



Universidad de Navarra

**CARACTERIZACIÓN, MEDIANTE ESTUDIO DE ASOCIACIÓN DEL  
GENOMA (*GENOME-WIDE ASSOCIATION STUDY, GWAS*), DE  
INDIVIDUOS CON FENOTIPOS EXTREMOS DE SENSIBILIDAD Y  
RESISTENCIA A DESARROLLAR CÁNCER DE PULMÓN NO  
MICROCÍTICO**

**JUAN PABLO FUSCO**

**Tesis doctoral**

**Pamplona, 13 de mayo de 2020**

**Director:**

José Luis Pérez Gracia  
Consultor Clínico  
Departamento de Oncología

**Co-directora:**

Ana Patiño García  
Consultor Técnico  
Catedrática de Genética

**A Dios, mis padres, mi esposa Cecilia,  
mis hijos Clara, Martín y mis hermanos.**

## **AGRADECIMIENTOS**

Al Dr. José Luis Pérez Gracia, creador de este proyecto, un gran oncólogo y motivador, con el que he pasado tardes enteras diseñando estrategias e hipótesis para identificar nuevos biomarcadores. Él me enseñó a entender la importancia de generar conocimiento.

A la Dra. Ana Patiño García, codirectora de esta tesis y parte fundamental de la misma. Sin la energía y espontaneidad que la caracteriza, este proyecto no habría salido adelante.

Al Dr. Alfonso Gúrpide Ayarra, por su colaboración en el proyecto, por su reconocida habilidad como clínico y por ser un gran ser humano.

A D<sup>a</sup> María Pilar Andueza, enfermera con gran dedicación al cuidado de los pacientes y a la recolección de muestras biológicas para estudios traslacionales.

A los Dres. Guillermo Pita y Anna González Neira por su dedicación al proyecto y al análisis de sus resultados desde el Centro Nacional de Investigaciones Oncológicas (CNIO).

A todos ellos les estaré siempre muy agradecido.

## INDICE:

<b>1. INTRODUCCIÓN.....</b>	<b>8</b>
1.1 EL TABACO COMO FACTOR DE RIESGO DEL CÁNCER DE PULMÓN.....	8
1.2 AGENTES CARCINÓGENOS DEL TABACO.....	10
1.3 MECANISMOS CARCINOGENICOS DEL TABACO.....	12
1.4 RIESGO INDIVIDUAL DE DESARROLLAR CÁNCER DE PULMÓN.....	13
1.4.1 <i>Modelos clínicos de predicción del riesgo de desarrollar cáncer de pulmón.....</i>	<i>14</i>
1.4.2 <i>Modelos de predicción del riesgo de desarrollar cáncer de pulmón que incorporan variables genéticas.....</i>	<i>19</i>
1.5 ESTUDIOS DE ASOCIACIÓN DEL GENOMA COMPLETO ( <i>GENOME-WIDE ASSOCIATION STUDIES, GWAS</i> ).....	20
1.5.1 <i>Genes relacionados con la reparación del ADN y las vías de señalización intracelular.....</i>	<i>23</i>
1.5.2 <i>Genes relacionados con el metabolismo de carcinógenos del tabaco.....</i>	<i>24</i>
1.5.3 <i>Genes relacionados con inflamación, estrés oxidativo y microambiente tumoral.....</i>	<i>27</i>
1.6 SELECCIÓN DE FENOTIPOS EXTREMOS.....	29
1.6.1 <i>Fenotipos de alto riesgo de desarrollar cáncer.....</i>	<i>30</i>
1.6.2 <i>Fenotipos de protección para desarrollar cáncer.....</i>	<i>31</i>
1.6.3 <i>Ejemplos de protección para el desarrollo de cáncer en modelos animales y experimentales.....</i>	<i>32</i>
<b>2 JUSTIFICACIÓN.....</b>	<b>35</b>
2.1 IMPACTO CLÍNICO.....	36
2.1.1 <i>De los resultados concretos del proyecto.....</i>	<i>36</i>
2.1.2 <i>De la validación de la metodología de selección de fenotipos extremos.....</i>	<i>37</i>
<b>3 HIPÓTESIS.....</b>	<b>38</b>
<b>4 OBJETIVOS.....</b>	<b>39</b>
4.1 OBJETIVO PRINCIPAL.....	39
4.2 OBJETIVOS SECUNDARIOS.....	39
<b>5 METODOLOGÍA.....</b>	<b>41</b>
5.1 DISEÑO.....	41
5.1.1 <i>Serie de identificación.....</i>	<i>42</i>
5.1.2 <i>Serie de validación.....</i>	<i>42</i>
5.2 CRITERIOS DE INCLUSIÓN.....	43
5.3 TAMAÑO MUESTRAL.....	44
5.4 PROCESAMIENTO Y ESTUDIO DE LAS MUESTRAS.....	45

5.4.1	<i>Extracción del ADN.....</i>	45
5.4.2	<i>Genotipado de la serie de identificación mediante GWAS.....</i>	45
5.4.1	<i>Validación de las variantes seleccionadas.....</i>	46
5.4.2	<i>Imputación genómica.....</i>	47
5.5	ANÁLISIS ESTADÍSTICO.....	47
5.6	CRITERIOS DE SELECCIÓN DE SNPs PARA LA VALIDACIÓN.....	48
5.7	VALOR PRONÓSTICO DE LA EXPRESIÓN DE ARNm Y EXPRESIÓN PROTEICA DE GENES DIANAS SELECCIONADOS.....	49
5.7.1	<i>Valor pronóstico de la expresión de ARNm de PDE10A y ATP10D en pacientes con CNMP en estadios I y II.....</i>	49
5.7.2	<i>Valor pronóstico de la expresión de la proteína PDE10A en pacientes con CNMP en estadios I y II.....</i>	50
5.8	CONSIDERACIONES ÉTICAS.....	51
<b>6</b>	<b>RESULTADOS.....</b>	<b>52</b>
6.1	CARACTERÍSTICAS DE LOS PACIENTES E INDIVIDUOS.....	52
6.2	IDENTIFICACIÓN DE SNPs ASOCIADOS A INDIVIDUOS CON FENOTIPOS EXTREMOS MEDIANTE GWAS.....	53
6.3	VALIDACIÓN DE LOS SNPs ASOCIADOS A INDIVIDUOS CON FENOTIPOS EXTREMOS.....	55
6.4	VALIDACIÓN FUNCIONAL CLÍNICA DE LOS SNPs IDENTIFICADOS.....	58
6.4.1	<i>Valor pronóstico de la expresión de ARNm de PDE10A y ATP10D en pacientes con CNMP en estadios I y II.....</i>	58
6.4.2	<i>Valor pronóstico de la expresión de la proteína PDE10A en el tumor en pacientes con CNMP en estadios I y II.....</i>	60
<b>7</b>	<b>DISCUSIÓN.....</b>	<b>64</b>
<b>8</b>	<b>CONCLUSIONES.....</b>	<b>71</b>
<b>9</b>	<b>BIBLIOGRAFÍA.....</b>	<b>72</b>
	<b>ANEXO I: HOJA DE INFORMACIÓN Y DOCUMENTO DE CONSENTIMIENTO INFORMADO.....</b>	<b>85</b>
	<b>ANEXO 2: PUBLICACIÓN DEL ESTUDIO.....</b>	<b>91</b>

## Tabla de abreviaturas

<b>8-OHdG</b>	8-hidroxiguanosina
<b>ADP</b>	Adenosina difosfato
<b>AMPc</b>	Adenosina monofosfato cíclico
<b>ARN</b>	Ácido ribonucleico
<b>ARNm</b>	Ácido ribonucleico mensajero
<b>AUC</b>	Área bajo la curva
<b>CaBIG</b>	Red de Informática Biomédica del Cáncer
<b>CARET</b>	Ensayo de la eficacia del $\beta$ -caroteno y retinol
<b>CCR5</b>	Receptor 5 de quimioquina
<b>CGEN</b>	Centro Nacional de Genotipado
<b>CI</b>	Intervalo de confianza
<b>CIMA</b>	Centro de investigación médica aplicada.
<b>CLPTM1</b>	Proteína de transmembrana asociada a fisura de labio y paladar
<b>CMP</b>	Carcinoma microcítico de pulmón
<b>CNIO</b>	Centro Nacional de Investigaciones Oncológicas
<b>CNMP</b>	Carcinoma no microcítico de pulmón
<b>CR</b>	Cromosoma
<b>CTLD</b>	Tomografía computarizada de baja dosis
<b>CXCR2</b>	Receptor de quimioquina tipo 2
<b>CYP2A6</b>	Citocromo P450 2A6
<b>CYP2E1</b>	Citocromo P450 2E1
<b>CHRNA3</b>	Subunidad alfa 3 del receptor colinérgico nicotínico
<b>CHRNA5</b>	Subunidad alfa 5 del receptor colinérgico nicotínico
<b>DECARPs</b>	Deficient cancer-risk phenotypes (fenotipos deficientes respecto al riesgo de cáncer)
<b>DPD</b>	Dihidropirimidina deshidrogenasa
<b>DPN</b>	Dinitrofenil
<b>EGF</b>	Factor de crecimiento epidérmico
<b>EGFR</b>	Receptor del factor de crecimiento epitelial
<b>EPIC</b>	Estudio prospectivo europeo sobre cáncer y nutrición
<b>EPOC</b>	Enfermedad Pulmonar Obstructiva Crónica
<b>FDA</b>	Food and Drug Administration (Administración de Medicamentos y Alimentos)
<b>FGF</b>	Factor de crecimiento de fibroblastos
<b>GEO</b>	Ómnibus de Expresión Génica
<b>GH</b>	Hormona de crecimiento
<b>GHR</b>	Receptor de la hormona de crecimiento
<b>GMPc</b>	Guanosina monofosfato cíclico
<b>GWAS</b>	Genome Wide Association Study (Estudio de asociación del genoma completo)
<b>HAS2</b>	Hialurónico sintetasa 2
<b>IARC</b>	International Agency for Research on Cancer (Agencia Internacional de Investigación del Cáncer)
<b>IGF1</b>	Factor de crecimiento similar a la insulina tipo I
<b>IHC</b>	Inmunohistoquímica
<b>IL-8</b>	Interleuquina 8
<b>LD</b>	Linkage disequilibrium (desequilibrio de ligamiento)
<b>LDL</b>	Lipoproteínas de baja densidad
<b>LEF</b>	Factor estimulador linfoide

<b>LLP</b>	Proyecto Pulmonar de Liverpool
<b>MAF</b>	Minor allele frequency (frecuencia del alelo menos frecuente)
<b>MEH/EPHX</b>	Epóxido hidrolasa microsomal
<b>MMP2</b>	Metaloproteinasa 2
<b>mTOR</b>	La diana de rapamicina en células de mamífero
<b>nAChR</b>	Receptores de acetilcolina
<b>NER</b>	Reparación por escisión de nucleótidos
<b>NF-κB</b>	Factor nuclear kappa
<b>NGS</b>	Next generation sequencing (Secuenciación de última generación)
<b>NLST</b>	National Lung Screening Trial (Ensayo nacional de cribado pulmonar)
<b>NNK</b>	Nitrosamina cetona derivada de la nicotina
<b>NNN</b>	N-nitrosornicotina
<b>OR</b>	Odds Ratio (razón de probabilidades)
<b>PAH</b>	Hidrocarburos aromáticos policíclicos
<b>PKA</b>	Proteína quinasa A
<b>PKG</b>	Proteína quinasa G
<b>PLCO</b>	Próstata, pulmón, colon y ovario
<b>PON 1</b>	Paraoxonosa 1
<b>PROCARPs</b>	Proficient cancer-risk phenotypes (fenotipos competentes respecto al riesgo de cáncer)
<b>SG</b>	Supervivencia global
<b>SLP</b>	Supervivencia libre de progresión
<b>SNP</b>	Single Nucleotide Polymorphism, (Polimorfismo de nucleótido único)
<b>TAC</b>	Tomografía axial computada
<b>TCF</b>	Factor de células T
<b>TCGA</b>	Atlas del genoma del cáncer
<b>TERT</b>	Telomerasa transcriptasa inversa
<b>TGF-β</b>	Factor de crecimiento transformante β
<b>TGF-α</b>	Factor de crecimiento transformante α
<b>UDP</b>	Uridinadifosfato
<b>UGT1A1</b>	Uridinadifosfato glucuronosiltransferasa 1A1
<b>UTR</b>	Región no traducida
<b>VEGF</b>	Factor de crecimiento del endotelio vascular
<b>VIH</b>	Virus de inmunodeficiencia humana
<b>VPN</b>	Valor predictivo negativo
<b>VPP</b>	Valor predictivo positivo

## **1. INTRODUCCIÓN.**

El cáncer de pulmón es la primera causa de muerte por tumores en el mundo. En 2018, en Europa, la incidencia fue de 470.000 nuevos casos y la tasa de muerte de 53 casos por cada 100.000 individuos, [1] en EE.UU se estimó una incidencia de 234.000 casos [2]. En España, la tasa de mortalidad en 2018 fue de 47.9 muertes por cada 100.000 habitantes [3].

Existen dos subtipos principales de cáncer de pulmón, el carcinoma no microcítico de pulmón (CNMP), que supone el 85% de los casos y el carcinoma microcítico de pulmón (CMP) con una frecuencia del 15%. A su vez, en el CNMP se distinguen diferentes tipos histológicos, siendo los más frecuentes el adenocarcinoma (40%), los tumores epidermoides (25-30%) y los carcinomas de células grandes (10-15%) [4].

### **1.1 El tabaco como factor de riesgo del cáncer de pulmón.**

El consumo de tabaco es el factor de riesgo más relevante para el desarrollo de múltiples cánceres y, en concreto, es responsable del 80 a 90% de los cánceres de pulmón [5]. El tabaco se fue popularizando en la segunda mitad del siglo XIX y no fue hasta mediados del siglo XX cuando sus efectos nocivos sobre la salud empezaron a ser conocidos.



En 1931, el patólogo argentino Ángel Roffo, de la Universidad de Buenos Aires, estudió en modelos animales la relación entre el tabaco y el cáncer. Para ello, aplicó sobre la piel de un grupo de conejos los productos obtenidos de la destilación del humo del tabaco (alquitrán, etc.), observando que se asociaban a una mayor incidencia de cáncer [6] [7].

En 1939, Müller estudió por primera vez la asociación entre el incremento del consumo de tabaco y la incidencia de cáncer de pulmón en Colonia, Alemania [8]. Tras comparar la información clínica (historia de consumo de tabaco, exposición ocupacional a tóxicos, etc.) de 86 pacientes con cáncer de pulmón con el mismo número de individuos sanos, observó que el consumo de tabaco en el grupo de pacientes con cáncer de pulmón era el doble que en los individuos sanos. El estudio concluyó que el consumo acumulado de más de 2.900 gramos de tabaco (1 cigarrillo = 0,8 gramos) se asociaba a mayor incidencia de cáncer de pulmón.

El abandono del hábito tabáquico impacta en la reducción del riesgo de muerte por cáncer de pulmón. Peto y cols. estimaron la reducción del riesgo de muerte por cáncer de pulmón tras el cese del hábito tabáquico a diferentes edades mediante el estudio de 2.400 casos y 4.600 controles, incluidos entre los años 1950 y 1990. Los individuos que abandonaban el tabaco a los 60, 50, 40 y 30 años, tuvieron un riesgo de desarrollar cáncer de pulmón a los 75 años del 10%, 6% 3%, y 2% respectivamente. Por tanto, el cese del hábito tabáquico a mediana edad, reduciría en más del 90% el riesgo de desarrollar cáncer de pulmón [9].

Brennan y cols. calcularon el riesgo acumulado de presentar cáncer de pulmón en un estudio de 2.633 casos y 2.884 controles en individuos fumadores y no fumadores de Europa central y del este [10]. El riesgo acumulado de muerte por cáncer de pulmón en hombres de 75 años fumadores activos osciló entre el 15 y el 28%; en los exfumadores, entre el 4 y el 7%; y entre los que nunca fumaron entre el 0,4 y el 1,1%. La reducción del riesgo en los individuos que abandonaron el hábito tabáquico antes de los 50 años osciló entre el 67% y el 83%. Los autores concluyeron que la mortalidad por cáncer de pulmón en las próximas tres décadas estará fuertemente condicionada por la capacidad de abandonar el hábito tabáquico.

## **1.2 Agentes carcinógenos del tabaco.**

La Agencia Internacional de Investigación sobre el Cáncer (International Agency for Research on Cancer, IARC) identificó 72 carcinógenos en el humo del tabaco e incluyó 19 de ellos dentro del grupo 1, lo que indica que son carcinógenos confirmados para los seres humanos (Tabla 1). El análisis del humo del tabaco mediante espectrometría de masas ha permitido identificar sus principales carcinógenos. Así, las N-nitrosaminas, el benceno, los hidrocarburos aromáticos policíclicos (PAH), el 1-3 butadieno, las aminas aromáticas, el cadmio o el polonio-210 se sitúan entre los compuestos con mayor concentración y capacidad carcinogénica [4]. La nitrosamina es el producto de la reacción del nitrito con el grupo amino. Existen principalmente dos compuestos presentes en el humo del tabaco con gran capacidad carcinogénica, la nitrosamina cetona derivada de la nicotina (NNK), generada a partir de la nitrosación de la nicotina, y la N-nitrosornicotina (NNN), producto de la nitrosación de la nornicotina. NNK y NNN se generan durante el

proceso de curación del tabaco [4]. Los hidrocarburos aromáticos policíclicos (PAH) están compuestos por dos o más anillos bencénicos y son el producto de la combustión incompleta del tabaco durante la inhalación. Existen 500 diferentes PAH en el humo del tabaco [11]. El benzopireno es el principal PAH y posee gran capacidad carcinogénica [12].

**Tabla 1. Carcinógenos del tabaco y del humo del tabaco identificados como nocivos por la FDA (Food and Drug Administration) y la IARC en 2013.**

---

<b>Carcinógenos categoría 1 (FDA-IARC).</b>
1-3 Butadieno
2-Aminoftaleno
Nitrosamina cetona derivada de la nicotina (NNK)
4-Aminobifenilo
Aflotoxina B1
Arsénico
Benceno
Benzopireno
Berilio
Cadmio
Cromo Hexavalente
Óxido de Etileno
Formaldehido
Níquel
Orto-toluidina
Polonio 210
Uranio (Isotopos 235,238)
Cloruro de Vinilo
N-Nitrosornicotina (NNN)

---

El mecanismo carcinógeno del polonio-210, un componente natural del tabaco, es la emisión de partículas alfa [13]. En 1964 Radford y cols. describieron la asociación del polonio-210 con el desarrollo de cáncer bronquial [14].

### **1.3 Mecanismos carcinogénicos del tabaco.**

En 2014 Warren y cols. revisaron los diferentes mecanismos por los cuales el tabaco genera cáncer [15]. Cuando el humo es inhalado, sus compuestos ingresan en el organismo y son metabolizados por enzimas como las de las familias del citocromo p-450, glutatión-S transferasa y UDP-glucoronil transferasa, que catalizan la transformación de los compuestos tóxicos en productos solubles para facilitar su eliminación. Durante este proceso se producen ciertos compuestos reactivos intermedios, como epóxidos o carbocationes, que se unen covalentemente a determinadas regiones del ADN. Estas uniones se denominan aductos. Los individuos fumadores tienen altos niveles de aductos, que producen una amplia gama de daños del ADN, como por ejemplo, alteraciones del azúcar en el sitio purínico / pirimidínico; modificaciones de las bases (ej. O6-mG and 8-oxoG); o alteraciones más relevantes como roturas de hebras del ADN [16]. Si los aductos producidos por el humo del tabaco o por sus metabolitos evaden el sistema de reparación y permanecen en el tiempo, pueden causar una codificación incorrecta del ADN, originando mutaciones que pueden ser replicadas. Si estas mutaciones ocurren en oncogenes como *KRAS* o en genes supresores de tumores como *TP53*, el resultado es la alteración de los mecanismos de control del crecimiento celular, lo que conlleva una proliferación descontrolada y cáncer [17].

Además, la nicotina actúa a través de receptores de crecimiento, como por ejemplo los receptores de acetilcolina (nAChR) y el receptor del factor de crecimiento epitelial (EGFR). La función de estos receptores es específica para cada tipo celular y su expresión y sensibilidad pueden ser modificados por la nicotina [18]. Por otro lado, el humo del tabaco induce efectos epigenéticos, como la metilación del promotor de genes que modulan la función celular, como el gen supresor *CDKN2A* [19].

La desregulación de la apoptosis y la activación de la angiogénesis son otros mecanismos que favorecen el desarrollo del cáncer. La nicotina inhibe la apoptosis por la vía del factor nuclear kappa (NF- $\kappa$ B) [19] y también induce la expresión de factores de crecimiento como el factor de crecimiento del endotelio vascular (VEGF), el factor de crecimiento de fibroblastos (FGF), el factor de crecimiento derivado de plaquetas (PDGF), el factor de crecimiento transformante  $\alpha$  (TGF- $\alpha$ ) o el factor de crecimiento transformante  $\beta$  (TGF- $\beta$ ) en células endoteliales y células del músculo liso [20].

#### **1.4 Riesgo individual de desarrollar cáncer de pulmón.**

Aunque el tabaco es el factor más importante para desarrollar cáncer de pulmón, no todos los individuos que fuman o han fumado tienen el mismo riesgo de desarrollar esta enfermedad, lo que sugiere que existen diferencias en la susceptibilidad individual. Estimar con mayor exactitud el riesgo individual de desarrollar cáncer de pulmón permitiría mejorar los programas de detección precoz, disminuyendo el número de falsos positivos y negativos de las pruebas, sus complicaciones, su coste y la ansiedad asociada a los estudios complementarios.

En esta sección revisaremos los principales factores relacionados en la actualidad con el riesgo individual de desarrollar cáncer de pulmón inducido por tabaco.

#### 1.4.1 Modelos clínicos de predicción del riesgo de desarrollar cáncer de pulmón.

El desarrollo de modelos de predicción que presenten una buena correlación (concordancia entre las probabilidades estimadas y la frecuencia del desenlace observado) y discriminación (capacidad de distinguir entre individuos que presentarán o no el desenlace) es esencial para mejorar los programas de detección precoz [21].

El riesgo de desarrollar cáncer de pulmón varía ampliamente entre individuos fumadores. Gray y cols. revisaron sistemáticamente 31 estudios que incluían 25 modelos de predicción para estimar el riesgo de desarrollar cáncer de pulmón [22]. En la tabla 2 se resumen algunos de los principales modelos, que se desarrollan a continuación.

Bach y cols. analizaron en 2003 los datos clínicos de los 18.172 pacientes incluidos en el estudio CARET (ensayo de la eficacia del  $\beta$ -caroteno y retinol) [23]. Diseñaron un modelo para estimar el riesgo absoluto de desarrollar cáncer de pulmón a 10 años en individuos fumadores. Se estimó el riesgo teniendo en cuenta la edad, el sexo, la historia del hábito tabáquico y la exposición al asbesto. El modelo muestra que el riesgo individual varía ampliamente entre diferentes individuos fumadores, oscilando por ejemplo entre un riesgo del 15% a 10 años para un hombre de 68 años que fumó activamente 2 paquetes

al día durante 10 años y continúa fumando, y un riesgo del 0,8% a 10 años para una mujer de 51 años que fumó 1 paquete al día durante 28 años hasta los 42 años [22].

**Tabla 2. Modelos de predicción epidemiológicos para estimar el riesgo de desarrollar cáncer de pulmón.** \*Punto de corte del riesgo de desarrollar CNMP: los individuos con valores superiores al punto de corte eran incluidos en el estudio de detección precoz; S: sensibilidad; E: especificidad.

<b>Modelo</b>	<b>Referencia</b>	<b>Punto de corte del riesgo* (%)</b>	<b>S</b>	<b>E</b>	<b>Tipo de estudio</b>
<b>Modelo Bach</b>	Bach [23]	2,5	0,32	0,88	Cohortes
	D`Amelio [24]	5	0,15	0,97	
		7,5	0,06	0,98	
<b>LLP (Proyecto pulmonar de Liverpool)</b>	D`Amelio [24]	2,5	0,62	0,70	Casos - controles
		2,5	0,66	0,66	
	Cassidy [25]	0,91	0,86	0,39	
	Raji [26]	2,5	0,26	0,94	
<b>Modelo Spitz</b>	D`Amelio [24]	5	0,06	0,99	Casos - controles
		7,5	0,02	1,0	
<b>Modelo afroamericano</b>	Etzel [27]	0,75	-	-	Casos - controles
<b>Modelo PLCO</b>	Tammemagi [28]	-	-	-	Cohortes
<b>Modelo PLCOM<sub>2012</sub></b>	Tammemagi [29]	1,34	0,83	0,62	Cohortes
<b>Modelo Hoggart</b>	Hoggart [30]	-	-	-	Cohortes

Maisonneuve y cols. diseñaron una ampliación del modelo de Bach con 5.203 individuos fumadores asintomáticos [31]. Combinaron la información epidemiológica y clínica con la información obtenida de la realización de una tomografía axial computada (TAC) inicial.

