

Curso 2011/12  
**CIENCIAS Y TECNOLOGÍAS/12**  
I.S.B.N.: 978-84-15910-07-7

**MARÍA DEL MAR DEL PINO YANES**

**Factores genéticos de susceptibilidad  
y gravedad del asma:  
estudio de genes candidatos y de la influencia  
del trasfondo genético poblacional**

**Directores**

**CARLOS A. FLORES INFANTE  
MARIANO N. HERNÁNDEZ FERRER**



**SOPORTES AUDIOVISUALES E INFORMÁTICOS**  
**Serie Tesis Doctorales**

Esta tesis se ha realizado gracias a la financiación de los proyectos EMER 07/001 y FIS PI081383 del Fondo de Investigación Sanitaria del Instituto de Salud Carlos III, el proyecto SAF 2004/06833 del Ministerio de Ciencia e Innovación, los proyectos FUNCIS 27/07 y 53/04 de la Fundación Canaria de Investigación y Salud y mediante el programa corporativo de Lesión Pulmonar Aguda del CIBERES. Mi contratación durante estos cuatro años ha sido financiada mediante un acuerdo entre el Gobierno de Canarias y el Instituto de Salud Carlos III.

## Agradecimientos

En primer lugar debo agradecer a mis directores Mariano Hernández y Carlos Flores el haberme dado la oportunidad de trabajar en este proyecto y el tiempo que han invertido en formarme durante estos cuatro años. De no ser por Mariano, que me transmitió el interés por la investigación durante mi beca de colaboración, y pensó en mí para realizar esta tesis, probablemente no habría trabajado en este tema. A Carlos tengo que agradecer el entusiasmo que ha mantenido en todo momento y el haberme hecho siempre partícipe de todas las facetas de la investigación.

Un estudio en el que se incluyen tantas muestras como las que se analizan aquí, requiere el esfuerzo de muchísimas personas. Por ello, en primer lugar debo reconocer el trabajo realizado por varios médicos y biólogos en la recolección de las muestras de asmáticos y su caracterización clínica: Inmaculada Sánchez, José Cumplido, Teresa Carrillo, Anselmo Sánchez, Ruperto González, Paloma Poza, Javier Figueroa, Orlando Acosta Fernández, José Carlos García Robaina, María José Torres y Tilman Klassert. En segundo lugar, a las personas implicadas en la recogida de muestras de controles: a Luisa Guerra y Milena Casula por las muestras del Banco de Sangre del Hospital Universitario Dr. Negrín y a David Comas y José M<sup>a</sup> Larruga por las muestras de individuos norteafricanos. No quiero olvidarme del personal de enfermería de los Servicios Hospitalarios participantes en este estudio, que se ha encargado de realizar las extracciones de sangre. Asimismo, mi más sincera gratitud a los más de dos mil donantes anónimos.

Muchos han sido los colaboradores implicados en los trabajos incluidos en esta tesis. Entre ellos, debo expresar mi agradecimiento a Jesús Villar por la oportunidad de participar en los proyectos de asociación de lesión pulmonar aguda y por contar conmigo en su grupo CIBERES. A Paula Tejera por la secuenciación de *IRAK3*, porque gracias a esos datos comenzó gran parte del trabajo de esta tesis. A Santiago Basaldúa por su ayuda programando herramientas como SNPing y SNPInfostats, que han facilitado mucho este trabajo. Además, no quiero dejar de lado a los colaboradores de Chicago (Shwu-Fan Ma, Xiaoguang Sun y Joe G. N. Garcia) por su ayuda con los ensayos de luciferasa y por dejarme colaborar en muchos de sus trabajos, ello ha permitido mejorar mi formación científica. Asimismo, me gustaría reconocer la labor de Alexis Hernández, por dedicar muchos fines de semana a la ayuda con el secuenciador.

Entre mis compañeros de laboratorio debo destacar a Almudena Corrales, no sólo por la ayuda que me ha prestado con las extracciones de ADN, la elaboración de la base de datos y la revisión de secuencias, sino más aún por su inestimable amistad: por ser desde el principio una de las personas que más me ha animado para que esta tesis saliese adelante.

A otros muchos compañeros de la Unidad de Investigación, por los buenos momentos que hemos pasado en el laboratorio, en el café y fuera de la Unidad: a Marialbert, Pablo, Teresa, Lina, Belinda, Lari, Bea, Isa, Violeta, Óliver, Diana, Jonay y muchos otros. También a los compañeros del Instituto de Enfermedades Tropicales y Salud Pública de Canarias, porque los meses que he pasado allí me han acogido como si fuese una más.

Esta tesis no habría sido posible sin mi familia, por ello quiero agradecer a mis padres y hermanos su cariño y su apoyo para seguir la carrera científica. A mi padre por inculcarme el perfeccionismo y meticulosidad, y a mi madre por servir siempre como ejemplo de capacidad de trabajo y tenacidad. A Menucha y Eladio, no sólo por su afecto y por los ánimos para continuar siempre adelante, sino también porque la mayor parte de esta tesis la he escrito en su casa. También a Jonatán, porque sin su ayuda no hubiese sido capaz de poner a funcionar el Linux.

Finalmente, pero quizá el más importante: gracias a Tobías. Porque me has apoyado durante estos siete maravillosos años de mi vida y has sido el mayor estímulo para que termine esta tesis. Siempre has estado conmigo animándome y transmitiéndome serenidad. Aunque no puedas estar el día de la lectura, espero que te sientas orgulloso de mí.

## Abreviaturas

1KGP	Proyecto 1000 genomas
<i>ADAM33</i>	Gen de la metalopeptidasa 33
<i>ADRB2</i>	Gen del receptor beta adrenérgico
ALI	Lesión pulmonar aguda
ANOVA	Análisis de la varianza
AUC	Área bajo la curva característica operativa del receptor
<i>CD14</i>	Gen del antígeno CD14
CEU	Individuos norteamericanos con ascendencia en el norte u oeste de Europa
CNV	Polimorfismo de variación en el número de copia
CI	Intervalo de confianza
CR	Tasa de completado de genotipos
EuroAIM	Marcador informativo de ascendencia europea
<i>FCER1B</i>	Cadena beta del receptor del factor cristalizable de la IgE
FDR	Tasa de falso descubrimiento
FEV1	Volumen expiratorio forzado en un segundo
FEV1 post	Volumen expiratorio forzado en 1 s tras exposición a un broncodilatador
GINA	Iniciativa global para el asma
GOA	Estudio de la genética del asma en la población española
GRS	Puntuaciones de riesgo genético
GWAS	Estudio de asociación genómico
HGDP	Panel mundial de la diversidad genética humana
HWE	Equilibrio de Hardy-Weinberg
Ig	Inmunoglobulina
<i>IL13</i>	Gen de la interleucina 13
<i>IL4</i>	Gen de la interleucina 4
<i>IL4R</i>	Gen del receptor de la interleucina 4
<i>IRAK3</i>	Gen de la quinasa 3 asociada al receptor de la interleucina 1
kb	kilobase
kD	kilodaltons
LBP	Proteína de unión a lipopolisacárido
LD	Desequilibrio de ligamiento
LPS	Lipopolisacárido
<i>LTA</i>	Gen de la linfotóxina $\alpha$
MAF	Frecuencia del alelo menos común
$M_D$	Medida de correlación multilocus entre polimorfismos
MHC	Complejo mayor de histocompatibilidad
<i>MS4A2</i>	Gen de la cadena beta del receptor del factor cristalizable de la IgE
OR	Razones de ventaja, medida del efecto de una asociación estadística
PAMP	Patrón molecular asociado a patógenos
pb	Pares de bases
PC1	Primera componente principal
PCA	Análisis de componentes principales
PCR	Reacción de la polimerasa en cadena
ROC	Característica Operativa del Receptor
SNP	Polimorfismo de un solo nucleótido
SPT	Prueba cutánea de alergia
<i>TLR</i>	Receptor de tipo Toll
<i>TNF</i>	Gen del factor de necrosis tumoral
tSNP	Tagging SNP (SNP informativo de una región del genoma)

## ***ÍNDICE***

---

<b>1. INTRODUCCIÓN</b> .....	<b>3</b>
<b>1.1. Definición de asma</b> .....	<b>3</b>
<b>1.2. Prevalencia mundial del asma</b> .....	<b>3</b>
<b>1.3. Prevalencia del asma en España</b> .....	<b>3</b>
<b>1.4. Patogenia del asma</b> .....	<b>4</b>
<b>1.5. Fisiopatología del asma</b> .....	<b>6</b>
<b>1.6. Genética del asma</b> .....	<b>6</b>
1.6.1. Análisis de ligamiento .....	7
1.6.2. Estudios de asociación en genes candidatos .....	9
1.6.3. Genes candidatos más firmemente replicados en asma .....	10
1.6.3.1. Gen de la interleucina 4 ( <i>IL4</i> ).....	13
1.6.3.2. Gen de la interleucina 13 ( <i>IL13</i> ).....	13
1.6.3.3. Gen del receptor de la interleucina 4 ( <i>IL4R</i> ) .....	14
1.6.3.4. Genes del factor de necrosis tumoral ( <i>TNF</i> ) y la linfotoxina $\alpha$ ( <i>LTA</i> ) .....	15
1.6.3.5. Gen del antígeno CD14 ( <i>CD14</i> ).....	17
1.6.3.6. Gen del receptor beta adrenérgico ( <i>ADRB2</i> ) .....	18
1.6.3.7. Gen de la cadena beta del receptor de la IgE ( <i>MS4A2</i> ).....	19
1.6.3.8. Gen de la metalopeptidasa 33 ( <i>ADAM33</i> ).....	20
1.6.4. <i>IRAK3</i> como gen candidato de susceptibilidad en asma .....	21
1.6.5. Estudios de la genética del asma en España .....	22
1.6.6. Estudios de asociación genómica (GWAS) .....	24
1.6.7. Meta-análisis de GWAS en asma .....	26
1.6.8. Perspectivas futuras en la genética del asma.....	27
<b>1.7. Relación entre ascendencia genética y asma</b> .....	<b>29</b>
1.7.1. Ascendencia genética africana y riesgo al asma en afroamericanos .....	29
1.7.2. Ascendencia genética amerindia y protección frente al asma en la población mexicana.....	30

## ÍNDICE

---

<b>1.8. Influencia norteafricana en la población española.....</b>	<b>30</b>
<b>2. HIPÓTESIS Y OBJETIVOS.....</b>	<b>36</b>
<b>3. METODOLOGÍA.....</b>	<b>41</b>
<b>3.1. Diseño y muestras para el estudio de estratificación poblacional .....</b>	<b>41</b>
3.1.1. Muestras españolas y norteafricanas.....	41
3.1.2. Datos de otras poblaciones.....	42
3.1.3. Selección de marcadores de ascendencia.....	42
<b>3.2. Diseño y muestras para el estudio de asociación.....</b>	<b>43</b>
3.2.1. Muestras de asmáticos y controles .....	43
3.2.2. Selección de SNPs en los genes candidatos.....	44
<b>3.3. Genotipado .....</b>	<b>47</b>
3.3.1. Preparación de las muestras para el genotipado.....	47
3.3.1.1. Extracción de ADN .....	47
3.3.1.2. Amplificación genómica .....	47
3.3.1.3. Cuantificación de la amplificación genómica.....	48
3.3.2. Técnicas de genotipado empleadas.....	49
3.3.2.1. Genotipado mediante MassARRAY® iPLEX Gold .....	49
3.3.2.2. Genotipado mediante reacciones de SNaPshot® .....	50
3.3.2.3. Genotipado mediante sondas KASPar® .....	51
<b>3.4. Análisis estadístico .....</b>	<b>53</b>
3.4.1. Estudio de estratificación poblacional y relación entre la ascendencia genética y el asma.....	53
3.4.1.1. Evaluación de la capacidad informativa de los marcadores de ascendencia genética .....	53
3.4.1.2. Relación entre poblaciones .....	53
3.4.1.3. Estimaciones de ascendencia poblacional e individual .....	54
3.4.1.4. Inferencia de ascendencia con distintos subconjuntos de polimorfismos ..	54



3.4.1.5. Diferencias entre las frecuencias alélicas entre muestras españolas .....	55
3.4.1.6. Estudio de la relación entre ascendencia genética y asma .....	55
3.4.2. Estudios de asociación de genes candidatos con el asma .....	56
3.4.2.1. Estimación del poder estadístico del estudio de asociación <i>a priori</i> .....	56
3.4.2.2. Análisis de los datos demográficos, de morbilidad y clínicos.....	56
3.4.2.3. Control de la calidad de los datos de genotipado .....	56
3.4.2.4. Asociación de tSNPs individuales con la enfermedad .....	56
3.4.2.5. Ajustes de la presencia de estratificación poblacional .....	57
3.4.2.6. Imputación de variantes no genotipadas: evaluación inicial y posterior análisis de asociación .....	57
3.4.2.7. Análisis de haplotipos.....	58
3.4.2.8. Adición secuencial de casos .....	58
3.4.2.9. Meta-análisis .....	59
3.4.2.10. Representación gráfica de las asociaciones .....	59
3.4.2.11. Evaluación de la cobertura de los genes candidatos en los <i>arrays</i> de genotipado empleados en los GWAS de asma.....	59
3.4.2.12. Cálculo de puntuaciones de riesgo genético conjunto .....	60
3.4.2.13. Medidas de la capacidad predictiva de las puntuaciones de riesgo genético.....	60
<b>4. RESULTADOS .....</b>	<b>65</b>
<b>4.1. Estudio de la estratificación poblacional en España originada por una distribución desigual de mezcla genética norteafricana.....</b>	<b>65</b>
<b>4.2. Estudio de la relación entre el trasfondo genético poblacional y la susceptibilidad y gravedad del asma y fenotipos relacionados en la población canaria .....</b>	<b>85</b>
<b>4.3. Asociación de variantes del gen <i>IRAK3</i> con la susceptibilidad al asma en la población española .....</b>	<b>109</b>
<b>4.4. Asociación de variantes de ocho de los genes candidatos más firmemente replicados en asma en la población española.....</b>	<b>141</b>

## ÍNDICE

---

<b>5. DISCUSIÓN .....</b>	<b>185</b>
<b>5.1. Estructura genética de la población española.....</b>	<b>185</b>
<b>5.2. Implicaciones de la estructura genética de la población española en los estudios de asociación de casos-control .....</b>	<b>187</b>
<b>5.3. Análisis de la relación entre la ascendencia norteafricana y el riesgo de asma en Canarias.....</b>	<b>189</b>
<b>5.4. Asociación de genes candidatos con el asma en la población española.....</b>	<b>190</b>
5.4.1. Replicación de la asociación de <i>IRAK3</i> con el asma .....	191
5.4.2. Replicación de la asociación de genes candidatos de susceptibilidad al asma más robustamente replicados .....	192
5.4.3. Efecto de la edad de comienzo del asma .....	195
5.4.4. Cobertura de la variación de los genes candidatos en las plataformas de genotipado en los GWAS de poblaciones europeas.....	196
5.4.5. Fortalezas y limitaciones del estudio de asociación.....	197
5.4.6. Capacidad predictiva de las variantes analizadas y varianza de la enfermedad explicada por la acumulación de variantes de riesgo.....	200
<b>6. CONCLUSIONES .....</b>	<b>205</b>
<b>7. BIBLIOGRAFÍA.....</b>	<b>209</b>

## ***1. INTRODUCCIÓN***

---

## 1. INTRODUCCIÓN

### 1.1. Definición de asma

El asma (OMIM 600807) es una enfermedad respiratoria heterogénea caracterizada por la inflamación crónica de las vías aéreas y asociada con sensibilización alérgica (atopia), eosinofilia, obstrucción del flujo aéreo reversible e hiperreactividad bronquial (Ober, 2005). Por la asociación del asma y la atopia, tanto a nivel individual como familiar, se ha sugerido una etiología común para ambas condiciones (Lester *et al.*, 2001). Por ello, actualmente el diagnóstico del asma está basado en la valoración clínica de las alteraciones funcionales pulmonares y del componente atópico.

### 1.2. Prevalencia mundial del asma

El asma afecta a todos los grupos de edad y es la enfermedad crónica más común durante la niñez. Siguiendo las mismas tendencias que las enfermedades autoinmunes y alérgicas, la prevalencia, morbilidad, mortalidad y carga económica asociadas al asma han aumentado en las dos últimas décadas en los países en desarrollo. Se estima que la prevalencia actual del asma en las distintas poblaciones mundiales oscila entre 1-18%, afectando a unas 300 millones de personas y con previsiones de aumentar en otros 100 millones en el año 2025 (Braman, 2006). La mortalidad asociada a esta enfermedad no supera actualmente el 1%, lo que supone la muerte de unas 250.000 personas cada año, según los datos de la iniciativa global para el asma (GINA, <http://www.ginasthma.org>).

### 1.3. Prevalencia del asma en España

En España, se calcula que existe una prevalencia de asma en torno al 5-6%, aunque se observa una variación considerable entre las distintas comunidades autónomas (5-12%) (The Spanish Group of the European Asthma Study, 1995). Un estudio de la prevalencia de síntomas asmáticos en adultos en Canarias, que empleó un cuestionario estandarizado similar al utilizado en otras regiones (*European Community Respiratory Health Survey*, ECRHS) y en el que participaron más de 7.000 personas, reveló que ésta es de un 17% (Juliá-Serdá *et al.*, 2005). Este valor de prevalencia es notablemente mayor que el de otras comunidades autónomas y que el de la mayoría de Europa (Braman, 2006). Otro estudio en adultos, con una definición de asma más concreta basada en pruebas de función pulmonar, aunque con un tamaño muestral más limitado (593 individuos), mostró una prevalencia de asma en Canarias de un 4,2% (intervalo de confianza [CI] al 95%: 2,5-5,9%) (Juliá-Serdá *et al.*, 2011). En el caso del asma en niños, la prevalencia de síntomas asmáticos en Canarias es del 18,4%, un valor que dobla la media española (9,9%), y que no es alcanzado por

## INTRODUCCIÓN

ninguna otra comunidad autónoma (valor mínimo de 7,1% en Pamplona y máximo en La Coruña con 12,9%) (Tabla 1) (Sánchez-Lerma *et al.*, 2009). De manera similar, la presencia de rinitis en niños canarios, basada en un cuestionario, también superó el valor observado en el resto de España (40,2% frente a un 17,3% de media nacional) (Arnedo-Pena *et al.*, 2005; Sánchez-Lerma *et al.*, 2009). Finalmente, la prevalencia de atopia en Canarias, estudiada mediante pruebas cutáneas de alergia y por medición de inmunoglobulina (Ig) E específica a alérgenos, también fue muy elevada 40,6% (95% CI: 35,9-45,6), principalmente causada por la sensibilización a ácaros (37,1%) (Juliá-Serdá *et al.*, 2011).

**Tabla 1.** Prevalencia de asma en niños de 6-7 años de edad en España.

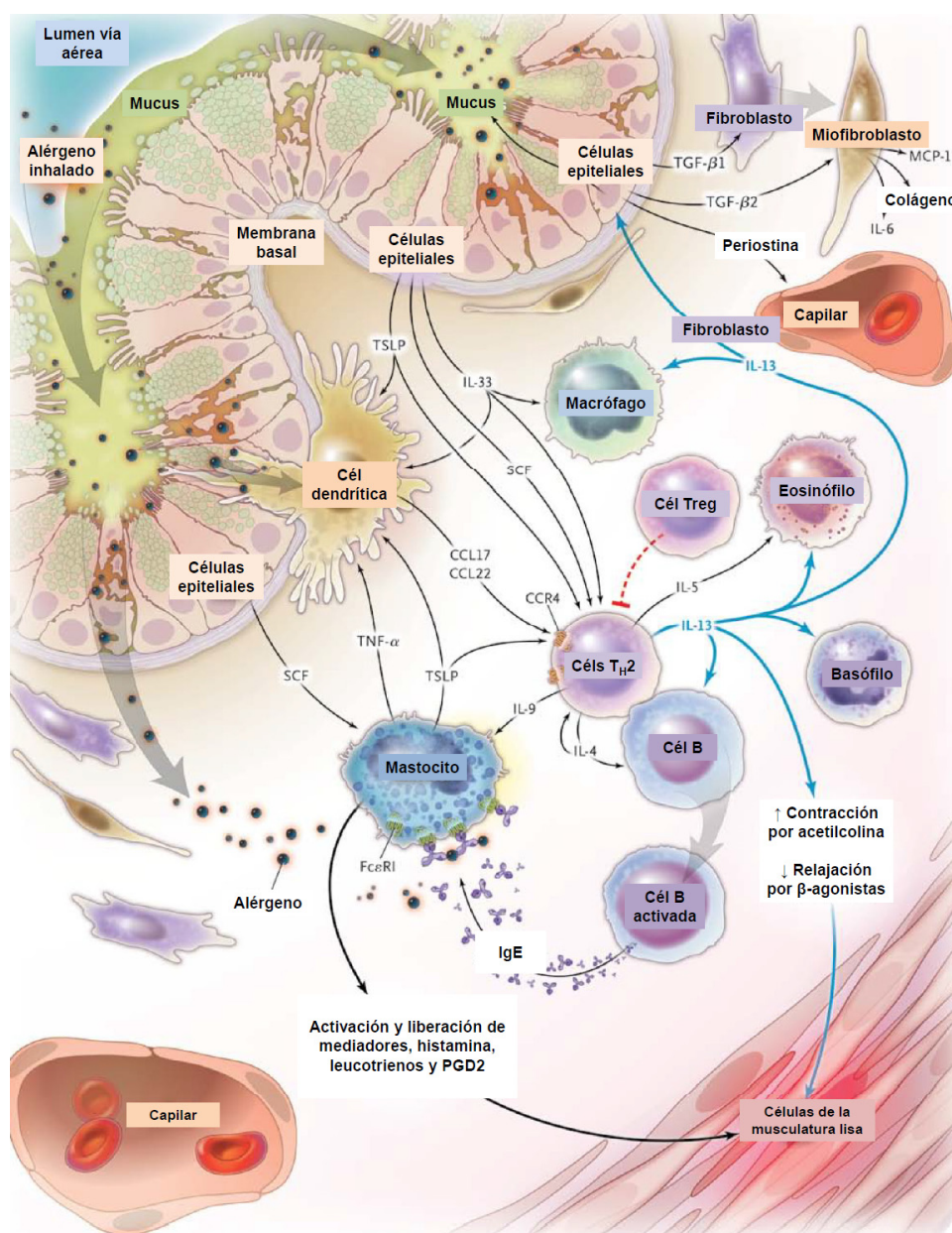
Población	Prevalencia (95% CI)
Media española <sup>a</sup>	9,9 (9,6-10,3)
Pamplona <sup>a</sup>	7,1 (6,2-8,0)
Castellón <sup>a</sup>	8,3 (7,5-9,2)
Barcelona <sup>a</sup>	8,5 (7,5-9,5)
San Sebastián <sup>a</sup>	8,6 (6,9-10,6)
Valencia <sup>a</sup>	9,3 (8,3-10,3)
Madrid <sup>a</sup>	9,4 (8,3-10,7)
Cartagena <sup>a</sup>	11,1 (9,9-12,3)
Asturias <sup>a</sup>	11,5 (10,4-12,7)
Bilbao <sup>a</sup>	12,2 (11,0-13,4)
La Coruña <sup>a</sup>	12,9 (11,7-14,2)
Las Palmas de Gran Canaria <sup>b</sup>	18,4 (16,7-20,2)

<sup>a</sup>Carvajal-Uruena *et al.* (2005); <sup>b</sup>Sánchez-Lerma *et al.* (2009).

