patente

Este documento es un justificante de que se ha recibido una solicitud española de patente por vía electrónica, utilizando la conexión segura de la O.E.P.M. Asimismo, se le ha asignado de forma automática un número de solicitud y una fecha de recepción, conforme al artículo 14.3 del Reglamento para la ejecución de la Ley 11/1986, de 20 de marzo, de Patentes. La fecha de presentación de la solicitud de acuerdo con el art. 22 de la Ley de Patentes, le será comunicada posteriormente.

Número de solicitud:	P201230250	
Fecha de recepción:	17 febrero 2012, 14:32 (CET)	
Oficina receptora:	OEPM Madrid	
Su referencia:	ES2604.1BIS	
Solicitante:	FUNDACIÓN PARA LA INVESTIGACIÓN BIOMÉDICA DEL HOSPITAL CLÍNICO SAN CARLOS	
Número de solicitantes:	1	
País:	ES	
Título:	Método de clasificación del carcinoma no microcítico de pulmón basado en la identificación de una respuesta inmune intratumoral.	
Documentos enviados:	Descripción-1.pdf (36 p.) Reivindicaciones-1.pdf (2 p.) Resumen-1.pdf (1 p.) Dibujos-1.pdf (5 p.) OLF-ARCHIVE.zip FEERCPT-1.pdf (1 p.) SEQLPDF.pdf (12 p.) SEQLTXT.txt	package-data.xml es-request.xml application-body.xml es-fee-sheet.xml feesheet.pdf request.pdf
Enviados por:	CN=ENTIDAD PONS CONSULTORES DE PROPIEDAD INDUSTRIAL SA - CIF A28750891 - NOMBRE PONS ARIÑO ANGEL - NIF 50534279J,OU=703015345,OU=fnmt clase 2 ca,O=FNMT,C=es	
Fecha y hora de recepción:	17 febrero 2012, 14:32 (CET)	
Codificación del envío:	99:E5:8D:D1:99:64:6F:B2:13:C0:FB:EA:B7:5C:BB:08:0A:AA:63:69	



JUSTIFICANTE DE PRESENTACIÓN ELECTRÓNICA

Este documento es un justificante del envío de documentos presentados con posterioridad relativos a una solicitud internacional PCT.

Número de solicitud:	PCT/ES2012/070919	
Número de envío:	300082112	
Fecha de recepción:	28 enero 2013, 11:21 (CET)	
Su referencia:	PCT2604.1 Trad	
Solicitante:		
Título:	MÉTODO DE CLASIFICACIÓN DEL CARCINOMA NO MICROCÍTICO DE PULMÓN BASADO EN LA IDENTIFICACIÓN DE UNA RESPUESTA INMUNE INTRATUMORAL	
Documentos enviados:	1) Traducción de la solicitud internacional 2) Carta	TRANAPP.pdf AGNTLET.pdf
Enviados por:	CN=ENTIDAD PONS CONSULTORES DE PROPIEDAD INDUSTRIAL SA - CIF A28750891 - NOMBRE PONS ARIÑO ANGEL - NIF 50534279J,OU=703015345,OU=fnmt clase 2 ca,O=FNMT,C=es	
Fecha y hora de recepción:	28 enero 2013, 11:21 (CET)	
Codificación del envío:	80:08:6A:32:AC:85:7E:90:DF:79:D1:08:4E:F7:64:EC:B3:31:4B:BF	

/Madrid, Oficina Receptora/