En el análisis multivariante las variables estadísticamente significativas asociadas a un mayor riesgo de desarrollar cáncer de pulmón fueron la edad mayor de 65 años, el consumo de tabaco mayor de 60 paquetes - año y la presencia de nódulos no sólidos mayores de 8 mm en la TAC. Los autores concluyeron que los hallazgos de la TAC podrían mejorar el modelo predictivo, aunque es necesario realizar una validación externa.

En 2008 Cassidy y cols. diseñaron el Proyecto Pulmonar de Liverpool (LLP) que estimaba la probabilidad individual de desarrollar cáncer de pulmón durante un período de 5 años [25]. Además de tener en cuenta los tres factores de riesgo más importantes para el cáncer de pulmón (edad, sexo y tabaquismo), el modelo LLP incorporó otros factores, como los antecedentes familiares de cáncer de pulmón, la exposición ocupacional al asbesto, el diagnóstico previo de neumonía y los antecedentes de diagnóstico de tumor maligno no pulmonar. Se incluyeron 579 pacientes con cáncer de pulmón (casos) y 1.157 individuos sin enfermedad (controles), de origen caucásico y en su mayoría fumadores (> 70%). Cuando el riesgo absoluto estimado era del 2,5%, el modelo mostró una sensibilidad del 62% y una especificidad del 70%, indicando una buena discriminación. Por tanto, el modelo LLP permite identificar subgrupos de poblaciones de fumadores y no fumadores que pueden beneficiarse más de la prevención o la vigilancia, aunque es necesaria una validación independiente para confirmar estos hallazgos.

En 2007 Spitz y cols. diseñaron un modelo para predecir el riesgo absoluto de presentar cáncer de pulmón en 1 año, mediante un estudio de casos y controles con un total de 3.852 individuos [32]. La tasa de verdaderos positivos en los individuos categorizados de alto riesgo de desarrollar cáncer de pulmón fue del 69% para fumadores y del 70% para



ex-fumadores. Los autores concluyeron que la capacidad de estimar el riesgo de este modelo fue modesta y similar a otros modelos de predicción.

D'Amelio y cols. compararon y validaron la precisión de los modelos de Bach, LLP y Spitz [24]. En un estudio con 3.197 casos y 1.703 controles, estimaron el riesgo de desarrollar cáncer de pulmón a 5 años en población caucásica. En general, los modelos de LLP y Spitz obtuvieron un valor de discriminación de 0,69 y el modelo de Bach obtuvo un valor de discriminación de 0,66. Además estimaron el valor predictivo positivo (VPP) y el valor predictivo negativo (VPN), con resultados globales para el VPP de entre 0,75 – 0,88 y VPN de entre 0,45 – 0,56 [24]. Los autores concluyeron que un elevado VPP en los tres modelos indicaba una elevada capacidad para identificar a individuos con alto riesgo de desarrollar cáncer de pulmón, mientras que los valores disminuidos de VPN indicaban que muchos individuos serán considerados de forma errónea de bajo riesgo a desarrollar cáncer de pulmón.

En 2011 Tammemagi y cols. diseñaron dos modelos de predicción de riesgo a 9 años. Utilizaron los datos prospectivos de 70.962 individuos del grupo control del modelo próstata, pulmón, colorrectal y ovario (modelo) [28]. El primer modelo incluyó la edad, el estado socioeconómico (educación), el índice de masa corporal, los antecedentes familiares de cáncer de pulmón, la enfermedad pulmonar obstructiva crónica (EPOC), la realización de una radiografía de tórax reciente, el tabaquismo (nunca, previo o actual), el índice paquetes - año y la duración del tabaquismo. En el segundo modelo además se incluyó el tiempo desde el cese del hábito tabáquico, definido como el tiempo en años desde que los individuos fumadores dejaron de fumar definitivamente. El modelo número

uno fue aplicado a una población general y el modelo número dos a una sub-cohorte de individuos fumadores. Los resultados del estudio mostraron en ambos modelos un área bajo la curva (AUC) en la prueba ROC (Característica Operativa del Receptor) de más del 0,80. Los autores concluyeron que los dos modelos de predicción de riesgo de desarrollar cáncer de pulmón demostraron una alta discriminación entre individuos de alto y bajo riesgo, independientemente del sexo y de la raza.

En 2013, Tammemagi y cols. modificaron el modelo PLCO de 2011 para crear un nuevo modelo de predicción de riesgo de desarrollar cáncer de pulmón, denominado modelo de predicción de riesgo de cáncer de próstata, pulmón, colon y ovario modificado (PLCO<sub>M2012</sub>) [29]. Se comparó el modelo PLCO<sub>M2012</sub> con el estudio nacional de cribado de cáncer de pulmón (National Lung Screening Trial, NLST) para determinar qué modelo seleccionaba con mayor precisión a los individuos que desarrollarán cáncer de pulmón en un periodo de 6 años. El estudio NLST incluyó 53.454 individuos de alto riesgo de desarrollar cáncer de pulmón y los aleatorizó a realizar TAC de tórax de baja dosis (CTLD) o radiografía de tórax, observando una disminución de la mortalidad del 20% para el grupo de CTLD [33]. El estudio PLCO<sub>M2012</sub> identificó y validó datos de 80.375 individuos en dos grupos, uno de control y otro de intervención, en el que se realizó vigilancia activa. El estudio PLCO<sub>M2012</sub>, comparado con NLST, identificó con mayor sensibilidad a los individuos en los que se debe iniciar el cribado de cáncer de pulmón. La sensibilidad fue respectivamente del 83% vs 71% ( $p < 0,001$ ) y la especificidad del 62,9% vs 62,7% ( $p = 0,54$ ) [29].

#### 1.4.2 Modelos de predicción del riesgo de desarrollar cáncer de pulmón que incorporan variables genéticas.

En esta sección se incluyen aquellos modelos que incorporan variables genéticas para predecir el riesgo individual de desarrollar cáncer de pulmón, con el objetivo de mejorar su sensibilidad y especificidad. Algunos de ellos se basan en los modelos descritos anteriormente.

Hoggart y cols. diseñaron un estudio prospectivo con 169.035 individuos fumadores y ex-fumadores [30] en individuos del estudio prospectivo europeo sobre cáncer y nutrición (EPIC) [34]. El estudio diseñó dos modelos de riesgo, un modelo para fumadores y exfumadores y otro para no fumadores. Se estratificó por variables epidemiológicas-clínicas. Se identificaron dos polimorfismos de nucleótido único (Single Nucleotide Polymorphisms, SNP) significativamente asociados al cáncer de pulmón en 2.359 (3%) individuos fumadores: rs8034191, en el cromosoma 15q25, relacionado con el gen del receptor de acetilcolina de tipo nicotínico (*CHRNA*); y rs402710, en el cromosoma 5p15, relacionado con el gen de la telomerasa (*TERT*) [35]. El modelo mostró una elevada capacidad de predicción a 1 y 5 años, con una correlación de 0,84 y 0,78 respectivamente. No obstante, el beneficio de añadir estos dos SNPs al modelo de predicción fue limitado, debido a que los individuos genotipados representaban solo un 3% de la cohorte en estudio

El estudio LLP desarrolló un modelo de riesgo para predecir el riesgo absoluto a 5 años

de desarrollar cáncer de pulmón [25]. Los autores recalibraron y ampliaron el modelo LLP incorporando un SNP localizado en el gen seizure 6-like (*SEZ6L*), para mejorar su precisión [36]. El SNP rs663048, variante polimórfica p.(Met430Ile), localizado en la región 22q12.2 del gen *SEZ6L* se asoció a un mayor riesgo de desarrollar cáncer de pulmón en estudios de casos y controles en población europea [26]. En el estudio de LLP ampliado, en 338 individuos se genotipó el SNP rs663048. Se observó una mejoría en la reclasificación del 12% en el modelo ampliado, en especial en individuos donde el modelo LLP no era capaz de clasificar en alto y bajo el riesgo de desarrollar cáncer de pulmón.

Aunque este tipo de modelos de predicción podrían ser muy relevantes para diseñar estrategias de detección precoz, no pueden aplicarse a la práctica clínica, porque no poseen suficiente sensibilidad y especificidad y no han sido suficientemente validados. Esto ha llevado a incorporar técnicas genómicas de alto rendimiento para intentar mejorar la capacidad de predecir el riesgo de desarrollar cáncer de pulmón, como se comentará en las siguientes secciones.

### **1.5 Estudios de asociación del genoma completo (*Genome-Wide Association Studies, GWAS*).**

Un SNP es una variante de la secuencia de ADN que afecta a una sola base nitrogenada localizada en cualquier región genética, dentro o fuera del gen y que puede o no modular su expresión. Por ejemplo, los SNPs implicados en el metabolismo celular, la regulación del ciclo celular y la inmunidad se asocian a susceptibilidad a desarrollar cáncer [37].

Los estudios GWAS utilizan plataformas de alto rendimiento que analizan simultáneamente varios millones de SNPs a lo largo de todo el genoma humano, permitiendo analizar de forma masiva la variabilidad genética entre diferentes grupos de individuos. Sus resultados se intentan asociar con el desarrollo de una enfermedad o rasgo determinado, para lo que se debe comparar un grupo de casos, que presentan la enfermedad en estudio, con un grupo control que no presenta la enfermedad.

Estas tecnologías de alto rendimiento de genotipado, no sólo permiten realizar estudios cualitativos de genotipado de marcadores genéticos en el genoma codificante y no codificante sino que además, aplicando los algoritmos adecuados, permiten identificar regiones de delección o duplicación. Aunque el GWAS es una herramienta de gran potencial, presenta limitaciones. Por ejemplo, las características fenotípicas complejas que componen factores genéticos y ambientales están determinadas muchas veces por un gran número de *loci* desconocidos. Además, los programas informáticos utilizados en la fase de identificación para estimar el riesgo de los efectos de un SNP no siempre reflejan la función biológica de esa alteración. Por otro lado, estos programas utilizan un nivel de significación estadística arbitraria, ya que se desconoce con exactitud el peso estadístico de cada hallazgo [38]. Otra limitación es que los estudios de GWAS se realizan en poblaciones concretas, por lo que no es posible extrapolar sus resultados a otras poblaciones. Por ejemplo, algunos SNP del receptor de acetilcolina de tipo nicotínico ( $nAChR\alpha3$ ) se han asociado a mayor riesgo de desarrollar CNMP en individuos caucásicos, pero no en asiáticos [39].

En los últimos 20 años se han publicado múltiples trabajos que asocian la susceptibilidad

de padecer cáncer de pulmón con SNPs identificados mediante GWAS. En la tabla 3 se describen algunos de los principales SNPs identificados en estos estudios.

**Tabla 3. SNPs asociados al riesgo de desarrollar cáncer de pulmón. Adaptado de los meta-análisis de Marshall [39] y Wang [40].** \* p de la prueba de Q para heterogeneidad. MAF: *frecuencia del alelo menos frecuente*.

Gen	SNP	MAF (%)	Número de casos / controles	Raza	p	Número de estudios
<b><i>APEX1</i> [41]</b>	rs1130409 (T/G)	-	1.807 / 3.065	Mixta	0,021	7
<b><i>AXIN2</i> [42]</b>	rs2240308 (C/T)	37.4	758 / 742	Mixta	6,39 × 10 <sup>-5</sup>	3
<b><i>CHRNA3</i> [40]</b>	rs6495309 (T/C)	38.4	3381 / 4244	Asiática	6,55 × 10 <sup>-8</sup>	4
<b><i>CXCR2</i> [43]</b>	rs1126579 (T/C)	55.4	942 / 964	Mixta	0,009	3
<b><i>CYP2E1</i> [44]</b>	rs6413432 (A/T)	22.1	2944 / 3347	Mixta	6,76 × 10 <sup>-8</sup>	14
<b><i>HYKK</i> [40]</b>	rs931794 (G/A)	32.8	2435 / 3180	Mixta	1,85 × 10 <sup>-7</sup>	5
<b><i>PON 1</i> [45]</b>	rs662 (A/G)	46.7	995 / 834	Mixta	2,02 × 10 <sup>-4</sup>	3
<b><i>CD3EAP</i> [46]</b>	rs967591 (A/G)	32.0	676 / 726	Mixta	0,036	3
<b><i>CYP2A6</i> [47]</b>	rs1801272 (A/T)	3.9	2411 / 2644	Caucásica	0,001	3
<b><i>HIF1A</i> [48]</b>	rs11549467 (A/G)	9.4	509 / 566	Mixta	1,62 × 10 <sup>-9</sup>	3
<b><i>PDCD5</i> [49]</b>	rs1862214 (G/C)	32.	737 / 683	Mixta	0,001	3
<b><i>TP53</i> [50]</b>	rs12951053 (G/T)	9.9	475 / 569	Mixta	0,011	3
<b><i>TP63</i> [51]</b>	rs10937405 (T/C)	42.6	4927 / 8794	Mixta	2,20 × 10 <sup>-4</sup>	4
<b><i>XRCC1</i> [52]</b>	rs3213255 (G/A)	38.1	1089 / 1506	Mixta	0,001	3
<b><i>ERCC2</i> [54]</b>	rs13181 (A>C)	--	6824 / 8683	Caucásica	0,35*	11
<b><i>MMP2</i> [55]</b>	rs2285053 (C>T)	--	859 / 86	Mixta	0,42*	2

En las siguientes secciones se resumen los principales genes que incluyen SNPs relacionados con el riesgo de desarrollar cáncer de pulmón, clasificándolos según su función.

1.5.1 Genes relacionados con la reparación del ADN y las vías de señalización intracelular.

- **Endonucleasa 1 apurínica / apirimidínica (*APEX1*):** es una enzima esencial en la reparación por escisión de nucleótidos (NER), un importante mecanismo de reparación del daño del ADN producido por la oxidación y alquilación. El Dr. Lo y cols. estudiaron 730 casos (pacientes con cáncer de pulmón) y 730 controles (individuos sin enfermedad) y observaron que el SNP 656T>G rs1760944, localizado en la región promotora del gen *APEX1*, se asociaba a un menor riesgo de cáncer de pulmón entre fumadores importantes [41].
- **Excision Repair Cross - Complementing 2 / Xeroderma Pigmentosum (*ERCC2 / XPD*)** es una proteína que participa en la vía de reparación NER. Zhan y cols. demostraron que el SNP del exón 13 del gen *ERCC2*, p.Lys751Gln (c.2251A>C), se asociaba a un incremento del riesgo de desarrollar cáncer de pulmón en fumadores caucásicos (OR: 1,26, CI 1,12-1,42) [54].
- **Axis inhibition protein 2 (*AXIN2*):** es una proteína implicada en la vía de señalización Wnt /  $\beta$ -catenina, que tiene un importante papel en la carcinogénesis al interaccionar con factores de transcripción (TCF / LEF) [56]. Un meta-análisis realizado por Gong y cols. describió que el SNP rs2240308 (c.148C>T) del gen *AXIN* se asociaba significativamente con menor riesgo global de diferentes tipos de cáncer, incluyendo el cáncer de pulmón [42]. Esta disminución se observaba en homocigosis (TT vs CC), heterocigosis (CT vs CC: OR = 0,52, 95% CI: 0,36-0,74,  $P < 0,001$ ) y en modelos dominantes (CT+TT vs CC: OR = 0,69, 95% CI:

0,56-0,85,  $P < 0,001$ ). Este mismo SNP no se asociaba al riesgo de desarrollar cáncer de pulmón en población japonesa, lo que podría sugerir que esta asociación varía entre diferentes poblaciones.

### 1.5.2 Genes relacionados con el metabolismo de carcinógenos del tabaco.

Son múltiples las enzimas implicadas en la bioactivación y detoxificación de carcinógenos. Las más relevantes incluyen:

- **La subunidad alfa-3 del receptor de acetilcolina de tipo nicotínico (nAChR $\alpha$ 3)**, una proteína codificada por el gen *CHRNA3*, presente en cerebro y pulmón. En el cerebro tiene un papel relevante en el control del hábito tabáquico. Por otro lado, en el pulmón los receptores de nicotina expresados en las células epiteliales interactúan con la nicotina y / o sus derivados (4-metilnitrosamina, butanona), lo que genera señales de proliferación celular y transformación neoplásica [57]. Los SNP del receptor de acetilcolina de tipo nicotínico podrían alterar la expresión del gen y su función proteica, con la consecuente mayor susceptibilidad para desarrollar cáncer de pulmón. Wang y cols. realizaron un estudio de casos y controles con aproximadamente 2.890 individuos en cada brazo que asoció el SNP rs6495309 (C>T) de forma estadísticamente significativa con un aumento del riesgo de desarrollar cáncer de pulmón en población china (OR 1,43, 95% IC: 1,27-1,61;  $p = 2,6 \times 10^{-9}$ ) [40] [57]. También encontraron una asociación significativa entre el SNP rs6495309 y el desarrollo de cáncer de



pulmón en poblaciones de raza mixta.

- **Citocromo P450 2A6 (CYP2A6):** es una enzima del complejo CYP2A6, implicada en la inactivación de la nicotina y la activación de carcinógenos como las nitrosaminas específicas del tabaco [58]. La nicotina es el principal alcaloide del humo del tabaco y es un compuesto psicoactivo responsable de la adicción al tabaco. La acción y la duración del efecto de la nicotina sobre el sistema nervioso central dependen de su metabolismo y de su aclaramiento. La inactivación y eliminación de la nicotina dependen principalmente de la actividad de la enzima CYP2A6, que convierte la nicotina en dos metabolitos principales denominados cotinina y 3 hidroxí-cotinina. Estos dos metabolitos reflejan la capacidad de aclaramiento de la nicotina por parte de la enzima CYP2A6 [59]. Varios estudios han mostrado que los individuos denominados metabolizadores lentos presentan menor función de CYP2A6, lo que genera un menor consumo de tabaco. Además, los metabolizadores lentos tendrían un menor riesgo de desarrollar cáncer de pulmón en comparación con aquellos individuos con función enzimática normal (OR 2,03, IC 95% 1,21 – 3,40;  $p = 0,007$ ), lo que podría explicarse por la activación más lenta de los carcinógenos del humo del tabaco [60].

Tanner y cols. estudiaron la importancia clínica de la enzima CYP2A6 y su relación con el cáncer, observando que el SNP rs1801272 del gen *CYP2A6* produce una disminución de la actividad enzimática [47]. Esta variación en la actividad podría tener efectos protectores indirectos contra el cáncer de pulmón, por la menor exposición a carcinógenos nocivos. La enzima modula la dependencia a la nicotina, la cantidad de tabaco consumido y la capacidad de un individuo para

dejar de fumar. Un estudio muestra que los fumadores con metabolismo lento de la enzima CYP2A6 tienen más probabilidades de dejar de fumar que aquellos con función enzimática normal [61]. Por otro lado, CYP2A6 puede actuar directamente mediante la activación metabólica de nitrosaminas procarcinógenas específicas del tabaco como NNN y NNK a través de la hidroxilación  $\alpha$ . Los individuos fumadores con actividad más lenta de CYP2A6 tendrían menor activación de nitrosaminas, lo que disminuiría la exposición a carcinógenos pulmonares activados [62]. Por lo tanto, ser un metabolizador lento de la enzima CYP2A6 reduciría el consumo de tabaco y la activación de carcinógenos, lo que se asociaría a un menor riesgo de desarrollar cáncer de pulmón.

- **El citocromo P450 2E1 (CYP2E1)** pertenece a una familia de enzimas con un importante papel en el metabolismo de diferentes carcinógenos como benceno, butadieno y nitrosaminas específicas del tabaco [63]. Existen diferentes SNPs del gen *CYP2E1* implicados en la susceptibilidad de desarrollar cáncer. En 2015 Ye y cols. publicaron un meta-análisis con más de 5.000 casos y controles que incluía 34 estudios en población asiática, en el que se observó que rs6413432 del gen *CYP2E1* se asociaba de forma significativa a una disminución del riesgo de desarrollar cáncer de pulmón en fumadores [44].

Liu y cols. estudiaron la relación de los SNPs del gen *CYP2E1* con el riesgo de desarrollar cáncer de cavidad oral en un estudio de casos y controles en individuos fumadores caucásicos y afroamericanos. Se analizó en ambos grupos el ADN germinal de muestras obtenidas de la cavidad oral, el grupo de casos el DNA germinal fue obtenido en el seguimiento tras la cirugía. Como hallazgo de interés

en el grupo de casos se identificaron variantes alélicas del gen *CYP2E1* asociadas a un incremento del riesgo de desarrollar este tumor [64].

### 1.5.3 Genes relacionados con inflamación, estrés oxidativo y microambiente tumoral.

- **Receptor de quimioquina tipo 2 (*CXCR2*)** y su ligando, IL-8 (interleuquina 8) generan una unión de alta afinidad que constituye un importante mediador de inflamación. IL-8 es producida por monocitos, macrófagos y células tumorales y parece promover la angiogénesis [65]. En 2015 Ryan y cols. utilizaron herramientas *in-silico* para diseñar un estudio de casos y controles en población europea-americana y japonesa y observaron que el SNP rs1126579 en gen *CXCR2* (3' UTR) se asociaba a reducción del riesgo de cáncer de pulmón en ambas poblaciones [43]. Los autores concluyeron que el SNP rs1126579 altera el sitio de unión del microRNA miR-516a-3p con el gen *CXCR2*, lo que modifica su expresión génica.
- **El gen paraoxonasa 1 (*PON1*)**, codifica una enzima antioxidante que atenúa los efectos del estrés oxidativo. *PON1* se expresa en diferentes tejidos, incluyendo las células endoteliales y las células del epitelio alveolar del pulmón, donde interactúa con el humo del tabaco. El consumo de tabaco induce estrés oxidativo, que puede causar daños graves a las moléculas celulares, incluyendo ADN, proteínas y lípidos y es un mecanismo carcinogénico relacionado con el desarrollo de cáncer de pulmón [66].

La actividad de *PON1* está parcialmente modulada por los SNPs del gen. Eom y cols. observaron que aproximadamente el 54% de la variación de la actividad de la enzima PON1 se asocia al SNP Arg192Glu (R192Q, rs662) en población coreana [67]. Un meta-análisis que analizaba 25 estudios mostró que los individuos de origen asiático que presentaban el SNP (rs662) en el gen presentaban menor riesgo de desarrollar cáncer [68]. En 2015 Eom y cols. genotiparon 416 casos con cáncer de pulmón y 416 controles para investigar la relación entre los SNPs de *PON1* y el estrés oxidativo producido por el hábito tabáquico [45]. Los niveles urinarios de 8-hidroxiguanosina (8-OHdG) y ácido tiobarbitúrico se utilizaron como indicadores biológicos del estrés oxidativo. El genotipo polimórfico para el SNP rs662 se asoció a una disminución de la incidencia de cáncer de pulmón en comparación con el genotipo normal (OR 0,60, 95% CI: 0,36 – 0,99). Los individuos no fumadores con genotipo *PON1* rs662 normal poseían bajos niveles urinarios de 8-OHdG. El estudio sugiere un posible efecto protector de rs662 sobre la carcinogénesis pulmonar y el estrés oxidativo, que podría estar modulado por la interacción entre los SNPs de *PON1* y el consumo de tabaco.

- **Metaloproteinasa 2 (*MMP2*):** es una proteasa involucrada en el proceso de escisión del colágeno IV, que constituye un componente fundamental de la membrana basal y está implicada en la iniciación y el desarrollo del cáncer [55]. Los SNPs de *MMP2* rs243865 y rs2285053 identificados en el promotor del gen han demostrado tener una función reguladora sobre su transcripción [69]. Peng y cols. publicaron un meta-análisis de 51 estudios incluyendo más de 40.000 individuos que investigaba la asociación entre polimorfismos de la región de los

promotores de la *MMP2* y el desarrollo de cáncer. [55] Se identificó que los SNPs *MMP2* rs243865 y rs2285053 se asociaron significativamente a menor riesgo de cáncer de pulmón.