### 1.4. Patogenia del asma

El patrón de inflamación que se observa en el asma es similar al de otras enfermedades alérgicas, con activación de mastocitos, eosinófilos, linfocitos T colaboradores de tipo 2 (T<sub>H</sub>2) y células asesinas naturales; y con un perfil de producción de citocinas tipo T<sub>H</sub>2 (IL-3, IL-4, IL-5, IL-9 y IL-13). Sin embargo, también participan en la patogenia del asma otros tipos de células estructurales de la vía aérea, tales como epiteliales, de musculatura lisa, fibroblastos o endoteliales. Todas estas células juegan un papel fundamental en la enfermedad, no sólo como diana, sino como parte activa en el proceso inflamatorio y de reparación de la vía aérea (Cookson, 2004) (Figura 1). Si bien el epitelio se consideró durante mucho tiempo sólo como una barrera física, las evidencias actuales destacan su papel central en la modulación de la actividad de las células dendríticas del pulmón por medio de la liberación de citocinas propias de la inmunidad innata, que a su vez promueven la respuesta de tipo T<sub>H</sub>2 (Hammad & Lambrecht, 2011). Por ello, actualmente se considera que la alteración del epitelio es uno de los estímulos iniciadores del proceso asmático, que contribuye a sus manifestaciones clínicas y a su posterior persistencia (Holgate, 2011).

Las interacciones celulares que hacen posible este proceso inflamatorio se realizan a través de mediadores celulares y de moléculas con funciones muy variadas (quimiocinas, citocinas e IgE). A nivel histológico, es frecuente constatar un engrosamiento de la capa reticular de la membrana basal, fibrosis subepitelial, hipertrofia e hiperplasia de la musculatura lisa bronquial, proliferación y dilatación de los vasos, hiperplasia de las glándulas mucosas e hipersecreción. Estos procesos se asocian con una pérdida progresiva de la función pulmonar que no se previene o no es del todo reversible mediante las terapias actuales. Este fenómeno, conocido como “remodelación”, ocasiona que el paciente responda sólo de manera parcial al tratamiento (Cookson, 2004).



**Figura 1.** Esquema simplificado de los procesos implicados en la patología del asma. Adaptado de Kraft (2011).

### 1.5. Fisiopatología del asma

El hecho fisiológico que desencadena la exacerbación asmática es el estrechamiento de la vía aérea y la subsiguiente obstrucción al flujo aéreo, que de forma característica es reversible (Cookson, 2004). Este fenómeno se produce por contracción del músculo liso bronquial, edema e hipersecreción mucosa. Además, diversos factores desencadenantes pueden ocasionar la exacerbación del asma (alérgenos, tabaco, frío, humedad, agentes infecciosos o contaminantes). La bronco-constricción aguda inducida por alérgenos es consecuencia de la liberación de mediadores por parte de los mastocitos, como la histamina, leucotrienos y prostaglandinas, así como una gran variedad de citocinas. Sin embargo, en algunos pacientes, los anti-inflamatorios no esteroideos pueden también causar una obstrucción aguda de la vía aérea por un mecanismo no dependiente de la IgE (Lemanske & Busse, 2010). Otros estímulos como el ejercicio, el aire frío o irritantes inespecíficos pueden causar obstrucción aguda de la vía aérea, que provocarán una respuesta cuya magnitud dependerá de la inflamación subyacente. Independientemente de los estímulos desencadenantes, una característica típica del asma es la variación o fluctuación de los síntomas y de la función pulmonar en el tiempo, incluso en un mismo día, más allá de los cambios fisiológicos circadianos.

Conforme la enfermedad se hace más persistente y la inflamación progresa, otros factores contribuyen a la limitación del flujo aéreo: el edema de la vía aérea, la hipersecreción de moco y la formación de tapones compuestos por exudados celulares y restos mucosos. Una circunstancia característica de la enfermedad, aunque no exclusiva, es el fenómeno de la hiperrespuesta bronquial. Definida como una “respuesta broncoconstrictora exagerada frente a una gran variedad de estímulos físicos, químicos o biológicos”, la inflamación es un factor fundamental para determinar el grado de hiperrespuesta bronquial, pero no es el único. Además, el grado de hiperrespuesta bronquial se correlaciona parcialmente con la gravedad clínica del asma y con marcadores de la inflamación, aunque no de forma muy estrecha (O'Byrne & Inman, 2003). Influyen también los cambios estructurales y la disfunción neuroreguladora. El tratamiento antiinflamatorio mejora el control del asma y reduce la hiperrespuesta bronquial, pero no la elimina del todo.

### 1.6. Genética del asma

A pesar del efecto notable del ambiente en el desarrollo del asma, son muchas y muy diversas las evidencias que apuntan a la existencia de un componente genético importante en la susceptibilidad a esta enfermedad compleja (Ober & Hoffjan, 2006). Los estudios en gemelos y los patrones de herencia familiar han puesto de manifiesto una influencia

considerable de factores genéticos (Holberg *et al.*, 1999; Thomsen *et al.*, 2010), con una contribución genética estimada en torno a un 60% (Duffy *et al.*, 1990).

### 1.6.1. Análisis de ligamiento

La primera aproximación empleada en el estudio de la genética del asma fueron los análisis de ligamiento. Estos estudios se basan en el análisis de cientos de polimorfismos de tipo microsatélite (pequeñas secuencias de ADN de hasta 6 nucleótidos que se repiten en número variable), en familias multi-generacionales, en busca de alelos marcadores que se transmitan de manera conjunta con la enfermedad en mayor medida que lo esperado por azar. Por su carácter no dirigido, debido a que se escogen polimorfismos distribuidos por todo el genoma, en estos estudios no son necesarias las hipótesis *a priori* sobre qué regiones o procesos estarían implicados en la patología (Ober & Yao, 2011).

Estos análisis han revelado más de una veintena de regiones cromosómicas extensas ligadas al asma y fenotipos relacionados. Muchas de estas señales de ligamiento se han replicado en estudios independientes, incluyendo las regiones cromosómicas 5q, 6p, 11q, 12q, 13q, 16p, 17q y 20p (Wills-Karp & Ewart, 2004). Recientemente, se han realizado dos análisis en los que se combinaron los datos de 11 estudios independientes de ligamiento al asma (Denham *et al.*, 2008) y 20 estudios de ligamiento con fenotipos relacionados (Bouzigon *et al.*, 2010). El primero de ellos incluyó en total 5.832 individuos procedentes de 1.267 familias con asma de ascendencia europea (Denham *et al.*, 2008) y el segundo 10.027 individuos pertenecientes a 3.024 familias con asma o fenotipos relacionados de ascendencia diversa (Bouzigon *et al.*, 2010). Los resultados de estos estudios revelaron varias regiones cromosómicas de ligamiento al asma y fenotipos afines (Tabla 2), algunas de ellas se han confirmado posteriormente como portadoras de genes importantes en la susceptibilidad al asma, como se verá más adelante: 2p12-q22.1 (incluye los genes *DPP10* e *IL1RN*), 6p22.3-21.1 (donde residen los genes de la región del complejo mayor de histocompatibilidad [MHC]) o 17p12-q25 (portadora de los genes *ORMDL3* y *GSDML*).



## INTRODUCCIÓN

**Tabla 2.** Regiones cromosómicas ligadas al asma o fenotipos relacionados mediante el análisis conjunto de varios estudios independientes.

Fenotipo	Región cromosómica
Asma	2p12-q22.1 y 6p22.3-21.1 <sup>a,b</sup>
Hiperrespuesta bronquial	6p22.3-21.1, 2q22.1-q23.3, 7q12.11-q31.1, y 5q23.2-q34 <sup>a,b</sup>
Prueba cutánea de alergia positiva	3p25.3-q24 y 17p12-q25 <sup>a,b</sup>
Resultado cuantitativo de la prueba cutánea de alergia	5q23-q33 <sup>b</sup>
Conteo de eosinófilos	2q32-q34 <sup>b</sup>
Niveles de IgE	5q11.2-q14.3 y 6pter-p22.3 <sup>a</sup>

<sup>a</sup>Denham *et al.* (2008); <sup>b</sup>Bouzigon *et al.* (2010).

Dado que las regiones cromosómicas identificadas mediante los análisis de ligamiento son extensas e incluyen en muchos casos cientos de genes, sólo algunos de ellos de susceptibilidad al asma, se sugirió que el asma representaría una enfermedad poligénica (Wills-Karp & Ewart, 2004). Así, por ejemplo, la región cromosómica situada en 5q es portadora de varios genes asociados con el asma (*IL4*, *IL13*, *CD14*, *ADRB2*, *SPINK5*, *LTC4S*, *CYFIP2*, *CD14*, *ADRB2* y *TIM1*) (Ober & Hoffjan, 2006). Por ello, muchos de los análisis de ligamiento se han acompañado de exploraciones más finas de las regiones cromosómicas identificadas, seleccionando marcadores genéticos dentro de las mismas, una estrategia denominada clonaje posicional, lo que ha permitido descubrir nueve genes de susceptibilidad al asma (Tabla 3).

**Tabla 3.** Genes de susceptibilidad al asma identificados a través de clonaje posicional.

Gen	Posición cromosómica	Estudio
<i>ADAM33</i>	20p13	Van Eerdewegh <i>et al.</i> , 2002
<i>PHF11</i>	13q14	Zhang <i>et al.</i> , 2003
<i>DPP10</i>	2q14.1	Allen <i>et al.</i> , 2003
<i>NPSR1 (GPR154)</i>	7p14.3	Laitinen <i>et al.</i> , 2004
<i>HLAG</i>	6p21.3	Nicolae <i>et al.</i> , 2005
<i>CYFIP2</i>	5q33.3	Noguchi <i>et al.</i> , 2005
<i>IRAK3 (IRAKM)</i>	12q14.3	Balaci <i>et al.</i> , 2007
<i>COL29A1</i>	3q22.1	Soderhall <i>et al.</i> , 2007
<i>OPN3/CHML</i>	1 qter	White <i>et al.</i> , 2008

Los análisis de ligamiento constituyen una herramienta prometedora en la identificación de nuevos genes y procesos biológicos relacionados con las enfermedades complejas. Sin embargo, su poder estadístico para detectar variantes de riesgo con efectos

leves es reducido. Precisamente en el asma, como en la mayoría de enfermedades complejas, se espera que las variantes genéticas con efectos leves expliquen una gran proporción de la enfermedad.

### **1.6.2. Estudios de asociación en genes candidatos**

Para un análisis más fino de las regiones de susceptibilidad al asma, ha sido necesario seleccionar genes candidatos y estudiar si el polimorfismo de los mismos se relaciona con la enfermedad en estudios de asociación de tipo caso-control. En contraposición a los análisis de ligamiento, ante un mismo tamaño de muestra, los estudios de tipo caso-control permiten detectar asociación entre genes con efectos leves y la enfermedad, y además no requieren el estudio de individuos procedentes de una misma familia.

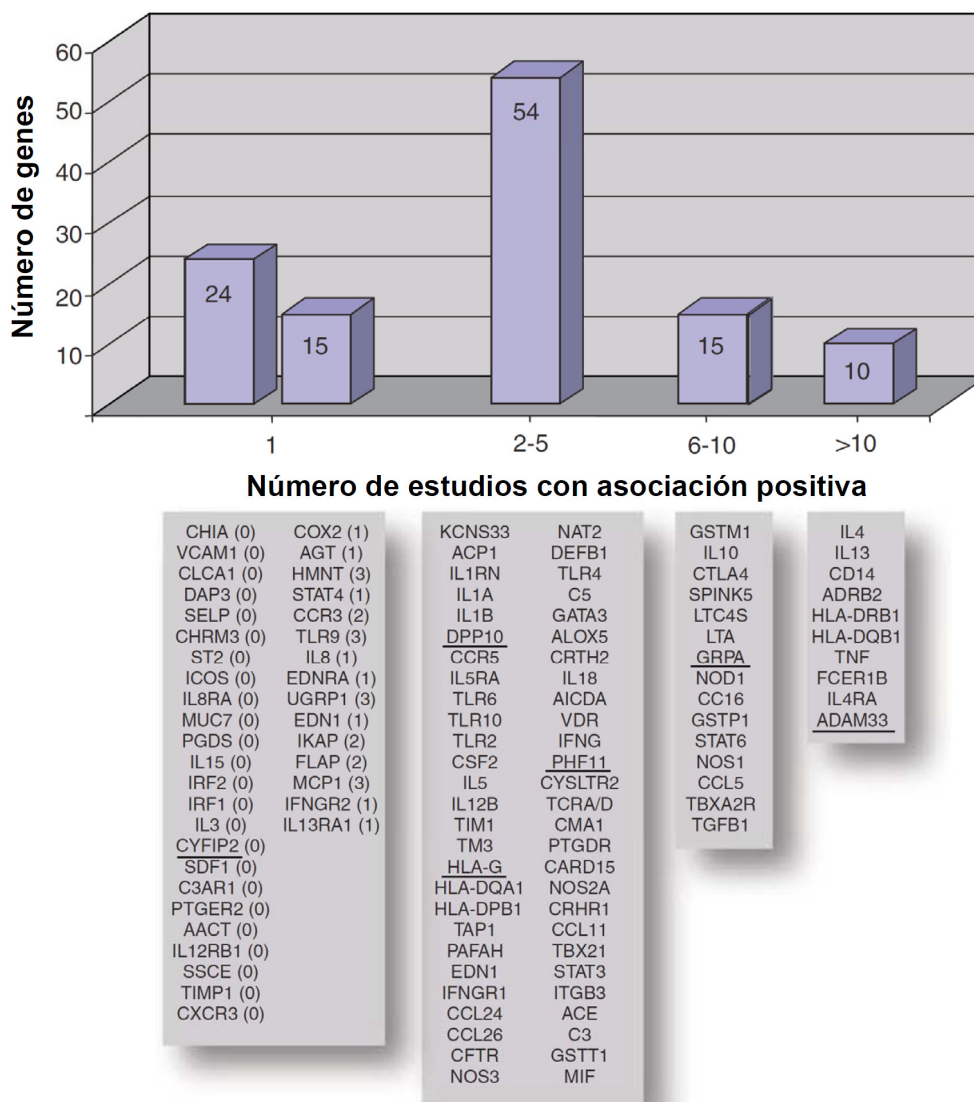
En los estudios de asociación de tipo caso-control se comparan muestras de individuos sin relación familiar pertenecientes a dos grupos, casos (individuos enfermos) y controles (individuos sin la enfermedad), en busca de diferencias estadísticamente significativas en la frecuencia de determinadas variantes situadas en uno o varios sitios del genoma. Comúnmente, las variantes génicas analizadas en estos estudios son los polimorfismos de un solo nucleótido ó SNPs (del inglés *single nucleotide polymorphism*), variantes en la secuencia de ADN que afectan a un solo nucleótido y que generalmente presentan sólo dos alelos (*loci* bialélicos). A diferencia de las mutaciones, cuya frecuencia es muy baja en la población general (<<1%), aunque con efectos drásticos en la enfermedad, la frecuencia del alelo menos común (MAF) de los SNPs es de al menos un 1% en la población y sus efectos en la función génica son nulos en la mayoría de casos. Sin embargo, como ya se ha demostrado en otras enfermedades y caracteres complejos, algunas de estas variantes pueden afectar a la función de los distintos elementos que componen los genes, pese a que en muchos casos se desconoce su efecto.

Se han publicado más de 200 estudios de asociación de tipo caso-control en asma y se han asociado más de 120 genes candidatos con la susceptibilidad a esta enfermedad o con alguno de los fenotipos relacionados en al menos un estudio (Ober & Hoffjan, 2006). No obstante, los efectos de estos genes en el asma resultaron leves (Bosse & Hudson, 2007), e incluso algunos de ellos mostraron inconsistencias entre estudios, de manera que ninguno de ellos puede ser considerado como un factor de riesgo indiscutible en esta enfermedad. Considerando la replicación de los resultados como un criterio de autenticidad de los mismos (Chanock *et al.*, 2007), menos de una treintena de asociaciones han mostrado efectos reproducibles en distintos estudios y poblaciones.

La complejidad inherente al asma y las limitaciones metodológicas podrían explicar tales inconsistencias. En la bibliografía son frecuentes los estudios con tamaños de muestra reducidos, diseños inadecuados, análisis sin los ajustes ni la estimación de efectos pertinentes, normalmente enfocados a un número de polimorfismos insuficiente para estudiar la variación total del gen analizado (1 ó 2 por gen), e ignorando las correlaciones o desequilibrio de ligamiento (LD) subyacentes en el mismo (Ioannidis *et al.*, 2006). Además, no se ha prestado atención a la presencia o no de subgrupos de población en las muestras de casos y controles (comúnmente denominado estratificación poblacional), lo que puede haber provocado que algunas de las asociaciones publicadas sean espurias. En este sentido, muchos estudios han demostrado que las estimas de ascendencia basadas en polimorfismos genéticos son capaces de proveer medidas adecuadas del fondo genético ancestral (Barnholtz-Sloan *et al.*, 2005), permitiendo realizar ajustes estadísticos por la presencia de estratificación poblacional.

### **1.6.3. Genes candidatos más firmemente replicados en asma**

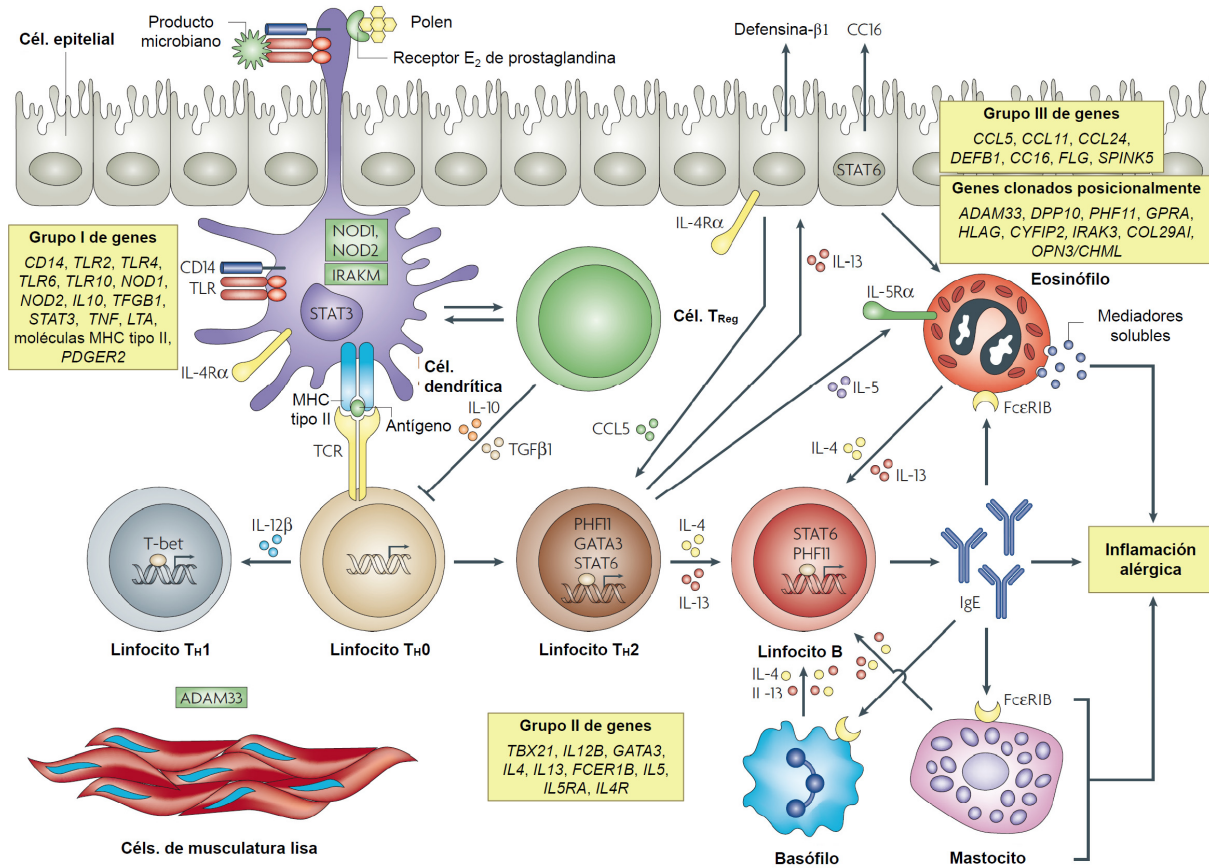
De los más de 200 artículos que se han publicado estudiando la genética del asma, varias revisiones han señalado que existen 10 genes cuya asociación con el asma o fenotipos relacionados se ha replicado en más de diez estudios independientes (Ober & Hoffjan, 2006; Vercelli, 2008) (Figura 2): los genes de la interleucinas 4 y 13 (*IL4* y *IL13*, en 5q23), el gen del antígeno CD14 (*CD14*, en 5q31), el gen del receptor beta 2 adrenérgico (*ADRB2*, en 5q32), los genes del MHC tipo II *HLA-DRB1* y *HLA-DQB1* (6p21), el gen del factor de necrosis tumoral (*TNF*, en 6p21), el gen codificante de la cadena beta del receptor del factor cristalizable de la IgE (*MS4A2*, también conocido como *FCER1B*, en 11q13), el gen del receptor de la interleucina 4 (*IL4RA*, alias *IL4R*, en 16p12) y el gen de la metalopeptidasa 33 (*ADAM33*, en 20p13). Puesto que la replicación de los resultados de los estudios de asociación gen-enfermedad en muestras independientes es en la actualidad la única herramienta de validación para demostrar la relevancia de un gen candidato en el desarrollo de enfermedades complejas (Chanock *et al.*, 2007), estos 10 genes han sido propuestos como firmes candidatos de susceptibilidad al asma (Ober & Hoffjan, 2006), aunque no constituyan factores de riesgo en todas las poblaciones.



**Figura 2.** Genes asociados con asma o fenotipos atópicos en al menos un estudio. Los genes que se han asociado en una sola población se encuentran separados entre aquellos sin estudios de replicación publicados y los que no han sido replicados en estudios posteriores. Los genes identificados por clonaje posicional están subrayados. Adaptado de Ober & Hoffjan (2006).

Todos ellos son genes claves en el asma, ya sea por las evidencias de ligamiento, por su función biológica o por resultados de clonaje posicional (Figura 3). Estos genes se agrupan por las funciones que desempeñan en: genes de inmunidad innata (grupo I), de diferenciación de linfocitos T<sub>H</sub>2 y efectores de esta línea linfocítica (grupo II) y genes expresados por las células epiteliales (grupo III) (Vercelli, 2008).

## INTRODUCCIÓN



**Figura 3.** Representación de la función de algunos de los genes más robustamente asociados con la susceptibilidad al asma. Adaptado de Vercelli (2008).

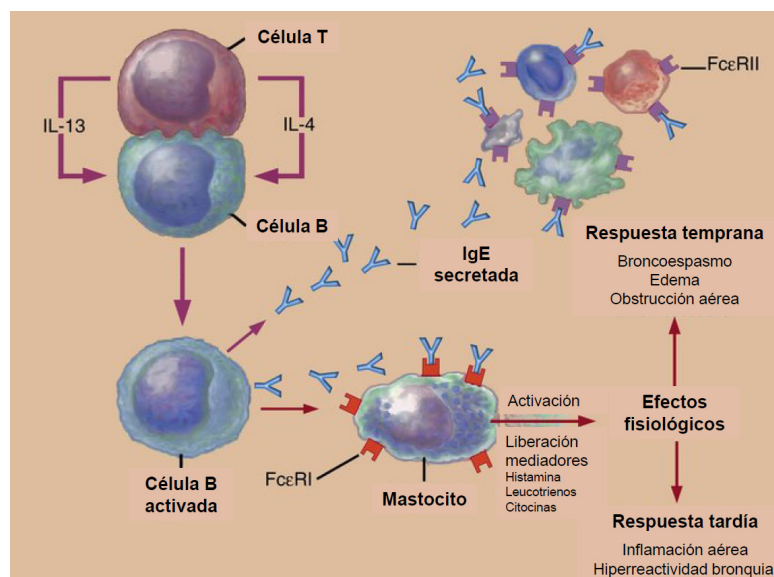
En esta tesis nos centraremos en el estudio de los genes que han sido propuestos como firmes candidatos de susceptibilidad al asma (*IL4*, *IL13*, *IL4R*, *TNF*, *CD14*, *ADRB2*, *MS4A2*, *ADAM33*), con la excepción de los genes de MHC tipo II, situados en una región altamente polimórfica. La variación genética de la región MHC de tipo II supera la encontrada en cualquier otro *locus*, ya que contiene un gran número de alelos de baja frecuencia. Por ello, su caracterización alélica se realiza habitualmente para los alelos clásicos, empleando sondas específicas de alelo, cuyos costes económicos son tan elevados como el estudio de medio millón de SNPs de todo el genoma con plataformas comerciales (Leslie *et al.*, 2008). Además de estos ocho genes, el gen *TNF* se estudiará conjuntamente con el gen de la linfotóxina  $\alpha$  (*LTA*), como una única región denominada *LTA-TNF* dado que: i) el gen *TNF* se localiza muy próximo al gen codificante de la *LTA*, a tan sólo 1.252 pares de bases (pb) en la región MHC tipo III, y ambos genes muestran un elevado LD (Belfer *et al.*, 2004), de manera que en muchos estudios se han asociado haplotipos que abarcan toda la región *LTA-TNF* con el asma (Migita *et al.*, 2005; Randolph *et al.*, 2005), ii) ambos pertenecen a la superfamilia de ligandos TNF, se unen a los mismos receptores y

median efectos pleiotrópicos similares (Nedwin *et al.*, 1985; Aggarwal, 2003) y iii) el gen de la *LTA* ha sido asociado con asma en más de seis estudios independientes (Figura 2).

#### 1.6.3.1. Gen de la interleucina 4 (IL4)

El gen *IL4* (número NCBI: 3565) consta de 4 exones y abarca 9 kilobases (kb) en la región 5q31.1. En esta misma región, y a poca distancia, están codificadas otras citocinas (*IL5* e *IL13*) y factores de crecimiento hematopoyéticos (Le Beau *et al.*, 1989). El gen *IL4* codifica para la citocina IL-4, de 153 aminoácidos, que es el principal estímulo para la síntesis de IgE al inducir el cambio de isotipo de las células B desde IgG (Gauchat *et al.*, 1990) (Figura 4), y de la diferenciación de los linfocitos T colaboradores desde  $T_H0$  a  $T_H2$  (Messi *et al.*, 2003; Ghoreschi *et al.*, 2003). Además, la IL-4 participa en la proliferación y diferenciación de las células B y tiene actividad como factor de crecimiento en células T y mastocitos. La IL-4 es la citocina característica de la subpoblación  $T_H2$  y actúa como inductora y efectora de estas células (Fields *et al.*, 2002).

Los primeros estudios que mostraron ligamiento del gen *IL4* con un fenotipo atópico se realizaron en relación con los niveles séricos de IgE (Kawashima *et al.*, 1998; Marsh *et al.*, 1994).



**Figura 4.** Representación de la acción de las citocinas IL-4 e IL-13 liberadas por células  $T_H2$  en la activación de la producción de IgE por parte de células B. Adaptado de Pernis & Rothman (2002).

#### 1.6.3.2. Gen de la interleucina 13 (IL13)

El gen *IL13* (número NCBI: 3596) consta de 4 exones y abarca 2,9 kb en la región 5q31.1. Este gen codifica la citocina interleucina 13, que es principalmente producida por linfocitos  $T_H2$ , aunque también por mastocitos, basófilos y eosinófilos (Schmid-Grendelmeier *et al.*, 2002). La actividad de la IL-13 es independiente de la IL-4, pese a que comparten

algunas vías de señalización (LaPorte *et al.*, 2008). Sin embargo, al igual que la IL-4, presenta un papel fundamental como efector de la vía  $T_H2$ , induce la expresión de CD23 en células B, aumenta la IgM de superficie y la expresión de moléculas de antígeno MHC II e induce la transcripción del gen de la cadena pesada de la IgE en células B (Figura 4) (Minty *et al.*, 1993; Punnonen *et al.*, 1993).

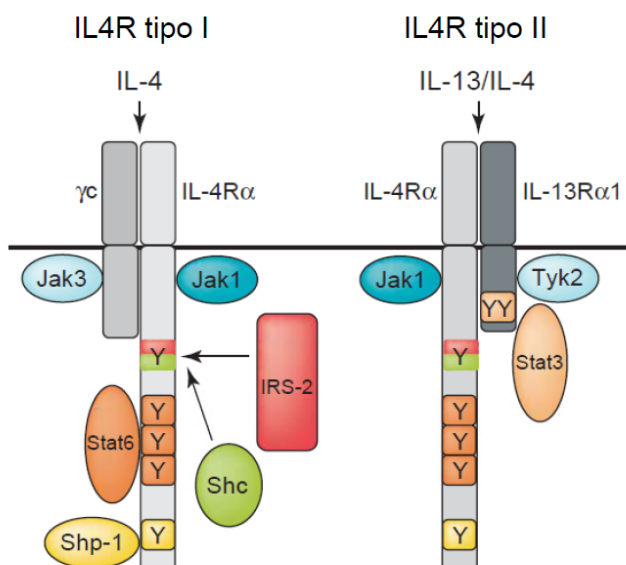
Tanto la IL-13 como la IL-4 interaccionan con elevada afinidad con la cadena  $\alpha$  del receptor de la IL-4 (IL4R) (Kelly-Welch *et al.*, 2003). Además, la IL-13 puede unirse a otros dos receptores, IL13RA1 o IL13RA2, este último actúa como receptor señuelo, dado que su unión no se traduce en ninguna señal intracelular. Las colas C-terminales de la IL-4 y la IL-13 interaccionan con tirosinas quinasas de la familia Janus, conduciendo a la interacción con STAT6, el cual se une a secuencias consenso en los genes de ambas citocinas que son reguladoras de su expresión génica (Kotanides & Reich, 1996).

En modelos animales se ha comprobado que la IL-13 produce respuestas  $T_H2$  en los pulmones (hiperrespuesta bronquial, infiltración de células inflamatorias, hipersecreción de moco y fibrosis de la vía aérea) (Zhu *et al.*, 1999). Además, se ha demostrado que la IL-13 y sus receptores se encuentran expresados en gran medida en el tracto respiratorio de asmáticos y pacientes con rinitis, lo que unido a su localización cromosómica en una región de ligamiento (Denham *et al.*, 2008; Bouzigon *et al.*, 2010), llevó a postularlo como un gen candidato en el asma. Actualmente, es uno de los pocos genes candidatos que ha mostrado asociación en estudios de asociación genómicos (GWAS) con asma, superando los estrictos niveles de significación impuestos en este tipo de análisis (Li *et al.*, 2010a; Moffatt *et al.*, 2010).

### 1.6.3.3. Gen del receptor de la interleucina 4 (IL4R)

El gen *IL4R* (número NCBI: 3566) consta de 12 exones y se extiende 51 kb en la región 16p12.1. Este gen codifica para la cadena  $\alpha$  del receptor de la IL-4.

El receptor de la IL-4 de células linfocíticas está formado por dos subunidades: la subunidad  $\alpha$ , que se une a la IL-4 y transduce sus funciones promotoras del crecimiento y activadoras de la transcripción, y una subunidad  $\gamma_c$ , que es común a muchos de los receptores de citocinas, formando un receptor de tipo I (Zurawski *et al.*, 1993) (Figura 5). La señalización de este receptor se lleva a cabo por la vía JAK/STAT (JAK3 ó 4/STAT6) y por una vía en la que participa el sustrato de respuesta a la insulina (IRS-2). En líneas no linfocíticas, la cadena alfa de este receptor no se asocia con  $\gamma_c$ , sino que forma un receptor de tipo II al unirse a una cadena de transducción de señales que también forma parte del receptor de la IL-13 (Figura 5) (Kelly-Welch *et al.*, 2003).



**Figura 5.** Representación de los distintos tipos de receptores IL4R: a la izquierda receptor de tipo I con la subunidad  $\alpha$  asociada a una subunidad  $\gamma_c$ , a la derecha el receptor de tipo II unido a una cadena del receptor de la IL-13. Tomado de Chatila (2004).

El dominio extracelular de este receptor está codificado por los exones 3 al 7 del gen, el dominio transmembrana en el exón 9, y los dominios intracelulares se codifican en los exones del 10 al 12 (Kruse *et al.*, 1999a). Además, existen formas alternativas de procesamiento del ARN de *IL4R* que producen una forma soluble del receptor, que es codificado por los exones 3 al 8 y carece de exones para el dominio transmembrana y las regiones intracelulares. La forma soluble de la IL4R no tiene capacidad de señalización, por lo que se ha postulado que podría funcionar como un antagonista para la unión del ligando IL-4 (Bergin *et al.*, 2006).

En 1997 se evaluó por primera vez su posible relación con la atopia (Hershey *et al.*, 1997). Posteriormente, un análisis genómico en busca de alelos de susceptibilidad en una población endogámica de origen europeo encontró ligamiento en la región cromosómica 16p (Ober *et al.*, 1999). Posteriormente, Ober *et al.* (2000b) examinaron el gen *IL4R* como *locus* candidato de susceptibilidad en dicha región cromosómica y encontraron asociación con asma, no sólo en esta población aislada, sino también en poblaciones panmíticas (Ober *et al.*, 2000a).

#### 1.6.3.4. Genes del factor de necrosis tumoral (TNF) y la linfotoxina $\alpha$ (LTA)

El gen *TNF* (número NCBI: 7124) presenta 4 exones y abarca 2,8 kb de la región cromosómica 6p21.3. Este gen codifica para el factor de necrosis tumoral, una citocina proinflamatoria multifuncional de 233 aminoácidos, secretada predominantemente por



## INTRODUCCIÓN

---

monocitos y macrófagos. El TNF presenta efectos en el metabolismo de los lípidos, coagulación, resistencia a la insulina y en la función endotelial (Aggarwal, 2003).

El TNF fue originalmente identificado en el suero de ratones tras una inyección de *Mycobacterium bovis* y endotoxina. El suero obtenido de estos animales resultó citotóxico y citostático en varias líneas celulares humanas y murinas, produjo hemorragias necróticas y en algunos casos la regresión de ciertos tumores transplantados en ratones (Shirai *et al.*, 1985). El TNF es una citocina que media la inmunidad innata y es el principal mediador de la respuesta inflamatoria aguda frente a bacterias gram negativas y otros microorganismos infecciosos, y es responsable de muchas de las complicaciones sistémicas de las infecciones graves.

El TNF es sintetizado como pro-TNF, una proteína unida a membrana de 26 kilodaltons (kD) y que posteriormente es escindida por enzimas procesadoras (como la metalopeptidasa ADAM17) para liberar su forma soluble de 17 kD (Black *et al.*, 1997). Esta forma soluble se une a sus dos receptores principales TNFRI y TNFRII, con distintas consecuencias en su unión (activación de caspasas y apoptosis o activación de NF- $\kappa$ B e inflamación) (Aggarwal, 2003).

Distintas variantes del gen *TNF* han sido implicadas en la susceptibilidad de varias enfermedades: sepsis, esclerosis múltiple, lupus eritematoso sistémico y enfermedad inflamatoria intestinal (Teuffel *et al.*, 2010; Clancy *et al.*, 2010; Yang *et al.*, 2011; Ng *et al.*, 2011). En el caso del asma, su localización cromosómica y su función han llevado a su análisis como gen candidato de susceptibilidad en numerosos estudios (Ober & Hoffjan, 2006). Sin embargo, su localización dentro del MHC, una región altamente polimórfica que contiene numerosos genes implicados en las respuestas inmunes, complica el esclarecimiento de su implicación en las distintas enfermedades.

Como se comentó con anterioridad, el gen *LTA* se estudiará en asociación con asma junto con el *TNF*. Este gen (también conocido como factor de necrosis tumoral beta; número NCBI: 4049) presenta 4 exones y abarca 2,2 kb de la región cromosómica 6p21.3. El gen *LTA* codifica una proteína homotrimérica soluble que es secretada por linfocitos activados y actúa como modulador de la respuesta inmune. Esta forma homotrimérica comparte receptor con el TNF, y se une tanto al receptor TNFR-1 como al receptor TNFR-2 (Aggarwal, 2003).

La linfotóxina  $\alpha$  se caracterizó por primera vez como un factor biológico en linfocitos estimulados por mitógenos y se observó que presentaba una actividad anticelular en líneas celulares neoplásicas. Gray *et al.* (1984) aislaron el ADN complementario (ADNc) humano codificante para la linfotóxina  $\alpha$  y lo expresaron en *E.coli*. Además, observaron efectos citotóxicos y necróticos en líneas celulares murinas y humanas y en sarcomas murinos *in*

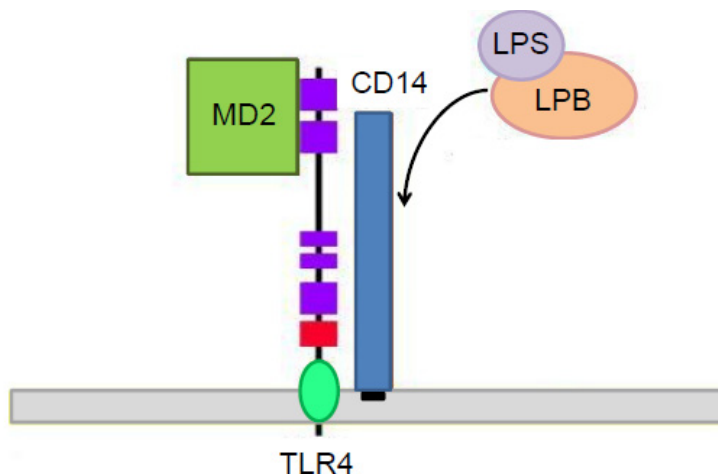
vivo. El gen de *LTA* es homólogo al *TNF* y las citocinas codificadas por ambos presentan un 50% de identidad en su secuencia de aminoácidos (Nedwin *et al.*, 1985).

La linfotoxina  $\alpha$  es producida principalmente por linfocitos T activados (Cuturi *et al.*, 1987), en forma de proteína secretada altamente inducible que media una gran variedad de procesos inflamatorios, inmunoestimulatorios y de respuestas antivirales, y se encuentra implicada en la formación de los órganos linfoides secundarios y en la apoptosis (Selmaj *et al.*, 1991; Matsumoto *et al.*, 1996; Drayton *et al.*, 2002). Algunas variantes genéticas de este gen se han asociado con la susceptibilidad a la lepra y a padecer infartos de miocardio (Ozaki & Tanaka, 2005; Alcais *et al.*, 2007), además de la predisposición al asma (Ober & Hoffjan, 2006).

#### 1.6.3.5. Gen del antígeno CD14 (CD14)

El gen del antígeno de diferenciación monocítica *CD14* (número NCBI: 929) se localiza en la región cromosómica 5q31.3, consta de 3 exones y abarca 2 kb. Este gen codifica dos versiones de la proteína CD14, una proteína anclada a la membrana por medio de glucofosfatidilinositol (mCD14) y una forma plasmática soluble (sCD14). Ambas moléculas son críticas para la transducción de señales dependientes del reconocimiento de lipopolisacárido bacteriano (LPS). La sCD14 confiere sensibilidad al LPS en células carentes de mCD14. Asimismo, los niveles de sCD14 se asocian con enfermedades infecciosas y con una elevada mortalidad debida a shock por bacterias gram negativas (LeVan *et al.*, 2001).

Los receptores de tipo Toll (TLRs) son una familia de proteínas de membrana que actúan como receptores de reconocimiento de patrones moleculares asociados a patógenos (PAMP), para así activar una respuesta inmunitaria innata frente al microorganismo reconocido. En el reconocimiento del LPS bacteriano, éste no se produce directamente por un TLR, sino que la molécula de LPS interacciona en primer lugar con la proteína de unión a lipopolisacárido (LBP) y luego con CD14, para presentar así LPS al miembro 4 de la familia de TLRs (TLR4) (Figura 6), el cual induce la expresión de genes inflamatorios a través del factor nuclear NF- $\kappa$ B y la vía de señalización de las proteínas cinasas activadas por mitógenos (MAPK, de las siglas en inglés *Mitogen-activated protein kinases*) (Wright *et al.*, 1990). La proteína CD14 también actúa iniciando un proceso inflamatorio en respuesta a peptidoglicanos de bacterias gram positivas y en la activación de TLRs que reconocen ácidos nucleicos bacterianos (Gupta *et al.*, 1996). Además, CD14 es importante en el reconocimiento de células apoptóticas para su eliminación por fagocitosis, sin promover la liberación de citocinas proinflamatorias (Devitt *et al.*, 1998).



**Figura 6.** Actuación de CD14 en el reconocimiento de LPS, facilitado por la unión a la LPB.

En 1999 una variante del promotor del gen *CD14* se relacionó por primera vez con fenotipos atópicos y con los niveles circulantes de sCD14 (Baldini *et al.*, 1999). La asociación de este gen con susceptibilidad al asma ha presentado una modulación ambiental (por ejemplo, dependiente de la posesión de animales o de la exposición a endotoxina bacteriana). Esto podría explicar que un mismo polimorfismo del gen haya mostrado efectos completamente opuestos, incluso en una misma población (Vercelli, 2008).

### 1.6.3.6. Gen del receptor beta adrenérgico (*ADRB2*)

El gen *ADRB2* (número NCBI: 154) abarca 2 kb en la región cromosómica 5q31-q32 y carece de intrones. Este gen codifica para el receptor beta adrenérgico, una proteína de 413 aminoácidos, con 7 regiones de aminoácidos hidrofóbicos que sugieren la presencia de dominios transmembrana. Tanto estas regiones como los lazos citoplasmáticos presentan una gran identidad con la proteína homóloga en hámsteres (95 y 93%, respectivamente), lo que sugiere que se trata de dominios funcionales importantes (Kobilka *et al.*, 1987).

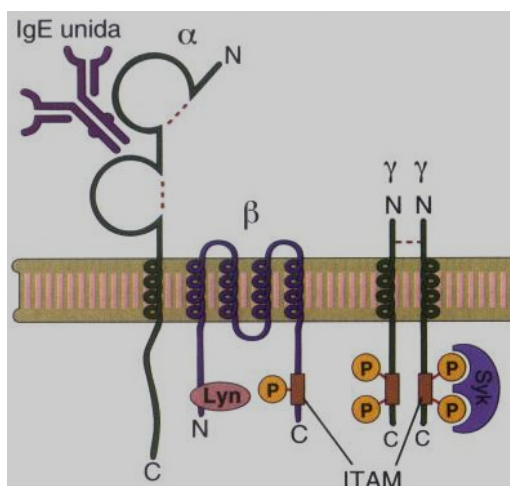
Un estudio reveló la asociación de una variante del gen *ADRB2* con el asma nocturno (Reihnsaus *et al.*, 1993). Los pacientes con asma nocturno representan un subconjunto de asmáticos que experimentan un marcado empeoramiento de la obstrucción de las vías aéreas y de otros síntomas mientras están durmiendo. Szeffler *et al.* (1991) encontraron que los neutrófilos circulantes y los receptores beta 2 adrenérgicos en los linfocitos disminuyen durante la madrugada en pacientes con asma nocturno, algo que no ocurre en el resto de asmáticos. Además, dado que el uso de agonistas del receptor beta adrenérgico es uno de los tratamientos más empleados en el asma según las guías GINA (<http://www.ginasthma.org>), varios estudios de asociación han analizado la relación entre variantes de este gen y la respuesta a tratamiento (Lee *et al.*, 2004; Finkelstein *et al.*, 2009).

Por tanto, la asociación de *ADRB2* con asma parece ser dependiente del fenotipo considerado.

#### 1.6.3.7. Gen de la cadena beta del receptor de la IgE (*MS4A2*)

El gen *MS4A2* (del inglés, *membrane-spanning 4-domains, subfamily A, member 2*, frecuentemente denominado *FCER1B*; número NCBI: 2206) consta de 7 exones y abarca unas 10 kb de la región cromosómica 11q12-q13. Este gen codifica la cadena  $\beta$  del receptor de alta afinidad de la IgE, una proteína de 28 kD.

La respuesta alérgica implica la unión de un alérgeno al complejo formado por la IgE unida a su receptor, presente en la membrana de mastocitos y basófilos. Como consecuencia de este reconocimiento, se produce la activación de las células portadoras del receptor, lo que a su vez promueve la liberación de compuestos responsables de las manifestaciones alérgicas, como la histamina o citocinas  $T_H2$  (Vercelli, 2008). La estructura del receptor de la IgE se ha modelizado como un tetrámero compuesto por una cadena  $\alpha$ , una cadena  $\beta$  y dos cadenas  $\gamma$  unidas por puentes disulfuro (Figura 7) (Kinet, 1999).



**Figura 7.** Estructura del receptor de la IgE, formado por una cadena  $\alpha$ , una cadena  $\beta$  (codificada por el gen *MS4A2*) y dos cadenas  $\gamma$  unidas por puentes disulfuro.

En experimentos de co-transfección en células de las cadenas  $\alpha$  y  $\gamma$ , se mostró la capacidad de constituir complejos triméricos formados sólo por estas dos cadenas, lo que llevó a sugerir que la cadena  $\beta$  sería prescindible. Además, se comprobó que los complejos triméricos se expresaban en células de Langerhans, células dendríticas y monocitos, lo que llevó a relacionarlos con la presentación de antígenos. Por su parte, el tetrámero se expresa en mastocitos y basófilos, lo que los relaciona con la liberación de mediadores de alergia. La presencia de la cadena  $\beta$ , que actúa como chaperona para el transporte del complejo FCER1 a la superficie celular, amplificaría la expresión del complejo en la membrana y la capacidad de señalización (Donnadieu *et al.*, 2003).

Sandford *et al.* (1993) demostraron que el gen *MS4A2* estaba ligado a la atopía, aunque sólo cuando eran tenidos en cuenta los alelos derivados por vía materna (los paternos no mostraron ligamiento). El papel funcional del receptor de alta afinidad de la IgE, unido a su localización en la región 11q13, hacen de éste un gen candidato en atopía. Además, Folster-Holst *et al.* (1998) también encontraron evidencias de ligamiento en familias con dermatitis atópica con un microsátelite próximo a este gen. Posteriormente, se ha evaluado su posible implicación en el asma en numerosos estudios dirigidos (Ober & Hoffjan, 2006; Vercelli, 2008).

### 1.6.3.8. Gen de la metalopeptidasa 33 (ADAM33)

El gen *ADAM33* (ADAM metalopeptidasa dominio 33; número NCBI: 80332) codifica la proteína 33 de la familia de desintegrinas y proteasas, consta de 22 exones y engloba unas 14 kb dentro de la región cromosómica 20p13. Yoshinaka *et al.* (2002) aislaron por primera vez el cDNA codificante para *ADAM33*. La proteína humana, de 813 aminoácidos, presenta un 70% de identidad con su homóloga murina. Esta proteína se encuentra anclada a la membrana y posee un dominio metalopeptidasa. El dominio metalopeptidasa actúa como mediador de las interacciones célula-célula y de las células con la matriz extracelular por su papel proteolítico, lo que facilita la liberación de proteínas de superficie (tales como citocinas y receptores) (White, 2003). Además, la región desintegrina modula la adhesión celular mediada por integrinas. Estas proteínas se han implicado en procesos biológicos tales como la fertilización, el desarrollo muscular y la neurogénesis (White, 2003).