En resumen, existe una amplia variación interindividual en el metabolismo del tabaco, lo que contribuye a generar diferente susceptibilidad a desarrollar cáncer en distintos individuos. Aunque los resultados de los estudios de GWAS son muy relevantes para generar hipótesis sobre el desarrollo del cáncer de pulmón, aún no son aplicables a la práctica clínica asistencial para estimar el riesgo de desarrollar esta enfermedad. Por ello, en este proyecto hemos desarrollado una estrategia alternativa para identificar SNPs asociados al riesgo de desarrollar cáncer de pulmón mediante GWAS, tal y como se describe en las siguientes secciones.

## **1.6 Selección de fenotipos extremos.**

Aunque los seres humanos presentamos grandes similitudes en nuestra información genética, las variaciones se asocian con una amplia gama de fenotipos diferentes. La importancia de caracterizar estas diferencias es reconocida por la comunidad científica y existen iniciativas muy relevantes para comprender su relación con el desarrollo de las diferentes enfermedades [70].

Una estrategia para identificar las características genéticas asociadas a los diferentes fenotipos clínicos consiste en seleccionar individuos que presentan fenotipos extremos

respecto a la característica estudiada, para analizar en dichos sujetos el sustrato molecular subyacente a dichos fenotipos. Un ejemplo de esta estrategia es la identificación de mutaciones activadoras del gen *EGFR* como biomarcador de actividad de los inhibidores de *EGFR* en pacientes con cáncer de pulmón, realizada en dos estudios retrospectivos que analizaron pacientes que presentaron respuestas marcadas a estos tratamientos. De forma llamativa, tan solo fue preciso estudiar respectivamente a 9 [71] y 16 pacientes, [72] para identificar estas mutaciones. Este y otros ejemplos, confirman que este tipo de diseños permiten mejorar la potencia estadística con tamaños muestrales reducidos, tal y como se ha revisado en estudios previos [73] [74].

#### 1.6.1 Fenotipos de alto riesgo de desarrollar cáncer.

Una de las estrategias de mayor éxito para identificar factores genéticos asociados con alto riesgo de desarrollar cáncer ha sido identificar y estudiar individuos o familias con fenotipos muy característicos, consistentes en presentar un riesgo muy elevado de desarrollar tumores. Algunos ejemplos bien conocidos son el hallazgo de mutaciones germinales de *TP53* en el síndrome de Li-Fraumeni [75], descrito mediante la identificación de un exceso de riesgo de muerte por rhabdomyosarcoma en hermanos [76] [77]; el descubrimiento de que los pacientes con retinoblastoma hereditario presentaban una inactivación de ambas copias del gen supresor tumoral del retinoblastoma, [78] [79], tal y como anticipó Knudson; [80] o la detección de mutaciones en *BRCA1*, un gen aislado en familias cuyos individuos presentaban riesgo muy elevado de desarrollar cáncer de mama a edades tempranas [81]. El factor común de estos hallazgos es que se describieron en un número relativamente pequeño de individuos con fenotipos muy

característicos, consistentes en un riesgo muy elevado de desarrollar tumores a edades tempranas. La identificación de estos fenotipos fue el factor primordial que permitió realizar estos descubrimientos y el que les confirió su gran validez biológica. Muchos de estos hallazgos constituyen la base del desarrollo de la disciplina del cáncer hereditario, y también han dado lugar a diferentes desarrollos terapéuticos, como los inhibidores de Poli ADP ribosa polimerasa (PARP-1) que han demostrado una gran eficacia clínica en el tratamiento de diferentes tumores [82].

#### 1.6.2 Fenotipos de protección para desarrollar cáncer.

La mayoría de los estudios que han utilizado la estrategia de selección de fenotipos extremos han analizado fenotipos perjudiciales, consistentes en un aumento del riesgo de desarrollar cáncer. Probablemente esto se debe a que dichos fenotipos son más sencillos de identificar, debido a las anomalías que presentan. Sin embargo, algunos trabajos realizados en otras áreas de la medicina han estudiado fenotipos beneficiosos, caracterizados por la ausencia de una enfermedad a pesar de la existencia de importantes factores de riesgo asociados a la misma. Uno de los ejemplos más llamativos es la identificación de individuos que no desarrollaban infección por algunas variantes del Virus de la Inmunodeficiencia Humana (VIH) a pesar de presentar una exposición muy alta al mismo [83]. El estudio de estos individuos permitió identificar distintas alteraciones en el gen que codifica para el correceptor 5 de quimiocina (*CCR5*), correceptor utilizado por el virus para penetrar en sus células diana [84-86]. Dado que las alteraciones de *CCR-5* no parecen producir ningún otro efecto, la observación de que dichos individuos no desarrollaban VIH a pesar de su elevada exposición al virus resultó fundamental para

identificar la causa de dicha protección, que se ha convertido en una diana terapéutica para esta enfermedad.

Otro ejemplo de selección de fenotipos extremos protectores es la identificación de mutaciones inactivadoras en el gen *PCSK9* (proteína convertasa subtilisina / kexina tipo 9) en 2 de 128 individuos con valores muy bajos de lipoproteínas de baja densidad (LDL) [87]. Este hallazgo ha llevado al diseño de estrategias farmacológicas para inhibir PCSK9, como evolocumab, un anticuerpo monoclonal que inhibe la proteína PCSK9 y que reduce significativamente los niveles de LDL y el número de eventos cardiovasculares en pacientes con enfermedad arterioesclerótica [88].

### 1.6.3 Ejemplos de protección para el desarrollo de cáncer en modelos animales y experimentales.

Aunque no se han descrito alteraciones genéticas que protejan del desarrollo de cáncer en seres humanos, algunos modelos de experimentación animales apoyan la existencia de este tipo de alteraciones. Por ejemplo, los modelos de ratón con copias adicionales del gen supresor p53 (ratones “súper TP53”) presentan resistencia al desarrollo de tumores inducidos por carcinógenos químicos, sin ninguna otra alteración fenotípica relevante [89]. Otro modelo de ratón con déficit inducido de ciclina D1 también presenta resistencia marcada al desarrollo de tumores inducidos por los oncogenes ERBB2 y HRAS [90]. Aunque estos modelos animales fueron creados artificialmente, su existencia



apoya la plausibilidad biológica de la existencia de alteraciones moleculares asociadas con resistencia al desarrollo de tumores.

Finalmente, se ha descrito una posible menor incidencia de cáncer en algunas especies de animales, como los elefantes, en los que se han estudiado posibles mecanismos protectores, como la existencia de una correlación negativa entre su tamaño y el número de retrovirus endógenos, potencialmente carcinogénicos [91]; o la refuncionalización del Factor Inhibidor de Leucemia (LIF6), una proteína pro-apoptótica regulada por TP53 en respuesta al daño del ADN [92].

También se ha descrito una menor incidencia de cáncer en la rata lampiña (*Heterocephalus glaber*), un roedor de vida longeva [93] y alta resistencia al desarrollo de tumores espontáneos e inducidos [94] [95] Esta resistencia se ha atribuido a que los fibroblastos de estos animales producen ácido hialurónico de muy alto peso molecular - posiblemente para permitir que su piel sea muy elástica, lo que facilita su vida en el medio subterráneo. Así, el bloqueo del ácido hialurónico mediante enzimas degradadoras, o mediante la generación modelos *knock-out*, revierte esta protección frente al desarrollo de tumores. [96] Otro trabajo ha identificado alteraciones en varios genes de esta especie relacionados con la función de los telómeros, con la señalización de TGF-beta mediante SMAD-3, proteína que tiene función de segundo mensajero. Además, otro posible mecanismo propuesto como resistencia al desarrollo de cáncer está relacionado con la inhibición por contacto mediada por el gen *CDKN2A* [97].

El objetivo de este proyecto es identificar factores genéticos asociados a un aumento y disminución de la susceptibilidad individual de desarrollar cáncer de pulmón inducido por consumo de tabaco. Para ello, hemos estudiado individuos con fenotipos extremos de alta y baja sensibilidad a desarrollar este tumor, seleccionando sujetos con elevado consumo de tabaco que, o bien desarrollaron cáncer de pulmón a edad muy temprana (casos extremos, con alta sensibilidad), o bien no lo desarrollaron a edad muy tardía (controles extremos, con baja sensibilidad). Esta estrategia puede ayudar a identificar las causas moleculares que subyacen a estos fenotipos de gran relevancia clínica, al concentrar el estudio en un número reducido de individuos que presentan los fenotipos clínicos relevantes de interés, siendo por tanto baja la posibilidad de que se deban al azar; y al excluir del análisis a los individuos con fenotipos intermedios, de significado incierto, que podrían confundir la interpretación de los resultados.

La selección de fenotipos extremos ha demostrado su utilidad en la interpretación de la gran cantidad de datos generados por las técnicas de alto rendimiento, como la secuenciación genómica masiva [98] [99]. Además, la realización de los estudios directamente en seres humanos implica una mayor relevancia clínica de los posibles hallazgos; y la realización de estudios en un número pequeño de individuos con fenotipos característicos, simplifica el proceso y disminuye el coste. Por último, hemos utilizado una técnica molecular de muy alto rendimiento, como GWAS, para aumentar las posibilidades de encontrar hallazgos de interés.

## 2 JUSTIFICACIÓN.

El tabaco es el principal factor de riesgo conocido para desarrollar CNMP. No obstante, el nivel de riesgo que confiere no es uniforme para todos los individuos expuestos, ya que algunos individuos fumadores desarrollan CNMP a edad muy temprana, mientras que otros no lo desarrollan a edades avanzadas. Esto sugiere la existencia de una diferente susceptibilidad individual al desarrollo de CNMP inducido por tabaco. La identificación de estas diferencias en enfermedades multifactoriales, como el CNMP, conlleva una gran dificultad metodológica por múltiples razones, entre las que destacan que el factor estudiado no es el único que produce la enfermedad; y que probablemente, las diferencias en la susceptibilidad tampoco residen en una única causa. Una estrategia adecuada para minimizar estos problemas puede basarse en:

- La identificación de individuos con una exposición muy elevada al tabaco que, o bien desarrollan CNMP a edad temprana; o bien no lo desarrollan a edad avanzada. Estos dos grupos de individuos probablemente presentan una mayor posibilidad de albergar mecanismos de aumento o disminución del riesgo de desarrollar CNMP asociado al consumo de tabaco que la población general.
- La caracterización genómica de estos individuos mediante técnicas de alto rendimiento que valoran un número muy elevado de variables, como GWAS, que permite analizar cientos de miles o incluso millones de SNPs simultáneamente.

Por tanto, el objetivo de este proyecto es describir perfiles genéticos asociados a fenotipos extremos de muy bajo y alto riesgo de desarrollar cáncer de pulmón, ante la exposición al factor de riesgo más relevante que se conoce, utilizando una técnica de alto rendimiento como GWAS.

## **2.1 Impacto clínico.**

### **2.1.1 De los resultados concretos del proyecto.**

La identificación de perfiles genéticos asociados a fenotipos de muy bajo y muy alto riesgo de desarrollar cáncer de pulmón inducido por el consumo de tabaco, que es el factor de riesgo más relevante conocido, conllevaría los siguientes beneficios:

1. Identificación de poblaciones de alto riesgo de desarrollar CNMP inducido por tabaco. Estos individuos serían excelentes candidatos para realizar programas de información y deshabituación tabáquica con carácter preferente, así como para participar en programas de cribado para la detección precoz de cáncer de pulmón. Esta selección de la población estudiada podría mejorar los resultados de los programas de intervención y disminuir su coste.
2. Identificación de mecanismos biológicos de protección frente al desarrollo de CNMP, en el grupo de fumadores que no desarrolla cáncer. El conocimiento de estos mecanismos podría generar estrategias terapéuticas relevantes para la

prevención del desarrollo de CNMP inducido por tabaco. Además, hay que destacar que los individuos del grupo sin cáncer de nuestro estudio no solo no desarrollaron CNMP, sino tampoco otros tumores, dependientes o no del consumo de tabaco.

3. Identificación de vías de carcinogénesis y oncogénesis en el grupo de fumadores que desarrollan cáncer a edad temprana. Estas vías podrían ser manipuladas mediante intervenciones farmacológicas o de otro tipo. Los mecanismos identificados podrían ser estudiados también en otros tumores y las posibles intervenciones podrían ser extrapoladas a los mismos.

#### 2.1.2 De la validación de la metodología de selección de fenotipos extremos.

La validación de esta nueva metodología de investigación traslacional en Oncología permitiría su aplicación a otros tipos de tumores, estudiando individuos con muy alto y muy bajo riesgo de desarrollo de cáncer ante la exposición a factores de riesgo relevantes conocidos. También sería posible su aplicación a otras enfermedades no tumorales con factores de riesgo conocidos.

### **3 HIPÓTESIS.**

El riesgo de desarrollar CNMP inducido por tabaco no es uniforme en todos los individuos y depende de factores genéticos que confieren protección y / o aumento de la susceptibilidad individual. La selección de individuos con fenotipos extremos de muy alto y bajo riesgo de desarrollar CNMP inducido por tabaco y el estudio de su ADN genómico mediante una técnica de muy elevado rendimiento como GWAS puede permitir identificar estos factores.

## 4 OBJETIVOS.

### 4.1 Objetivo principal.

Identificar factores genéticos relacionados con la susceptibilidad a desarrollar CNMP inducido por tabaco. Para ello, se seleccionaron pacientes con CNMP y controles que presentaban fenotipos extremos respecto al riesgo de desarrollar CNMP inducido por tabaco y se estudió su ADN germinal mediante GWAS. Se estudiaron los grupos que se describen a continuación:

- **Casos extremos:** pacientes jóvenes fumadores diagnosticados de CNMP.
- **Controles extremos:** individuos mayores fumadores importantes no diagnosticados de CNMP ni de otros tumores.

### 4.2 Objetivos secundarios.

1. Validar los resultados del GWAS en una serie independiente de individuos con fenotipos similares.
2. Generar hipótesis que permitan validar funcionalmente en modelos preclínicos los SNP seleccionados.

3. Establecer por primera vez la prueba de concepto de que existen perfiles genéticos de protección frente al desarrollo de cáncer.
  
4. Validar la metodología de selección de fenotipos extremos.

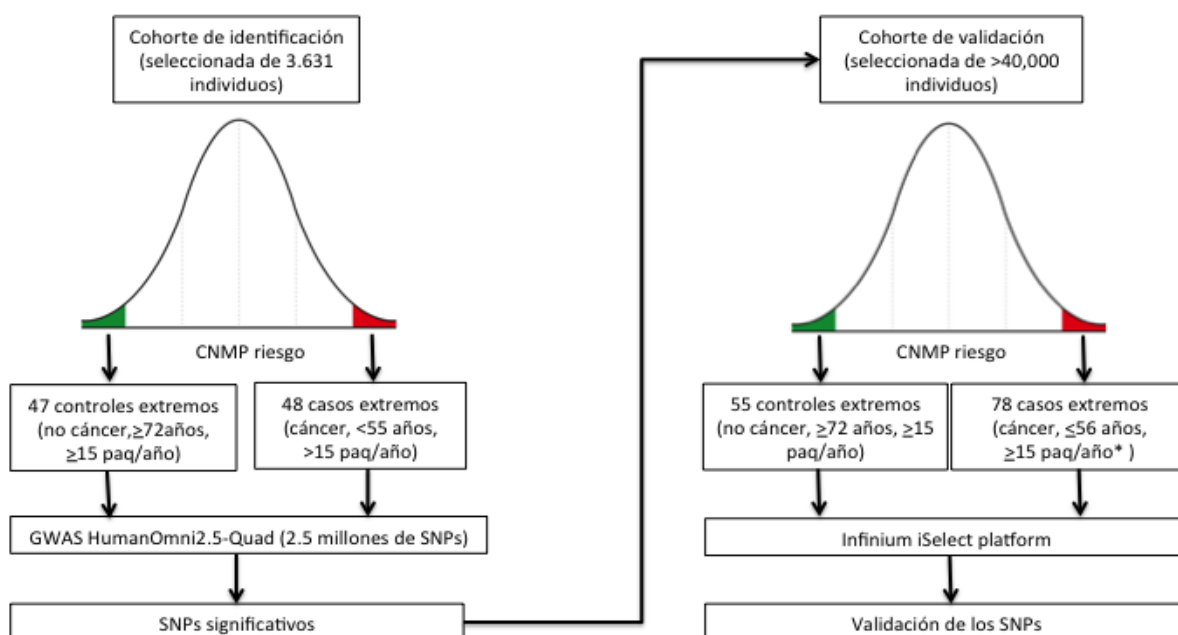


## 5 METODOLOGÍA.

### 5.1 Diseño.

Este es un estudio traslacional de casos y controles. Se seleccionaron individuos que presentaban fenotipos extremos respecto al riesgo de desarrollar CNMP inducido por tabaco. El estudio consta de dos series, de identificación y validación (Figura 1).

**Figura 1. Diseño del estudio:** seleccionamos de nuestras series individuos que presentaban fenotipos extremos de alto y bajo riesgo de desarrollar CNMP inducido por tabaco. Los fumadores importantes que desarrollaron CNMP a edad temprana se consideraron casos extremos y aquellos individuos que no desarrollaron CNMP a edad avanzada a pesar del elevado consumo de tabaco se consideraron controles extremos. Generamos una serie de identificación en la que estudiamos los SNPs mediante GWAS; y una serie de validación, en la que analizamos los SNP más significativos de la serie de identificación. **Abreviaturas:** CNMP: carcinoma no microcítico de pulmón; SNPs: polimorfismos de nucleótido único.



### 5.1.1 Serie de identificación.

En esta serie se analizaron mediante GWAS individuos que presentaban fenotipos de alto y bajo riesgo de desarrollar CNMP inducido por tabaco.

La serie de identificación estaba formada por casos extremos: pacientes muy fumadores que habían desarrollado CNMP a una edad temprana y por controles extremos: individuos muy fumadores que no habían desarrollado CNMP ni otros tumores a edad avanzada. Se establecieron como puntos de corte de edad y consumo de tabaco aquellos que permitían seleccionar a los individuos más extremos de nuestra serie.

### 5.1.2 Serie de validación.

En esta serie se validaron los resultados más significativos de la serie de identificación, en una población de individuos de características similares a la serie de identificación.

## 5.2 Criterios de inclusión.

Los sujetos seleccionados cumplían los siguientes criterios:

1. Fumadores importantes, definidos por un consumo de tabaco mayor o igual a 15 paquetes - año (15 o más años fumando un paquete de 20 cigarrillos al día o una cantidad equivalente), con las siguientes características:

- Casos extremos: pacientes menores de 55 años con diagnóstico confirmado de CNMP en cualquier estadio.
- Controles extremos: individuos mayores de 72 años no diagnosticados de CNMP ni de otros tumores.

2. Consentimiento informado por escrito.

3. Ausencia de diagnóstico de otros tumores en los últimos 5 años (excepto tumores cutáneos de tipo no melanoma) y de síndromes familiares de cáncer.

### 5.3 Tamaño muestral.

En este estudio utilizamos GWAS, una técnica que analiza un número tan elevado de variables que por definición, en su diseño no es posible establecer una hipótesis alternativa [100]. Es decir, no es posible definir *a priori* qué alteraciones moleculares van a ser el objetivo principal de estudio, lo que complica el cálculo del tamaño muestral. Por esta razón, en los estudios de GWAS suele ser habitual incluir un número muy elevado de pacientes para alcanzar la máxima significación estadística posible [101].

Por otro lado, en este estudio empleamos la metodología de selección de fenotipos extremos, que consiste en estudiar grupos reducidos de individuos con fenotipos muy característicos [73] [74]. Por el momento, no se han descrito métodos de cálculo del tamaño muestral para este tipo de estudios, si bien numerosos ejemplos confirman que es posible realizar hallazgos moleculares muy relevantes mediante esta estrategia estudiando muestras reducidas (en algunas ocasiones, incluso estudiando un solo individuo) [72] [102]. No obstante, es evidente que es preferible emplear muestras mayores, siempre que se disponga del suficiente número de individuos que presenten los fenotipos extremos de interés [103].

Por todo ello, en este estudio incluimos todos los individuos disponibles de nuestra serie que presentaban fenotipos muy característicos, dando prioridad a la selección de casos que presentaran fenotipos verdaderamente extremos sobre el tamaño muestral absoluto.

## 5.4 Procesamiento y estudio de las muestras.

### 5.4.1 Extracción del ADN.

La extracción del ADN genómico de células mononucleares obtenidas de sangre periférica se realizó en el Centro de Investigación Médica Aplicada (CIMA) de la Universidad de Navarra, en Pamplona. Para la extracción se empleó el Kit QIAGEN: Qiamp ADN Mini Kit REF. 51306, utilizando el protocolo recomendado para el mismo. Las muestras se conservaron a -20°C hasta su estudio. El ADN se preparó en placas de PCR de 96 pocillos (Axigen PCR 96 ref: PCR-96-SG-C) a una concentración 50 ng / µl. Se incluyeron controles internos. La concentración de ADN se cuantificó mediante espectrofotometría (Nanodrop).

### 5.4.2 Genotipado de la serie de identificación mediante GWAS.

El genotipado en la serie de identificación se realizó mediante la plataforma Illumina HumanOmni2.5-Quad BeadChip utilizando el protocolo del fabricante (Illumina, San Diego, EE.UU.). El chip utilizado (Array HumanOmni2.5-Quad BeadChip) permite genotipar un total de 2.5 millones de SNPs en todo el genoma. La técnica se realizó en la Unidad de Genotipado del Centro Nacional de Investigaciones Oncológicas (CNIO), que pertenece al Centro Nacional de Genotipado (CEGEN, [www.cegen.org](http://www.cegen.org)). En primer lugar, se lleva a cabo la amplificación global de la muestra y su fragmentación. El ADN se hibrida en un array de alta densidad donde se produce la extensión con bases

marcadas de manera simultánea para un total de 2.5 millones de posiciones polimórficas. La fluorescencia se detecta mediante un scanner y un software convierte la señal en genotipo.

#### 5.4.1 Validación de las variantes seleccionadas.

El genotipado en la serie de validación se realizó mediante la plataforma Infinium (Illumina, San Diego, CA, EE.UU.) siguiendo el protocolo del fabricante. Se diseñó un array que incluía los SNPs seleccionados en la fase de identificación.

La tecnología Infinium consta de varias fases. Inicialmente se utiliza una microplaca donde el ADN es desnaturalizado y luego amplificado en una reacción isotérmica seguida de fragmentación. El objetivo de este paso es obtener fragmentos de tamaño óptimo (300 - 600 pb) para su hibridación. La muestra es purificada con isopropanol y resuspendida con buffers que generan condiciones ideales para la hibridación. A continuación, la muestra se retira de la microplaca y se aplica sobre una matriz (microarray), un soporte sólido que tiene adheridas microesferas en su superficie que contienen sondas complementarias a los SNPs que se van a interrogar. Una vez que el ADN ha hibridado en el microarray, se lleva a cabo la extensión y marcaje de estas secuencias con moléculas nucleotídicas marcadas. Los nucleótidos T (timina) y A (adenina) están marcados con dinitrofenil (DNP) y los nucleótidos C (citosina) y G (guanina) con biotina. Tras un proceso de lavado quedan las sondas marcadas, los anticuerpos anti-DNP de color rojo y los anticuerpos anti-biotina (estreptovidina) de color verde, que se unen

específicamente a estas sondas. Posteriormente, las señales son amplificadas y la fluorescencia producida se detecta mediante el escaneo del microarray para su posterior análisis ([www.illumina.com](http://www.illumina.com)).

#### 5.4.2 Imputación genómica.

La imputación genómica es una herramienta esencial en el análisis de GWAS, ya que permite estimar con mayor precisión los efectos de muchas variantes genéticas que no son genotipadas directamente, aumentando así el poder de la prueba [104].

En nuestro estudio se realizó un proceso de imputación genómica con el objetivo de identificar variantes de SNP adicionales que no fueron genotipadas directamente, basándose en los resultados del estudio. La imputación genómica se realizó mediante el programa IMPUTE v2.3.2 usando el panel de haplotipos de 1000 genomas fase 3 (b37) [105] [106]. Los genotipos con índices  $> 0,3$  fueron analizados por asociación.

### **5.5 Análisis estadístico.**

En las series de identificación y de validación se estimaron para cada SNP su razón de riesgo (OR) y su error estándar. Se utilizaron modelos de regresión logística multivariante, ajustados por sexo. Las covariables edad y consumo de tabaco no se incluyeron en el análisis estadístico porque se utilizaron en el diseño.

Para el análisis de genotipado de la serie de validación, se utilizó GenTrain v2.0 en GenomeStudio 2011.1 (Illumina, San Diego, EE.UU.). Se excluyeron del análisis las variantes genéticas con más del 2 % de pérdida de genotipado con un *call rate* < 0,95% y aquellas que no cumplían con el equilibrio de Hardy-Weinberg ( $p \leq 1 \times 10^{-5}$ ).