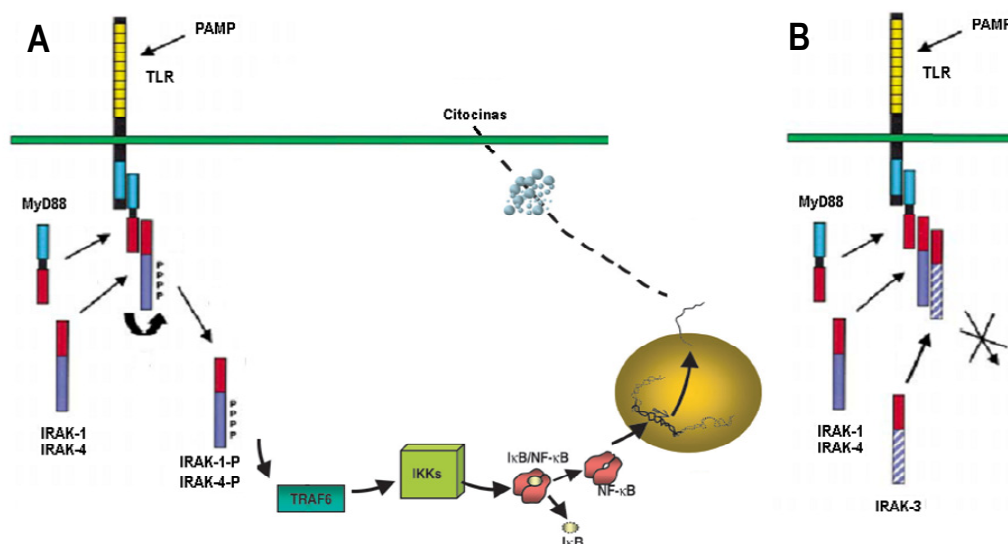
Se ha sugerido que el papel de la proteína *ADAM33* en el asma estaría relacionado con el modelado de las vías aéreas en estadios prenatales y en el posterior remodelado en etapas posteriores de la vida (Zou *et al.*, 2004). La proteína codificada por *ADAM33* se expresa en fibroblastos y en células de la musculatura lisa bronquial de las vías aéreas pero, a diferencia de la mayoría de genes asociados con asma, no se expresa en células epiteliales o inmunes. Además, los fibroblastos expresan seis formas alternativas de *ADAM33* mediante procesamiento del ARN, todas ellas carentes del dominio proteolítico (Powell *et al.*, 2004).

*ADAM33* fue el primer gen clonado posicionalmente en asma (Van Eerdewegh *et al.*, 2002). Mediante un análisis de ligamiento en 460 familias caucásicas, los autores detectaron una señal de ligamiento con asma e hiperrespuesta bronquial en la región cromosómica 20p, para después proponer a *ADAM33* como gen candidato de la región (Van Eerdewegh *et al.*, 2002). Tras este primer hallazgo, muchos SNPs y haplotipos distintos han sido asociados con asma en diferentes estudios (Ober & Hoffjan, 2006).

#### 1.6.4. *IRAK3* como gen candidato de susceptibilidad en asma

Además de las ocho regiones génicas candidatas más firmemente replicadas, en esta tesis se estudiará la asociación del gen *IRAK3* (kinasa 3 asociada al receptor de IL-1; número NCBI: 11213) con la susceptibilidad al asma. El gen *IRAK3* consta de 12 exones y abarca aproximadamente 50 kb de la región cromosómica 12q14.3. Este gen se detectó por primera vez en una librería de ADNc de leucocitos de sangre periférica, prediciendo una proteína de 596 aminoácidos con un 30-40% de identidad con IRAK-1 e IRAK-2, y compartiendo con ellas un dominio de muerte N-terminal y un dominio kinasa central (Wesche *et al.*, 1999). Mediante análisis de *Northern blot* se detectaron transcritos de *IRAK3* en células monocíticas, lo que llevó a su denominación inicial como *IRAKM*. En la actualidad este nombre se conserva, pese que se ha detectado la expresión de IRAK-3 en otros tipos celulares, tales como células epiteliales de bronquios y alveolos (Balaci *et al.*, 2007; Villar *et al.*, 2010).

La proteína IRAK-3 pertenece a la familia de kinasas asociadas al receptor de la IL-1, que participan en la vía de señalización de los TLRs, así como en la señalización de la IL-1. IRAK-3 podría estar implicada en la tolerancia al LPS y se le ha atribuido una actividad como regulador negativo de la vía de señalización de los TLRs (Kobayashi *et al.*, 2002). En ausencia de IRAK-3, el reconocimiento de los PAMP por parte de los TLRs conduce al reclutamiento de una serie de proteínas adaptadoras que a su vez facilitan la unión de IRAK-1 o IRAK-4, las cuales tras fosforilarse producen la traslocación del factor de transcripción NF- $\kappa$ B y, en última instancia, la transcripción de genes de citocinas inflamatorias (Cohen, 2002). En este proceso, IRAK-3 podría impedir la fosforilación de IRAK-1 e IRAK-4, reprimiendo por tanto la señalización corriente abajo (Figura 8). Esta misma vía de activación de NF- $\kappa$ B se induce por la IL-1 y la IL-18.



**Figura 8.** Esquema simplificado de la vía de señalización de los TLRs inducida por un PAMP en ausencia (A) y en presencia (B) de IRAK-3. IRAK, kinasa asociada al receptor de la IL-1; MyD88, respuesta primaria de diferenciación mieloide 88; TRAF6, factor 6 asociado al receptor del TNF; IKKs, kinasas IκB; NF-κB, factor nuclear kappa B; IκB, inhibidor kappa B. Adaptado de Kobayashi *et al.* (2002) y de Villar *et al.* (2004).

A pesar de que *IRAK3* se localiza en una de las regiones cromosómicas ligadas al asma y fenotipos atópicos, su evaluación como gen candidato para la susceptibilidad en asma es reciente. Balaci *et al.* (2007) realizaron el clonaje posicional de este gen en asma en poblaciones del sur de Europa. Tras un análisis de ligamiento empleando microsatélites de la región cromosómica 12q13-24 en 100 familias de asmáticos sardos, los autores detectaron ligamiento con asma de comienzo temprano (0-13 años), y posteriormente refinaron la asociación. Empleando 22 SNPs de una región de casi 400 kb que incluía a *IRAK3*, encontraron asociación de varios SNPs y haplotipos de SNPs con asma en individuos sardos, tanto en estudios familiares como en un estudio de casos y controles sin relación de parentesco, realizando siempre una estratificación de la edad de comienzo de la enfermedad. Para la validación de esta asociación, los autores estudiaron además 6 SNPs en una muestra italiana de tamaño limitado, obteniendo replicación de la asociación de 2 de los SNPs (Balaci *et al.*, 2007).

### 1.6.5. Estudios de la genética del asma en España

En España se han llevado a cabo hasta el momento 10 estudios de asociación de genes candidatos con el asma, en distintas regiones (Barcelona, Badajoz, Canarias, Murcia, Valladolid, Salamanca y Vitoria). En total, en ellos se han analizado 29 genes candidatos y sus resultados han sido publicados en 22 estudios hasta la fecha (Tabla 4). Las limitaciones de los estudios realizados en la población española son comunes a las ya mencionadas con anterioridad respecto a los estudios de asociación en genes candidatos: i) Los tamaños

muestrales han sido reducidos, incluyendo 271±164 individuos como media: 144±75 casos y 127±164 controles; ii) El número de polimorfismos analizados ha sido siempre limitado 1,83±1,77 y sólo en un caso se ha asegurado una cobertura adecuada de la variación del gen de estudio (Klassert *et al.*, 2010); y por último iii) Ningún estudio ha considerado la posible existencia de estratificación poblacional en sus muestras.

**Tabla 4.** Estudios de asociación entre polimorfismos genéticos y asma realizados en la población española.

Estudio	Casos (n)	Controles (n)	Genes	Polimorfismos
Soriano <i>et al.</i> , 1997	203	168	<i>HLA-DRB1</i>	Alelos clásicos, 2 dígitos
			<i>HLA-DQB1</i>	Alelos clásicos, 4 dígitos
Lázaro <i>et al.</i> , 1999	144	184	<i>CFTR</i>	DF508, G542X, IVS8-6(5T), R75Q, G576A, L997F, 470M/V, R668C
Laitinen <i>et al.</i> , 2000	145	46	<i>FCER2</i>	Varios microsatélites
Soriano <i>et al.</i> , 2000	146	50	<i>MS4A2</i>	Ile181Leu, Gly237Glu, STR intron 5
			<i>TRA@</i>	Microsatélite próximo al gen
			<i>TRB@</i>	Microsatélite
Torio <i>et al.</i> , 2000	189	150	<i>5q31-q33</i>	Varios microsatélites
			<i>HLA-DRB1</i>	Alelos clásicos, 2 dígitos
			<i>HLA-DQB1</i>	Alelos clásicos, 4 dígitos
Cárdaba <i>et al.</i> , 2001	31	53	<i>HLA-DRB1</i>	Alelos clásicos, 2 dígitos
			<i>HLA-DQB1</i>	Alelos clásicos, 4 dígitos
			<i>TRA@</i>	<i>Ava II</i>
			<i>MS4A2</i>	<i>Rsa I</i> int 2, <i>Rsa I</i> ex 7
			<i>LTA</i>	<i>Nco I</i>
Isidoro-García <i>et al.</i> , 2003	44	68	<i>CD14</i>	<i>Ava II</i>
			<i>AID</i>	7888 C/T
Torio <i>et al.</i> , 2003	213	150	<i>HLA-DRB1</i>	Alelos clásicos, 2 dígitos
			<i>HLA-DQB1</i>	Alelos clásicos, 4 dígitos
Isidoro-García <i>et al.</i> , 2005a	133	79	<i>IL4</i>	-33C>T
			<i>IL4RA</i>	576Q>R
Isidoro-García <i>et al.</i> , 2005b	123	103	<i>LTC4S</i>	-444 A/C
Arriba-Méndez <i>et al.</i> , 2006	87	79	<i>CYSLTR1</i>	927 T>C
Sanz <i>et al.</i> , 2006	130	68	<i>CYSLTR1</i>	927 T>C
			<i>LTC4S</i>	-444 A>C
Tellería <i>et al.</i> , 2006	80	64	<i>ADRB2</i>	Arg16Gly, Gln27Glu
García-Martin <i>et al.</i> , 2007	270	295	<i>HNMT</i>	Thr105Ile
			<i>ABP1</i>	His645Asp
Arriba-Méndez <i>et al.</i> , 2008	109	79	<i>CYSLTR1</i>	927 T>C
			<i>LTC4S</i>	-444 A>C
Muro <i>et al.</i> , 2008	306	242	<i>CCL5</i>	-403A/G, -28C/G
Pascual <i>et al.</i> , 2008	150	98	<i>NOS2A</i>	Microsatélite
Tellería <i>et al.</i> , 2008	61	61	<i>ALOX5</i>	-176/-147
Sanz <i>et al.</i> , 2009	86	114	<i>PTGDR</i>	- 197T>C, - 441C>T, - 549T>C, - 613C>T
Torres-Galván <i>et al.</i> , 2009	111	82	<i>LTC4S</i>	-444 A/C
			<i>ALOX5</i>	-176/-147
			<i>ALOX5AP</i>	-169/-146
Cantero-Recasens <i>et al.</i> , 2010	301	470	<i>TRPV1</i>	I585V
			<i>TRPV4</i>	P19S
Klassert <i>et al.</i> , 2010	102	100	<i>TAC1</i>	rs2072099, rs1229434, rs3779470, rs28475844, rs4526299, rs1510300, rs1397202
			<i>TAC3</i>	rs11609825, rs733629, rs2122694, rs17119327, rs17119330, rs2291855
			<i>TAC4</i>	rs883010, rs4794068



### 1.6.6. Estudios de asociación genómica (GWAS)

El carácter dirigido de los estudios de asociación a genes candidatos no permite la detección de procesos biológicos nunca antes implicados en la susceptibilidad a padecer una determinada enfermedad. Sin embargo, los recientes avances tecnológicos y los estudios genéticos de consorcios internacionales han permitido ampliar los estudios a todo el genoma por medio de los estudios de asociación genómicos (GWAS), incrementando la solidez de las asociaciones detectadas y permitiendo descubrir nuevos genes de susceptibilidad.

Los GWAS retienen la eficacia de los estudios de asociación en genes candidatos, al permitir detectar alelos de riesgo con efectos leves en la enfermedad, pero tienen la ventaja de explorar el genoma sin la necesidad de disponer de una hipótesis sobre los procesos biológicos implicados. Esto es importante si tenemos en cuenta que para muchos de los genes identificados mediante los análisis de ligamiento en enfermedades de herencia simple no se suponía que tuviesen relación con la enfermedad.

En la actualidad, los GWAS utilizan micromatrices de oligonucleótidos (*arrays*) comerciales para genotipar simultáneamente un contenido estándar de cientos de miles de SNPs. Además, debido a que las mutaciones causantes de enfermedad aparecerán en un contexto genético concreto, y dada la baja tasa de recombinación del genoma, se espera que las variantes genéticas de riesgo permanezcan en la población durante bastantes generaciones, en asociación con variantes cercanas debido al LD, lo que permite la realización de imputación. Esto es, la inferencia estadística del genotipo de un SNP determinado a partir del genotipo de otros SNPs próximos altamente informativos, denominados tagSNPs o tSNPs (Marchini & Howie, 2010), con la consiguiente reducción en los costes de determinación de genotipos, en adelante denominado genotipado. Los tSNPs informan del estado de los SNPs que no son genotipados usando la información de LD disponible de datos referencia propios de cada laboratorio o de consorcios internacionales, como puede ser del Proyecto HapMap (Frazer *et al.*, 2007; Altshuler *et al.*, 2008). El proyecto HapMap (Frazer *et al.*, 2007), que ha caracterizado más de 3,5 millones de SNPs del genoma en distintas poblaciones, ha demostrado que la determinación de medio millón de SNPs permite inferir la variación en >80% de las variantes comunes (con frecuencia >5%) en las poblaciones europeas. Además, los algoritmos de imputación existentes permiten obtener los genotipos de más de 2,5 millones de SNPs del HapMap con poco margen de error, lo que proporciona una mayor resolución de la región asociada. Esto aumentará las posibilidades de identificar la variante causal, así como estudiar de forma más sencilla otros tipos de variación genética más compleja, como por ejemplo la variación estructural, la cual es más difícil de determinar con técnicas rutinarias de genotipado.

En contrapartida, los GWAS necesitan utilizar umbrales de significación más restrictivos para reducir resultados espurios, debidos simplemente al azar por las múltiples comparaciones realizadas, y requieren tamaños muestrales mayores para poder alcanzar tales niveles de significación (Altshuler *et al.*, 2008). A pesar de ello, los GWAS han permitido la identificación de genes de susceptibilidad nunca anticipados en algunas enfermedades complejas, sugiriendo inmediatamente nuevas hipótesis para el desarrollo de la enfermedad (Sladek *et al.*, 2007; Radstake *et al.*, 2010).

La reciente aplicación de los GWAS al asma ha revelado la existencia de variantes comunes de riesgo en unos pocos genes/regiones de susceptibilidad y con efectos no siempre consistentes en todas las poblaciones estudiadas (Tabla 5).

**Tabla 5.** GWAS de asma, genes identificados y *arrays* comerciales empleados.

Estudio	Población	Genes	Array(s) utilizado(s)
Moffatt <i>et al.</i> , 2007	Europea	<i>ORMDL3</i> *	Illumina HumanHap300
Hancock <i>et al.</i> , 2009	Mexicana	<i>TLE4</i>	Illumina HumanHap550
Himes <i>et al.</i> , 2009	Europeo-americana	<i>PDE4D</i>	Illumina HumanHap550
Mathias <i>et al.</i> , 2010	Afroamericana/afro-caribeña	<i>ADRA1B, PRNP, DPP10, GNA13</i>	Illumina HumanHap650Y
Sleiman <i>et al.</i> , 2010	Europeo-americana	<i>DENND1B</i> *, <i>ORMDL3</i> *	Illumina HumanHap300, 550, 650Y
Li <i>et al.</i> , 2010a	Europea	<i>RAD50, IL13, HLA-DR-DQ, LRPB1, SNX10, CA10, KCNJ2</i>	Illumina HumanCNV370
Moffatt <i>et al.</i> , 2010	Europea	<i>IL18R1</i> *, <i>HLA-DRB1</i> *, <i>HLA-DQ</i> *, <i>IL33</i> *, <i>ORMDL3/GSDML</i> *, <i>SMAD3</i> *, <i>IL2RB</i> *, <i>SLCA22A5, IL13, RORA</i>	Illumina Human610
Ferreira <i>et al.</i> , 2011b	Australiana (europea)	<i>ORMDL3, IL1RL1</i> , delección en 17q21	Illumina HumanHap300, CNV370 <sup>a</sup> , 610
Torgerson <i>et al.</i> , 2011	Europeo-americana, afroamericana/afro-caribeña, latinoamericana	<i>ORMDL3/GSDML</i> *, <i>IL1RL1</i> *, <i>TSLP</i> *, <i>RTP2</i> *, <i>IL33, PYHIN1</i>	Illumina Human 1M, 550, 610 <sup>b</sup> y 650Y; Affymetrix 500K y 6.0
Hirota <i>et al.</i> , 2011	Japonesa	<i>HLA-DRA</i> *, <i>NOTCH4</i> *, <i>C6orf10</i> *, <i>BTNL2, PBX2, HLA-DQB1, HLA-DQA2, HLA-DOA, IKZF4, LOC729675, GAB1, TSLP</i>	Illumina HumanHap550 y 610
Ferreira <i>et al.</i> , 2011a	Australiana (europea)	<i>IL6R</i> * y región 11q13.5*	Illumina HumanHap300, CNV370 <sup>a</sup> , y 610

\* Genes asociados con significación genómica en cada estudio.

Muchos de los genes que han sido señalados en los GWAS de asma se relacionan con procesos biológicos previamente implicados en la enfermedad (*FCER1A*, *IL1RL1*, *IL6R*) (Moffatt *et al.*, 2010; Ferreira *et al.*, 2011a), otros son completamente inesperados, entre los que destacan *ORMDL3* (relacionado con el metabolismo de esfingolípidos) (Moffatt *et al.*, 2010), *TLE4* (involucrado en la regulación de la actividad de las células B) (Hancock *et al.*, 2009), o *DENND1B* (que se expresa en las células dendríticas e interactúa con TNF- $\alpha$ ) (Sleiman *et al.*, 2010). A pesar de ello, el solape entre los genes revelados por los distintos estudios es sólo parcial, lo que sugiere que futuros estudios seguirán revelando nuevos genes de susceptibilidad al asma.

### 1.6.7. Meta-análisis de GWAS en asma

Los meta-análisis son un procedimiento estadístico para combinar los resultados de diferentes estudios y obtener una nueva aproximación a la realidad. Este tipo de estudios son de gran utilidad cuando los resultados previos son contradictorios o cuando se sospecha que éstos carecen de un tamaño de muestra adecuado, puesto que permiten aumentar el poder estadístico y por tanto detectar asociaciones con efectos más leves en la enfermedad (Ober & Yao, 2011). Hasta el momento se han llevado a cabo dos meta-análisis con datos de GWAS en asma (Moffatt *et al.*, 2010, Torgerson *et al.*, 2011).

El primero de estos meta-análisis de GWAS en asma se realizó por el consorcio GABRIEL y se centró en población europea, lo que proporcionó una muestra de 10.365 casos y 16.110 controles procedentes de 23 estudios de toda Europa en los que se genotiparon 582.892 SNPs (Moffatt *et al.*, 2010). En este estudio, se alcanzaron umbrales de significación genómicos de  $p \leq 7,2 \times 10^{-8}$  en SNPs cercanos o pertenecientes a 6 genes (*HLA-DQ*, *ORMDL3/GSDML*, *IL33*, *IL18R1*, *IL2RB*, y *SMAD3*). Otros 3 genes mostraron valores próximos a la significación (*SLCA22A5*, *IL13*, y *RORA*), es decir, no alcanzaron los valores de significación genómicos aunque éstos fueron elevados ( $p \leq 5,0 \times 10^{-7}$ ). Asimismo, en este estudio se observó que, generalmente, las asociaciones fueron más significativas entre los asmáticos de comienzo temprano (0-16 años) y especialmente en SNPs de la región 17q21 próxima a *ORMDL3/GSDML*, cuya asociación resultó sólo significativa entre los casos de comienzo temprano, pero no en los de comienzo en edad adulta.

El segundo meta-análisis de GWAS en asma, realizado por el consorcio EVE, consideró muestras de distintas poblaciones americanas, no sólo europeo-americanas, sino también afroamericanas, afro-caribeñas y latinoamericanas, que se analizaron tanto conjuntamente como de manera estratificada para más de 2 millones de SNPs (Torgerson *et al.*, 2011). Estableciendo un umbral de significación de  $p \leq 2 \times 10^{-8}$  encontraron asociación de SNPs cercanos o pertenecientes a 4 genes: *ORMDL3/GSDML*, *IL1RL1* y *TSLP* al analizar

todas las poblaciones conjuntamente; y *RTP2*, que resultó específicamente asociado en latinos. Otras 11 regiones mostraron valores próximos a la significación ( $p \leq 1,0 \times 10^{-6}$ ), incluyendo algunas específicas en determinadas poblaciones. Un análisis posterior de replicación de estas 15 regiones, destacó la asociación del gen *PYHIN1* en la región 1q23.1 sólo en individuos de ascendencia africana (Torgerson *et al.*, 2011).

Considerados conjuntamente los resultados de ambos meta-análisis, 4 *loci* mostraron asociación en las muestras de distintas etnias: *ORMDL3/GSDML*, *IL1RL1/IL18R1*, *TSLP* e *IL33*, los cuales pueden considerarse genes de susceptibilidad al asma. Además, tres de estas asociaciones resaltan la importancia de las citocinas derivadas de células epiteliales (*TSLP* e *IL-33*), que promueven la diferenciación y activación de linfocitos  $T_{H2}$ , y de sus receptores (*IL1RL1* codifica ST2, receptor de *IL-33*).

Por tanto, los genes que se estudiarán en este trabajo, exceptuando *IL13*, no han sido revelados en los GWAS o meta-análisis de GWAS llevados a cabo hasta este momento (Tabla 5). Aunque no hay una explicación clara al respecto, existe la posibilidad de que estas regiones no estén adecuadamente representadas en los *arrays* empleados en los GWAS o la necesidad de definir subgrupos de fenotipos dentro del asma en los que estos genes tengan un mayor efecto.

#### 1.6.8. Perspectivas futuras en la genética del asma

Los GWAS han demostrado ser una herramienta muy eficaz para la identificación de genes y procesos biológicos que subyacen a la susceptibilidad a distintas enfermedades complejas, incluida el asma. Sin embargo, los alelos de riesgo identificados hasta este momento mediante los GWAS no explican todo el componente genético de las enfermedades complejas (Manolio *et al.*, 2009). Dado que se estima que quedan muchos más genes de susceptibilidad por descubrir, surgen nuevos retos en el estudio de la genética de la enfermedad, como la necesidad de ampliar los estudios al espectro de frecuencia alélica baja ( $MAF < 5\%$ ), que quizás tengan efectos mayores en la enfermedad.

Otra posible explicación a la imposibilidad de predecir el riesgo a la enfermedad mediante datos genéticos es que los GWAS no estén revelando variantes causales sino variantes correlacionadas con ellas, cuyos efectos sean inferiores a los reales. Alternativamente, la causalidad podría deberse a otro tipo de variación más compleja, por ejemplo los polimorfismos de variación en el número de copia (llamados CNVs) o a cambios epigenéticos, poco relacionados con la variación de SNPs (Manolio *et al.*, 2009). Sin embargo, las evidencias actuales demuestran que en general la gran mayoría de los CNVs mantienen una gran correlación con los SNPs. Por tanto, con los GWAS realizados también

## INTRODUCCIÓN

---

ha sido posible asociar de manera indirecta los CNVs más frecuentes con la enfermedad (Craddock *et al.*, 2010).

Por otro lado, como se ha mostrado con anterioridad, la mayoría de los GWAS de asma se han realizado en poblaciones con ascendencia en el Norte de Europa (8 de los 11 estudios; Tabla 5). Dado que es muy probable que existan genes de susceptibilidad y/o alelos de riesgo con especificidad poblacional, más aún si las variantes implicadas en la etiología son de baja frecuencia, es necesario que se realicen GWAS en otras poblaciones, sobre todo más diversas (Manolio *et al.*, 2009). Es más, aunque se mantengan los efectos en dos poblaciones distintas, una variante de riesgo puede detectarse en una población simplemente porque alcance una frecuencia mayor. La interacción de los genes con el ambiente también contribuiría a las diferencias en los efectos detectados en distintas poblaciones y a la detección de las variantes de riesgo en los estudios, lo que puede incluso traducirse en que un mismo alelo tenga efectos opuestos en poblaciones diferentes (Choudhry *et al.*, 2005).