## **5.6 Criterios de selección de SNPs para la validación.**

Los SNPs para el estudio de validación se seleccionaron a partir de los resultados del estudio de GWAS en la serie de identificación, empleando los siguientes criterios:

1. Mayor significación estadística en el análisis de regresión logística. Se evitó incluir SNPs redundantes (en desequilibrio de ligamiento). Cuando existían dos SNPs con una alta correlación ( $r^2 > 0,8$ ), solo se seleccionó aquel que obtuvo el mejor “score” de Illumina.
2. SNPs relacionados con genes con interés biológico, teniendo en cuenta los datos publicados en la literatura.
3. SNPs monomórficos en alguno de los grupos, es decir, aquellos cuyos alelos variantes estaban presentes exclusivamente en uno de los dos grupos de individuos con fenotipos extremos, incluso si no se cumplían los criterios previos de forma estricta.



## 5.7 Valor pronóstico de la expresión de ARNm y expresión proteica de genes dianas seleccionados.

Para estimar el valor funcional y biológico de los SNPs validados en un modelo clínico, evaluamos la expresión de ARNm de *PDE10A* y *ATP10D* y la expresión proteica de *PDE10A* en tejido tumoral de pacientes con CNMP en estadios I y II tratados mediante resección quirúrgica.

### 5.7.1 Valor pronóstico de la expresión de ARNm de *PDE10A* y *ATP10D* en pacientes con CNMP en estadios I y II.

Se correlacionaron los datos transcriptómicos con la supervivencia libre de progresión (SLP) y la supervivencia global (SG) mediante herramientas *in silico* (KM-plotter) en pacientes que presentaban CNMP en estadios I y II.

Utilizamos la base de datos (<http://kmplot.com>) [107] para analizar la expresión génica de pacientes con CNMP con estadios I y II. Los datos de los pacientes analizados fueron descargados de la Red de Informática Biomédica del Cáncer (CaBIG), el Ómnibus de Expresión Génica (GEO) y el Atlas del Genoma del Cáncer (TCGA).

Calculamos la SLP y la SG desde la cirugía hasta la fecha de recurrencia o muerte, respectivamente. Evaluamos las diferencias significativas entre los grupos mediante pruebas de log-rank.

### 5.7.2 Valor pronóstico de la expresión de la proteína PDE10A en pacientes con CNMP en estadios I y II.

Evaluamos el valor pronóstico de la expresión proteica de PDE10A mediante análisis inmunohistoquímico en pacientes con CNMP en estadios I y II operados en la Clínica Universidad de Navarra, en Pamplona. El análisis inmunohistoquímico y el grado de expresión se realizaron según el protocolo descrito previamente [108].

Para el análisis inmunohistoquímico de la proteína PDE10A se desparafinaron los cortes histológicos, se redujo la actividad de peroxidasa endógena y se recuperó el antígeno como se describe en Grbesa y cols. [109]. Posteriormente, los cortes histológicos se mantuvieron a humedad y temperatura constantes durante una noche, a 4°C, junto a los anticuerpos humanizados anti-PDE10A (Genetex) a dilución 1 / 500 (diluyente Dako). Tras la aplicación del polímero marcado (Dako) durante 30 minutos a temperatura ambiente, los cortes histológicos se trataron con diaminobencidina y fueron teñidos con hematoxilina. La especificidad del anticuerpo PDE10A se confirmó previamente mediante un estudio de Western Blot e inmunocitoquímica en líneas celulares. La especificidad se expresó como los diferentes niveles de proteína. Además, se realizaron controles de isotipo y controles negativos para mostrar la especificidad del anticuerpo. Las escalas de extensión y de tinción fueron establecidas por análisis semicuantitativo.

Los cortes fueron evaluados por dos observadores independientes que establecieron las variables de extensión (porcentaje de células positivas; 0-100%) y de intensidad de la tinción (1: leve; 2: moderada; y 3: intensa). Se seleccionó la mediana de expresión de

PDE10A en el tumor como punto de corte para discriminar entre expresión baja versus alta. Las lecturas discordantes independientes fueron revisadas simultáneamente por ambos observadores

Se utilizaron los modelos de riesgos proporcionales de Cox univariantes y multivariantes para determinar los efectos de las variables en la SG y la SLP. Las variables con  $p < 0,1$  en el análisis univariante se incluyeron en el análisis multivariante. El análisis estadístico se realizó con SPSS v15.0 (Chicago, Illinois, EE. UU.) y Stata12 (College Station, TX: EE. UU.).

## **5.8 Consideraciones éticas.**

El estudio fue aprobado por el Comité Ético de Investigación de la Clínica Universidad de Navarra [110]. Antes de ser incluidos en el estudio, todos los pacientes firmaron el consentimiento informado (Anexo 1). En este documento se explica la naturaleza del estudio y se especifica la información previa requerida por la Ley 14/2007, del 3 de julio, de Investigación Biomédica.

## 6 RESULTADOS.

### 6.1 Características de los pacientes e individuos.

La serie de identificación se seleccionó a partir de 3.631 individuos de las bases de datos de la Clínica Universidad de Navarra (Pamplona, Navarra), del CIMA y del hospital Universitario Nuestra señora de la Candelaria (Tenerife).

La serie de validación independiente se seleccionó a partir de 39.880 individuos incluidos en el brazo español del (EPIC, [www.epic-spain.com](http://www.epic-spain.com)). También se incluyeron casos adicionales de la Clínica Universidad de Navarra.

Se analizaron 228 individuos con fenotipos extremos, 95 en la fase de identificación y 133 en la fase de validación (Tabla 4). En la serie de identificación se incluyeron 47 casos extremos con una edad media de 49 años (rango: 38 - 55) y una media de consumo de tabaco de 41 paquetes - año (rango: 15 - 49). Los controles extremos fueron 48 individuos con edad media de 76 años (rango: 72 - 84) y una media de consumo de tabaco de 69 paquetes - año (rango: 40 - 150). Un individuo del grupo control de esta primera fase fue excluido por bajo *call rate* en el genotipado ( $< 0,98$ ).

En la serie de validación se incluyeron 78 casos extremos con una media de edad de 49 años (rango: 35 - 56) y consumo medio de tabaco de 38 paquetes - año (rango: 4 - 101); y 55 controles extremos con una edad media de 77 años (rango 72 - 85) con consumo

medio de tabaco de 48 paquetes - año (rango 16 - 123). En la tabla 4 se presentan características adicionales de los sujetos incluidos en el estudio.

**Tabla 4. Características de los individuos incluidos en el estudio en las fases de identificación y validación.** NA: No aplicable.

	Serie de identificación (n = 95)		Serie de validación (n = 133)	
	Casos extremos (n = 47)	Controles extremos (n = 48)	Casos extremos (n = 78)	Controles extremos (n = 55)
<b>Sexo (hombres / mujeres)</b>	31 / 16	43 / 5	52 / 26	51 / 4
<b>(%)</b>	(66 / 34)	(89 / 11)	(67 / 33)	(93 / 7)
<b>Media de edad, (años)</b>	49	76	49	77
<b>(rango)</b>	(38 - 55)	(72 - 84)	(35 - 56)	(72 - 85)
<b>Media de consumo, (paquetes - año)</b>	41	69.4	38	48
<b>(rango)</b>	(15 - 99)	(40 - 150)	(4 - 101)	(16 - 123)
<b>Histología</b>		NA		NA
<b>Adenocarcinoma</b>	29		44	
<b>Carcinoma escamoso</b>	14		12	
<b>Otros</b>	4		21	

## 6.2 Identificación de SNPs asociados a individuos con fenotipos extremos mediante GWAS.

En la fase de identificación se genotiparon 2.379.855 SNPs. Se excluyeron 61.061 SNPs debido a bajo *call rate* (< 0,95). Otros 370 SNP fueron excluidos por no cumplir el equilibrio

de Hardy-Weinberg. En total se seleccionaron 2.318.553 SNPs para el análisis. De estos 2.318.553 SNPs se seleccionaron 36 SNPs para ser estudiados en la fase de validación, siguiendo los criterios descritos previamente (principalmente su alta asociación en el estudio de GWAS,  $p < 5 \times 10^{-4}$ , Tabla 5).

**Tabla 5. SNPs seleccionados para la fase de validación.** Las posiciones de los cromosomas se basan en Genome Reference Consortium Human Build 37 (GRCh37 / hg19). SNP ID; identificación de polimorfismo, Cr; cromosoma, MAF; Frecuencia del alelo menos frecuente.

SNP ID	Cr.	Posición	Gen	MAF	P valor
rs1189107	14	56917786	37kb5' de C14orf101	0,30	4.46E-05
rs1300661	14	56919803	35kb5' de C14orf101	0,30	7.41E-05
rs2825830	21	21197762	68kb3'def AP000946.2	0,44	7.46E-05
rs17405280	5	31384619	16kb3' de DROSHA	0,43	7.91E-05
rs9983664	21	21091587	174kb3' de AP000946.2	0,45	9.22E-05
rs2235967	14	68249499	ZFYVE26	0,33	1.18E-04
rs4811157	20	49949839	44kb3' de AL079339.1	0,37	1.22E-04
rs6727285	2	5488456	201kb3' de AC107057.1	0,39	1.32E-04
rs12660420	6	166193932	PDE10A	0,28	1.48E-04
rs12853031	23	6126513	NLGN4X	0,50	1.49E-04
rs17064225	5	174417049	CTC-281M20.1	0,24	1.78E-04
rs6693447	1	2330190	RER1	0,43	2.26E-04
rs656661	5	31384205	16kb3' de DROSHA	0,40	2.30E-04
rs73160755	12	132177959	18kb5'def SFSWAP	0,29	2.37E-04
rs13049227	21	28097013	112kb3' de ADAMTS1	0,47	2.51E-04
rs1225147	11	76555560	16kb5' de ACER3	0,20	2.53E-04
rs12504035	4	131121249	176kb5' de RP11-521E5.1	0,17	2.65E-04
rs11095019	23	6124767	NLGN4X	0,49	2.69E-04
rs56845266	7	64081538	45kb5' de ZNF107	0,17	2.82E-04
rs2306911	17	80613588	RAB40B	0,48	2.94E-04
rs7098190	10	125457952	3.8kb3' de GPR26	0,30	2.98E-04
rs10281505	7	20716244	ABC5	0,18	2.99E-04
rs2577254	2	11975372	1.7kb3' de MIR4262	0,47	3.19E-04
rs7672688	4	113001776	7.9kb5' de RP11-269F21.3	0,46	3.23E-04
rs11980286	7	50043518	ZPBP	0,17	3.40E-04
rs2913366	5	5417809	3kb5' de KIAA0947	0,46	3.56E-04
rs6835978	4	47500814	ATP10D	0,24	3.69E-04
rs1225132	11	76551635	20kb5' de ACER3	0,20	3.77E-04
rs7278106	21	21075420	190kb3' de AP000946.2	0,39	3.78E-04
rs11236557	11	75552129	UVRAG	0,34	3.87E-04
rs1371121	7	149375086	37kb5' de KRBA1	0,27	4.07E-04
rs1034331	21	21141056	125kb3' de AP000946.2	0,39	4.39E-04
rs6833846	4	113924858	RP11-650J17.1	0,22	4.43E-04
rs7946600	11	99925167	CNTN5	0,28	4.44E-04
rs13232564	7	18952433	HDAC9	0,16	4.76E-04
rs2407254	21	21071070	195kb3' de AP000946.2	0,31	4.81E-04

### 6.3 Validación de los SNPs asociados a individuos con fenotipos extremos.

Se replicaron dos SNPs en la fase de validación (Tabla 6). El SNP que alcanzó mayor significado estadístico fue rs12660420 en el locus 6q27 ( $p$  combinada =  $5,66 \times 10^{-5}$ ; OR combinada 2,88, IC 95%: 1,69 – 4,61). Este SNP se localiza en la región exónica de un transcrito no codificante del gen *PDE10A*.

**Tabla 6. SNP validados.** La asociación entre los SNPs y el riesgo de NSCLC se analizó mediante regresión logística ajustada por sexo. Se consideró un modelo de herencia aditivo. Los OR son por copia del alelo menor. Las posiciones de los cromosomas se basan en Genome Reference Consortium Human Build 37 (GRCh37 / hg19). SNP: polimorfismo de nucleótido único; Cr: Cromosoma; MAF: frecuencia del alelo menos frecuente; OR: odds ratio; CI: intervalo de confianza; Ref: alelo de referencia; Alt: alelo alternativo.

SNP	Cr.	Posición	Ref.	Alt.	Serie	P	MAF	OR	95% IC
rs12660420	6	166193932	C	T	Identificación	1,48E-04	0,28	5,29	2,2 -12,51
					Validación	0,04805	0,32	1,88	1,01 – 3,5
					Combinadas	5,66E-05	0,31	2,80	1,69 – 4,61
rs6835978	4	47500814	A	G	Identificación	3,69E-04	0,24	5,15	2,09 – 12,7
					Validación	0,04303	0,32	1,80	1,02 – 3,19
					Combinadas	1,02E-04	0,31	2,57	1,6 – 4,1

El SNP rs12660420 se asoció a un bloque de desequilibrio de ligamiento de 85,8 Kb en el cromosoma 6 (coordenadas 166.109.230-166.195.063, GRCh37 - hg19), definido por los SNPs más alejados en ambos sentidos y que poseen una correlación detectable ( $r^2 > 0,20$ ) con el SNP rs12660420.

Además, se imputaron a este locus todas las variantes conocidas que presentaban una MAF > 5%. Se imputaron y analizaron de forma fiable 275 SNP (escala de imputación  $r^2 > 0,3$ ), pero ninguno de ellos alcanzó una  $p$  más significativa que la alcanzada por rs12660420.

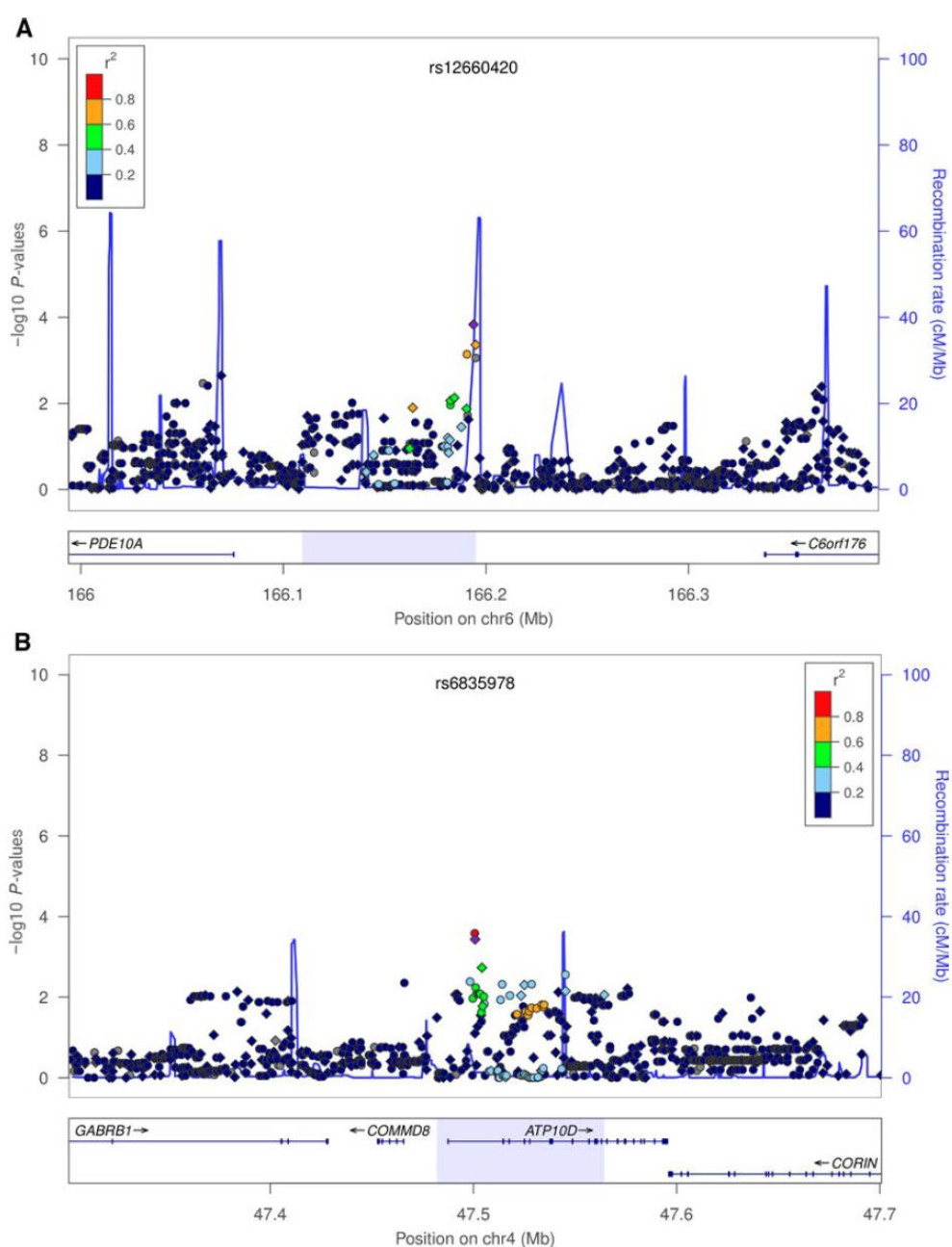
El segundo SNP replicado en la fase de validación fue rs6835978, localizado en el cromosoma 4p12 ( $p$  combinada =  $1,02 \times 10^{-4}$ , OR combinada 2,57, IC 95%: 1,6 – 4,1, Tabla 6). Este SNP es una variante intrónica del gen *ATP10D* (ATPase phospholipid transporting 10D) [111]. Este gen no ha sido relacionado con cáncer de pulmón ni con otros tumores.

El SNP rs6835978 se asoció a un tramo de desequilibrio de ligamiento de 82 Kb en el cromosoma 4 (coordenadas 47.481.971 - 47.564.368) (GRCh37 - hg19). Realizamos imputación genómica de 228 SNPs asociados y rs6835978 se mantuvo como el SNP más significativo (junto con rs12510653, en desequilibrio de ligamiento completo con rs6835978 [ $r^2 = 1$ ]).

Las gráficas con los desequilibrios de ligamiento relacionados con los SNP rs12660420 y rs6835978 se muestran en la Figura 2.



**Figura 2.** Asociación y recombinación del bloque de desequilibrio de ligamiento (LD) que contiene los SNPs validados y límites de  $\pm 200$  Kb. El gráfico muestra la región genómica que contiene (A) rs12660420 (B) rs6835978, SNPs asociados con el riesgo de cáncer de pulmón y los valores p (asociación  $-\log_{10}$ ) de los SNPs genotipados e imputados en la serie de identificación (MAF > 5%). Los círculos y los diamantes indican los SNPs genotipados e imputados, respectivamente. También se muestran las tasas de recombinación. El color de los SNPs indica la fuerza de LD ( $r^2$ ) para el SNP rs12660420 (A) y para el SNP rs6835978 (B). El rectángulo debajo de la gráfica muestra el mapa de genes en la región. La imagen sombreada en azul claro muestra el bloque LD asociados a los SNPs validados. Las tasas de recombinación se basan en el Proyecto 1000 Genomas, y las coordenadas genómicas se basan en GRCh37 / hg19. Para la realización de los gráficos se utilizó el software LocusZoom [112].



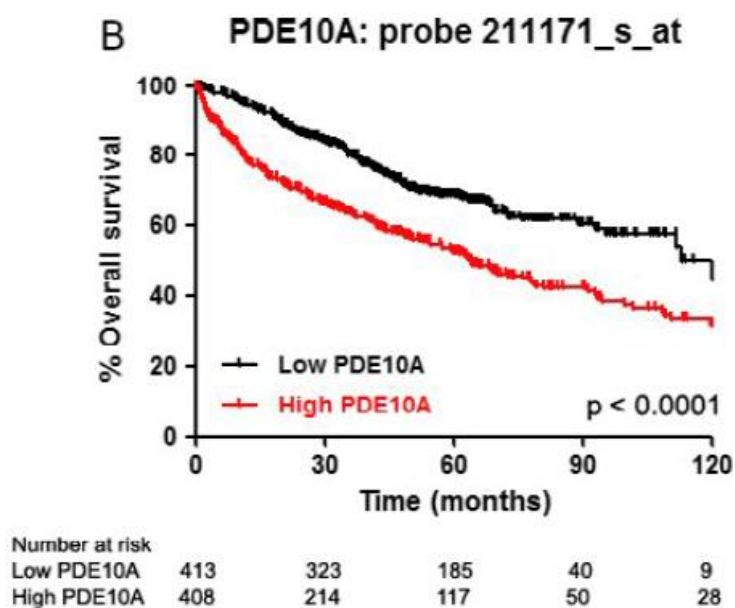
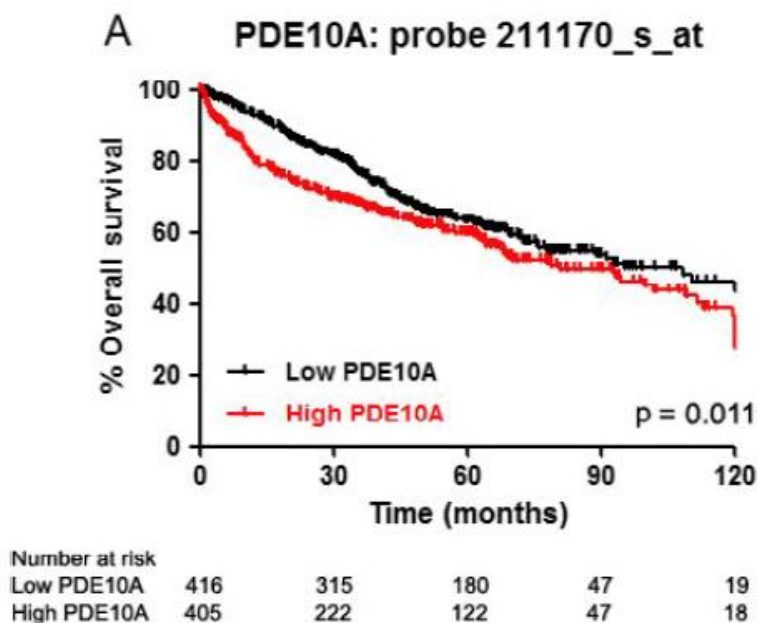
## 6.4 Validación funcional clínica de los SNPs identificados.

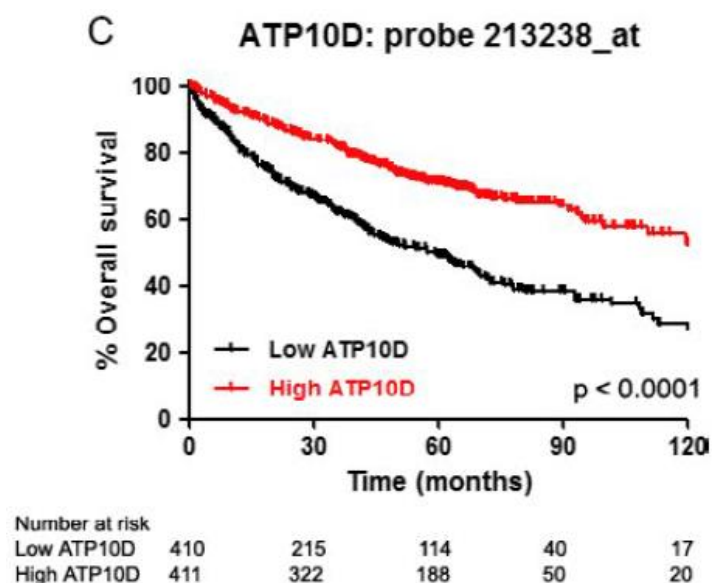
### 6.4.1 Valor pronóstico de la expresión de ARNm de PDE10A y ATP10D en pacientes con CNMP en estadios I y II.

Evaluamos el valor pronóstico de la expresión de ARNm de *PDE10A* y *ATP10D* con la herramienta *in silico* KM-plotter [107] en 821 pacientes que presentan CNMP en estadios I y II. Analizamos la expresión de *PDE10A* utilizando las sondas 211170\_s\_at y 211171\_s\_at y la expresión de *ATP10D* utilizando la sonda 213238\_s\_at. Clasificamos a los pacientes en dos grupos, según el valor medio de la expresión de cada marcador biológico.

Los pacientes con mayor expresión de ARNm de *PDE10A* utilizando ambas sondas, mostraron una disminución de la SG (Figura 3, paneles A y B,  $p = 0,011$  y  $p < 0,0001$ ). La baja expresión de m de *ATP10D* se asoció con menor SG (Figura 3, panel C,  $p < 0,0001$ ).