Otro aspecto importante es que, pese a la enorme información que proporcionan los GWAS, en la predicción de riesgo a enfermedad sólo se ha utilizado la información de los SNPs más significativamente asociados, sin considerar el resto de información del genoma. Los GWAS identifican unas pocas variantes de susceptibilidad que alcanzan los umbrales de significación genómicos, pero también desvelan muchas otras que son sugerentes, y que en muchos casos se descartan por la dificultad de discernir si son reales o no. Sin embargo, su utilización podría ayudar a identificar más genes de susceptibilidad al asma si los datos se integran con los resultados de otro tipo de información biológica, redes de interacción de proteínas en vías funcionales conocidas, etc (Wei *et al.*, 2009).

Otra posibilidad es que la causalidad implique variación menos frecuente. El Proyecto 1000 Genomas (1KGP) (The 1000 Genomes Project Consortium, 2010) ha revelado la existencia de millones de variantes, SNPs e inserciones/deleciones (indels) cuya frecuencia en la población es baja (1-5%). Dado que los GWAS actuales tienen un poder estadístico limitado para detectar asociaciones con SNPs de baja frecuencia (<1%), al haber empleado *arrays* de genotipado diseñados a partir de la información del proyecto HapMap (3,5 millones de variantes con  $(MAF \geq 5\%)$ ) (Frazer *et al.*, 2007; Altshuler *et al.*, 2008), y no basados en los más de 17 millones de variantes genéticas reveladas por 1KGP (The 1000 Genomes Project Consortium, 2010), la mayoría de estas variantes nunca han sido estudiadas en relación al desarrollo del asma ni con ninguna otra enfermedad compleja. En el supuesto de que estas variantes tuvieran efectos mayores, lo cual es biológicamente plausible, éstas podrían explicar el componente genético del asma no revelado hasta el momento. Se espera que este tipo de variantes de baja frecuencia pueda estudiarse

adecuadamente empleando *arrays* de nueva generación, y usando además algoritmos de imputación a partir de los datos del 1KGP (The 1000 Genomes Project Consortium, 2010). Sin embargo, se prevé que serán necesarios proyectos de secuenciación masiva de genes candidatos, exomas o genomas completos para un mejor entendimiento de la genética del asma.

### **1.7. Relación entre ascendencia genética y asma**

Un aspecto importante en la susceptibilidad al asma lo constituye la composición genética de la población estudiada, puesto que se observan disparidades étnicas en la prevalencia y gravedad de la enfermedad que no pueden explicarse únicamente por diferencias ambientales, sociales, culturales o por factores económicos. Ello se encuentra además reforzado por el hecho de que la variación genética de la población humana se encuentra estructurada en las distintas poblaciones mundiales (Li *et al.*, 2008).

A continuación se detallan dos ejemplos de relación entre ascendencia genética y asma, en poblaciones afroamericanas y mexicanas, que muestran que la asociación entre la ascendencia genética y el riesgo o protección frente al asma puede ser indicativa de la existencia de genes o regiones genómicas que confieren una distinta susceptibilidad a la enfermedad en la población, y que son específicas de una de las poblaciones parentales. Algo similar podría ocurrir en la población canaria, que muestra una mayor prevalencia de asma que el resto de España y es resultado de una mezcla reciente entre poblaciones europeas y norteafricanas.

#### **1.7.1. Ascendencia genética africana y riesgo al asma en afroamericanos**

Una población en la que la ascendencia genética presenta un papel fundamental en el riesgo al asma es la población afroamericana. Esta población se ha originado por la mezcla reciente de poblaciones de Europa y del oeste de África (Zakharia *et al.*, 2009). Se ha sugerido que los alelos de riesgo de genes que promueven directa o indirectamente una respuesta linfocitaria de tipo  $T_H2$  se encuentran en mayor frecuencia en poblaciones cuyos ancestros son originarios de regiones tropicales, puesto que podrían haber conferido cierta ventaja selectiva en la defensa frente a los parásitos (Le Souef *et al.*, 2006).

Dado que en el oeste africano la prevalencia del asma es similar a la de Europa, y que la población afroamericana presenta una mayor prevalencia de la enfermedad que la población europeo-americana con la que comparte condiciones ambientales similares, según esta hipótesis, la mayor prevalencia del asma en afroamericanos podría ser resultado de la interacción entre los alelos de riesgo heredados de los ancestros africanos y los factores de riesgo ambientales presentes en las sociedades con estilo de vida occidental. En

este sentido, se ha demostrado la influencia de la ascendencia genética en el asma en afroamericanos, al asociarse la ascendencia africana con un mayor riesgo de padecer asma (Flores *et al.*, 2011), y con la presencia de mayores niveles de IgE totales en suero en poblaciones afro-caribeñas (Vergara *et al.*, 2009). Asimismo, se ha encontrado relación entre la ascendencia africana y la disminución de la función pulmonar en afroamericanos (Tsai *et al.*, 2006; Kumar *et al.*, 2010), incluso sugiriéndose que el diagnóstico del asma grave podría mejorarse si la información de la ascendencia genética se incluyese en los modelos de cálculo de riesgo (Kumar *et al.*, 2010).

Además, en la población afroamericana se han encontrado recientemente factores genéticos de riesgo heredados de sus ancestros africanos cuya asociación con la enfermedad es exclusiva de la población afroamericana. Este es el caso del gen *PYHIN1*, cuyo alelo de riesgo no está presente en las poblaciones de ascendencia europea y se presenta en baja frecuencia en poblaciones latinoamericanas (Torgerson *et al.*, 2011).

### **1.7.2. Ascendencia genética amerindia y protección frente al asma en la población mexicana**

Otro ejemplo de relación entre ascendencia y asma lo representa la población mexicana, que muestra los valores de ascendencia amerindia más elevados y la menor prevalencia de asma entre las poblaciones hispanas (Hunninghake *et al.*, 2006).

En contraste con la población afroamericana, un estudio asoció la ascendencia amerindia con la presencia de asma leve (Salari *et al.*, 2005). Además, recientemente se ha revelado que determinadas regiones cromosómicas portadoras de alelos protectores para el desarrollo del asma presentan una mayor ascendencia amerindia que la esperada (Hancock *et al.*, 2009). Por tanto, estos estudios sugieren que la ascendencia amerindia se relaciona con un menor riesgo de asma. Al ser la población mexicana portadora de alelos protectores provenientes de la población nativa, éstos disminuyen la susceptibilidad y gravedad de la enfermedad en esta población.

### **1.8. Influencia norteafricana en la población española**

Las poblaciones que habitan la Península Ibérica muestran una relativa homogeneidad genética (Cavalli-Sforza *et al.*; 1994, González *et al.*, 2003; Flores *et al.*, 2004; Laayouni *et al.*, 2010), salvo algunas excepciones como la población vasca (Rodríguez-Ezpeleta *et al.*, 2010). Esto se debe a que las poblaciones peninsulares se han visto afectadas por los mismos grandes fenómenos migratorios prehistóricos que el resto de poblaciones europeas, pero de una manera uniforme (Flores *et al.*, 2004), con independencia de la magnitud de la influencia debida a las expansiones neolíticas desde el

Oriente Próximo (Cavalli-Sforza *et al.*, 1994; Richards *et al.*, 2000; Semino *et al.*, 2000; Barbujani & Goldstein, 2004; Balaesque *et al.*, 2010). A pesar de ello, varias evidencias indican la existencia de influencia africana desigual entre distintas poblaciones españolas, tanto en la Península Ibérica como en las Islas Canarias. En este sentido, se ha encontrado una mayor diversidad haplotípica autosómica a nivel genómico en las poblaciones del suroeste de Europa, lo que se ha interpretado como resultado de las migraciones directas desde el norte de África hacia la cuenca mediterránea (Auton *et al.*, 2009). Además, los estudios basados en el ADN mitocondrial y en la porción no recombinante del cromosoma Y han encontrado una influencia norteafricana del 8-10% en la Península Ibérica (González *et al.*, 2003; Pereira *et al.*, 2005; Flores *et al.*, 2004; Adams *et al.*, 2008; Capelli *et al.*, 2009), con estimas algo mayores en las regiones del noroeste y del sur (González *et al.*, 2003; Maca-Meyer *et al.*, 2003; Flores *et al.*, 2004; Adams *et al.*, 2008). Esta influencia norteafricana podría deberse al legado genético musulmán de su reinado en la Península Ibérica que comenzó en el año 711 y se prolongó durante tres siglos (Chejne, 1974).

En el caso particular de las Islas Canarias, se ha demostrado que tales influencias son aún mayores, entre 22-38% de influencia norteafricana y menos de un 5% de mezcla subsahariana (Rando *et al.*, 1999; Flores *et al.*, 2001; Flores *et al.*, 2003; Maca-Meyer *et al.*, 2004; Fregel *et al.*, 2005; Fregel *et al.*, 2009). Esto se explicaría por la influencia genética de la población aborígen y por el tráfico de esclavos de origen subsahariano (Flores *et al.*, 2001). Las Islas Canarias se encontraban habitadas por una población aborígen en el momento de la conquista española en el siglo XV. Desde el principio de la conquista, los cronistas relacionaron las costumbres y el dialecto de los aborígenes canarios con los de las poblaciones bereberes del noroeste africano. Desde entonces, se han acumulado numerosas evidencias, tanto antropológicas como arqueológicas, que apoyan dicho origen. De la misma manera, los estudios genéticos llevados a cabo han encontrado influencia norteafricana en la población canaria, usando marcadores clásicos (Flores *et al.*, 2001), autosómicos (Maca-Meyer *et al.*, 2004; Fregel *et al.*, 2005), ADN mitocondrial (Rando *et al.*, 1999) y marcadores de la porción no recombinante del cromosoma Y (Flores *et al.*, 2003).

Estas diferencias en la mezcla africana presente en las poblaciones españolas podrían ser una fuente de estratificación poblacional con efectos potencialmente negativos en los estudios de asociación (Price *et al.*, 2006; Chanock *et al.*, 2007). Además, éstas podrían ser causantes de la distinta incidencia que se observa entre la población canaria y peninsular para algunas enfermedades como el asma (Juliá Serda *et al.*, 2005), la diabetes tipo II (Boronat *et al.*, 2006), y la hipertensión (de Pablos-Velasco *et al.*, 2002), entre otras.



## ***2. HIPÓTESIS Y OBJETIVOS***

---

### 2. HIPÓTESIS Y OBJETIVOS

Las **hipótesis de partida** para este trabajo son:

1) La población española presenta una mezcla genética africana desigualmente distribuida entre regiones generando la presencia de estructura poblacional.

2) La ascendencia africana, estimada a partir de polimorfismos genéticos, se asocia con la susceptibilidad y/o gravedad del asma o fenotipos alérgicos en la población canaria.

3) Asumiendo un modelo de variante común enfermedad común, en el que variantes genéticas frecuentes en la población constituyen factores de riesgo en el desarrollo de asma, variantes comunes en genes de gran relevancia en la patogénesis del asma en varias poblaciones constituyen también factores de susceptibilidad y/o gravedad de esta enfermedad en la población española.

Los **objetivos** de esta tesis doctoral son:

1) Evaluar un conjunto de polimorfismos informativos de ascendencia que permitan distinguir entre ascendencia genética europea y africana, con los que estudiar si existen diferencias en la mezcla africana entre distintas regiones de España.

2) Determinar si las diferencias en mezcla norteafricana entre regiones españolas podrían sesgar los resultados de los estudios de asociación de tipo caso-control en individuos españoles no relacionados, y evaluar distintas metodologías para la corrección eficiente de estos efectos.

3) Estudiar la relación entre el trasfondo genético poblacional de la población canaria y la susceptibilidad y gravedad del asma y fenotipos relacionados.

4) Analizar la asociación de variantes comunes del gen *IRAK3* con la susceptibilidad y gravedad del asma y fenotipos relacionados en la población española.

5) Estudiar la asociación de variantes comunes de ocho de los genes candidatos con mayor número de replicaciones en estudios independientes en la bibliografía (*IL4*, *IL13*, *IL4R*, *LTA-TNF*, *CD14*, *ADRB2*, *MS4A2* y *ADAM33*) con la susceptibilidad y gravedad del asma y fenotipos relacionados en la población española.

6) Explorar los efectos de la edad de comienzo del asma en la asociación de los genes candidatos con la susceptibilidad a la enfermedad.

7) Estudiar la magnitud del riesgo conjunto que proporcionaría la acumulación de variantes de riesgo en los distintos genes candidatos de susceptibilidad al asma y su relación con la edad de comienzo de la enfermedad.

### ***3. METODOLOGÍA***

---

### 3. METODOLOGÍA

#### 3.1. Diseño y muestras para el estudio de estratificación poblacional

En los estudios de asociación de tipo caso-control realizados con individuos sin relación de parentesco, es muy importante que los individuos de ambos grupos representen una muestra homogénea de la población en cuanto a su origen étnico y geográfico, evitando diferencias en la composición genética entre grupos. La existencia de estratificación poblacional podría originar asociaciones falsas si existiese una mayor prevalencia de la enfermedad en los distintos subgrupos o si éstos no estuviesen igualmente representados en los casos y controles (Price et al., 2006; Chanock *et al.*, 2007). Puesto que ningún trabajo previo ha estudiado la posible presencia de estratificación poblacional en la población española, debida a una mezcla africana desigual entre regiones, en esta tesis se realizó un estudio previo a los estudios de asociación para explorar dicha posibilidad. Para ello, se realizó un muestreo de población peninsular, canaria y norteafricana, y se obtuvo información de otras poblaciones de interés, procedente de bases de datos de acceso libre y de datos publicados.

##### 3.1.1. Muestras españolas y norteafricanas

Se obtuvieron 77 muestras de ADN representativas de la población peninsular, del Banco Nacional de ADN ([www.bancoadn.org](http://www.bancoadn.org)), con origen en distintas regiones: Andalucía (n=15), Murcia (n=5), Extremadura (n=5), Castilla La Mancha (n=4), Valencia (n=7), Castilla y León (n=9), Madrid (n=1), Cataluña (n=2), Galicia (n=4), Cantabria (n=3), Navarra (n=4), La Rioja (n=4), Asturias (n=1), y muestras con un origen peninsular heterogéneo (n=13). La población canaria estaba representada por 104 muestras procedentes de donantes sanos y recogidas en bancos de sangre (El Hierro, n=7; La Palma, n=7; La Gomera, n=7; Tenerife, n=30; Gran Canaria, n=30; Lanzarote, n=13 y Fuerteventura, n=10). Éstas se compararon además con 15 muestras procedentes del Banco Nacional de ADN, todas ellas originarias de Gran Canaria. Además, se incluyeron 68 muestras de ADN de donantes nacidos en el Norte de África: 25 de Casablanca y 25 de Rabat (cedidas por el Dr. David Comas, Institut de Biología Evolutiva, Barcelona), y 18 procedentes de otras regiones de Marruecos (cedidas por los Dres. José Larruga, Ana González y Vicente Cabrera, Universidad de La Laguna).

En todos los casos se trató de muestras de sangre o ADN de donantes sanos sin relación de parentesco, que declararon que sus cuatro abuelos habían nacido en el mismo lugar. Así, la asignación de la localidad de origen se realizó en función de la procedencia de sus cuatro abuelos.

### 3.1.2. Datos de otras poblaciones

Se utilizaron datos genómicos de más de 640.000 SNPs procedentes del panel mundial de la diversidad genética humana (HGDP) (Li *et al.*, 2008), correspondientes a 29 individuos argelinos de etnia mozabite y 24 individuos vascos franceses, ambos grupos escogidos como representativos de las poblaciones norteafricana y española respectivamente, por ser las poblaciones más cercanas a estas ascendencias recogidas en el HGDP. Además, se emplearon datos depositados en el proyecto HapMap (Frazer *et al.*, 2007), correspondientes a individuos subsaharianos de ascendencia nigeriana del grupo étnico yoruba (YRI).

Con el fin de comparar las poblaciones analizadas dentro de un contexto europeo, se consideraron además datos publicados para muestras de individuos de distintos orígenes: 163 suecos, 57 polacos, 76 ingleses, 119 italianos, 68 griegos, y 55 españoles (Price *et al.*, 2008), así como muestras de 60 individuos norteamericanos con ascendencia en el norte u oeste de Europa (CEU) depositadas la base de datos del HapMap (Frazer *et al.*, 2007).

### 3.1.3. Selección de marcadores de ascendencia

En el momento en que se comenzó este estudio no existía en la bibliografía ningún conjunto de polimorfismos informativos de ascendencia específicamente escogidos para distinguir entre población peninsular y norteafricana, con los que estudiar las diferencias en mezcla norteafricana entre Canarias y la Península Ibérica. Sin embargo, un trabajo previo en el que se analizó la estratificación poblacional presente en Europa había descrito un conjunto de 100 marcadores, denominados EuroAIMs, capaces de distinguir adecuadamente entre poblaciones del noroeste y sureste de Europa (Price *et al.*, 2008). De estos 100 EuroAIMs, se seleccionaron 93 para los que hubo datos disponibles de todas las poblaciones estudiadas en dicho trabajo (Dr. Alkes Price, Harvard School of Public Health, Boston, comunicación personal) para genotiparlos en las muestras de este estudio.

## 3.2. Diseño y muestras para el estudio de asociación

### 3.2.1. Muestras de asmáticos y controles

Siguiendo un diseño de tipo caso-control de individuos sin relación de parentesco, se incluyeron en el estudio de asociación muestras de individuos con dos generaciones de ancestros nacidos en España, que formaron parte del estudio GOA (estudio de la genética del asma en la población española).

Grupo de casos: incluyó 607 muestras de sangre o saliva de sujetos diagnosticados de asma con más de 5 años de edad y con una caracterización clínica detallada realizada por los Servicios de Neumología y Alergia del Hospital Universitario Nuestra Señora de Candelaria (Tenerife) y por los Servicios de Alergia del Hospital Universitario Insular y del Hospital Universitario Dr. Negrín (ambos en Gran Canaria). En todos los casos se requirió el consentimiento informado del paciente o de sus familiares y se registraron distintas variables demográficas y clínicas de interés para el estudio: edad, sexo, diagnóstico, función pulmonar medida mediante pruebas espirométricas (determinación del volumen expiratorio forzado en un segundo [FEV1] y del volumen expiratorio forzado en un segundo tras exposición a un broncodilatador [FEV1 post]), edad de comienzo de la enfermedad, niveles de IgE totales y específicas, atopía, rinitis, resultado para la prueba cutánea de alergia (SPT), etc (Tabla 6).

**Tabla 6.** Características clínicas y demográficas de la muestra de casos y controles

Variable	Casos (n=607)	Controles (n=1.271)	p-valor
Género (varones, %)	21,9	59,1	<0,001 <sup>a</sup>
Edad (P <sub>25</sub> -P <sub>75</sub> )	25-44	32-48	<0,001 <sup>b</sup>
Hábitos tabáquicos (alguna vez fumadores, %)	27,3	48,1	<0,001 <sup>a</sup>
Características clínicas:			
log <sub>10</sub> IgE en suero	2,02-2,64		
Gravedad del asma (%):			
Grave	47,3		
Moderado	28,6		
Leve	24,1		
Edad de diagnóstico (años, P <sub>25</sub> -P <sub>75</sub> )	21-28		
FEV1 (% valor predicho, P <sub>25</sub> -P <sub>75</sub> )	68-95		
FEV1 post (litros, P <sub>25</sub> -P <sub>75</sub> )	2,1-3,9		
Rinitis (%)	74,2		
Dermatitis (%)	20,0		
SPT <sup>c</sup> (% positivo)	76,9		
IgE específica (% positivo)	51,9		
Atopia <sup>d</sup> (%)	76,9		
Tratamiento de síntomas asmáticos (%)	55,6		
Índice de masa corporal (kg/m <sup>2</sup> , P <sub>25</sub> -P <sub>75</sub> )	21,0-27,5		

<sup>a</sup>Estadístico de  $\chi^2$ ; <sup>b</sup>Estadístico de Mann-Whitney; <sup>c</sup>Positivo para al menos uno de los siete alérgenos comunes evaluados, <sup>d</sup>Positivo para SPT o IgE específica.

En el grupo de asmáticos, la definición de atopía se basó en el resultado positivo en al menos una de las pruebas cutáneas de alergia, o por IgE específica elevada en el suero. La prueba cutánea de alergia se consideró positiva si el paciente desarrolló un habón con un diámetro 3 mm mayor que un control salino en respuesta al pinchazo subcutáneo con

extractos de alérgenos. En el caso de la IgE específica, para considerarse elevada el paciente debió presentar niveles  $>0,35$  UI mL<sup>-1</sup> frente a algún alérgeno. En los dos tipos de pruebas, los alérgenos considerados pertenecieron a 7 grandes grupos: ácaros, epitelios, polen, hongos, alimentos, látex y otros.

Grupo de controles: se incluyeron 1.271 muestras de sangre o ADN de individuos con dos generaciones de ancestros de origen español representativos de la población general, procedentes de bancos de sangre del territorio nacional (Hospital Universitario Dr. Negrín y Banco Nacional de ADN). Todos los controles cumplieron el requisito de no presentar enfermedades respiratorias, diagnóstico de asma u otros fenotipos atópicos relacionados (alergia, rinitis, dermatitis atópica, conjuntivitis atópica), ni tampoco presentaron antecedentes familiares conocidos de ninguna de estas condiciones.

### 3.2.2. Selección de SNPs en los genes candidatos

Con el fin de realizar un estudio eficiente de la asociación de variantes comunes de los genes candidatos (*IL4*, *IL13*, *IL4R*, *LTA-TNF*, *CD14*, *ADRB2*, *MS4A2*, *ADAM33* e *IRAK3*), se seleccionó para cada gen un conjunto mínimo de tSNPs, polimorfismos que por su elevada correlación con otras variantes comunes del mismo permiten resultar informativos para predecir su variación. En el proceso de selección, se forzó que los algoritmos incluyesen como parte del conjunto de tSNPs a aquellos SNPs que en estudios anteriores hubiesen resultado asociados con asma y fenotipos relacionados, o que presentasen funcionalidad deducida de estudios *in silico*, *in vitro* o *in vivo*. Por medio de la aplicación informática PupaSuite (Conde *et al.*, 2006) se exploró la posible funcionalidad asociada a las variantes comunes de los genes candidatos, por su presencia en sitios de unión de factores de transcripción, de unión de microRNAs, de regulación del procesamiento de formas alternativas del ARN o por predecir cambios de aminoácido.

La información de los genotipos de los SNPs presentes en los genes candidatos en muestras de referencia de la población europea procedió de bases de datos de re-secuenciación públicas o privadas (Tabla 7), y se incluyeron los polimorfismos situados en al menos 2 kb flanqueantes al gen. Esta elección se debió a que la información de re-secuenciación proporciona una información más completa de las variantes comunes de los genes candidatos que la información de genotipado del proyecto HapMap que, como media, sólo incluye información de un SNP por cada kb (Frazer *et al.*, 2007). Como se comentó anteriormente, dada la proximidad y el elevado LD entre los genes *LTA* y *TNF*, ambos fueron estudiados como una sola región. Del mismo modo, a pesar de que los diversos estudios de asociación anteriores han estudiado los genes *IL13* e *IL4* de manera

independiente (Ober & Hoffjan, 2006), dada su escasa separación física (12,6 kb), en este estudio también se consideraron como una sola región.