Figura 3. Correlación de la expresión de ARNm de *PDE10A* y de *ATPD10D* con la supervivencia global (SG) en 821 pacientes con CNMP en estadios I y II, mediante la herramienta *in silico* (Kaplan-Meier-Plotter). Paneles A y B: SG en relación con la expresión de ARNm de *PDE10A* utilizando las sondas 211170\_s\_at (A) y 211171\_s\_at (B). Figura C: SG en relación con la expresión de mRNA de *ATPD10D* utilizando la sonda 213238\_s\_at.





#### 6.4.2 Valor pronóstico de la expresión de la proteína PDE10A en el tumor en pacientes con CNMP en estadios I y II.

Validamos el valor pronóstico de la expresión tumoral de *PDE10A* en un grupo de pacientes diagnosticados de CNMP en estadios I y II tratados con resección quirúrgica en la Clínica Universidad de Navarra, con un seguimiento medio de 46,5 meses (rango intercuartil: 91,6 – 22,0). Las características de los pacientes están representadas en la Tabla 7.

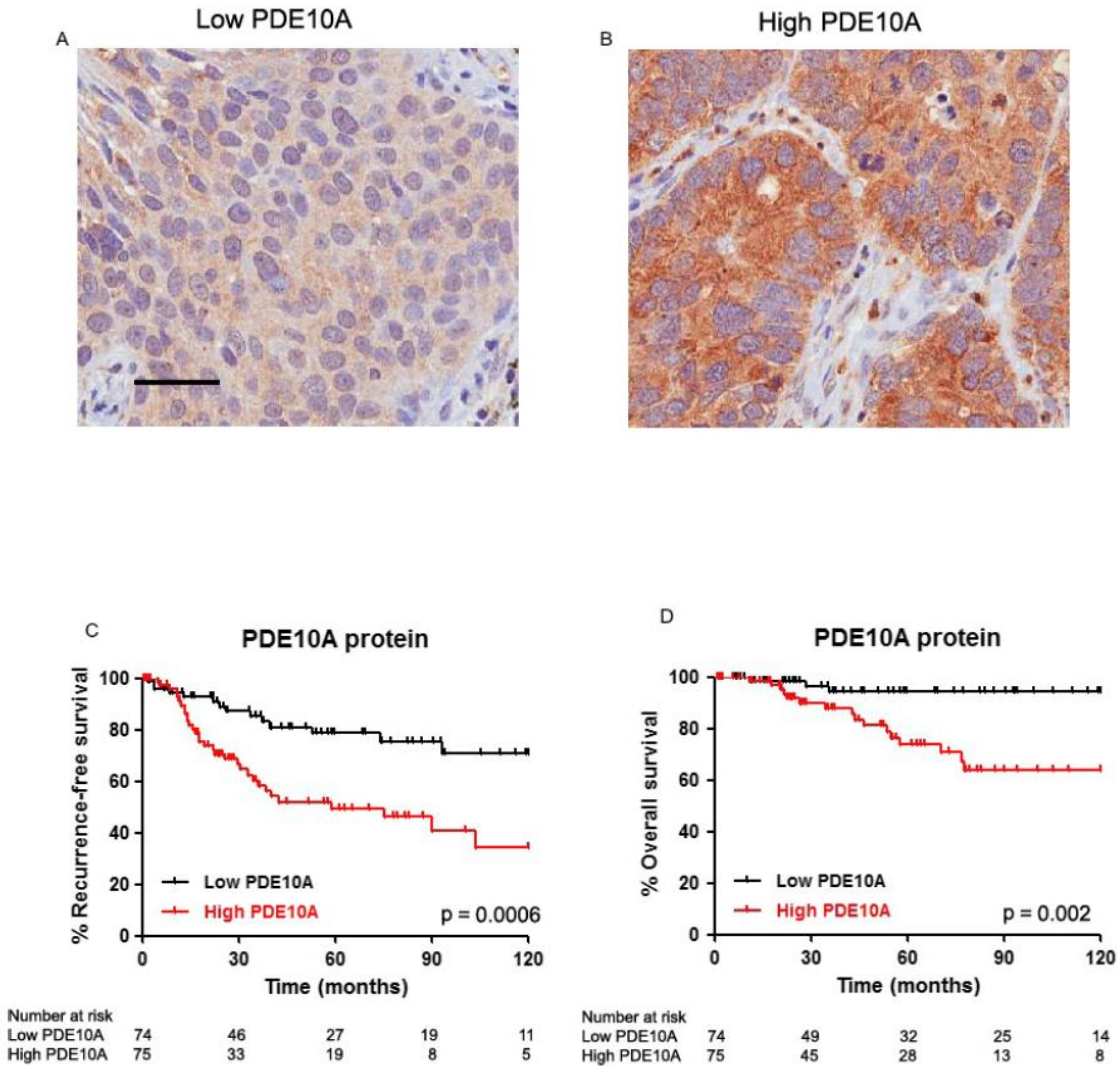
La expresión elevada de la proteína PDE10A en el tumor se asoció con una disminución de SLP ( $p = 0,0006$ ) y SG ( $p = 0,002$ ) (Figura 4). Además, la expresión de PDE10A se correlacionó significativamente con la SLP y la SG en los análisis univariante y multivariante (Tabla 8). En el análisis multivariante, la expresión de PDE10A fue un factor pronóstico independiente, superior al estadio y a la edad. PDE10A fue la única proteína

que mantuvo la significación estadística con respecto a la SLP (OR 2,64; 95% CI 1,38 – 5,04; p = 0,003) y la SG (OR 2,91; 95% CI 1,38 – 6,14; p = 0,005).

**Tabla 7. Características clínico-patológicas de 149 pacientes con CNMP en estadios I y II tratados con cirugía.**

<b>Características de los pacientes (n = 149)</b>	<b>n (%)</b>
<b>Edad</b>	
≤ 65 años	74 (49,7)
> 65 años	75 (50,3)
<b>Género</b>	
Hombre	119 (79,9)
Mujer	30 (20,1)
<b>Histología</b>	
Adenocarcinoma	73 (49,0)
Escamoso	59 (39,6)
Otros	17 (11,4)
<b>Estadio</b>	
I	101 (67,8)
II	48 (32,2)
<b>Historia de tabaquismo</b>	
Nunca	15 (10,1)
Ex-fumador	101 (67,8)
Fumador actual	33 (22,1)
<b>Tratamiento</b>	
Cirugía	36 (24)
Cirugía más quimioterapia adyuvante	113 (76)

**Figura 4. Correlación entre la expresión de la proteína PDE10A en el tumor y la supervivencia libre de progresión (SLP) y la supervivencia global (SG) en 149 pacientes diagnosticados de CNMP en estadios I y II tratados con resección quirúrgica. Paneles A y B: Imágenes histológicas de inmunohistoquímica correspondientes a casos representativos con expresión baja (A) y alta (B) de PDE10A. Escala: 50 micras. Panel C: SLP en relación con la expresión de la proteína PDE10A. Panel D: SG en relación con la expresión de la proteína PDE10A.**



**Tabla 8. Análisis de regresión univariante y multivariante de Cox de la expresión de proteína PDE10A en relación con la supervivencia libre de progresión (SLP) y supervivencia global (SG) en 149 pacientes con CNMP con estadios I y II tratados con resección quirúrgica. \*Ajustado por edad, sexo y estadio. OR: OR: razón de probabilidades CI: intervalo de confianza.**

Factor	n	SLP				SG			
		Univariante OR (95% CI)	p	Multivariante OR (95% CI)*	p	Univariante OR (95% CI)	p	Multivariante OR (95% CI)*	p
<b>PDE10A</b>									
<b>Bajo</b>	74	-	-	-	-	-	-	-	-
<b>Alto</b>	75	2,83 (1,50 - 5,34)	0,001	2,64 (1,38 - 5,04)	0,003	2,86 (1,37 - 5,97)	0,005	2,91 (1,38 - 6,14)	0,005
<b>Estadio</b>									
<b>I</b>	101	-	-	-	-	-	-	-	-
<b>II</b>	48	1,74 (0,96 - 3,15)	0,066	1,425 (0,78 - 2,61)	0,250	1,97 (1,011 - 3,84)	0,046	1,700 (0,86 - 3,31)	0,125
<b>Edad</b>									
<b>&lt; 65</b>	74	-	-	-	-	-	-	-	-
<b>&gt; 65</b>	75	1,36 (0,76 - 2,46)	0,305	-	-	1,82 (0,93 - 3,58)	0,082	2,07 (1,04 - 4,09)	0,038

## 7 DISCUSIÓN.

En este estudio hemos identificado y validado dos variantes nuevas asociadas con individuos que presentan fenotipos extremos de riesgo de desarrollar CNMP inducido por el tabaco. En nuestro conocimiento, este es el primer estudio que utiliza esta estrategia para identificar nuevas variantes genéticas relacionadas con el riesgo de desarrollar CNMP. Esta metodología, que consiste en estudiar grupos reducidos de individuos con fenotipos muy característicos y clínicamente relevantes, lo que contribuye a enriquecer la expresión de biomarcadores en los individuos estudiados, ha permitido la identificación de biomarcadores muy relevantes en oncología, como se ha revisado en otros trabajos [73]

En la fase de identificación, empleamos la plataforma HumanOmni2.5-Quad BeadChip, un chip de alta densidad que permite genotipar 2,5 millones de SNP en todo el genoma, incluyendo SNPs con MAF de hasta el 1%. En nuestro conocimiento, este es el primer estudio que utiliza una plataforma de GWAS tan potente para valorar SNPs relacionados con el riesgo de desarrollar cáncer.

La asociación validada con mayor valor estadístico fue el SNP rs12660420, que está localizado en una región no codificante del gen PDE10A. Este gen codifica una proteína que pertenece a la familia de las fosfodiesterasas de nucleótidos cíclicos. Las fosfodiesterasas representan una gran superfamilia de hidrolasas que controlan los niveles intracelulares de nucleótidos cíclicos, hidrolizando cAMP y cGMP a 5-AMP y 5-



GMP, y son reguladores críticos de multitud de procesos fisiológicos y fisiopatológicos que están bajo el control de nucleótidos cíclicos [113]. PDE4, una diana muy relevante para la terapia de la enfermedad pulmonar obstructiva crónica [114], también se expresa en tumores de pulmón y promueve la progresión del cáncer de pulmón [115]. Algunos datos preclínicos indican que los inhibidor de PDE4 también pueden tener un papel relevante en el tratamiento de pacientes con CNMP que expresan mutaciones de EGFR y que presentan niveles altos de BIM y mTOR antes del tratamiento [116]. Li y cols. demostraron que la expresión de PDE10A, está aumentada tanto a nivel de ARNm como de proteína en células tumorales de colon, en comparación con los colonocitos normales [117]. También demostraron que el bloqueo de PDE10A suprime el crecimiento de células tumorales de colon; y que la sobreexpresión de PDE10A promueve la proliferación tanto de colonocitos como de adenomas [117]. El mismo grupo comunicó recientemente que Pf-2545920, un inhibidor altamente específico de PDE10A, inhibe el crecimiento de células tumorales de colon a concentraciones que aumentan los niveles de cGMP y cAMP y activan PKG y PKA; y que Pf-2545920 reduce la transcripción de survivina mediada por  $\beta$ -catenina, lo que resulta en la activación de caspasas y de la apoptosis [118]. Shen y cols. desarrollaron un modelo predictivo que indica que las mutaciones de PDE10A pueden influir en su regulación alostérica en el adenocarcinoma de pulmón [119]. También comunicaron que los niveles altos de expresión de PDE10A se correlacionan con peor pronóstico el adenocarcinoma de pulmón; y que Pf-2545920 también suprime el crecimiento de líneas celulares de CNMP [119]. Finalmente, se han descrito mutaciones somáticas en PDE10A en hasta el 19% de los cánceres de próstata, cuya presencia se correlaciona con niveles aumentados de la proteína de unión al elemento de respuesta a c-AMP fosforilado (p-CREB) [120].

Todos estos datos sugieren que PDE10A podría ser una diana relevante para el tratamiento y la prevención del cáncer.

La segunda variante validada es el SNP rs6835978, que es una variante intrónica de ATP10D, un gen perteneciente a una subfamilia de ATPasas de tipo P implicadas en la translocación de fosfolípidos a través de las membranas biológicas celulares, desde el exoplasma al citoplasma. Aunque ATP10D no se ha relacionado previamente ni con el desarrollo ni con el pronóstico del CNMP, la variabilidad genética en este gen sí se ha asociado con los niveles de esfingolípidos en plasma [121]. Los esfingolípidos bioactivos, como la esfingosina-1-fosfato (S1P) y las ceramidas, son moléculas de señalización involucradas en la activación de vías directamente relevantes para la carcinogénesis [122]. La ceramida desempeña un papel importante en el desarrollo de enfermedades pulmonares crónicas, y también se ha relacionado con el desarrollo del cáncer de pulmón inducido por el humo del tabaco [123]. Alberg y cols. han comunicado que las concentraciones totales elevadas de S1P y ceramida en plasma se asocian a un mayor riesgo de desarrollar cáncer de pulmón [124].

Además de identificar y validar estos dos SNP en individuos con fenotipos de alto y bajo riesgo de desarrollar CNMP inducido por tabaco, en este trabajo hemos realizado una validación funcional clínica de la relevancia de los genes relacionados con los SNPs validados. Para ello, decidimos explorar la expresión de ambos genes en tumores de pacientes diagnosticados de CNMP en estadios iniciales. Decidimos centrarnos en este grupo de pacientes porque consideramos que era más probable que los genes relacionados con el riesgo de desarrollar CNMP tuvieran relevancia clínica en tumores

en estadios iniciales que en casos avanzados, que presentan mayor complejidad genética.

En primer lugar, estudiamos la expresión del ARNm de ambos genes en los tumores de 821 pacientes con CNMP en estadios I y II, utilizando una herramienta *in silico* [107]. Nuestro análisis confirmó que la expresión de ARNm elevada de *PDE10A* y disminuida de *ATP10D* se correlacionaban significativamente con menor SG, utilizando varias sondas diferentes de expresión de ARNm. Esta es la primera asociación de *ATP10D* con el pronóstico del cualquier tipo de tumor.

En segundo lugar, analizamos los niveles de expresión de la proteína *PDE10A* en 149 pacientes con CNMP en estadios I y II tratados con resección quirúrgica en la Clínica Universidad de Navarra, confirmando que la expresión elevada de *PDE10A* se asociaba de manera muy significativa a menor SG y menor SLP.

Todos los resultados de correlación de la expresión de *PDE10A* en el tumor, tanto a nivel de ARNm como de proteína, son consistentes entre ellos y también con resultados previos comunicados en pacientes con CNMP [119]. Por tanto, *PDE10A* y *ATP10D* pueden ser dianas útiles para determinar el pronóstico del CNMP, así como para desarrollar nuevas estrategias terapéuticas.

El CNMP inducido por el tabaco representa uno de los desafíos más relevantes para la salud pública. Por ello, la identificación de factores genéticos que puedan conferir a los seres humanos tanto un riesgo mayor como, muy especialmente, una protección intrínseca para desarrollar este tipo de tumores podría tener una enorme relevancia clínica y social. En primer lugar, podría mejorar la identificación de poblaciones de alto riesgo, lo que podría aumentar la eficacia, la eficiencia y la efectividad de las estrategias de detección y seguimiento de cáncer con un determinado componente genético hereditario; de los programas de educación y deshabituación tabáquica; y de los programas de detección precoz de tumores en estadios iniciales, cuando el tratamiento tiene altas posibilidades de ser curativo. También podría mejorar el conocimiento de los mecanismos de carcinogénesis y -lo que podría ser más novedoso y relevante- de protección natural contra el desarrollo de cáncer. En este sentido, hay que destacar que, aunque nuestro estudio fue diseñado específicamente para CNMP, los individuos de las cohortes libres de cáncer (controles extremos), tanto de la serie de identificación como la de validación, no desarrollaron ningún otro tumor. Por lo tanto, nuestros resultados también podrían estar relacionados con otras neoplasias, relacionadas o no con el tabaco.

En nuestro conocimiento, este es el primer estudio que define la existencia de fenotipos de menor riesgo de desarrollar CNMP inducido por tabaco y que intenta caracterizar las causas genéticas de dicha protección. El estudio de estos fenotipos competentes respecto al riesgo de cáncer (Proficient cancer-risk phenotypes, PROCARPs) que presentan un riesgo de cáncer reducido, puede permitir identificar los mecanismos de protección subyacentes y puede revelar hallazgos muy relevantes para la prevención

del cáncer. No obstante, la identificación de PROCARPs supone un gran reto metodológico, ya que en caso de existir, probablemente sean fenotipos no aparentes - es decir, no presentan ninguna alteración evidente asociada- y por ello es necesario diferenciarlos de individuos “normales” que simplemente presentan ausencia de la enfermedad [74]. La hipótesis de nuestro trabajo es que la selección de individuos sobreexuestos a factores de riesgo bien reconocidos de cáncer que, sin embargo, no desarrollan la enfermedad, puede permitir la identificación y el estudio de estos PROCARPs [73] [125]. En contraste, los individuos con mayor riesgo de desarrollar tumores (fenotipos deficientes respecto al riesgo de cáncer, DECARPs) presentan fenotipos aparentes, más sencillos de identificar, porque además de desarrollar el tumor, lo hacen de una manera muy característica, por ejemplo a edades tempranas, con agregación familiar, etc., como por ejemplo ocurre en los síndromes familiares de cáncer [74].

La limitación principal de nuestro estudio es que el tamaño de la muestra es limitado, debido a la dificultad para identificar individuos con fenotipos extremos. Esto complica la posibilidad de alcanzar valores de probabilidad en el rango de la significación estadística habitualmente requerida para los estudios que utilizan GWAS. Sin embargo, para superar esta limitación, hemos realizado una validación funcional clínica de la relevancia de la expresión tumoral de ARNm de *PDE10A* y *ATP10D* y de la expresión tumoral de *PDE10A* en dos series independientes de pacientes de CNMP en estadios I y II, y encontramos correlaciones muy significativas con la SG y la SLP. La validación funcional en modelos preclínicos también será necesaria para confirmar la relevancia de nuestros hallazgos. Además, a pesar de nuestros esfuerzos para reclutar individuos

altamente seleccionados, puede persistir heterogeneidad fenotípica con respecto a algunas características demográficas y clínicas. No obstante, esta limitación constituye, de hecho, la mayor oportunidad para mejorar esta estrategia, mediante la aplicación de criterios de selección más estrictos para identificar fenotipos más homogéneos con respecto a la edad, el consumo de tabaco u otras variables relevantes. Nuestra estrategia también puede aplicarse utilizando diferentes técnicas de alto rendimiento, como la secuenciación del exoma o del genoma, la evaluación del transcriptoma, epigenoma, etc. Por último, serán necesarios nuevos estudios para definir de manera más precisa los criterios óptimos de selección de los diferentes fenotipos extremos.

En resumen, hemos identificado y validado dos nuevas variantes genéticas en *PDE10A* y *ATP10D* asociadas con individuos que presentan fenotipos extremos de muy alto y muy bajo riesgo de desarrollar CNMP inducido por tabaco y hemos confirmado la relevancia pronóstica de la expresión en el tumor de las respectivas proteínas asociadas a estas variantes en pacientes con CNMP en estadios I y II. Nuestros hallazgos pueden tener implicaciones relevantes para definir poblaciones de cáncer de alto riesgo y para caracterizar mecanismos moleculares de carcinogénesis y de resistencia para desarrollar cáncer, apoyando así la viabilidad y la utilidad de la metodología de selección de fenotipos extremos en la investigación del cáncer.

## 8 CONCLUSIONES.

1. Hemos definido por primera vez la existencia de individuos con fenotipos extremos de alto y bajo riesgo de desarrollar CNMP asociado al consumo de tabaco y hemos caracterizado su perfil genómico mediante GWAS.
2. Los SNP rs12660420 y rs6835978 localizados en los genes *PDE10A* y *ATP10D* se asocian significativamente a individuos que presentan fenotipos extremos de alto y bajo riesgo de desarrollar CNMP asociado al consumo de tabaco. Estos hallazgos han sido replicados en la serie de validación.
3. La sobreexpresión de ARNm de *PDE10A* y *ATP10D* en el tumor de pacientes con CNMP en estadios I y II, se correlaciona respectivamente con menor y mayor supervivencia global.
4. La sobreexpresión de la proteína PDE10A en el tumor de pacientes con CNMP en estadios I y II, se correlaciona con menor supervivencia global y supervivencia libre de progresión.
5. Esta metodología podría permitir desarrollar estrategias de prevención y diagnóstico precoz e identificar nuevos mecanismos de carcinogénesis y de protección frente al desarrollo de cáncer.

## 9 BIBLIOGRAFÍA.

1. Organization), I.A.o.R.o.C.W.H., *EuropeSource: Globocan2018*. 2018.
2. *NCCN Guidelines Non-Small Cell Lung Cancer*. Version 1.2019.
3. *Statistics Explained* (<http://ec.europa.eu/eurostat/statisticsexplained/>). 2018.
4. DeVita, H., and Rosenberg's, *Cancer*. 2015. **10**: p. 64-66.
5. Warren, G.W. and K.M. Cummings, *Tobacco and lung cancer: risks, trends, and outcomes in patients with cancer*. Am Soc Clin Oncol Educ Book, 2013: p. 359-64.
6. A.Roffo, "*El tabaco como cancerígeno*", . Boletín del Instituto de Medicina Experimental para el estudio y el tratamiento del cáncer,Boletín del Instituto de Medicina Experimen tal para el estudio y el tratamiento del cáncer., 1937: p. pp. 287-336.
7. Roffo., A., *Durch Tabak beim Kaninchen entwickeltes Carcinom*. J Cancer Res Clin Oncol, 1931.
8. Müller, F.H., *Tabakmißbrauch und Lungencarcinom*. 1939. **85**.: p.;49:57.
9. Peto, R., et al., *Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies*. BMJ, 2000. **321**(7257): p. 323-9.
10. Brennan, P., et al., *High cumulative risk of lung cancer death among smokers and nonsmokers in Central and Eastern Europe*. Am J Epidemiol, 2006. **164**(12): p. 1233-41.
11. Talhout, R., et al., *Hazardous compounds in tobacco smoke*. Int J Environ Res Public Health, 2011. **8**(2): p. 613-28.



12. Humans, I.W.G.o.t.E.o.C.R.t., *Some non-heterocyclic polycyclic aromatic hydrocarbons and some related exposures*. IARC Monogr Eval Carcinog Risks Hum, 2010. **92**: p. 1-853.
13. Zaga, V., et al., *Polonium and lung cancer*. J Oncol, 2011. **2011**: p. 860103.
14. Radford, E.P., Jr. and V.R. Hunt, *Polonium-210: A Volatile Radioelement in Cigarettes*. Science, 1964. **143**(3603): p. 247-9.
15. Warren, G.W., et al., *The 2014 Surgeon General's report: "The health consequences of smoking--50 years of progress": a paradigm shift in cancer care*. Cancer, 2014. **120**(13): p. 1914-6.
16. Hang, B., *Formation and repair of tobacco carcinogen-derived bulky DNA adducts*. J Nucleic Acids, 2010. **2010**: p. 709521.
17. Greenman, C., et al., *Patterns of somatic mutation in human cancer genomes*. Nature, 2007. **446**(7132): p. 153-8.
18. Chen, R.J., et al., *Epigenetic effects and molecular mechanisms of tumorigenesis induced by cigarette smoke: an overview*. J Oncol, 2011. **2011**: p. 654931.
19. in *How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General*. 2010: Atlanta (GA).
20. Conklin, B.S., et al., *Nicotine and cotinine up-regulate vascular endothelial growth factor expression in endothelial cells*. Am J Pathol, 2002. **160**(2): p. 413-8.
21. Toll, D.B., et al., *Validation, updating and impact of clinical prediction rules: a review*. J Clin Epidemiol, 2008. **61**(11): p. 1085-94.
22. Gray, E.P., et al., *Risk Prediction Models for Lung Cancer: A Systematic Review*. Clin Lung Cancer, 2016. **17**(2): p. 95-106.

23. Bach, P.B., et al., *Variations in lung cancer risk among smokers*. J Natl Cancer Inst, 2003. **95**(6): p. 470-8.
24. D'Amelio, A.M., Jr., et al., *Comparison of discriminatory power and accuracy of three lung cancer risk models*. Br J Cancer, 2010. **103**(3): p. 423-9.
25. Cassidy, A., et al., *The LLP risk model: an individual risk prediction model for lung cancer*. Br J Cancer, 2008. **98**(2): p. 270-6.
26. Raji, O.Y., et al., *Incorporation of a genetic factor into an epidemiologic model for prediction of individual risk of lung cancer: the Liverpool Lung Project*. Cancer Prev Res (Phila), 2010. **3**(5): p. 664-9.
27. Etzel, C.J., et al., *Development and validation of a lung cancer risk prediction model for African-Americans*. Cancer Prev Res (Phila), 2008. **1**(4): p. 255-65.
28. Tammemagi, C.M., et al., *Lung cancer risk prediction: Prostate, Lung, Colorectal And Ovarian Cancer Screening Trial models and validation*. J Natl Cancer Inst, 2011. **103**(13): p. 1058-68.
29. Tammemagi, M.C., et al., *Selection criteria for lung-cancer screening*. N Engl J Med, 2013. **368**(8): p. 728-36.
30. Hoggart, C., et al., *A risk model for lung cancer incidence*. Cancer Prev Res (Phila), 2012. **5**(6): p. 834-46.
31. Maisonneuve, P., et al., *Lung cancer risk prediction to select smokers for screening CT--a model based on the Italian COSMOS trial*. Cancer Prev Res (Phila), 2011. **4**(11): p. 1778-89.
32. Spitz, M.R., et al., *A risk model for prediction of lung cancer*. J Natl Cancer Inst, 2007. **99**(9): p. 715-26.

33. National Lung Screening Trial Research, T., et al., *Reduced lung-cancer mortality with low-dose computed tomographic screening*. N Engl J Med, 2011. **365**(5): p. 395-409.
34. Amos, C.I., et al., *Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1*. Nat Genet, 2008. **40**(5): p. 616-22.
35. McKay, J.D., et al., *Lung cancer susceptibility locus at 5p15.33*. Nat Genet, 2008. **40**(12): p. 1404-6.
36. Gorlov, I.P., et al., *Seizure 6-like (SEZ6L) gene and risk for lung cancer*. Cancer Res, 2007. **67**(17): p. 8406-11.
37. Deng, N., et al., *Single nucleotide polymorphisms and cancer susceptibility*. Oncotarget, 2017. **8**(66): p. 110635-110649.
38. Wray, N.R., et al., *Pitfalls of predicting complex traits from SNPs*. Nat Rev Genet, 2013. **14**(7): p. 507-15.
39. Marshall, A.L. and D.C. Christiani, *Genetic susceptibility to lung cancer--light at the end of the tunnel?* Carcinogenesis, 2013. **34**(3): p. 487-502.
40. Wang, J., et al., *Genetic predisposition to lung cancer: comprehensive literature integration, meta-analysis, and multiple evidence assessment of candidate-gene association studies*. Sci Rep, 2017. **7**(1): p. 8371.
41. Lo, Y.L., et al., *A polymorphism in the APE1 gene promoter is associated with lung cancer risk*. Cancer Epidemiol Biomarkers Prev, 2009. **18**(1): p. 223-9.
42. Gong, J., et al., *Quantitative assessment of the association between AXIN2 rs2240308 polymorphism and cancer risk*. Sci Rep, 2015. **5**: p. 10111.
43. Ryan, B.M., et al., *Identification of a functional SNP in the 3'UTR of CXCR2 that is associated with reduced risk of lung cancer*. Cancer Res, 2015. **75**(3): p. 566-75.

44. Ye, X.H., et al., *Association between the CYP2E1 polymorphisms and lung cancer risk: a meta-analysis*. Mol Genet Genomics, 2015. **290**(2): p. 545-58.
45. Eom, S.Y., et al., *Interactions between paraoxonase 1 genetic polymorphisms and smoking and their effects on oxidative stress and lung cancer risk in a Korean population*. PLoS One, 2015. **10**(3): p. e0119100.
46. Yin, J., et al., *Fine-mapping markers of lung cancer susceptibility in a sub-region of chromosome 19q13.3 among Chinese*. Oncotarget, 2016. **7**(38): p. 60929-60939.
47. Tanner, J.A. and R.F. Tyndale, *Variation in CYP2A6 Activity and Personalized Medicine*. J Pers Med, 2017. **7**(4).
48. Gladek, I., et al., *HIF1A gene polymorphisms and human diseases: Graphical review of 97 association studies*. Genes Chromosomes Cancer, 2017. **56**(6): p. 439-452.
49. Spinola, M., et al., *Association of the PDCD5 locus with lung cancer risk and prognosis in smokers*. J Clin Oncol, 2006. **24**(11): p. 1672-8.
50. Mechanic, L.E., et al., *Common genetic variation in TP53 is associated with lung cancer risk and prognosis in African Americans and somatic mutations in lung tumors*. Cancer Epidemiol Biomarkers Prev, 2007. **16**(2): p. 214-22.
51. Rasool Bhat, G.H., et al., *Genetic variant rs10937405 of TP63 and susceptibility to lung cancer risk in north Indian population*. J Genet, 2019. **98**(2).
52. David-Beabes, G.L. and S.J. London, *Genetic polymorphism of XRCC1 and lung cancer risk among African-Americans and Caucasians*. Lung Cancer, 2001. **34**(3): p. 333-9.
53. Kiyohara, C., et al., *EPHX1 polymorphisms and the risk of lung cancer: a HuGE review*. Epidemiology, 2006. **17**(1): p. 89-99.

54. Zhan, P., et al., *ERCC2/XPD Lys751Gln and Asp312Asn gene polymorphism and lung cancer risk: a meta-analysis involving 22 case-control studies*. J Thorac Oncol, 2010. **5**(9): p. 1337-45.
55. Peng, B., et al., *Meta-analysis of association between matrix metalloproteinases 2, 7 and 9 promoter polymorphisms and cancer risk*. Mutagenesis, 2010. **25**(4): p. 371-9.
56. Akiyama, T., *Wnt/beta-catenin signaling*. Cytokine Growth Factor Rev, 2000. **11**(4): p. 273-82.
57. Wu, C., et al., *Genetic variants on chromosome 15q25 associated with lung cancer risk in Chinese populations*. Cancer Res, 2009. **69**(12): p. 5065-72.
58. Hecht, S.S., *Progress and challenges in selected areas of tobacco carcinogenesis*. Chem Res Toxicol, 2008. **21**(1): p. 160-71.
59. Messina, E.S., R.F. Tyndale, and E.M. Sellers, *A major role for CYP2A6 in nicotine C-oxidation by human liver microsomes*. J Pharmacol Exp Ther, 1997. **282**(3): p. 1608-14.
60. Wassenaar, C.A., et al., *Relationship between CYP2A6 and CHRNA5-CHRNA3-CHRNA4 variation and smoking behaviors and lung cancer risk*. J Natl Cancer Inst, 2011. **103**(17): p. 1342-6.
61. Chenoweth, M.J., et al., *CYP2A6 slow nicotine metabolism is associated with increased quitting by adolescent smokers*. Pharmacogenet Genomics, 2013. **23**(4): p. 232-5.
62. Zhu, A.Z., et al., *Alaska Native smokers and smokeless tobacco users with slower CYP2A6 activity have lower tobacco consumption, lower tobacco-specific nitrosamine exposure and lower tobacco-specific nitrosamine bioactivation*. Carcinogenesis, 2013. **34**(1): p. 93-101.

63. Lieber, C.S., *Cytochrome P-450E1: its physiological and pathological role*. *Physiol Rev*, 1997. **77**(2): p. 517-44.
64. Liu, S., et al., *Elucidation of CYP2E1 5' regulatory RsaI/PstI allelic variants and their role in risk for oral cancer*. *Oral Oncol*, 2001. **37**(5): p. 437-45.
65. Alfaro, C., et al., *Tumor-Produced Interleukin-8 Attracts Human Myeloid-Derived Suppressor Cells and Elicits Extrusion of Neutrophil Extracellular Traps (NETs)*. *Clin Cancer Res*, 2016. **22**(15): p. 3924-36.
66. Valko, M., et al., *Free radicals, metals and antioxidants in oxidative stress-induced cancer*. *Chem Biol Interact*, 2006. **160**(1): p. 1-40.
67. Eom, S.Y., et al., *Effects of intronic and exonic polymorphisms of paraoxonase 1 (PON1) gene on serum PON1 activity in a Korean population*. *J Korean Med Sci*, 2011. **26**(6): p. 720-5.
68. Fang, D.H., et al., *Differential effects of paraoxonase 1 (PON1) polymorphisms on cancer risk: evidence from 25 published studies*. *Mol Biol Rep*, 2012. **39**(6): p. 6801-9.
69. Zhou, G., et al., *Functional polymorphisms and haplotypes in the promoter of the MMP2 gene are associated with risk of nasopharyngeal carcinoma*. *Hum Mutat*, 2007. **28**(11): p. 1091-7.
70. International HapMap, C., *A haplotype map of the human genome*. *Nature*, 2005. **437**(7063): p. 1299-320.
71. Lynch, T.J., et al., *Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib*. *N Engl J Med*, 2004. **350**(21): p. 2129-39.
72. Paez, J.G., et al., *EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy*. *Science*, 2004. **304**(5676): p. 1497-500.

73. Perez-Gracia, J.L., et al., *The role of extreme phenotype selection studies in the identification of clinically relevant genotypes in cancer research*. *Cancer*, 2002. **95**(7): p. 1605-10.
74. Perez-Gracia, J.L., et al., *Selection of extreme phenotypes: the role of clinical observation in translational research*. *Clin Transl Oncol*, 2010. **12**(3): p. 174-80.
75. Malkin, D., et al., *Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms*. *Science*, 1990. **250**(4985): p. 1233-8.
76. Miller, R.W., *Deaths from childhood cancer in sibs*. *N Engl J Med*, 1968. **279**(3): p. 122-6.
77. Li, F.P. and J.F. Fraumeni, Jr., *Rhabdomyosarcoma in children: epidemiologic study and identification of a familial cancer syndrome*. *J Natl Cancer Inst*, 1969. **43**(6): p. 1365-73.
78. Benedict, W.F., et al., *Patient with 13 chromosome deletion: evidence that the retinoblastoma gene is a recessive cancer gene*. *Science*, 1983. **219**(4587): p. 973-5.
79. Cavenee, W.K., et al., *Expression of recessive alleles by chromosomal mechanisms in retinoblastoma*. *Nature*, 1983. **305**(5937): p. 779-84.
80. Knudson, A.G., Jr., *Mutation and cancer: statistical study of retinoblastoma*. *Proc Natl Acad Sci U S A*, 1971. **68**(4): p. 820-3.
81. Hall, J.M., et al., *Linkage of early-onset familial breast cancer to chromosome 17q21*. *Science*, 1990. **250**(4988): p. 1684-9.
82. Moore, K., et al., *Maintenance Olaparib in Patients with Newly Diagnosed Advanced Ovarian Cancer*. *N Engl J Med*, 2018. **379**(26): p. 2495-2505.
83. Rowland-Jones, S., et al., *HIV-specific cytotoxic T-cells in HIV-exposed but uninfected Gambian women*. *Nat Med*, 1995. **1**(1): p. 59-64.

84. Liu, R., et al., *Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection*. Cell, 1996. **86**(3): p. 367-77.
85. Samson, M., et al., *Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene*. Nature, 1996. **382**(6593): p. 722-5.
86. Quillent, C., et al., *HIV-1-resistance phenotype conferred by combination of two separate inherited mutations of CCR5 gene*. Lancet, 1998. **351**(9095): p. 14-8.
87. Cohen, J., et al., *Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9*. Nat Genet, 2005. **37**(2): p. 161-5.
88. Sabatine, M.S., et al., *Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease*. N Engl J Med, 2017. **376**(18): p. 1713-1722.
89. Garcia-Cao, I., et al., *"Super p53" mice exhibit enhanced DNA damage response, are tumor resistant and age normally*. EMBO J, 2002. **21**(22): p. 6225-35.
90. Yu, Q., Y. Geng, and P. Sicinski, *Specific protection against breast cancers by cyclin D1 ablation*. Nature, 2001. **411**(6841): p. 1017-21.
91. Katzourakis, A., et al., *Larger mammalian body size leads to lower retroviral activity*. PLoS Pathog, 2014. **10**(7): p. e1004214.
92. Vazquez, J.M., et al., *A Zombie LIF Gene in Elephants Is Upregulated by TP53 to Induce Apoptosis in Response to DNA Damage*. Cell Rep, 2018. **24**(7): p. 1765-1776.
93. Buffenstein, R., *Negligible senescence in the longest living rodent, the naked mole-rat: insights from a successfully aging species*. J Comp Physiol B, 2008. **178**(4): p. 439-45.



94. Liang, S., et al., *Resistance to experimental tumorigenesis in cells of a long-lived mammal, the naked mole-rat (Heterocephalus glaber)*. Aging Cell, 2010. **9**(4): p. 626-35.
95. Seluanov, A., et al., *Hypersensitivity to contact inhibition provides a clue to cancer resistance of naked mole-rat*. Proc Natl Acad Sci U S A, 2009. **106**(46): p. 19352-7.
96. Tian, X., et al., *High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat*. Nature, 2013. **499**(7458): p. 346-9.
97. Kim, E.B., et al., *Genome sequencing reveals insights into physiology and longevity of the naked mole rat*. Nature, 2011. **479**(7372): p. 223-7.
98. Emond, M.J., et al., *Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis*. Nat Genet, 2012. **44**(8): p. 886-9.
99. Iyer, G., et al., *Genome sequencing identifies a basis for everolimus sensitivity*. Science, 2012. **338**(6104): p. 221.
100. Hardy, J. and A. Singleton, *Genomewide association studies and human disease*. N Engl J Med, 2009. **360**(17): p. 1759-68.
101. Hosking, F.J., S.E. Dobbins, and R.S. Houlston, *Genome-wide association studies for detecting cancer susceptibility*. Br Med Bull, 2011. **97**: p. 27-46.
102. Van Kuilenburg, A.B., et al., *Heterozygosity for a point mutation in an invariant splice donor site of dihydropyrimidine dehydrogenase and severe 5-fluorouracil related toxicity*. Eur J Cancer, 1997. **33**(13): p. 2258-64.
103. Perez-Gracia, J.L., et al., *Strategies to design clinical studies to identify predictive biomarkers in cancer research*. Cancer Treat Rev, 2017. **53**: p. 79-97.

104. Li, Y., et al., *Genotype imputation*. *Annu Rev Genomics Hum Genet*, 2009. **10**: p. 387-406.
105. Howie, B.N., P. Donnelly, and J. Marchini, *A flexible and accurate genotype imputation method for the next generation of genome-wide association studies*. *PLoS Genet*, 2009. **5**(6): p. e1000529.
106. Genomes Project, C., et al., *A global reference for human genetic variation*. *Nature*, 2015. **526**(7571): p. 68-74.
107. Gyorfy, B., et al., *Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer*. *PLoS One*, 2013. **8**(12): p. e82241.
108. Pajares, M.J., et al., *Expression of tumor-derived vascular endothelial growth factor and its receptors is associated with outcome in early squamous cell carcinoma of the lung*. *J Clin Oncol*, 2012. **30**(10): p. 1129-36.
109. Grbesa, I., et al., *Expression of sirtuin 1 and 2 is associated with poor prognosis in non-small cell lung cancer patients*. *PLoS One*, 2015. **10**(4): p. e0124670.
110. *Comité Etico de investigación clínica (CEIC)*.
111. Consortium, G.T., *The Genotype-Tissue Expression (GTEx) project*. *Nat Genet*, 2013. **45**(6): p. 580-5.
112. Pruim, R.J., et al., *LocusZoom: regional visualization of genome-wide association scan results*. *Bioinformatics*, 2010. **26**(18): p. 2336-7.
113. Azevedo, M.F., et al., *Clinical and molecular genetics of the phosphodiesterases (PDEs)*. *Endocr Rev*, 2014. **35**(2): p. 195-233.
114. Martinez, F.J., et al., *Effect of roflumilast on exacerbations in patients with severe chronic obstructive pulmonary disease uncontrolled by combination therapy*

- (REACT): a multicentre randomised controlled trial. *Lancet*, 2015. **385**(9971): p. 857-66.
115. Pullamsetti, S.S., et al., *Phosphodiesterase-4 promotes proliferation and angiogenesis of lung cancer by crosstalk with HIF*. *Oncogene*, 2013. **32**(9): p. 1121-34.
  116. Karachaliou, N., et al., *BIM and mTOR expression levels predict outcome to erlotinib in EGFR-mutant non-small-cell lung cancer*. *Sci Rep*, 2015. **5**: p. 17499.
  117. Li, N., et al., *Phosphodiesterase 10A: a novel target for selective inhibition of colon tumor cell growth and beta-catenin-dependent TCF transcriptional activity*. *Oncogene*, 2015. **34**(12): p. 1499-509.
  118. Lee, K., et al., *beta-catenin nuclear translocation in colorectal cancer cells is suppressed by PDE10A inhibition, cGMP elevation, and activation of PKG*. *Oncotarget*, 2016. **7**(5): p. 5353-65.
  119. Shen, Q., et al., *Proteome-Scale Investigation of Protein Allosteric Regulation Perturbed by Somatic Mutations in 7,000 Cancer Genomes*. *Am J Hum Genet*, 2017. **100**(1): p. 5-20.
  120. de Alexandre, R.B., et al., *Phosphodiesterase sequence variants may predispose to prostate cancer*. *Endocr Relat Cancer*, 2015. **22**(4): p. 519-30.
  121. Hicks, A.A., et al., *Genetic determinants of circulating sphingolipid concentrations in European populations*. *PLoS Genet*, 2009. **5**(10): p. e1000672.
  122. Furuya, H., Y. Shimizu, and T. Kawamori, *Sphingolipids in cancer*. *Cancer Metastasis Rev*, 2011. **30**(3-4): p. 567-76.
  123. Goldkorn, T., S. Filosto, and S. Chung, *Lung injury and lung cancer caused by cigarette smoke-induced oxidative stress: Molecular mechanisms and therapeutic*

- opportunities involving the ceramide-generating machinery and epidermal growth factor receptor. Antioxid Redox Signal, 2014. 21(15): p. 2149-74.*
124. Alberg, A.J., et al., *Plasma sphingolipids and lung cancer: a population-based, nested case-control study. Cancer Epidemiol Biomarkers Prev, 2013. 22(8): p. 1374-82.*
125. Perez-Gracia, J.L., and M. G. Ruiz-Ilundain., *Cancer protective mutations: looking for the needle in the haystack. Clin. Transl. Oncol. 3:169–171., 2001.*

**ANEXO I: HOJA DE INFORMACIÓN Y DOCUMENTO  
DE CONSENTIMIENTO INFORMADO**

## HOJA DE INFORMACIÓN DEL PACIENTE

Estimado señor/ a:

Se le ha invitado a participar en un estudio de investigación sobre el cáncer titulado *Determinación de polimorfismos asociados al cáncer de pulmón no microcítico*. Su participación en el estudio es voluntaria. Si no desea participar, esto no tendrá ningún impacto sobre los cuidados médicos que usted reciba.

En cumplimiento de la legislación vigente sobre Investigación Biomédica (Ley 14/2007, del 3 de julio), le informamos de los siguientes puntos:

- a) Finalidad de la investigación: en este estudio participan pacientes con cáncer y también individuos sanos que no tienen cáncer. Nuestro objetivo es analizar que factores pueden ser importantes en la aparición del cáncer y en su evolución una vez que ya ha aparecido. Aunque la estructura general del código genético es muy similar en todos los seres humanos, en realidad cada persona presenta varios millones de pequeñas diferencias con las demás, llamadas polimorfismos. La función de la mayoría de estos polimorfismos es aún desconocida, pero sabemos que algunos se asocian con distintas enfermedades, características de cada individuo, etc. En este estudio recogeremos muestras de sangre para obtener ADN (ácido desoxirribonucleico, el material del que están hechos los genes) y analizaremos los polimorfismos de los pacientes e individuos sanos participantes para intentar establecer asociaciones entre su presencia o ausencia y las características de las personas estudiadas. También recogeremos información sobre su historial médico y sobre sus hábitos (consumo de tabaco, etc.) para obtener información que pudiera tener relevancia sobre nuestro estudio.

- b) Beneficios esperados: Su colaboración puede ayudar a avanzar en la investigación del cáncer, aunque también es posible que no se obtenga ningún resultado concluyente.
- c) Posibles inconvenientes vinculados con la donación y obtención de la muestra: se obtendrán 20 ml de sangre. Las posibles complicaciones asociadas son la posibilidad de aparición de dolor, hematoma y más raramente infección en el punto de extracción. Se intentará obtener la muestra aprovechando una extracción de sangre que se le tenga que hacer por otro motivo.
- d) El responsable de la investigación en su centro es el Dr. / Dra. \_\_\_\_\_.
- e) Tiene usted derecho a revocar el consentimiento y sus efectos, incluida la posibilidad de solicitar la destrucción o la anonimización de la muestra, aunque dichos efectos no se extenderán a los datos resultantes de las investigaciones que ya se hayan llevado a cabo. En ese caso, esto no repercutirá sobre la asistencia que usted reciba.
- f) Los estudios se realizarán en el Centro de Investigación Médica Aplicada de la Universidad de Navarra y en el Centro Nacional de Investigaciones Oncológicas. Al terminar la investigación, las muestras se conservarán para poder realizar estudios dentro de la misma línea de investigación, bajo la responsabilidad de la Clínica Universidad de Navarra. Estas investigaciones se realizarán cumpliendo los requerimientos previstos en la legislación vigente.
- g) Tiene usted derecho a conocer los datos genéticos que se obtengan a partir del análisis de las muestras donadas.
- h) Sus datos serán tratados con absoluta confidencialidad de acuerdo a la legislación vigente, de manera que será imposible asociarle a usted con los resultados del estudio. Solamente los médicos que le atienden tendrán acceso a sus datos personales. Su nombre e iniciales no aparecerán en ningún documento del estudio. Su nombre será sustituido por un código. El código que figure en el documento será guardado por su médico, para evitar que terceras personas tengan acceso a los datos.
- i) Es poco probable que se obtengan datos relevantes para su salud como consecuencia de los estudios que se realizarán sobre sus muestras.
- j) No obstante, en el caso de que sí se obtengan datos relevantes para su salud, puede indicarnos si desea ser contactado para que se le comuniquen dichos

resultados. Para ello debe indicar su decisión en la hoja que se le presentará más adelante para firmar.

- k) En caso de que se obtenga información relevante para su salud y que manifieste su interés en que dicha información le sea comunicada, usted deberá valorar la conveniencia de transmitir dicha información a sus familiares.

En el caso de que usted haya sido diagnosticado de un tumor, nos gustaría pedir su autorización para solicitar y estudiar las biopsias/muestras que se le han realizado previamente, siempre que esto no interfiera con los procesos diagnósticos que pudieran ser precisos para el tratamiento de su enfermedad. Esto puede ser muy importante para confirmar los posibles hallazgos del estudio. Para ello debe indicar su decisión en la hoja que se le presentará más adelante para firmar

Si tiene alguna duda o pregunta sobre su participación en este estudio y sobre sus derechos, debe contactar con:

Dr.: ..... Teléfono: .....



## CONSENTIMIENTO DEL PACIENTE POR ESCRITO

### *DETERMINACIÓN DE POLIMORFISMOS ASOCIADOS AL CÁNCER DE PULMÓN NO MICROCÍTICO*

Por favor, si está de acuerdo en participar en el estudio cumplimente el siguiente consentimiento

Yo, D/Dña.....

He recibido información clara, de forma oral y por escrito del Dr. /Dra. .... sobre los métodos, riesgos, complicaciones y fines del estudio. He tenido la ocasión de preguntar libremente sobre todas las cuestiones relacionadas con el proyecto. He entendido perfectamente el propósito del estudio, así como los procedimientos que incluye y sus posibles incomodidades, riesgos directos, remotos o complicaciones. El / la médico que me atiende me ha explicado de forma satisfactoria la utilización que se dará a los datos de mi historia médica y a las muestras obtenidas, y que dicha información será siempre mantenida de forma codificada. Sé que el estudio se hace con fines de investigación y que puedo abandonarlo en cualquier momento del mismo sin repercusión alguna sobre mi tratamiento

Por tanto, voluntariamente, quiero participar y DOY MI CONSENTIMIENTO para ser incluido en el estudio autorizando al equipo investigador para que obtenga la sangre y realice las determinaciones analíticas y procedimientos de los que consta el estudio. En caso de que existiera material sobrante de dicha muestra tras la realización de este estudio, DOY MI CONSENTIMIENTO para que dicho material pueda ser empleado en otros estudios del mismo equipo investigador y dentro de la misma línea de investigación, en las condiciones señaladas. También autorizo a realizar estudios sobre mis muestras de biopsia, siempre que dichos estudios no interfieran sobre procesos diagnósticos que pudieran ser necesarios en el futuro.