**Tabla 7.** Fuente de los datos de secuenciación de los genes candidatos utilizados para la selección de tSNPs.

Gen	Fuente de los datos	Población	Nº de cromosomas
<i>IL13-IL4</i>	SeattleSNPs database <sup>a</sup>	Europeo-americana	46
<i>CD14</i>	Innate Immunity NHLBI Program for PGA <sup>b</sup>	Europeo-americana	46
<i>ADRB2</i>	SeattleSNPs database <sup>a</sup>	Europeo-americana	46
<i>LTA-TNF</i>	SeattleSNPs database <sup>a</sup>	Europeo-americana	46
<i>MS4A2</i>	T1D Database <sup>c</sup>	Inglesa	192
<i>IL4R</i>	SeattleSNPs database <sup>a</sup>	Europeo-americana	46
<i>ADAM33</i>	NIEHS Environmental Genome Project <sup>d</sup>	Europeo-americana	44
<i>IRAK3</i>	Resultados de proyecto SAF 2004-06833	Española	64

<sup>a</sup>Disponible en <http://pga.gs.washington.edu>

<sup>b</sup>Disponible en <https://regepi.bwh.harvard.edu/IIPGA2>

<sup>c</sup>Datos compartidos por el Prof. John Todd (University of Cambridge) y descritos en Smyth *et al.* (2006).

<sup>d</sup>Disponible en <http://egp.gs.washington.edu>.

En el caso del gen *IRAK3*, la información de partida procedió de datos de secuenciación del gen en una muestra de 32 individuos sanos sin relación de parentesco de la población española, disponible en nuestro laboratorio como resultado de un proyecto anterior. Las regiones secuenciadas comprendieron un total de 22,6 kb del gen e incluyeron: los 12 exones, al menos 100 pb de las regiones intrónicas flanqueantes de los mismos, 1,5 kb de la región 5', y 4,2 kb de la región 3'. Como parte de esta tesis se procedió a la comparación manual de las secuencias y a la validación de las variantes nuevas y/o las variantes exclusivas de las 32 muestras, mediante secuenciaciones independientes. Las reacciones de secuenciación de ambas cadenas se llevaron a cabo con los mismos cebadores que se habían empleado en las amplificaciones por reacción de la polimerasa e cadena (PCR), utilizando el BigDye<sup>®</sup> v3.1 Terminator Cycle Sequencing kit (Applied Biosystems, Foster City, CA) y siguiendo las recomendaciones del fabricante. Posteriormente, las secuencias fueron analizadas en un secuenciador automático ABI PRISM 310 Genetic Analyzer (Applied Biosystems) e interpretadas mediante el programa Sequence Analysis (Applied Biosystems). Por último, la denominación de las variantes se realizó siguiendo las guías establecidas internacionalmente (Antonarakis, 1998).



Independientemente de la fuente de información de genotipos, los datos de re-secuenciación fueron inicialmente filtrados de manera que sólo se tuvieron en cuenta aquellas variantes con  $MAF \geq 5\%$  en la muestra de referencia. Posteriormente se reconstruyeron los haplotipos, y por último se procedió a la selección de los tSNPs. A continuación se detalla cada uno de los pasos llevados a cabo en el proceso:

1. En primer lugar, a partir de los datos de re-secuenciación se calcularon las frecuencias alélicas de los polimorfismos, mediante el programa Haploview 4.2 (Barrett *et al.*, 2005), y se seleccionaron aquellos que presentaron una  $MAF \geq 5\%$  en la población.

2. A continuación, se infirieron los haplotipos a partir de los genotipos mediante el programa PHASE 2.1 (Stephens *et al.*, 2001), el cual ofrece una reconstrucción bayesiana de haplotipos utilizando la información obtenida *a priori* de haplotipos ya conocidos (todos excepto las muestras dobles heterocigóticas) empleando cadenas de Markov-Monte Carlo. Tras comprobar que distintas carreras independientes con parámetros diferentes convergían en los mismos valores de probabilidad, los haplotipos fueron obtenidos de la carrera, de un total de cinco, que proporcionase un mejor ajuste con los datos observados.

3. Por último, los haplotipos con frecuencia superior al 1% se emplearon para la selección de un conjunto mínimo de tSNPs que proporcionase una cobertura eficiente del gen, utilizando el programa TagIT (Ahmadi *et al.*, 2005) para MATLAB® (The MathWorks Inc., Natick, MA). En este paso, los SNPs con funcionalidad demostrada o predicha y aquellos que se hubiesen asociado con asma con anterioridad fueron forzados a estar incluidos como parte de los tSNPs seleccionados. El conjunto final de tSNPs se evaluó mediante simulaciones en las que se realizó un remuestreo de SNPs (eliminando en cada ronda un SNP del conjunto). Con ello se comprobó que los SNPs comunes del gen eran capturados con una cobertura haplotípica de  $r^2 \geq 0,8$  como media (Ahmadi *et al.*, 2005).

Además, se comprobó que los conjuntos de tSNPs obtenidos para cada gen representasen adecuadamente la variación recogida en la base de datos del HapMap (Frazer *et al.*, 2007), llegando en algunos casos a tener que aumentarse el número de SNPs. Este fue el caso concreto del gen de la *IL4R*, para el cual la re-secuenciación depositada en la base de datos no se había realizado para el gen completo, debido al desconocimiento de la presencia de dos exones reguladores adicionales en la región 5' del gen. La Tabla 8 recoge el número de SNPs seleccionados para cada gen.

**Tabla 8.** Resumen del conjunto de tSNPs seleccionados para cada gen.

Gen	Región cromosómica (posición en Megabases)	Tamaño del gen (kb)	Región cubierta (kb)	Nº tSNPs	Cobertura ( $r^2$ )
<i>IL13-IL4</i>	5q31.1 (132,0)	12	29,0	10	1,00
<i>CD14</i>	5q31.1 (140,0)	2	7,0	6	1,00
<i>ADRB2</i>	5q31 (148,2)	2	9,5	8	1,00
<i>LTA-TNF</i>	6p21.3 (31,5)	6	9,3	11	1,00
<i>MS4A2</i>	11q13 (59,9)	10	15,3	7	0,97
<i>IL4R</i>	16p12.1 (27,3)	51	56,0	21	1,00
<i>ADAM33</i>	20p13 (3,6)	14	15,2	19	1.00
Total		97	141,3	82	

### 3.3. Genotipado

#### 3.3.1. Preparación de las muestras para el genotipado

##### 3.3.1.1. Extracción de ADN

De los participantes a los que se extrajo sangre periférica, se obtuvieron 4 ml en un tubo con ácido etilendiaminotetraacético (EDTA) que fue conservado a -20 °C hasta la extracción del ADN genómico mediante métodos convencionales basados en columnas de afinidad GFX™ Genomic Blood DNA Purification Kit (GE Healthcare, Salt Lake City, UT). En el caso de las muestras de saliva, principalmente obtenidas en el caso de que los pacientes fuesen niños asmáticos, se recogieron 2 ml en tubos Oragene® (DNA Genotek, Ottawa, ON) y el ADN se purificó siguiendo las recomendaciones del fabricante.

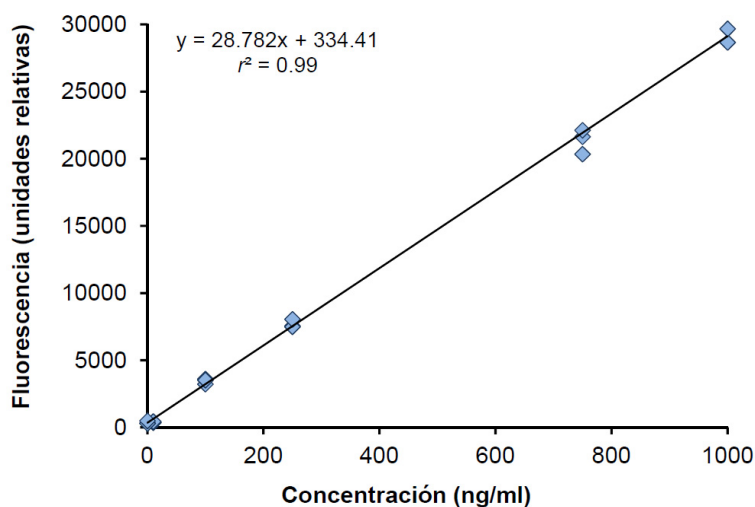
##### 3.3.1.2. Amplificación genómica

Con el fin de obtener ADN suficiente para el genotipado de todos los marcadores a estudiar, una alícuota de la extracción de ADN de aproximadamente 20 ng se empleó para una amplificación genómica con Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare). Esta técnica se basa en la amplificación por desplazamiento múltiple de una hebra de ADN llevada a cabo por la polimerasa  $\Phi$ 29, en presencia de oligonucleótidos hexámeros con secuencias aleatorias. Así, a partir de 20 ng de ADN fue posible obtener entre 4 y 7  $\mu$ g de amplificado genómico en forma de fragmentos >10 kb, en un volumen final de 20  $\mu$ l. Para las sucesivas aplicaciones se diluyó el producto amplificado previa cuantificación.

### 3.3.1.3. Cuantificación de la amplificación genómica

La cuantificación de ADN mediante técnicas espectrofotométricas (como el NanoDrop) no proporciona una medida fiable del resultado de la amplificación genómica sin purificar, donde existe una mezcla de ADN, cebadores hexámeros que no han sido utilizados, proteínas, etc, que enmascaran la emisión en el espectrofotómetro. Por ello, para verificar que las reacciones de amplificación genómica se habían realizado exitosamente y poder llevar a cabo su cuantificación, se desarrolló un método rápido basado en la lectura de fluorescencia de moléculas de unión específica a ADN bicatenario a tiempo real. En base a la cuantificación con PicoGreen® (Molecular Probes, Eugene, OR), desarrollamos una técnica rápida y económica para la cuantificación de ADN en placa de 96 pocillos empleando SYBR Green I (Molecular Probes).

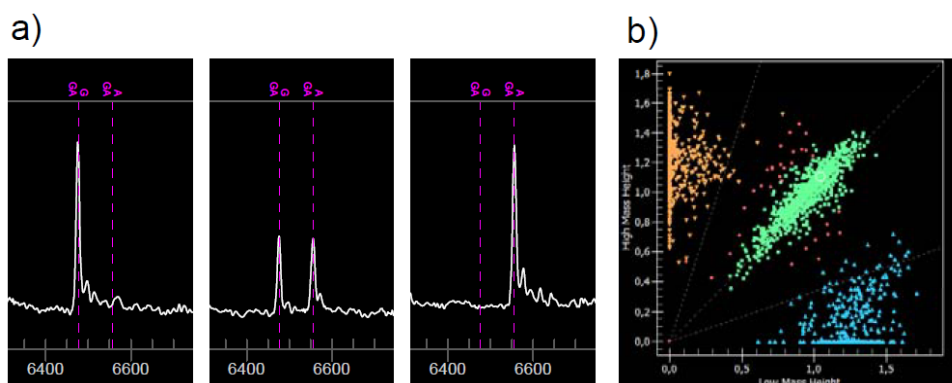
Para ello, se emplearon diluciones 1/5.000 de SYBR Green I (Molecular Probes) y diluciones 1/200 de los productos procedentes de la amplificación genómica, en un volumen de 20  $\mu$ l. Con ello, se registraron los valores de fluorescencia emitida a 520 nm, bajo una excitación de 490 nm y a 25°C, utilizando un sistema de detección de PCR en tiempo real iCycler iQ Real-Time (Bio-Rad Laboratories, Hercules, CA). En cada placa de lectura se incluyó una curva estándar con distintas concentraciones de ADN, preparada a partir de diluciones seriadas del marcador de peso molecular XIV (Roche, Basilea, Suiza) y que incluyó los siguientes puntos: 0, 10, 100, 250, 750 y 1.000 ng/ml. Todas las diluciones se prepararon en el momento de la cuantificación usando Tris 5 mM y EDTA 0,5 mM (pH=7,5). Las concentraciones de ADN se obtuvieron a partir de la fluorescencia emitida (en términos de unidades de fluorescencia relativa) por interpolación en la curva estándar (Figura 9). De esta forma, la lectura de cada placa de 96 pocillos se pudo realizar en 5 segundos.



**Figura 9.** Ejemplo de curva estándar obtenida mediante distintas diluciones del marcador de peso molecular para la interpolación de la muestra de amplificación genómica a cuantificar.



Este espectro es analizado mediante el programa informático MassARRAY Typer 3.4, que realiza la asignación automática de genotipos, al agrupar los individuos para cada uno de los SNPs según sus patrones espectrales (Figura 11).

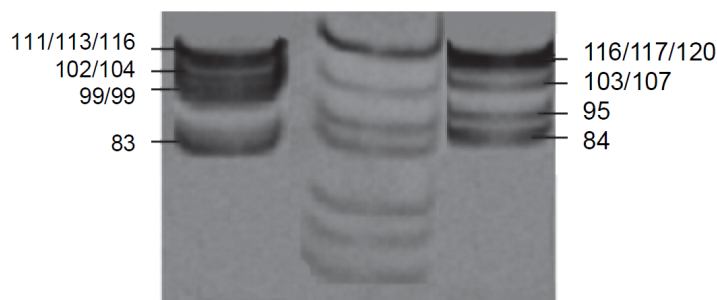


**Figura 11.** a) Ejemplo de espectros para cada uno de los genotipos posibles de un SNP. b) Agrupamiento de las muestras según su espectro en los tres genotipos posibles para ese SNP.

### 3.3.2.2. Genotipado mediante reacciones de SNaPshot<sup>®</sup>

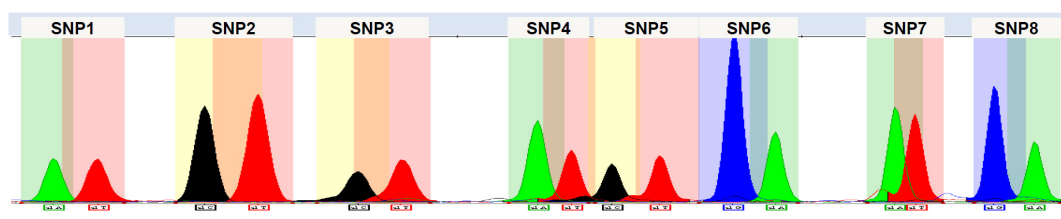
Al igual que la técnica anterior, el SNaPshot<sup>®</sup> consta de una reacción de PCR *multiplex* seguida de una reacción de extensión de una sola base, pero en este caso el tamaño de los fragmentos se resuelve mediante una electroforesis capilar en un secuenciador automático, ya que la mezcla de reacción contiene los cuatro dideoxinucleótidos marcados con fluoróforos diferentes.

Las PCRs *multiplex* se realizaron empleando la enzima HotStartTaq<sup>®</sup> Master Mix Kit (QIAGEN, Valencia, CA) en un volumen final de reacción de 2,5  $\mu$ l en un termociclador Veriti™ 96-Well Fast Thermal Cycler (Applied Biosystems), siguiendo las condiciones: 94°C 15 min para la desnaturalización inicial, seguido de 94°C 20 s, 56°C 30 s, 72°C 1 min repetido durante 45 ciclos y con un paso final de extensión de 72°C 5 min. El diseño de estas PCRs *multiplex* se centró en la amplificación de fragmentos muy pequeños, en muchos casos de tamaño solapante, por lo que su visualización por medio de geles de acrilamida no desnaturalizante al 10% sólo permitió tener indicios de una amplificación exitosa (Figura 12).



**Figura 12.** Ejemplo de un gel de acrilamida correspondiente a dos reacciones distintas de PCR *multiplex*. A la izquierda del marcador de peso molecular V (Roche) se muestra una PCR de 8 amplicones, y a la derecha otra de 7.

Las productos de PCR se purificaron usando ExoSAP-IT® (USB Corp., Cleveland, OH). Diluciones de los mismos se sometieron a reacciones de extensión de una sola base usando SNaPshot® Multiplex Kit reactions (Applied Biosystems), para posteriormente resolver las diferencias de tamaño de los productos de extensión mediante electroforesis capilar en equipos ABI 310 o ABI 3500 (Applied Biosystems). La asignación de genotipos se realizó de manera automática empleando el programa GeneMapper v3.2 (Applied Biosystems) (Figura 13).



**Figura 13.** Ejemplo de genotipado simultáneo de 8 SNPs mediante reacciones *multiplex* de SNaPshot®. Eje x, tiempo de carrera (s); Eje y, intensidad (unidades de fluorescencia relativas).

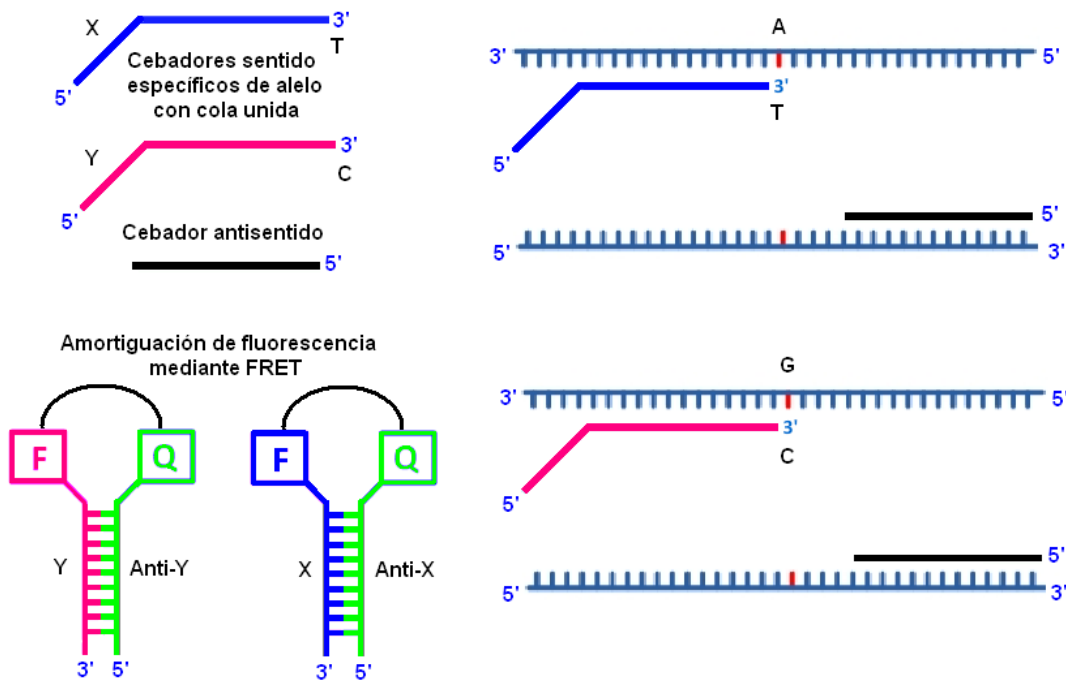
### 3.3.2.3. Genotipado mediante sondas KASPar®

El genotipado mediante sondas KASPar® (KBioscience) consiste en una PCR competitiva con cebadores específicos de alelo, en la que los productos de PCR se detectan mediante un sistema de sondas marcadas mediante dos fluorocromos en un equipo de PCR en tiempo real.

En esta técnica se diseñan dos cebadores sentido (5'-3') distintos mediante el programa PrimerPicker software ([www.kbioscience.co.uk/primer-picker](http://www.kbioscience.co.uk/primer-picker)), uno por cada alelo que se pretende detectar. Estos cebadores específicos de alelo presentan en su extremo 5' una cola de nucleótidos específica y distinta para cada uno de ellos (X e Y en la Figura 14). Además, se diseña un solo cebador antisentido (3'-5') sin modificación alguna. La mezcla de reacción contiene otros 4 cebadores universales. Por un lado un cebador marcado con FAM, un fluoróforo denominado 6-carboxyfluoresceína con absorción máxima a 494 nm y emisión a 518 nm, que es idéntico a la cola X y un cebador antisentido para éste (anti-X) que está

## METODOLOGÍA

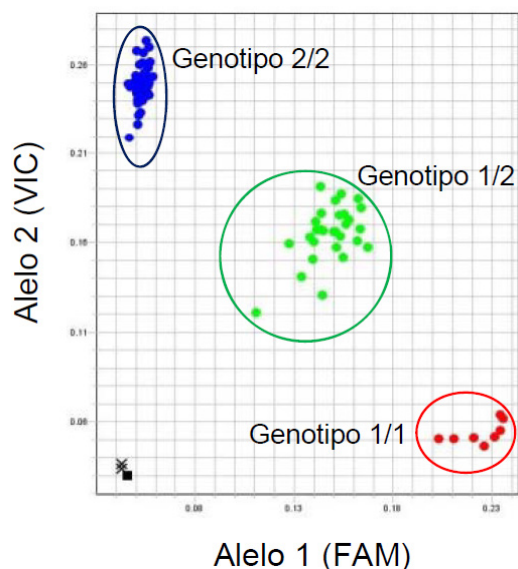
marcado con un compuesto que bloquea la emisión de fluorescencia (denominado en inglés *quencher*). Este bloqueo de la fluorescencia se produce mediante un fenómeno de transferencia de energía de resonancia fluorescente (FRET). Por otro lado, la mezcla contiene un cebador marcado con un fluoróforo cuyo espectro de absorción tiene su máximo a 538 nm y su emisión a 554 nm, denominado VIC<sup>®</sup>. Este cebador marcado con VIC<sup>®</sup> es idéntico a la cola Y, y un cebador antisentido para éste (anti-Y) marcado con un compuesto *quencher* (Figura 14). En los primeros ciclos de la PCR, el ADN es amplificado usando los cebadores específicos de alelo con colas X ó Y, momento en el que no hay emisión de fluorescencia porque los cebadores universales están hibridados por parejas, de manera que el *quencher* amortigua la emisión de fluorescencia por la cercanía con el fluoróforo. El nuevo ADN sintetizado incluirá la cola X o Y en cadena inversa, lo que hace que se una el cebador marcado X ó Y con el compuesto fluorescente correspondiente y que se emita fluorescencia al liberarse del *quencher*.



**Figura 14.** Elementos que intervienen en una reacción de genotipado mediante sondas KASPar.

En el proceso de genotipado mediante sondas KASPar<sup>®</sup> (KBioscience), la detección de la señal de fluorescencia resultante al final de la reacción se realizó mediante un sistema de PCR en tiempo real 7500 Fast Real-Time PCR System (Applied Biosystems). Posteriormente, un análisis de agrupamiento en función de la fluorescencia permitió clasificar a los individuos en uno de los tres grupos de genotipos, mediante el programa SDS 2.0.1 (Applied Biosystems). Este análisis se basa en los distintos espectros de emisión de fluorescencia de los fluoróforos VIC y FAM, dando lugar a tres grupos de individuos: individuos homocigóticos del alelo 1 (marcados sólo con FAM), individuos homocigóticos del alelo 2

(marcados sólo con VIC<sup>®</sup>) e individuos heterocigóticos (con emisión de fluorescencia por VIC<sup>®</sup> y FAM) (Figura 15).



**Figura 15.** Agrupamiento de muestras de acuerdo con la señal de fluorescencia en un ensayo de genotipado mediante sondas KASPar. Eje x, intensidad de fluorescencia del alelo 1, marcado con FAM (unidades de fluorescencia relativas); Eje y, intensidad de fluorescencia del alelo 2, marcado con VIC<sup>®</sup> (unidades de fluorescencia relativas).

### 3.4. Análisis estadístico

#### 3.4.1. Estudio de estratificación poblacional y relación entre la ascendencia genética y el asma

##### 3.4.1.1. Evaluación de la capacidad informativa de los marcadores de ascendencia

Para realizar una evaluación de la capacidad informativa de los polimorfismos de ascendencia mediante distintos estadísticos y calcular los recuentos de alelos y los estadísticos exactos de desviación del equilibrio de Hardy-Weinberg (HWE) (Wigginton *et al.*, 2005), se desarrolló en nuestro laboratorio un programa informático denominado SNPInfostats (Pino-Yanes *et al.*, 2011). Para cada uno de los polimorfismos, SNPInfostats permitió estimar la diferencia absoluta de frecuencia alélica ( $\delta$ ), la distancia genética entre poblaciones medidas mediante  $F_{ST}$  de Weir & Cockerham (Weir, 1996), así como el índice de información  $I_n$  (Rosenberg *et al.*, 2003).

##### 3.4.1.2. Relación entre poblaciones

El programa EIGENSOFT (Price *et al.*, 2006) se utilizó para realizar un análisis de componentes principales (PCA) entre individuos y para realizar un análisis de la varianza (ANOVA) de las diferencias poblacionales de las puntuaciones individuales obtenidas para cada componente principal. Asimismo, se empleó un análisis multidimensional para



representar los  $F_{ST}$  entre pares de poblaciones en dos dimensiones, mediante el programa SPSS ver.15 (SPSS Inc., Chicago, IL).

### 3.4.1.3. Estimaciones de ascendencia poblacional e individual

Para conocer la magnitud de la posible contribución subsahariana al trasfondo genético de la población canaria se realizaron estimaciones de mezcla a nivel poblacional con el programa LEADMIX (Wang, 2003), que permitió calcular dos tipos de medidas: un estimador de mínimos cuadrados ( $m_R$ ), que no tiene en cuenta el muestreo ni la deriva genética entre poblaciones (Roberts & Hiorns, 1965), y el estimador de la máxima probabilidad ( $m_W$ ), que tiene en cuenta los efectos del muestreo y de la deriva genética entre poblaciones y el grado de diferenciación entre las poblaciones parentales antes del evento de mezcla (Wang, 2003).

La ascendencia de cada uno de los individuos se calculó como la probabilidad de pertenecer a una determinada población, mediante una asignación no supervisada de individuos a grupos poblacionales empleando el programa STRUCTURE 2.2 (Pritchard *et al.*, 2000). Para inferir el número de poblaciones ( $K$ ) se realizaron tres análisis independientes con un calentamiento de 50.000 iteraciones y 200.000 repeticiones desde  $K=1$  a  $K=7$ , asumiendo un modelo de frecuencias alélicas no correlacionadas y sin mezcla (tal como se sugiere en el manual del programa STRUCTURE 2.2 para detectar estructuras poblacionales sutiles) y fijando el valor de la frecuencia alélica *a priori* o  $\lambda$  en 1. Con ello, el número de poblaciones se dedujo empleando el método de Evanno *et al.* (2005) que se basa en la tasa de cambio en la estimación de la probabilidad posterior de los datos respecto a los distintos valores de  $K$ . Además, la consistencia de los resultados se valoró con sucesivos análisis con mayor número de iteraciones.

### 3.4.1.4. Inferencia de ascendencia con distintos subconjuntos de polimorfismos

Con el propósito de explorar si al reducir el número de EuroAIMs la información obtenida resultaba similar a la que proporcionaba el conjunto completo de 93 EuroAIMs, se seleccionaron distintos subconjuntos que incluyesen un menor número de marcadores: 23, 46 y 69 marcadores. Para ello, los EuroAIMs se ordenaron en base al promedio informativo de cada conjunto, mediante puntuaciones obtenidas al considerar los valores de  $I_n$  y  $\bar{d}$ , pero sin tener en cuenta el  $F_{ST}$  debido a su redundancia en información con respecto al  $I_n$  ( $r^2=0.99$ ). El programa STRUCTURE se empleó con cada uno de los subconjuntos de EuroAIMs, con los mismos parámetros ya descritos, y en todos los casos identificó correctamente  $K=2$ . Para clasificar a cada individuo como perteneciente a una población u otra se adoptó como límite un valor de mezcla de 0,70. Además, para cada subconjunto de

EuroAIMs se estimaron correlaciones de *Pearson* entre las estimas de mezcla obtenidas con cada subconjunto comparados con el conjunto total de 93 EuroAIMs.

#### 3.4.1.5. Diferencias entre las frecuencias alélicas entre muestras españolas

Para explorar si las frecuencias alélicas de los EuroAIMs diferían significativamente entre muestras españolas, cada uno de estos polimorfismos se consideró como un *locus* candidato de asociación en una simulación de un estudio de asociación caso-control en la que los individuos canarios se consideraron como casos y los peninsulares como controles.

En un primer paso, las frecuencias alélicas se evaluaron mediante el estadístico de tendencias de Cochran-Armitage, que es similar a un estadístico de contraste de diferencias alélicas, pero sin asumir la existencia de HWE (Balding, 2006), y se contabilizó el número de comparaciones que resultaron significativas al nivel  $p=0,05$  para evaluar el desvío de lo esperado por azar.

Con el objetivo de evaluar si la realización de ajustes de estratificación poblacional en los estadísticos de asociación resultaría beneficiosa, se estimó la significación de las diferencias entre frecuencias alélicas incorporando estos ajustes. Para ello, la variable que simulaba el estado caso-control de cada individuo se consideró como una variable dependiente y se emplearon como covariables los genotipos de cada uno de los EuroAIMs y los valores de la primera componente principal (PC1) obtenidos mediante el programa EIGENSOFT. Para ello, los genotipos se codificaron con 0, 1 y 2, en función del número de alelos de frecuencia menor que portasen. Además, se evaluó la alternativa de incluir como covariable la estima de ascendencia norteafricana obtenida para cada individuo mediante el programa STRUCTURE. Estos ajustes se realizaron tanto para el conjunto completo de 93 EuroAIMs como para los distintos subconjuntos reducidos, en todos los casos con modelos de regresión calculados con el programa SNPAssoc (González *et al.*, 2007).

#### 3.4.1.6. Estudio de la relación entre ascendencia genética y asma

Para explorar la asociación entre la ascendencia genética y el asma o fenotipos alérgicos en la población canaria, se analizó un subconjunto de 734 individuos de la muestra total del estudio de asociación que declaró tener dos generaciones de ancestros nacidos en Canarias: 391 asmáticos y 343 controles. En este estudio se empleó únicamente el programa EIGENSOFT (Price *et al.*, 2006; Price *et al.*, 2008) puesto que presenta un mayor poder discriminatorio para detectar diferencias entre poblaciones estrechamente relacionadas (Heath *et al.*, 2008; Li & Yu, 2008) que algoritmos alternativos que tratan de clasificar a los individuos en poblaciones discretas (como por ejemplo STRUCTURE). Éste se utilizó para realizar los PCA y para las comparaciones de las diferencias poblacionales de las puntuaciones individuales obtenidas para la PC1 mediante un ANOVA. Para el análisis

de fenotipos cuantitativos se emplearon modelos de regresión lineal mediante la utilización del programa SPSS 15.0 (SPSS Inc.).

### 3.4.2. Estudios de asociación de genes candidatos con el asma

#### 3.4.2.1. Estimación del poder estadístico del estudio de asociación a priori

El poder estadístico se estimó empleando el programa informático Power 3 (García-Closas & Lubin, 1999), obteniéndose que el tamaño de la muestra utilizada (N=1.878 individuos) proporcionaría un poder estadístico de un 80% para detectar efectos, estimados en forma de razones de ventaja (ORs), de al menos 1,45 veces en variantes génicas con frecuencias de al menos un 20% en la población.

#### 3.4.2.2. Análisis de los datos demográficos, de morbilidad y clínicos

Para evaluar la existencia de posibles diferencias entre los grupos de casos y controles en cuanto a los datos demográficos, de morbilidad y clínicos, se empleó el paquete estadístico SPSS 15.0 (SPSS Inc.). Se utilizó el estadístico de  $\chi^2$  para las variables categóricas, y el estadístico no paramétrico de Mann-Whitney para la comparación de aquellas variables continuas.

#### 3.4.2.3. Control de la calidad de los datos de genotipado

Para cada SNP genotipado se estimó su posible desviación del HWE mediante el estadístico exacto de Wigginton *et al.* (2005) en casos y en controles por separado. Se consideró un umbral para descartar SNPs que se desviasen de HWE en controles en función del número de SNPs analizados, adoptando una corrección de tipo Bonferroni para ajustar por el número de estadísticos realizados. Además se estimó una tasa de completado de genotipos (CR, del inglés *completion rate*) por SNP e individuo, fijando un umbral de descarte de SNPs o individuos en un CR<85%. Finalmente, se calculó la tasa de discordancia entre genotipos duplicados y su CI al 95%.

#### 3.4.2.4. Asociación de tSNPs individuales con la enfermedad

Las asociaciones de SNPs individuales con la enfermedad se analizaron bajo un modelo de herencia aditivo por medio del estadístico de tendencias de Cochran-Armitage, que asume que la adición de cada alelo de riesgo al genotipo del individuo aumenta proporcionalmente el riesgo de padecer la enfermedad (Sasieni, 1997). El nivel de significación se obtuvo con el programa SNPing para STATISTICA, desarrollado en nuestro laboratorio (Sun *et al.*, 2010).

Las asociaciones detectadas se juzgaron en el contexto de las múltiples comparaciones realizadas mediante la estimación de tasas de falso descubrimiento (FDR)

usando el programa QVALUE (Storey & Tibshirani, 2003) en el entorno R 2.8.1 (R Development Core Team, 2008). El FDR indicó la tasa de falsos positivos (fijada en un 5%) en los que se incurre cuando un resultado fue considerado significativo.

Para obtener una medida del efecto de las asociaciones detectadas, se determinaron los ORs y sus 95% CI. En este contexto, el OR se define como la razón entre la probabilidad de que se produzca la enfermedad en presencia de un alelo de riesgo frente a que no ocurra al no portar dicho alelo. Para ello, y para la realización de los ajustes por covariables que pudieran confundir la asociación, se emplearon regresiones logísticas utilizando el programa SNPAssoc (González *et al.*, 2007). Éstas también fueron utilizadas para determinar la independencia de las asociaciones de distintos SNPs de un mismo gen, utilizando una eliminación hacia atrás.

#### 3.4.2.5. Ajustes de la presencia de estratificación poblacional

Todas las muestras del estudio de asociación, tanto casos como controles, se genotiparon para los 93 EuroAIMs. Los genotipos se emplearon para realizar un PCA con el programa EIGENSOFT (Price *et al.*, 2006) usando también los datos de las muestras analizadas como parte del estudio de estratificación inicial. Las puntuaciones correspondientes a la PC1 se emplearon en el ajuste de estratificación poblacional mediante su introducción como una covariable en un modelo de regresión logística calculado con el programa SNPAssoc (González *et al.*, 2007).

#### 3.4.2.6. Imputación de variantes no genotipadas: evaluación inicial y posterior análisis de asociación

Además de evaluar la asociación de variantes genotipadas, la información de los tSNPs se empleó en la imputación de variantes no genotipadas y su análisis de asociación con la enfermedad utilizando el programa TUNA (Wen & Nicolae, 2008), para proporcionar así una mayor resolución de los genes candidatos analizados.

El programa TUNA (Wen & Nicolae, 2008) permite estimar las frecuencias alélicas de variantes no genotipadas y estudiarlas en asociación con la enfermedad, siempre que se disponga de los patrones de LD del gen en una muestra de referencia. El programa estima inicialmente un parámetro que mide la correlación multilocus entre marcadores ( $M_D$ ), una medida similar al  $r^2$  entre parejas de SNPs y que representa el LD existente entre varios *loci*. Este valor indica si el conjunto de tSNPs genotipados permite extraer información fiable de una variante no genotipada, o lo que es lo mismo, el contenido informativo que posee el conjunto de tSNPs para un SNP dado. Wen & Nicolae (2008) sugirieron analizar en asociación indirecta con enfermedad sólo aquellos SNPs con  $M_D \geq 0,7$ . Para los SNPs

escogidos, el programa estima la frecuencia de su alelo menos frecuente, a partir de la frecuencia acumulada de haplotipos que están en alto LD con ese alelo, ponderados por un peso calculado a partir de la base de LD de la población de referencia.

Con el fin de seleccionar qué parámetros podrían indicar que un SNP podría ser predicho con un error aceptable (menor al 5%), antes de aplicar este método en los estudios de asociación se realizó una evaluación de la eficacia del método sobre los datos de re-secuenciación del gen *IRAK3* en la población española. Además, dado que no siempre es posible secuenciar los genes candidatos en los estudios de asociación, sino que frecuentemente se recurre a datos de referencia de bases de datos como el HapMap para la selección de tSNPs (Frazer *et al.*, 2007), también se evaluaron los errores en las estimas alélicas al emplear este tipo de datos como muestra de referencia para los patrones de LD.

Tras elegir los parámetros más adecuados para seleccionar qué variantes no genotipadas podían ser fiablemente predichas (con errores menores del 5%), éstas fueron evaluadas indirectamente estimando sus frecuencias a partir de los datos genotípicos de los tSNPs en las muestras de casos y controles. Para ello se empleó como referencia los datos de re-secuenciación y/o del HapMap (Frazer *et al.*, 2007) transformados en haplotipos mediante el programa PHASE 2.1 (Stephens *et al.*, 2001).

### 3.4.2.7. Análisis de haplotipos

Para cada gen se estudió si existían evidencias de asociación en forma de haplotipos de tSNPs usando un estadístico global con el programa WHAP (Purcell *et al.*, 2007). La comparación de frecuencias de haplotipos dirigidos a una región de interés, ya fuese por el interés en la replicación de resultados de algún estudio previo o porque hubiesen mostrado funcionalidad, se hizo con el programa HAPSTAT (Lin *et al.*, 2008), que tiene la ventaja de permitir considerar la posible asociación del haplotipo más frecuente en la población con la enfermedad. La estimación de los efectos de los haplotipos en forma de ORs y su 95% CI se realizó mediante regresión logística con el programa THESIAS (Tregouet *et al.*, 2004).

### 3.4.2.8. Adición secuencial de casos

Para evaluar si las asociaciones encontradas podrían tener un mayor efecto al considerar variables cuantitativas en la definición de los casos, se empleó un análisis de adición secuencial de casos por orden ascendente, en función de la variable cuantitativa considerada. En éste, se compararon distintos subconjuntos de casos definidos por la variable cuantitativa con respecto a los controles, realizando 1.000 permutaciones para ajustar por las comparaciones múltiples realizadas. En este análisis se empleó un algoritmo específico (Macgregor *et al.*, 2006) para el entorno R (R Development Core Team, 2008). Este algoritmo trabaja en primer lugar ordenando los casos de forma ascendente en función

de la variable cuantitativa considerada. Seguidamente los casos con los menores valores de la variable cuantitativa se añaden en el análisis y se calculan los estadísticos de asociación para el subgrupo de casos considerados frente a la totalidad de controles, usando un estadístico  $\chi^2$  a partir de una tabla de contingencia. Esto se repite para distintos subconjuntos de casos, reteniéndose de todos los estadísticos realizados el de menor  $p$ -valor. Finalmente, para establecer la significación de los  $p$ -valores obtenidos con el procedimiento anterior, los genotipos se permutan en el conjunto de la muestra entre casos y controles mediante 1.000 permutaciones. El  $p$ -valor empírico, corregido por las múltiples comparaciones realizadas, se obtiene de la proporción de réplicas permutadas que rinden  $p$ -valores menores que los observados en el conjunto de datos real.

#### 3.4.2.9. Meta-análisis

Con el fin de evaluar si las asociaciones encontradas en este estudio para alguno de los genes eran consistentes con las descritas en la bibliografía, esto es, mismos alelos asociados y similares efectos (protección o riesgo), se realizó un análisis conjunto incluyendo datos de nuestro estudio y los datos disponibles de estudios anteriores. Para ello, se empleó un meta-análisis de efectos fijos o aleatorios con la librería *rmeta* 2.16 para el entorno R (R Development Core Team, 2008).

#### 3.4.2.10. Representación gráfica de las asociaciones

Para la representación gráfica de la asociación de SNPs individuales se empleó la librería LocusZoom (Pruim *et al.*, 2010) para el entorno R (R Development Core Team, 2008), que permite la representación del gen de interés frente a los valores de asociación en escala logarítmica, además de representar el LD de un SNP determinado con respecto al resto de SNPs del gen que han sido estudiados en asociación. Por el reducido número de muestras en los que se basó la secuenciación de estos genes, para estas representaciones se optó por la utilización de los valores de LD procedentes de 1KGP (The 1000 Genomes Project Consortium, 2010).

#### 3.4.2.11. Evaluación de la cobertura de los genes candidatos en los arrays de genotipado empleados en los GWAS de asma

Dado que, salvo por la *IL13*, los genes candidatos estudiados en este trabajo no han sido revelados como genes de susceptibilidad en los GWAS de asma realizados hasta la fecha, se evaluó si la cobertura de estos genes en los *arrays* de genotipado empleados hasta este momento había sido adecuada. Para ello, se procedió a descargar el contenido de SNPs de los *arrays* de la base de datos UCSC Genome Bioinformatics (<http://genome.ucsc.edu>) para las mismas regiones cromosómicas consideradas en la

selección de tSNPs. Además, se obtuvieron los genotipos para los SNPs contenidos en las mismas regiones génicas para individuos CEU de la base de datos del HapMap (Frazer *et al.*, 2007; Altshuler *et al.*, 2008) y de 1KGP, estos últimos mediante la utilización de la herramienta ENGINES (<http://spsmart.cesga.es/engines>) (Amigo *et al.*, 2011).

Mediante el programa Haploview 4.2 (Barrett *et al.*, 2005) se evaluó la cobertura de los SNPs con  $MAF \geq 5\%$  contenidos en los *arrays* sobre los datos de HapMap y 1KGP, tanto por el LD por parejas de SNPs como por haplotipos de 2 ó 3 SNPs, considerando un umbral de cobertura de  $r^2 \geq 0,8$ . El valor de porcentaje de cobertura se obtuvo como la relación entre el número de SNPs con  $MAF \geq 5\%$  podrían analizarse con el *array* (ya sea por estar genotipados o por poder ser estudiados indirectamente), respecto al número total de SNPs con  $MAF \geq 5\%$  en esa región.

### 3.4.2.12. Cálculo de puntuaciones de riesgo genético conjunto

Finalmente, se determinó si la acumulación de alelos de riesgo en los distintos genes asociados proporcionaría un mayor riesgo a padecer asma que cada variante considerada por separado. Para ello, se calcularon distintas puntuaciones de riesgo genético (GRS, del inglés *genetic risk score*). En este cálculo se codificaron los genotipos de los SNPs asociados independientemente de cada gen con el asma, según los alelos de riesgo portados por cada individuo, asignando valores de 0, 1 o 2, que posteriormente fueron sumados. Todos los SNPs incluidos en el modelo se consideraron de manera independiente y con un mismo peso. Por último, se aplicaron modelos de regresión logística o lineal para estudiar la asociación del GRS con la susceptibilidad del asma y la edad de comienzo empleando el programa SPSS ver.15 (SPSS Inc.).

### 3.4.2.13. Medidas de la capacidad predictiva de las puntuaciones de riesgo genético

Para cuantificar no sólo el efecto de la asociación del GRS en términos de razón de riesgos sino para evaluar su poder predictivo, se determinaron distintos parámetros. En diagnóstico, la capacidad discriminativa de un biomarcador se evalúa principalmente por medio de dos parámetros: sensibilidad y especificidad (Kraft *et al.*, 2009). Mientras que la sensibilidad es la probabilidad de obtener un resultado positivo en un individuo realmente afectado por la enfermedad, la especificidad es la probabilidad de obtener un resultado negativo en un individuo que no padece la enfermedad. Una prueba diagnóstica perfecta presenta una sensibilidad de uno (todos los enfermos dan un resultado positivo) y una especificidad de uno (todos los individuos sanos dan negativo) (Evans *et al.*, 2009). En una prueba compuesta por varias determinaciones, como lo es la elaboración de un perfil genético de riesgo, los resultados negativos y positivos se evalúan estableciendo un punto de corte para la probabilidad de desarrollar la enfermedad. Es posible determinar la

sensibilidad y la especificidad en función de distintos puntos de corte y realizar una representación gráfica por medio de curvas ROC (acrónimo de *Receiver Operating Characteristic*, o Característica Operativa del Receptor), cuyo origen se debe a las telecomunicaciones y los problemas asociados a la recepción de señales. Las curvas ROC son una representación gráfica de la sensibilidad frente a (1-especificidad) para la clasificación de individuos en enfermos o sanos en función del umbral de discriminación (Pepe, 2003).

En el caso de las puntuaciones de riesgo genético, se representó en cada punto de la curva ROC la probabilidad de desarrollar la enfermedad ante un punto de corte de alelos de riesgo definido ( $k$ ) al asignar a aquellos individuos portadores de más de  $k$  alelos de riesgo en el grupo de alto riesgo y a los portadores de  $k$  alelos o menos en el grupo de bajo riesgo. El área bajo la curva ROC (AUC) es una medida de la capacidad discriminativa del marcador, siendo  $AUC=0,5$  un valor que indica poca capacidad discriminativa y  $AUC=1$  un poder discriminatorio perfecto (Kraft *et al.*, 2009). Este área puede interpretarse como la probabilidad de que ante dos individuos, uno sano y otro enfermo, el enfermo sea identificado. Por ello, un  $AUC=0,5$  implica que sólo la mitad de individuos serían clasificados correctamente, algo que no difiere de lo esperado por azar (Janssens *et al.*, 2006). Se considera que un  $AUC=0,8$  podría ser útil para la identificación de individuos con alto riesgo de desarrollar la enfermedad (Janssens *et al.*, 2006).

Finalmente, se determinó la proporción de la varianza de la enfermedad explicada por el modelo de GRS por medio del coeficiente de determinación de Nagelkerke a partir de la regresión logística (Nagelkerke, 1991). Tanto las curvas ROC como el coeficiente de Nagelkerke se determinaron empleando el programa SPSS ver.15 (SPSS Inc.).



## ***4. RESULTADOS***

---

## 4. RESULTADOS

### 4.1. Estudio de la estratificación poblacional en España originada por una distribución desigual de mezcla genética norteafricana

En el primer artículo incluido en esta tesis se llevó a cabo una evaluación de la mezcla norteafricana presente en la población española previo al estudio de asociación de genes candidatos de susceptibilidad al asma. De esta manera, se determinó si la existencia de estratificación poblacional podría sesgar los resultados de los estudios posteriores. Además, en este trabajo determinamos la capacidad de un conjunto de marcadores para distinguir entre ascendencia norteafricana y peninsular, corregir los estadísticos de asociación y explorar si existe relación entre el riesgo al asma en Canarias y la ascendencia norteafricana.

Tras genotipar 93 marcadores informativos de ascendencia en individuos españoles (peninsulares y canarios) e individuos norteafricanos y realizar distintos análisis a nivel poblacional e individual, no detectamos mezcla subsahariana, pero sí una influencia norteafricana desigual en la población española. En el caso de los individuos peninsulares, la mezcla genética norteafricana se estimó en un 5%, mientras que en individuos canarios resultó ser de un 17% como media, aunque con una gran variación entre distintos individuos (desde 0% hasta 96%). Como consecuencia de estas diferencias, una comparación de frecuencias alélicas entre muestras de individuos peninsulares y canarios, simulando un estudio de asociación con enfermedad, mostró un exceso de marcadores con diferencias significativas. Sin embargo, este exceso de  $p$ -valores significativos se pudo disminuir adecuadamente al corregir los estadísticos con un pequeño número de marcadores de ascendencia.

Los resultados de este estudio han sido publicados en la revista *PLoS One* con el título “*North African influences and potential bias in case-control association studies in the Spanish population*”.

# North African Influences and Potential Bias in Case-Control Association Studies in the Spanish Population

María Pino-Yanes<sup>1,2</sup>, Almudena Corrales<sup>1,2</sup>, Santiago Basaldúa<sup>2</sup>, Alexis Hernández<sup>3</sup>, Luisa Guerra<sup>4</sup>, Jesús Villar<sup>1,5,6</sup>, Carlos Flores<sup>1,2\*</sup>

**1** CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain, **2** Research Unit, Hospital Universitario N.S. de Candelaria, Tenerife, Spain, **3** Instituto Nacional de Toxicología y Ciencias Forenses, Delegación de Canarias, Tenerife, Spain, **4** Hematology Service, Hospital Universitario Dr. Negrín, Las Palmas de Gran Canaria, Spain, **5** Multidisciplinary Organ Dysfunction Evaluation Research Network, Research Unit, Hospital Universitario Dr. Negrín, Las Palmas de Gran Canaria, Spain, **6** Keenan Research Center, St. Michael's Hospital, Toronto, Canada

## Abstract

**Background:** Despite the limited genetic heterogeneity of Spanish populations, substantial evidences support that historical African influences have not affected them uniformly. Accounting for such population differences might be essential to reduce spurious results in association studies of genetic factors with disease. Using ancestry informative markers (AIMs), we aimed to measure the African influences in Spanish populations and to explore whether these might introduce statistical bias in population-based association studies.