En el caso de que en este estudio se obtengan datos relevantes para mi salud:

- Deseo ser contactado e informado.
  
- No deseo ser informado.

Respecto a la utilización de las biopsias previas, siempre que esto no interfiera con los procesos diagnósticos que pudieran ser precisos para el tratamiento de mi enfermedad.

-Doy mi conformidad para que se estudien mis muestras.

-No doy mi conformidad para que se estudien mis muestras.

Comprendo que mi participación es voluntaria y que puedo retirarme del estudio:

- Cuando quiera
- Sin tener que dar explicaciones
- Sin que esto repercuta en mis cuidados médicos

El médico


Firma y Fecha:

El paciente

Firma y Fecha:

## **ANEXO 2: PUBLICACIÓN DEL ESTUDIO**

# Genomic characterization of individuals presenting extreme phenotypes of high and low risk to develop tobacco-induced lung cancer

Juan Pablo Fusco<sup>1,2,\*</sup>, Guillermo Pita<sup>3,\*</sup>, María José Pajares<sup>2,4,5,\*</sup>, Maria Pilar Andueza<sup>1,2</sup>, Ana Patiño-García<sup>2,6</sup>, Juan P. de-Torres<sup>2,7</sup>, Alfonso Gurrpide<sup>1,2</sup>, Javier Zulueta<sup>2,5,7</sup>, Rosario Alonso<sup>3</sup>, Nuria Alvarez<sup>3</sup>, Ruben Pio<sup>2,4,5</sup>, Ignacio Melero<sup>2,5,8</sup>, Miguel F. Sanmamed<sup>1,2</sup>, Maria Rodriguez Ruiz<sup>1,2</sup>, Ignacio Gil-Bazo<sup>1,2,5</sup>, Jose María Lopez-Picazo<sup>1,2</sup>, Ciro Casanova<sup>9</sup>, Rebeca Baz Davila<sup>10</sup>, Antonio Agudo<sup>11,12</sup>, Maria Dolores Lozano<sup>2,5,13</sup>, Alvaro Gonzalez<sup>2,14</sup>, Nuria Sala<sup>11,12</sup>, Eva Ardanaz<sup>2,15</sup>, Javier Benitez<sup>3</sup>, Luis Montuenga<sup>2,4,5,\*</sup>, Anna Gonzalez-Neira<sup>3,\*</sup> & Jose Luis Perez-Gracia<sup>1,2,5,\*</sup> 

<sup>1</sup>Department of Oncology, Clinica Universidad de Navarra, Pamplona, Spain

<sup>2</sup>Health Research Institute of Navarra (IDISNA), Pamplona, Spain

<sup>3</sup>Human Genetics Group, Spanish National Cancer Centre (CNIO), Madrid, Spain

<sup>4</sup>Program in Solid Tumors and Biomarkers, Center for Applied Medical Research (CIMA), Pamplona, Spain

<sup>5</sup>Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Spain

<sup>6</sup>Department of Pediatrics and Clinical Genetics, Clinica Universidad de Navarra, Pamplona, Spain

<sup>7</sup>Pulmonary Department, Clinica Universidad de Navarra, Pamplona, Spain

<sup>8</sup>Departments of Immunology and Oncology, Clinica Universidad de Navarra and Center for Applied Medical Research (CIMA), Pamplona, Spain

<sup>9</sup>Pulmonary Department and Research Department, Hospital Universitario La Candelaria, Santa Cruz de Tenerife, Spain

<sup>10</sup>Research Unit, Hospital Universitario La Candelaria, Santa Cruz de Tenerife, Spain

<sup>11</sup>Unit of Nutrition and Cancer, Cancer Epidemiology Research Program, Catalan Institute of Oncology-ICO, IDIBELL, Barcelona, Spain

<sup>12</sup>Translational Research Laboratory, Catalan Institute of Oncology-ICO, IDIBELL, Barcelona, Spain

<sup>13</sup>Pathology Department, Clinica Universidad de Navarra, Pamplona, Spain

<sup>14</sup>Department of Biochemistry, Clinica Universidad de Navarra, Pamplona, Spain

<sup>15</sup>Navarra Public Health Institute, CIBER Epidemiology and Public Health (CIBERESP), Pamplona, Spain

## Keywords

ATP10D, cancer risk, extreme phenotypes, genome-wide association study, non-small cell lung cancer, PDE10A, single nucleotide polymorphism, tobacco

## Correspondence

Jose Luis Perez-Gracia, Department of Oncology, University Clinic of Navarra, Avda. Pio XII, 36, 31008 Pamplona, Spain.  
Tel: +34 948 255 400; Fax: +34 948 255 500;  
E-mail: jlgracia@unav.es

## Funding Information

Spanish Society of Medical Oncology, Fundación SEOM and Fundación Salud 2000, Government of Navarra.

Received: 4 January 2018; Revised: 3 March 2018; Accepted: 21 March 2018

doi: 10.1002/cam4.1500

\*These authors contributed equally to this work.

## Abstract

Single nucleotide polymorphisms (SNPs) may modulate individual susceptibility to carcinogens. We designed a genome-wide association study to characterize individuals presenting extreme phenotypes of high and low risk to develop tobacco-induced non-small cell lung cancer (NSCLC), and we validated our results. We hypothesized that this strategy would enrich the frequencies of the alleles that contribute to the observed traits. We genotyped 2.37 million SNPs in 95 extreme phenotype individuals, that is: heavy smokers that either developed NSCLC at an early age (extreme cases); or did not present NSCLC at an advanced age (extreme controls), selected from a discovery set ( $n = 3631$ ). We validated significant SNPs in 133 additional subjects with extreme phenotypes selected from databases including >39,000 individuals. Two SNPs were validated: rs12660420 ( $p_{\text{combined}} = 5.66 \times 10^{-5}$ ;  $OR_{\text{combined}} = 2.80$ ), mapping to a noncoding transcript exon of *PDE10A*; and rs6835978 ( $p_{\text{combined}} = 1.02 \times 10^{-4}$ ;  $OR_{\text{combined}} = 2.57$ ), an intronic variant in *ATP10D*. We assessed the relevance of both proteins in early-stage NSCLC. *PDE10A* and *ATP10D* mRNA expressions correlated with survival in 821 stage I–II NSCLC patients ( $p = 0.01$  and  $p < 0.0001$ ). *PDE10A* protein expression correlated with survival in 149 patients with stage I–II NSCLC ( $p = 0.002$ ). In conclusion, we validated two variants associated with extreme phenotypes of high and low risk of developing tobacco-induced NSCLC. Our findings may allow to identify individuals presenting high and low risk to develop tobacco-induced NSCLC and to characterize molecular mechanisms of carcinogenesis and resistance to develop NSCLC.

## Introduction

In the recent years, genetic susceptibility to lung cancer has been explored by many genome-wide association studies (GWAS), which have reported several genomic loci and candidate genes that exert moderate effects on lung cancer risk. The most well-known loci and genes identified are the *CHRNA3* and *CHRNA5* subunits of the  $\alpha 7$  nicotinic acetylcholine receptor (nAChR), located on the 15q25 region [1–3]; *TERT* and *CLPTM1L* [4–7], which are closely located at 5p15.33; *BAT3*, mapped to 6p21.23 [8]; and *GPC5* at 13q31.3 [9]. Most of these studies follow a random sampling design, including smoking habits as a covariate. Several noteworthy associations are found exclusively in subgroups of subjects defined upon their smoking habits. For example, the 15q25 locus is associated with lung cancer in former and current smokers, while the 13q31 locus is associated with susceptibility to lung cancer in never smokers [7, 9].

However, despite tobacco smoke being the most relevant known risk factor for lung cancer, no association studies have analyzed the individual risk that it confers through the study of subjects presenting extreme phenotypes of high

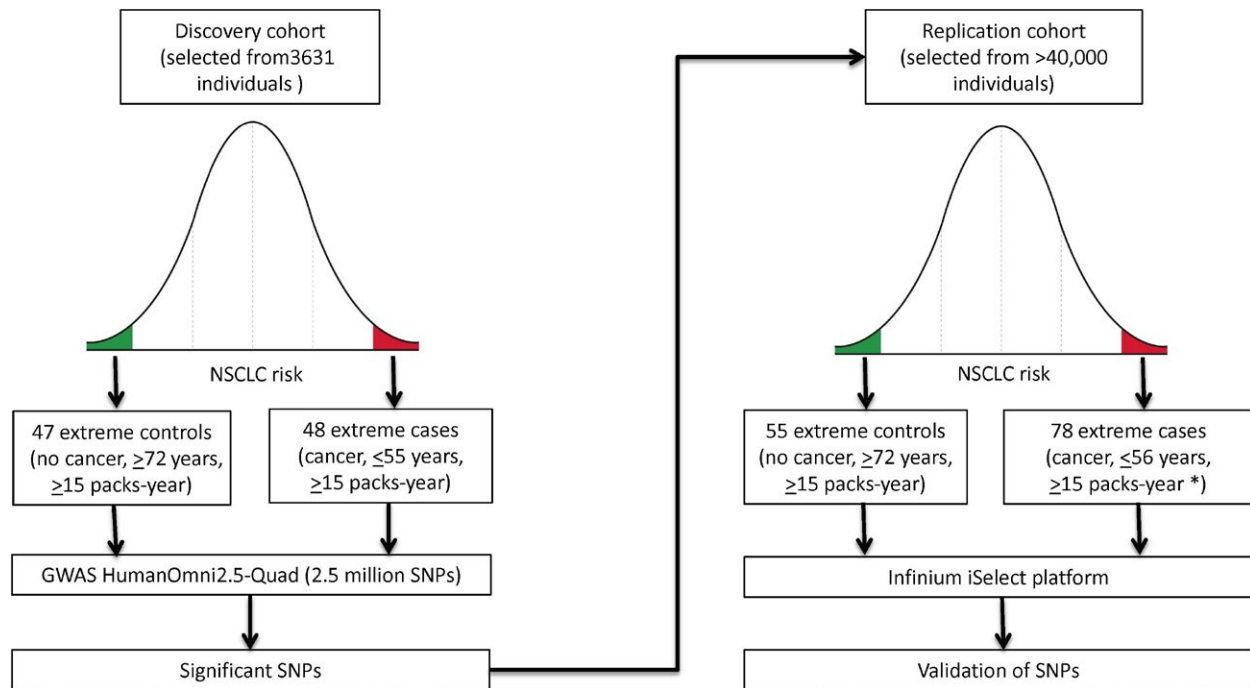
and low risk to develop non-small-cell lung cancer (NSCLC) induced by tobacco. Such designs may help to identify high-risk populations that could benefit from prevention and screening programs and to identify the molecular mechanisms that underlie such clinically relevant phenotypes.

We conducted a GWAS applying an extreme phenotype sampling in heavy smokers presenting high and low risk of developing NSCLC, and we validated our results using the same approach. We selected individuals who developed tobacco-induced NSCLC at early onset and healthy subjects who did not develop NSCLC at an advanced age, despite having a long smoking history. We aimed to identify new susceptibility variants related to the respective extreme phenotypes.

## Patients and Methods

### Study design

We performed a two-stage extreme phenotype study to increase the efficiency of discovering single nucleotide polymorphisms (SNPs) associated with the risk of developing tobacco-induced NSCLC. We hypothesized that risk



**Figure 1.** Study design. From our series, we selected the individuals presenting extreme phenotypes of high and low risk of developing NSCLC induced by tobacco. Heavy smokers that either developed NSCLC at an early age were selected as extreme cases, and individuals that did not develop NSCLC at an advanced age, despite heavy tobacco consumption, were selected as extreme controls. The specific thresholds to define these populations were set to select the most extreme phenotypes in our series. \*We included selected cases that developed NSCLC at extremely young ages but presented tobacco consumption <15 pack-years, given their phenotypic relevance and because we assumed that they were too young to achieve the threshold of smoke consumption.

alleles would be strongly enriched in the phenotypic extremes, and therefore, a limited number of carefully selected individuals with extreme phenotypes might be sufficient to identify novel candidate genes and/or alleles [10].

We enrolled subjects into a discovery and a validation set (Fig. 1). The cancer cohort subjects (extreme cases) were selected from heavy smokers ( $\geq 15$  packs-years) with histologically confirmed diagnosis of NSCLC at an early age ( $\leq 55$  years). In the validation series, we included selected cases that developed NSCLC at extremely young ages but presented tobacco consumption  $< 15$  packs-years, given their phenotypic relevance and because we assumed that they were too young to achieve the threshold of smoke consumption. We also included some borderline cases for age ( $\leq 56$  years). The cancer-free cohort individuals (extreme controls) were selected from heavy smokers ( $\geq 15$  packs-years) that had not developed NSCLC or any other type of cancer at an advanced age ( $\geq 72$  years). The thresholds for tobacco consumption and age were set with the aim to select from our series the individuals presenting the most extreme phenotypes regarding individual risk of developing NSCLC induced by tobacco.

The discovery set was recruited among 3631 patients included in the databases of the University Clinic of Navarra (Pamplona, Spain), Center for Applied Medical Research (CIMA, Pamplona, Spain), and Hospital Universitario Nuestra Señora de La Candelaria (Tenerife, Spain).

The independent validation set was recruited from the Spanish branch of the European Prospective Investigation in Cancer (EPIC) Project ([www.epic-spain.com](http://www.epic-spain.com)), which includes genomic DNA samples and clinical data from 39,880 individuals, and from additional new cases from the University Clinic of Navarra.

Samples and data from patients were processed following standard operating procedures approved by the respective Ethical and Scientific Committees. All patients gave written informed consent to allow the use of their biological samples and clinical data for research purposes. The study protocol was approved by the Ethics Committee of the University Clinic of Navarra.

## DNA genotyping

Genomic DNA was obtained from peripheral blood mononuclear cells using the QIAamp DNA Mini Kit (Qiagen Iberia, Madrid, Spain) according to the manufacturer's instructions and stored at  $-20^{\circ}\text{C}$  until use. Genotyping in the discovery set was performed using the Illumina HumanOmni2.5-Quad BeadChip according to the manufacturer's protocols (Illumina, San Diego, CA, USA). Genotyping in the validation set was performed using the

Infinium assay following the manufacturer's instructions (Illumina).

## Statistical analysis

Discovery and validation analyses, per-allele odds ratios (OR), and standard errors were estimated for each SNP using a multivariate logistic regression model, adjusting for sex. The covariates age and tobacco consumption were used for the design, and therefore, they were not included in the statistical analysis.

Genotype clustering and calling were carried out using GenTrain v2.0 in GenomeStudio v2011.1 (Illumina). Samples were excluded from further analyses if they had more than 2% missing genotype data SNPs with a call rate  $< 0.95$  or deviated from the Hardy–Weinberg equilibrium (HWE) ( $p$ -value  $\leq 1 \times 10^{-5}$ ) were also excluded. From the discovery analysis, SNPs with a  $p$ -value  $\leq 5 \times 10^{-4}$  and minor allele frequency (MAF)  $> 5\%$  in the single SNP analysis were selected for validation. When two SNPs presented a very high correlation ( $r^2 > 0.8$ ), they were considered redundant, and only the one that presented the highest Illumina score was selected.

## Genomic imputation

To find map causing variants, data imputation for the untyped SNPs at all the validated SNPs was performed using IMPUTE v2.3.2 [11] for samples from the discovery set using phased haplotypes from the 1000 Genomes Phase 3 (b37) panel as reference [12]. Those imputed genotypes with scores  $> 0.3$  were tested for association.

**Table 1.** Characteristics of patients.

	Discovery set ( $n = 95$ )		Validation set ( $n = 133$ )	
	Extreme cases	Extreme controls	Extreme cases	Extreme controls
<i>N</i>	47	48	78	55
Sex M/F	31/16	43/5	52/26	51/4
(%M/%F)	(66/34)	(89.4/10.6)	(66.7/33.3)	(92.7/7.3)
Mean age, years	49	76	49	77
(range)	(38–55)	(72–84)	(35–56)	(72–85)
Mean tobacco consumption, packs-year (range)	41	69.4	38	48
Histological subtypes				
Adenocarcinoma	29	NA	44	NA
Squamous cell carcinoma	14		12	
Other	4		21	

NA, not applicable.

## Assessment of the prognostic value of mRNA expression and protein expression of target genes

We attempted to validate the functional relevance of the genes associated with the validated SNPs in NSCLC, assessing the expression of the related mRNA and proteins in patients with early-stage NSCLC. We correlated the transcriptomic data of *ATP10D* and *PDE10A* with recurrence-free survival (RFS) and overall survival (OS) using the KM-Plotter application [13] in patients with early-stage NSCLC. The database used (<http://kmplot.com>) includes gene expression data and survival information on 821 patients with stage I–II NSCLC, downloaded from the Cancer Biomedical Informatics Grid (CaBIG), the Gene Expression Omnibus (GEO), and the Cancer Genome Atlas (TCGA). Survival curves were estimated using Kaplan–Meier curves. Significant differences between groups were tested using the log-rank test. RFS and OS were calculated respectively from the date of surgery to the date of recurrence or death.

We assessed the prognostic value of PDE10A protein expression in tumor samples from 149 patients with stage I–II NSCLC resected at University Clinic of Navarra. PDE10A immunohistochemical assay and staining scores established by semiquantitative analysis were performed as previously described [14]. We used an anti-human PDE10A antibody (Genetex) diluted 1:500. The specificity of PDE10A antibody was demonstrated by Western blot analysis and immunocytochemistry of cell lines expressing different levels of the protein. Isotype and negative (omission of the primary antibody) controls were performed [14].

The univariate and multivariate Cox proportional hazards modeling were used to determine the effects of variables on cancer-specific OS and RFS. Variables with  $p < 0.1$  in the univariate analysis were included in the multivariable analysis. The proportional hazards assumption was examined by testing interactions between the covariables of the final model and time. Statistical analysis was performed using SPSS 15.0 (Chicago, Ill., USA) and Stata12 (College Station, TX: USA).

## Results

### Identification and validation of SNPs related to individuals presenting extreme phenotypes of developing NSCLC induced by tobacco

We analyzed 228 individuals with extreme phenotypes, 95 in the discovery set and 133 in the validation set (Table 1). The discovery set included 47 extreme cases that developed NSCLC at an early age (mean: 49 years) and were heavy smokers (mean: 41 packs-years). Extreme controls were 48 individuals that did not develop NSCLC at an advanced age (mean: 76 years) despite high tobacco consumption (mean: 69 packs-years). One control from the discovery set was excluded due to low call rate (call rate  $< 0.98$ ). The validation set included 78 extreme cases (mean age: 49 years; and 38 packs-years) and 55 extreme controls (mean age: 77 years; and 48 packs-years).

We successfully genotyped 2,379,855 variants in the discovery set, and 61,061 were excluded from the analysis due to low genotyping call-rate (0.95), and 370 were deviated from HWE in controls, leaving 2,318,553 for further analysis. Thirty-six SNPs presenting high association in the GWAS ( $p < 5 \times 10^{-4}$ , Table S1) were analyzed in the validation set.

Two SNPs were successfully replicated in the validation set (Table 2). The strongest validated association was with SNP rs12660420 at the 6q27 locus (combined  $p = 5.66 \times 10^{-5}$ ; combined OR = 2.80, 95% CI: 1.69–4.61), which maps to a non-coding transcript of the *PDE10A* gene. This SNP tags a linkage disequilibrium (LD) block of 85.8Kb in chromosome 6 (coordinates 166,109,230–166,195,063) (NCBI build 37 assembly), as defined by the furthest upstream and downstream SNPs displaying a detectable correlation ( $r^2 > 0.20$ ) with rs12660420.

Subsequently, genotypes of all other known variants in the locus with minor allele frequency  $> 5\%$  were imputed to this locus. In total, 275 SNPs were reliably imputed (imputation  $r^2$  score  $> 0.3$ ) and analyzed but

**Table 2.** Single nucleotide polymorphisms validated in extreme cases versus extreme controls.

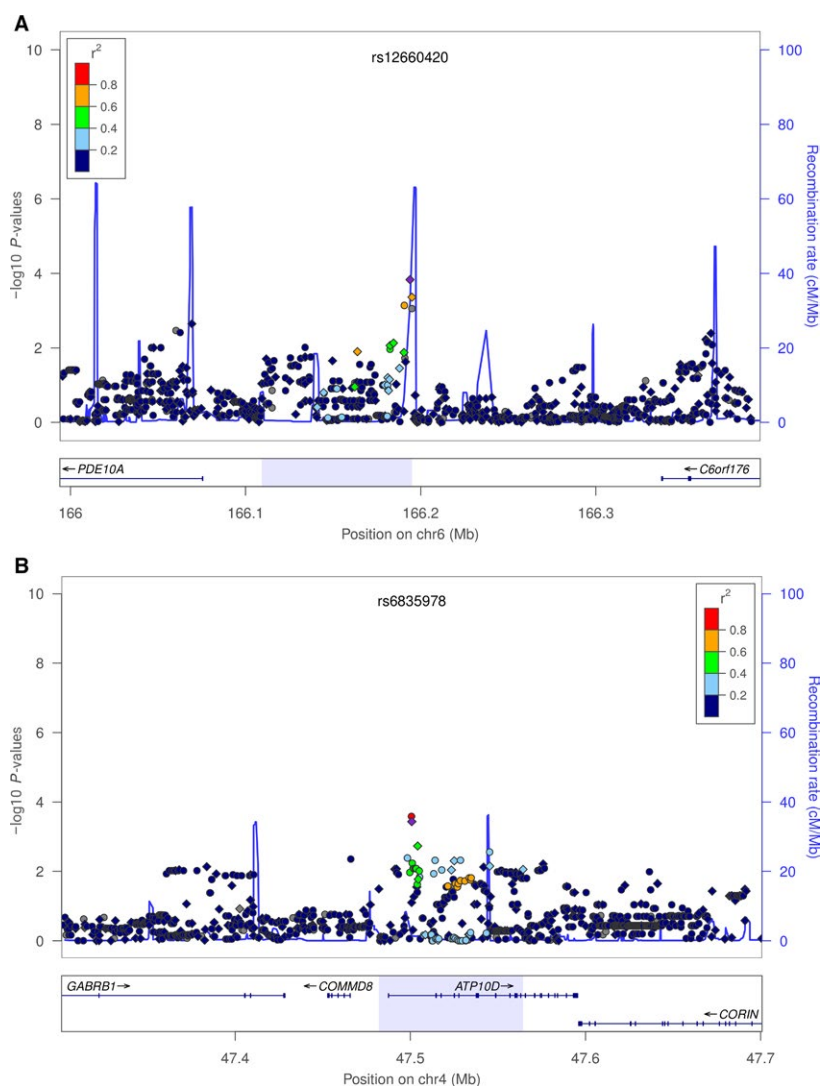
SNPs	Chromosome	Position	References	Alternate	Study	<i>p</i> -value	MAF	OR	95% CI
rs12660420	6	166193932	C	T	Discovery	1.48E-04	0.28	5.29	2.24–12.51
					Validation	0.04805	0.32	1.88	1.01–3.5
					Combined	5.66E-05	0.31	2.80	1.69–4.61
rs6835978	4	47500814	A	G	Discovery	3.69E-04	0.24	5.15	2.09–12.7
					Validation	0.04303	0.32	1.80	1.02–3.19
					Combined	1.02E-04	0.31	2.57	1.6–4.1

CI, confidence interval; MAF, minor allele frequency; OR, odds ratio; SNP, single nucleotide polymorphisms.

Associations between SNPs and NSCLC risk were assessed by logistic regression analyses adjusted for sex. Additive model of inheritance was considered. ORs are per copy of the specified minor allele. Chromosome positions are based on Genome Reference Consortium Human Build 37 (GRCh37/hg19).

none showed a lower  $p$ -value than rs12660420. The LD plot showing the  $p$ -values of these SNPs is shown in Figure 2. The related gene, *PDE10A*, encodes a protein that belongs to the cyclic nucleotide phosphodiesterase family. *PDE10A* expression correlates with lung cancer prognosis, and its pharmacological inhibition suppresses growth of NSCLC cell lines [15], so it seems a plausible candidate gene for lung cancer risk. Although the validated SNP lies in a non-coding transcript of the *PDE10A* gene, genetic variation in regulatory elements may influence expression.

The second validated SNP was rs6835978 (chr4:47498797) in 4p12 (combined  $p = 1.02 \times 10^{-4}$ ; combined OR = 2.57, 95% IC: 1.6–4.1, Table 2), an intronic variant in the *ATP10D* gene (ATPase phospholipid transporting 10D) [16]. *ATP10D* has not previously been related to lung cancer or other tumors. rs6835978 tags to a LD block of 82Kb in chromosome 4 (47,481,971–47,564,368), defined as described above. A total of 228 SNPs were reliably imputed in the region, and rs6835978 remained the most significantly associated SNP (along with rs12510653, in complete LD with it [ $r^2 = 1$ ]).



**Figure 2.** Association and recombination plot of the LD block containing the validated SNPs and  $\pm 200$  kb boundaries. Plot shows the genomic region containing (A) rs12660420 (B) rs6835978 SNPs associated with lung cancer and the  $-\log_{10}$  association  $p$ -values of genotyped and imputed SNPs in the discovery set (MAF > 5%). Circles and diamonds indicate genotyped and imputed SNPs, respectively. Recombination rates are also shown. SNP color indicates the strength of LD ( $r^2$ ) with rs12660420 (A) or rs6835978 (B). The rectangle below the plot shows the genes mapping in the region as well as, shaded in light blue, the LD block tagged by the associated SNPs. Recombination rates are based on the 1000 Genomes Project, and genomic coordinates are based on Genome Reference Consortium Human Build 37 (GRCh37/hg19). Plots were generated using LocusZoom software [31].



## Assessment of the prognostic value of mRNA expression and protein expression of target genes in patients with early-stage NSCLC

We evaluated the prognostic value of mRNA expression of *PDE10A* and *ATP10D* with the in silico tool KM-plotter [13] in 821 patients presenting stage I–II NSCLC. We analyzed *PDE10A* (probes 211170 and 211171) and *ATP10D* (probe 213238). We classified patients into two groups according to the median value of expression of each biomarker. Patients with higher *PDE10A* mRNA expression showed decreased OS using both probes ( $p = 0.011$  and  $p < 0.0001$ ; Fig. 3, panels A and B). Low mRNA expression of *ATP10D* was associated with shorter OS ( $p < 0.0001$ ; Fig. 3, panel C).

We assessed the prognostic value of PDE10A protein expression in 149 patients with stage I–II NSCLC (Table S2), followed during a median period of 46.5 months (interquartile range: 91.6–22.0). High tumor expression of PDE10A correlated with decreased RFS ( $p = 0.0006$ ) and OS ( $p = 0.002$ ) (Fig. 4). PDE10A expression correlated significantly with RFS and OS in the univariate analysis (Table 3). In the multivariate analysis, PDE10A expression was the strongest independent prognostic factor, over stage and age, and the only one that retained significance for both RFS (HR = 2.637; 95% CI = 1.381–5.036;  $p = 0.003$ ) and OS (HR = 2.909; 95% CI = 1.376–6.144;  $p = 0.005$ ).

## Discussion

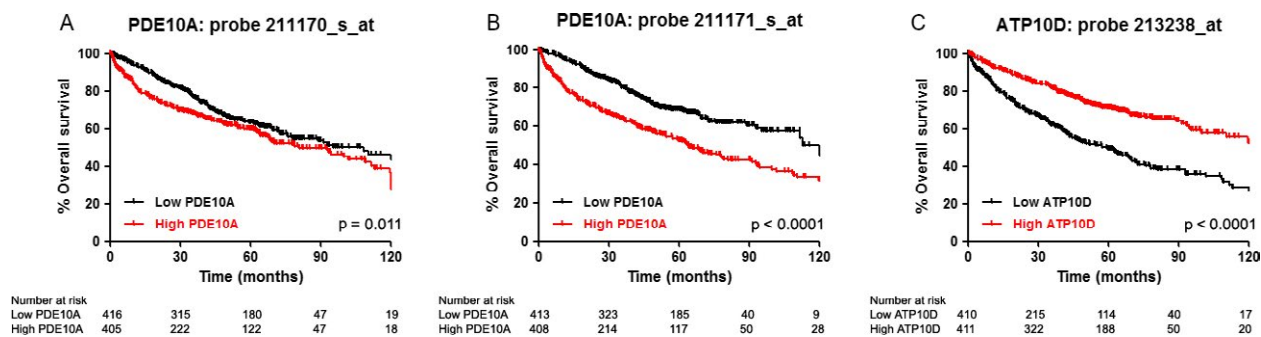
We identified and validated two novel variants associated with individuals presenting extreme phenotypes of risk to develop NSCLC induced by tobacco. To our knowledge, this is the first study that uses this strategy to identify novel genetic variants related to NSCLC cancer risk. This methodology consists of studying reduced groups of individuals with very characteristic and clinically relevant phenotypes, thus enriching biomarker expression in the individuals studied, and it has allowed outstanding cancer biomarkers to be identified, as reviewed elsewhere [17]. In the discovery phase, we used the HumanOmni2.5-Quad BeadChip array, a high-density chip that allows genotyping of 2.5 million SNPs across the whole genome, targeting genetic variation down to 1% MAF. To our knowledge, this is the first study that assesses cancer risk using this powerful platform.

The strongest validated association was SNP rs12660420, which maps to a non-coding transcript of *PDE10A*. This gene encodes a protein that belongs to the cyclic nucleotide phosphodiesterase family. Phosphodiesterases (PDE) represent a large superfamily of hydrolases that control

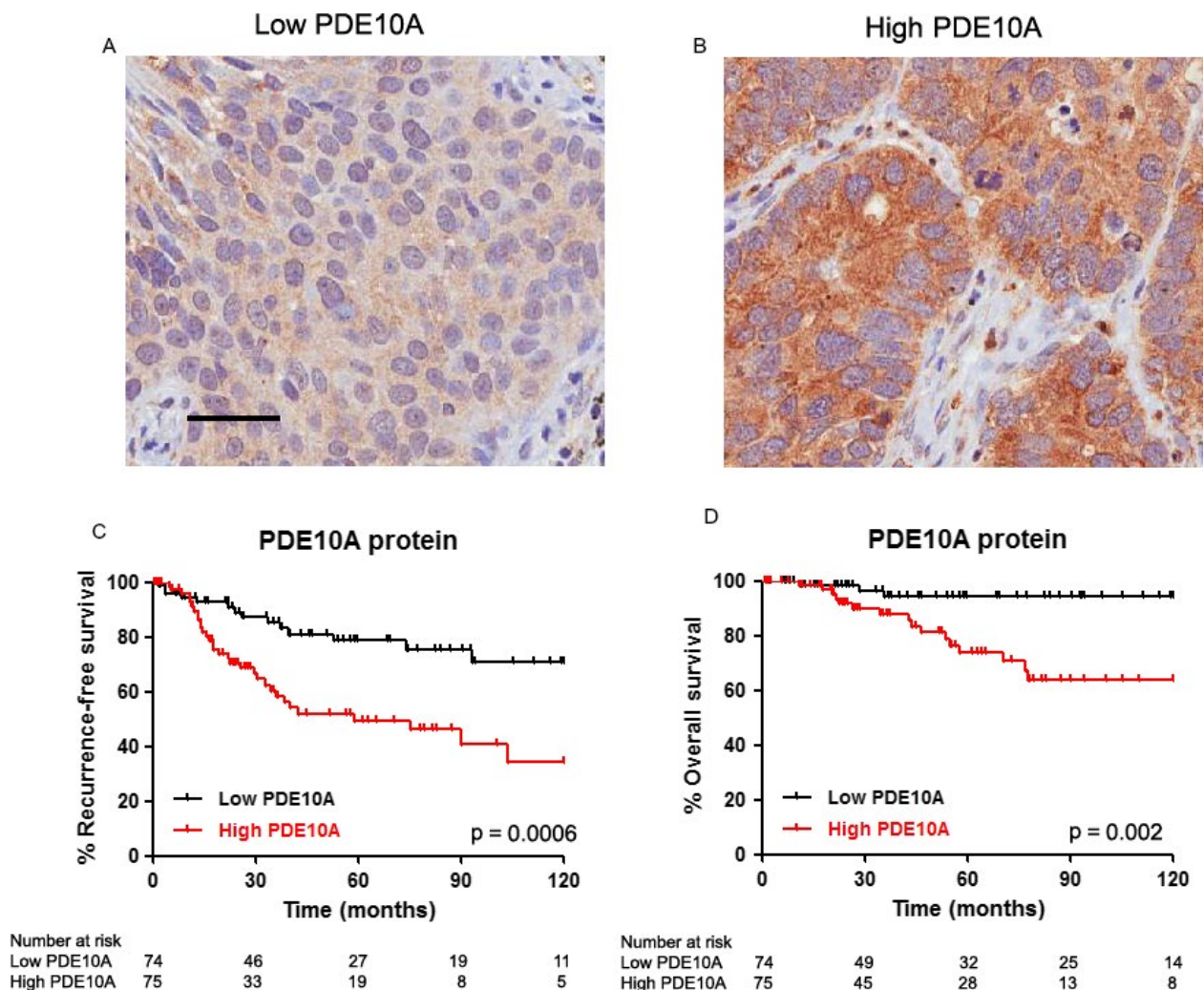
the intracellular levels of cyclic nucleotides by hydrolyzing cAMP and cGMP to 5\_AMP and 5\_GMP and are critical regulators of the myriad physiological and pathophysiological processes under cyclic nucleotide control [18]. PDE4, a well-known target for therapy of chronic obstructive pulmonary disease [19], is also expressed in lung cancer and promotes lung cancer progression [20]. Preclinical data indicate that PDE4 inhibitors have also a role in the treatment for EGFR-mutant NSCLC patients with high pretreatment levels of BIM and mTOR [21]. Li et al. showed that PDE10A mRNA and protein levels are overexpressed in colon tumor cells as compared to normal colonocytes; that silencing *PDE10A* suppressed the growth of colon tumor cells; and that overexpression of *PDE10A* promotes growth of colonocytes and adenoma cells [22]. The same group recently reported that Pf-2545920, a highly specific *PDE10A* inhibitor, inhibits colon tumor cell growth at concentrations that increase cGMP and cAMP levels and activate PKG and PKA; and that Pf-2545920 reduces  $\beta$ -catenin-mediated transcription of survivin, resulting in caspase activation and apoptosis [23]. Shen et al. [15] have reported that *PDE10A* mutations are predicted to influence *PDE10A* allosteric regulation in lung adenocarcinoma; that high levels of *PDE10A* expression correlate with worsened lung adenocarcinoma prognosis; and that Pf-2545920 also suppresses the growth of NSCLC cell lines. Finally, *PDE10A* somatic mutations have been described in up to 19% of prostate cancers and correlated with increased levels of phosphorylated C-AMP responsive element binding protein (p-CREB) [24]. These data suggest that PDE10A might be a relevant target for cancer therapy and prevention.

The second validated variant, SNP rs6835978, is an intronic variant of *ATP10D*, which belongs to a subfamily of P-type ATPases implicated in the translocation of phospholipids from the exoplasmic to the cytoplasmic leaflet of the cellular biological membranes. Although *ATP10D* has not previously been related to NSCLC, genetic variation in this gene has been associated with levels of plasma sphingolipids [25]. Bioactive sphingolipids, such as sphingosine-1-phosphate (S1P), and ceramides, are signaling molecules involved in the activation of pathways that are directly relevant to carcinogenesis [26]. Ceramide plays a major role in the development of chronic pulmonary diseases and has also been involved in lung cancer development induced by cigarette smoke [27]. Alberg et al. reported that higher concentrations of S1P and total ceramide in plasma were associated with an increased future risk of lung cancer [28].

We analyzed the prognostic value of *PDE10A* and *ATP10D* mRNA and PDE10A protein levels in patients with stage I–II NSCLC. We confirmed that high and low mRNA expression, respectively, of *PDE10A* and



**Figure 3.** Correlation of the transcriptomic data of *PDE10A* and *ATP10D* with overall survival (OS) using the Kaplan–Meier Plotter application in several databases including 821 patients with stage I–II NSCLC. Panels A and B: OS according to *PDE10A* mRNA expression levels using probes 211170 (A) and 211171 (B). Panel C: OS according to *ATP10D* mRNA expression levels using probe 213238.



**Figure 4.** Correlation of *PDE10A* protein expression with overall survival (OS) and relapse-free survival (RFS) in 149 patients with stage I–II NSCLC. Panels A and B: representative examples of tumors presenting low (A) and high (B) *PDE10A* staining. Scale bar: 50 micras. Panel C: RFS according to *PDE10A* staining. Panel D: OS according to *PDE10A* staining.

**Table 3.** Univariate and multivariate Cox regression analyses of PDE10A protein expression for recurrence-free survival (RFS) and overall survival (OS) in patients with stage I–II NSCLC.

Factor	n	RFS				OS			
		Univariate HR (95% CI)	p	Multivariate HR (95% CI) <sup>1</sup>	p	Univariate HR (95% CI)	p	Multivariate HR (95% CI) <sup>1</sup>	p
PDE10A									
Low	74	–	–	–	–	–	–	–	–
High	75	2.831 (1.501–5.340)	0.001	2.637 (1.381–5.036)	0.003	2.86 (1.371–5.970)	0.005	2.909 (1.376–6.144)	0.005
Stage									
I	101	–	–	–	–	–	–	–	–
II	48	1.742 (0.963–3.151)	0.066	1.425 (0.779–2.607)	0.250	1.971 (1.011–3.840)	0.046	1.700 (0.864–3.331)	0.125
Age									
<65	74	–	–	–	–	–	–	–	–
>65	75	1.361 (0.755–2.457)	0.305	–	–	1.822 (0.927–3.579)	0.082	2.065 (1.042–4.091)	0.038

OS, overall survival; RFS, relapse-free survival.

<sup>1</sup>Adjusted by gender, stage, and age.

*ATP10D* significantly correlated with shorter survival. To our knowledge, this is the first association of *ATP10D* with cancer prognosis. We confirmed that PDE10A protein levels in stage I–II NSCLC patients correlated with decreased survival. The results that we obtained for mRNA and protein expression of PDE10A are consistent among them and also with previous results reported in NSCLC [15]. These results render *PDE10A* and *ATP10D* as potentially useful prognostic and therapeutic targets in NSCLC.

Tobacco-induced NSCLC represents one of the most relevant challenges to public health. Therefore, the identification of genetic factors that confer to individuals either an increased risk or an intrinsic protection to develop the disease would be of critical relevance, allowing the identification of high-risk populations, in which tobacco cessation and screening programs may be most beneficial. Most importantly, it may lead to further understanding of mechanisms of carcinogenesis and natural protection against cancer, which might inspire new therapeutic strategies. Interestingly, even though our study was specifically designed for NSCLC, the individuals in the cancer-free cohort did not develop any other tumors. Therefore, our results may also relate to other neoplasms, especially those related to tobacco.

To our knowledge, this is the first study that defines and targets phenotypes of decreased risk to develop tobacco-induced NSCLC and that attempts to characterize the genetic causes of such protection. The study of these proficient cancer risk phenotypes (PROCARPs) may allow to identify the underlying protective mechanisms and may provide valuable targets for cancer prevention. The identification of PROCARPs is challenging, because it requires to distinguish individuals presenting a truly decreased cancer risk from others which just present absence of the disease (i.e., they are non-apparent phenotypes) [29]. We

hypothesized that the selection of individuals that are overexposed to either extrinsic or intrinsic cancer risk factors and yet do not develop cancer may allow to identify and study PROCARPs [17, 30]. On the other hand, phenotypes of increased cancer risk (deficient cancer risk phenotypes, DECARPs) are apparent phenotypes that are readily identifiable (e.g., cancer familiar syndromes, cancer development at young ages, etc.) [29].

The main limitation of our study is that the sample size is limited due to the difficulty to enroll extreme individuals. This may impair to reach *p*-values of genome-wide significance. Nevertheless, to overcome this limitation, we explored mRNA and protein expression of *PDE10A* and *ATP10D* in two different patient populations of early-stage NSCLC, and we found very significant correlations with survival. Functional validation will also be required to confirm the relevance of our findings. Also, despite our efforts to recruit highly selected individuals, phenotypic heterogeneity may persist regarding demographical and clinical characteristics. Nevertheless, this limitation constitutes in fact the greatest opportunity to improve this strategy, through the application of more stringent selection criteria to identify more homogeneous phenotypes with regard to age, tobacco consumption, or other relevant variables. Our strategy may also be pursued using different high-throughput techniques, such as exome or genome sequencing or assessment of the transcriptome, and epigenome. Finally, additional studies will be required to further elucidate the optimal definitions of extreme phenotypes.

In summary, we identified and validated two new genetic variants in *PDE10A* and *ATP10D* associated with individuals presenting extreme phenotypes of high and low risk of developing tobacco-induced NSCLC, and we confirmed the prognostic relevance of the associated proteins in early-stage NSCLC. Our findings may have relevant implications to define high-risk cancer populations and to

characterize molecular mechanisms of carcinogenesis and resistance to develop cancer, thus supporting the feasibility of extreme phenotype selection studies in cancer research.

## Acknowledgment

This work was supported by the Spanish Society of Medical Oncology; Fundación SEOM and Fundación Salud 2000; and Government of Navarra.

## Conflicts of Interest

The authors declare no conflicts of interest.

## References

- Hung, R. J., J. D. McKay, V. Gaborieau, P. Boffetta, M. Hashibe, D. Zaridze, et al. 2008. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 452:633–637.
- Amos, C. I., X. Wu, P. Broderick, I. P. Gorlov, J. Gu, T. Eisen, et al. 2008. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat. Genet.* 40:616–622.
- Thorgeirsson, T. E., F. Geller, P. Sulem, T. Rafnar, A. Wiste, K. P. Magnusson, et al. 2008. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 452:638–642.
- Rafnar, T., P. Sulem, S. N. Stacey, F. Geller, J. Gudmundsson, A. Sigurdsson, et al. 2009. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat. Genet.* 41:221–227.
- McKay, J. D., R. J. Hung, V. Gaborieau, P. Boffetta, A. Chabrier, G. Byrnes, et al. 2008. Lung cancer susceptibility locus at 5p15.33. *Nat. Genet.* 40:1404–1406.
- Landi, M. T., N. Chatterjee, K. Yu, L. R. Goldin, A. M. Goldstein, M. Rotunno, et al. 2009. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am. J. Hum. Genet.* 85:679–691.
- Wang, Y., P. Broderick, A. Matakidou, T. Eisen, and R. S. Houlston. 2010. Role of 5p15.33 (TERT-CLPTM1L), 6p21.33 and 15q25.1 (CHRNA5-CHRNA3) variation and lung cancer risk in never-smokers. *Carcinogenesis* 31:234–238.
- Wang, Y., P. Broderick, E. Webb, X. Wu, J. Vijaykrishnan, A. Matakidou, et al. 2008. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat. Genet.* 40:1407–1409.
- Li, Y., C.-C. Sheu, Y. Ye, M. de Andrade, L. Wang, S. C. Chang, et al. 2010. Genetic variants and risk of lung cancer in never smokers: a genome-wide association study. *Lancet Oncol.* 11:321–330.
- Perez-Gracia, J. L., M. F. Sanmamed, A. Bosch, A. Patiño-García, K. A. Schalper, V. Segura, et al. 2017. Strategies to design clinical studies to identify predictive biomarkers in cancer research. *Cancer Treat. Rev.* 53:79–97.
- Howie, B. N., P. Donnelly, and J. Marchini. 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 5:e1000529.
- 1000 Genomes Project Consortium A, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, and H. M. Kang, et al. 2015. A global reference for human genetic variation. *Nature* 526:68–74.
- Györfy, B., P. Surowiak, J. Budczies, and A. Lánczky. 2013. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One* 8:e82241. doi: 10.1371/journal.pone.0082241. eCollection 2013.
- Pajares, M. J., J. Agorreta, M. Larrayoz, A. Vesin, T. Ezponda, I. Zudaire, et al. 2012. Expression of tumor-derived vascular endothelial growth factor and its receptors is associated with outcome in early squamous cell carcinoma of the lung. *J. Clin. Oncol.* 30:1129–1136.
- Shen, Q., F. Cheng, H. Song, W. Lu, J. Zhao, X. An, et al. 2017. Proteome-scale investigation of protein allosteric regulation perturbed by somatic mutations in 7,000 cancer genomes. *Am. J. Hum. Genet.* 100:5–20.
- Flamant, S., P. Pescher, B. Lemerrier, M. Clément-Ziza, F. Képès, M. Fellous, et al. 2003. Characterization of a putative type IV aminophospholipid transporter P-type ATPase. *Mamm. Genome* 14:21–30.
- Perez-Gracia, J. L., M. G. Ruiz-Ilundain, I. Garcia-Ribas, and E. M. Carrasco. 2002. The role of extreme phenotype selection studies in the identification of clinically relevant genotypes in cancer research. *Cancer* 95:1605–1610.
- Azevedo, M. F., F. R. Faucz, E. Bimpaki, A. Horvath, I. Levy, R. B. de Alexandre, et al. 2014. Clinical and molecular genetics of the phosphodiesterases (PDEs). *Endocr. Rev.* 35:195–233.
- Martinez, F. J., P. M. A. Calverley, U.-M. Goehring, M. Brose, L. M. Fabbri, and K. F. Rabe. 2015. Effect of roflumilast on exacerbations in patients with severe chronic obstructive pulmonary disease uncontrolled by combination therapy (REACT): a multicentre randomised controlled trial. *Lancet* 385:857–866.
- Pullamsetti, S. S., G. A. Banat, A. Schmall, M. Szibor, D. Pomagruk, J. Hänze, et al. 2013. Phosphodiesterase-4 promotes proliferation and angiogenesis of lung cancer by crosstalk with HIF. *Oncogene* 32:1121–1134.

21. Karachaliou, N., J. Codony-Servat, C. Teixidó, S. Pilotto, A. Drozdowsky, C. Codony-Servat, et al. 2015. BIM and mTOR expression levels predict outcome to erlotinib in EGFR-mutant non-small-cell lung cancer. *Sci. Rep.* 5:17499.
22. Li, N., K. Lee, Y. Xi, B. Zhu, B. D. Gary, V. Ramírez-Alcántara, et al. 2015. Phosphodiesterase 10A: a novel target for selective inhibition of colon tumor cell growth and  $\beta$ -catenin-dependent TCF transcriptional activity. *Oncogene* 34:1499–1509.
23. Lee, K., A. S. Lindsey, N. Li, B. Gary, J. Andrews, A. B. Keeton, et al. 2016.  $\beta$ -catenin nuclear translocation in colorectal cancer cells is suppressed by PDE10A inhibition, cGMP elevation, and activation of PKG. *Oncotarget* 7:5353–5365.
24. de Alexandre, R. B., A. D. Horvath, E. Szarek, A. D. Manning, L. F. Leal, F. Kardauke, et al. 2015. Phosphodiesterase sequence variants may predispose to prostate cancer. *Endocr. Relat. Cancer* 22:519–530.
25. Hicks, A. A., P. P. Pramstaller, Å. Johansson, V. Vitart, I. Rudan, P. Ugozai, et al. 2009. Genetic determinants of circulating sphingolipid concentrations in European populations. *PLoS Genet.* 5:e1000672.
26. Furuya, H., Y. Shimizu, and T. Kawamori. 2011. Sphingolipids in cancer. *Cancer Metastasis Rev.* 30:567–576.
27. Goldkorn, T., S. Filosto, and S. Chung. 2014. Lung injury and lung cancer caused by cigarette smoke-induced oxidative stress: molecular mechanisms and therapeutic opportunities involving the ceramide-generating machinery and epidermal growth factor receptor. *Antioxid. Redox Signal.* 21:2149–2174.
28. Alberg, A. J., K. Armeson, J. S. Pierce, J. Bielawski, A. Bielawska, K. Visvanathan, et al. 2013. Plasma sphingolipids and lung cancer: a population-based, nested case-control study. *Cancer Epidemiol. Biomarkers Prev.* 22:1374–1382.
29. Pérez-Gracia, J. L., A. Gúrpide, M. G. Ruiz-Ilundain, C. Alfaro Alegría, R. Colomer, J. García-Foncillas, et al. 2010. Selection of extreme phenotypes: the role of clinical observation in translational research. *Clin. Transl. Oncol.* 12:174–180.
30. Perez-Gracia, J. L., and M. G. Ruiz-Ilundain. 2001. Cancer protective mutations: looking for the needle in the haystack. *Clin. Transl. Oncol.* 3:169–171.
31. Pruim, R. J., R. P. Welch, S. Sanna, T. M. Teslovich, P. S. Chines, T. P. Gliedt, et al. 2010. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26:2336–2337.

## Supporting Information

Additional supporting information may be found in the online version of this article:

**Table S1.** SNPs presenting high association in GWAS ( $p < 5 \times 10^{-4}$ ).

**Table S2.** Clinicopathological features of 149 stage I–II NSCLC patients.

**Appendix S1.** Immunohistochemical analysis.