**Methodology/Principal Findings:** We genotyped 93 AIMs in Spanish (from the Canary Islands and the Iberian Peninsula) and Northwest Africans, and conducted population and individual-based clustering analyses along with reference data from the HapMap, HGDP-CEPH, and other sources. We found significant differences for the Northwest African influence among Spanish populations from as low as  $\approx 5\%$  in Spanish from the Iberian Peninsula to as much as  $\approx 17\%$  in Canary Islanders, whereas the sub-Saharan African influence was negligible. Strikingly, the Northwest African ancestry showed a wide inter-individual variation in Canary Islanders ranging from 0% to 96%, reflecting the violent way the Islands were conquered and colonized by the Spanish in the XV century. As a consequence, a comparison of allele frequencies between Spanish samples from the Iberian Peninsula and the Canary Islands evidenced an excess of markers with significant differences. However, the inflation of  $p$ -values for the differences was adequately controlled by correcting for genetic ancestry estimates derived from a reduced number of AIMs.

**Conclusions/Significance:** Although the African influences estimated might be biased due to marker ascertainment, these results confirm that Northwest African genetic footprints are recognizable nowadays in the Spanish populations, particularly in Canary Islanders, and that the uneven African influences existing in these populations might increase the risk for false positives in association studies. Adjusting for population stratification assessed with a few dozen AIMs would be sufficient to control this effect.

**Citation:** Pino-Yanes M, Corrales A, Basaldúa S, Hernández A, Guerra L, et al. (2011) North African Influences and Potential Bias in Case-Control Association Studies in the Spanish Population. PLoS ONE 6(3): e18389. doi:10.1371/journal.pone.0018389

**Editor:** Dennis O'Rourke, University of Utah, United States of America

**Received:** October 27, 2010; **Accepted:** March 4, 2011; **Published:** March 30, 2011

**Copyright:** © 2011 Pino-Yanes et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was funded by grants to CF from Instituto de Salud Carlos III (PI081383) and Fundación Canaria de Investigación y Salud (FUNCIS, 23/07), and by a specific agreement between Instituto de Salud Carlos III and Gobierno de Canarias (EMER07/001). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: cflores@ull.es

## Introduction

Populations inhabiting the Iberian Peninsula have been influenced by the same major human prehistoric migrations that have affected the rest of European populations, regardless of the extent to which Neolithic expansions from the Near East have influenced their genetic makeup [1–5]. Y-chromosome studies have indicated that such migrations have influenced uniformly the Iberian genetic background [6]. Thus, with the exception of a few isolates, the Basques being the best characterized representatives [7], populations inhabiting the Iberian Peninsula show a substantial genetic homogeneity [1,6,8,9].

Despite this, several lines of evidence support the existence of identifiable unequal African influences in populations from Spain,

both from the Iberian Peninsula and the Canary Islands. The excess of diversity observed in Southwestern Europeans for genome-wide autosomal haplotypes was interpreted as resulting from direct migrations from North Africa across the Mediterranean [10]. In addition, mitochondrial DNA (mtDNA) and Y-chromosome studies have revealed geographically clustered North African influences of about 8–10% in Iberia [6,8,11–13], with somewhat higher estimates for populations of the Northwestern and Southern regions [6,8,12,14] that have been typically reconciled with a main historical migration from the nearby Northwest Africa as part of the Islamic rule starting in 711 CE and lasting seven centuries [15]. Besides, genetic footprints of important African influences have been demonstrated for particular Spanish populations such as the Canary Islanders, for

which many studies have suggested a genetic influence of 22–38% from Northwest Africans, and less than 5% from sub-Saharan Africans [16–21]. The historical admixture of Spanish colonizers with aborigines related to Northwest African Berbers, and with sub-Saharan Africans introduced as a result of slave trade, has been postulated as an explanation to account for the peculiarities of this population [17].

These major documented historical influences from divergent populations such as those from the nearby North Africa [22,23] are likely to have introduced subtle population differences among Spanish that might be considered in genetic epidemiology studies [13]. Accounting for such genetic differences (i.e. population stratification) is imperative to reduce false positive or negative results in case-control and cohort studies of association of genetic variants with disease [24,25]. In addition, a better characterization of the different genetic strata that are present in the population would enable future studies to address if genetic factors might underlie disparities among Spanish populations for the incidence of complex illness such as asthma [26], type 2 diabetes [27], and hypertension [28].

Ancestry informative markers (AIMs), i.e. genetic loci showing large allele frequency differences between populations, allow accurate apportioning of genetic influences in populations [29] and are useful to efficiently account for population stratification in genetic epidemiology studies with unrelated individuals where dense genotype data is not available [30]. In studies where samples from different European populations are compared, population stratification effects can be controlled by using a few hundred autosomal AIMs, termed EuroAIMs [31], which recover the largest features of the north-northwest (NNW) to south-southeast (SSE) axis of genetic differentiation in Europe [30,31]. A set of AIMs ascertained to specifically distinguish North African and European ancestries is not available in the literature. However, given that AIMs tend to be informative for ancestry inference within several world regions [29], we reasoned that previously selected EuroAIMs were likely to be informative for ancestry as well in these populations. Based on this principle, here we first assessed the ability of EuroAIMs to distinguish Spanish, Northwest and sub-Saharan Africans, and next utilized them to identify African influences in Spanish populations. We finally compared allele frequencies between Spanish to illustrate the potential effects of including samples with uneven African influences in population-based association studies.

**Results**

The potential ancestry informativeness of EuroAIMs for African and Spanish populations was initially evidenced by comparing genome-wide data from reference samples from the Human Genome Diversity Panel (HGDP) [32] (see explanations in Text S1). Subsequently, samples from outbred populations from Northwest Africa and Spain (from the Iberian Peninsula and the Canary Islands) were genotyped for 93 EuroAIMs. In the following, samples and populations from any part of Spain, either from the Iberian Peninsula or from the Canary Islands, will be referred to as Spanish. We reserved the term Iberian to refer to samples and populations from the Iberian Peninsula. Out of the 93 EuroAIMs, seven markers departed significantly from Hardy-Weinberg equilibrium (HWE) in at least one of these populations (Table S1). However, after considering the multiple comparisons using a Bonferroni-like correction (significance at  $p$ -value = 0.00054), only two markers departed significantly from HWE: rs1073321 in Canary Islanders and rs7277342 in Northwest Africans. Given that not a single marker deviated from

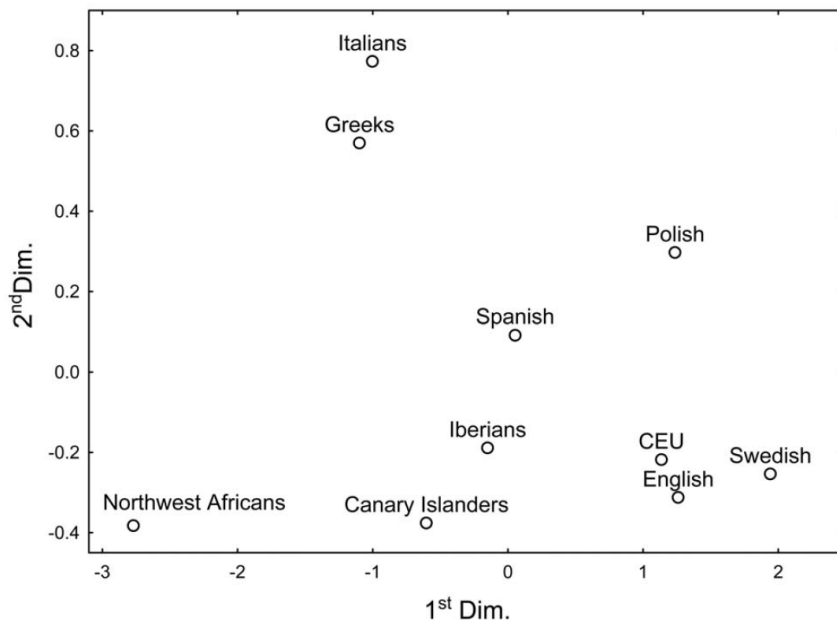
HWE after this correction in all three genotyped samples, and that genotyping was performed simultaneously for all samples obtaining similarly high completion rates ( $\geq 97\%$ ), we interpreted that HWE departures were more likely to be related to chance and retained all 93 EuroAIMs for further studies. On average, genetic differentiation levels ( $F_{ST}$ ) obtained for the comparison of the 93 EuroAIMs between Northwest Africans and Iberians, Northwest Africans and Yoruba Nigerians (YRI), and Iberians and YRI were 0.0422, 0.100 and 0.255, respectively (further information in Table S1). The three estimates were slightly lower when Canary Islanders were considered for comparisons instead of Iberians. These differentiation levels strongly support that there is enough information on the EuroAIMs set to dissect the sub-Saharan Africans from Iberians and Northwest Africans, and also to distinguish the latter two populations.

**Population and individual clustering using EuroAIMs**

To explore to which extent population groups were separated from each other, population and individual-based analyses were performed. For these analyses, sub-Saharan African populations were not included due to their considerable divergence to the rest of populations. Multidimensional scaling analysis of pairwise population  $F_{ST}$  genetic distances revealed three well separated clusters of populations (Figure 1): a clear separation of Northwest Africans from European populations, and of Southern from Northern European populations in agreement with previous observations [30,33]. Principal component analysis (PCA) of individuals revealed a similar clustering pattern (Figure 2) with two significant principal components (PCs) accounting for 68.5% of variance and clearly separating Northwest Africans from Europeans: PC1 distinguishing Northwest Africans from NNW Europeans ( $p = 6.23E-17$ ) and PC2 differentiating Northwest Africans from SSE Europeans ( $p = 3.88E-17$ ). Spanish populations were assigned at intermediate positions in the NNW-SSE axis of European differentiation, with Canary Islanders clustering in their vicinity, albeit showing a slight shift towards Northwest Africans (Figure 2). As a support for the admixed origin of Canary Islanders, restricting the PCA to Iberians, Northwest Africans and Canary Islanders only revealed one significant axis of variation separating Iberians from Northwest Africans ( $p = 8.93E-17$ ) (Figure 2), as would be expected for a typical admixed population because of the linear mixing of allele frequencies in the parental populations [34].

Among Spanish, the  $F_{ST}$  distances of Canary Islanders to Iberians and to a Spanish sample without information of sampling location [30] were at least four times larger than the distance between the latter samples (Table S2). Notably, small  $F_{ST}$  distances were obtained for different partitions of the Iberian sample (not shown), including a separation on an East-West axis as noted in a previous study [12]. Thus, according to these markers, Spanish populations show a reduced genetic heterogeneity, with Canary Islanders constituting the main source of population stratification among the samples analyzed. Although these markers were ascertained to recover the largest features of the NNW-SSE axis of genetic differentiation in Europe and not dissect the genetic heterogeneity of Spanish populations, these results are in agreement with results from previous studies using alternative marker sets [9].

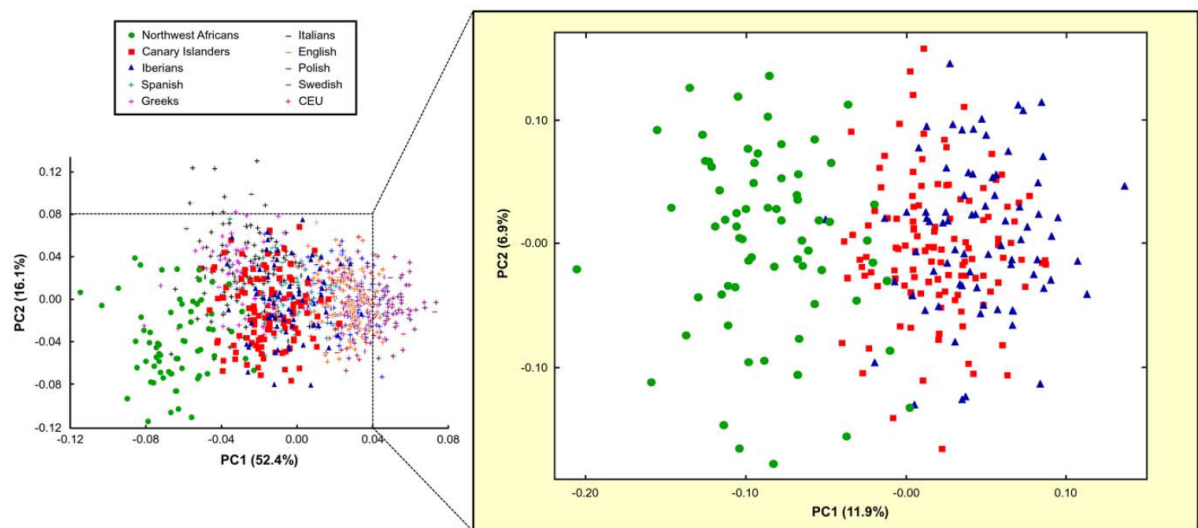
To verify that population relationships were not affected by selection processes, we re-assessed population and individual-based analyses excluding the EuroAIMs with potential effects on gene function according to the SNP Function Portal [35] (i.e. predicting coding non-synonymous changes, disrupting predicted miRNA target sites or being located within gene UTRs that can



**Figure 1.  $F_{ST}$ -based multidimensional scaling plot.** The  $r^2$  of  $F_{ST}$  distances to plot-derived distances was 0.995. The stress value was 0.036, indicating that additional dimensions were not necessary. Spanish, samples from Price et al. [30]; CEU, Utah residents with ancestry from northern and western Europe from HapMap [43]. doi:10.1371/journal.pone.0018389.g001

affect mRNA stability) (Table S3). The results with or without the EuroAIMs with predicted function were largely similar, albeit the distinction of Northwest Africans from Southern European populations was not as apparent when these markers were excluded from the analyses (not shown). Thus, given that we

aimed to accurately apportion the African influences in Spanish populations, we retained all 93 EuroAIMs for further studies. Taken together, these observations support the genetic discontinuity observed between Northwest Africans and their closest European neighbors [23,36], and the usefulness of



**Figure 2. Plot of the top two principal components from the analysis of populations.** Results from the analysis based on the 93 EuroAIMs restricted to Iberians, Northwest Africans and Canary Islanders are represented on the right panel. The percentage of explained variation is indicated in each axis. Spanish, samples from Price et al. [30]; CEU, Utah residents with ancestry from northern and western Europe from HapMap [43]. doi:10.1371/journal.pone.0018389.g002

previously ascertained EuroAIMs to recover the genetic ancestry of Northwest Africans and Europeans.

**African influences in Spanish populations**

As there is not a formal way to recognize if minor individual ancestries are accurate measurements or artifacts of the methods used, we first assessed the influences at population level with the only purpose of deciding which populations to include in STRUCTURE analyses. For that, we evaluated African influences in Canary Islanders as, among Spanish populations, it would be more likely to detect such influences in this population based on the historical records. Given that results from a three-population model of genetic contributions in Canary Islanders indicated that a direct sub-Saharan African influence in this population was negligible (Table 1), and that sub-Saharan Africans were clearly differentiated from the other two parental populations (Iberians and Northwest Africans) by means of an unsupervised assignment of individuals to populations (Figure S1), we excluded sub-Saharan Africans from further analyses. The Northwest African influence in Canary Islanders was estimated in ≈23% (Table 1), which was in the range of previous estimates [17,21]. Nevertheless, it is likely that this estimate is biased upwards. Note that a portion of the Northwest African component in Canary Islanders might be due to the Northwest African influences in the Iberian colonizers, given the gene flow between Northwest Africans and Iberians (see results below).

After excluding sub-Saharan Africans from the study, STRUCTURE identified two populations corresponding to Northwest Africans and Spanish (Figure 3). With few exceptions, which may represent unreported recent admixture, Northwest African and Iberian samples showed overwhelming contribution from a single population. Similar results were observed using subsets of as few as 23 of the markers ranking higher for different measures of ancestry informativeness, albeit at the cost of as much as 11% of individuals being classified in the incorrect population (see Text S2 for further details). Strikingly, the Northwest African influence in Canary Islanders was 17.4% on average (within the 95% confidence interval of population-based methods) while it was significantly lower in Iberians (≈5%; two-tailed t-test  $p = 1.00E-6$ ) (Table 2). In addition, a large inter-individual variation was observed for the Northwest African ancestry among Canary Islanders (range 0.0–95.7%). Surprisingly, about 9% of Canary Islanders showed ≥70% of Northwest African ancestry. The average Northwest African ancestry was neither different between the Canary Islanders collected for this study and those from the National Spanish DNA Bank (two-tailed t-test  $p = 0.488$ ), nor among the samples from the different islands (ANOVA  $p = 0.213$ ) (Table 2). Note, however, that estimates by island must be interpreted with caution given the small sample sizes. Comparable results were obtained with subsets of as few as 69 of the markers ranking higher

for different measures of ancestry informativeness (Text S2). With fewer markers, the estimated Northwest African influences in Canary Islanders resulted severely biased upwards as a consequence of the reduction of information.

**Controlling population stratification in population-based association studies in Spanish populations using EuroAIMs**

To limit population stratification effects in a typical case-control association study with unrelated individuals, samples are usually matched by political boundaries [37]. Because of the estimated disparities in Northwest African influences among the Spanish populations, we expected that multicentric studies, with country-wide sampling schemes, would be at risk of an increased rate of false positives. As a proof of principle, a comparison of the allele frequency differences of the 93 EuroAIMs between Iberians and Canary Islanders using the Cochran-Armitage trend test resulted in 12 tests with  $p < 0.05$  (Table 3). At 5% type-I error rate, only 4.65 markers were expected to be significant by chance, indicating the existence of an excess of markers with large differences between both samples ( $\chi^2$  test  $p = 0.0375$ ). To explore if this effect was controlled by using the ancestry information derived from EuroAIMs, we adjusted these comparisons using logistic regression models including as covariates either the STRUCTURE Northwest African individual ancestries or the PCA scores. Adjusting for STRUCTURE estimates, based on the full marker set or any of the subsets of markers ranking higher for different measures of ancestry informativeness, barely decreased the number of significant tests (Table 3), while significance levels dropped considerably (Figure S2). As an example, the smallest  $p$ -value, which was obtained for the marker with the largest  $F_{ST}$  between Iberians and Northwest Africans (0.4153), increased from  $p = 0.0002$  with the unadjusted test (the Cochran-Armitage trend test) to  $p = 0.006$  after the adjustment. The first  $p$ -value will remain significant after a Bonferroni-like correction for the multiple comparisons, but the second will not. Nevertheless, the adjustment of comparisons using the PC1 scores not only resulted in reduced significance levels but also in an effective control of false positives at 5% rate (Table 3; Figure S2). Results did not improve if the scores from the PC2 were also used for the adjustments (not shown).

**Discussion**

We have illustrated that a few dozen EuroAIMs ascertained to distinguish the NNW-SSE axis of genetic differentiation in Europe could also be utilized to distinguish Northwest Africans from European populations. Their use allowed us to accurately distinguish that Northwest African influences were minor in populations from the Iberian Peninsula, while these were

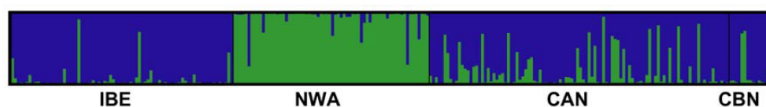
**Table 1.** Population-based estimates (95% confidence interval) of genetic contributions in Canary Islanders using 93 EuroAIMs.

Model	Estimator <sup>a</sup>	Iberian	Northwest African	Sub-Saharan African
3-populations	$m_R$	0.733 (0.666, 0.809)	0.294 (0.190, 0.391)	-0.027 (-0.064, 0.012)
3-populations	$m_W$	0.765 (-, -) <sup>b</sup>	0.234 (-, -) <sup>b</sup>	3.8E-5 (-, -) <sup>b</sup>
2-populations	$m_R$	0.751 (0.696, 0.818)	0.249 (0.182, 0.304)	-
2-populations	$m_W$	0.772 (0.564, 0.898)	0.228 (0.102, 0.436)	-

<sup>a</sup> $m_R$ : Moment estimator [47],  $m_W$ : Maximum-likelihood estimator [46];

<sup>b</sup>Computational burden precluded the adequate estimation of boundaries when the number of parental populations was set to three.

doi:10.1371/journal.pone.0018389.t001



**Figure 3. STRUCTURE results based on EuroAIMs.** Each vertical line represents an individual where colors indicate the proportion of the individual's genome derived from each of the two inferred populations. CAN, Canary Islanders collected for this study; CBN, Canary Islanders available from the National Spanish DNA Bank; IBE, Iberians; NWA, Northwest Africans.  
doi:10.1371/journal.pone.0018389.g003

substantially greater in Canary Islanders, ranging from as low as 0% to as high as 96% among individuals. We finally demonstrated that these African influences might constitute a potential source of population stratification in population-based association studies conducted in Spanish populations, and that this effect was appropriately controlled using a reduced number of EuroAIMs.

It was not unexpected to find a low level of bidirectional gene flow (5%) between Northwest Africans and Iberians given the results from previous studies [22,23]. However, while the North African influence in Iberian populations has been estimated in 8–10% based on markers with uniparental inheritance [6,12,13], we found slightly lower levels using autosomal markers, ranging from as low as 2.1% (SD = 4.6) in Eastern Iberians to as much as 9.0% (SD = 23.8) in Western Iberian samples, although these differences did not reach statistical significance (two-tailed t-test  $p = 0.156$ ). Strikingly, a parallel geographical clustering of the Northwest African influence in Iberian populations was recently revealed with Y-chromosome binary markers [12]. Given that in this study, Iberians were represented by a small number of samples from different mainland localities, it remains interesting to explore if such geographical patterning of the Northwest African influence is confirmed in reasonably-sized population samples from different localities.

The substantial Northwest African ancestry found for Canary Islanders supports that, despite the aggressive conquest by the Spanish in the XV century and the subsequent immigration, genetic footprints of the first settlers of the Canary Islands persist in the current inhabitants. Paralleling mtDNA findings [16], the largest average Northwest African contribution was found for the samples from La Gomera. Remarkably, the sub-Saharan African influence was unnoticeably in this study, despite our results

support that the EuroAIMs set contains enough information to distinguish sub-Saharan Africans from the other populations analyzed. This result contrasts with the previous evidences provided by mtDNA [16] and Y-chromosome [18], and the documented early introduction of sub-Saharan Africans after the conquest as a result of slave trade [17]. However, given the differences in inheritance among these loci, and the complex history of Canary Islanders [17], it is not surprising to reach different but complementary conclusions examining different loci [17,21]. On this basis, we found little support for a sub-Saharan African influence on Canary Islanders other than the introduced through the first settlers, given that genetic studies in aboriginal remains demonstrate the existence of a Saharan substrate [21] and that modern Western Saharan populations show substantial sub-Saharan African influences [22,23].

In view of the disparate estimates of the African influences obtained for the two Spanish populations analyzed, we anticipated an increase of false positive results in population-based association studies of genes with disease simply due to systematic differences in such influences. Using PC scores and, alternatively, individual admixtures obtained by means of STRUCTURE, we have shown that there is a potential benefit of correcting for population stratification in Spanish studies including samples from Canary Islanders. If no adjustment was done, the number of false positive associations resulted in more than twice the expected amount under the null hypothesis. Thus, even small levels of population admixture can undermine an association study and lead to false positive results [37]. On the contrary, at an affordable cost, genotyping a few dozen EuroAIMs would allow adjusting the association tests for population stratification to control the excess

**Table 2.** Average Northwest African ancestry estimates.

Population	Mean (SD)
Iberians (n = 77)	0.050 (0.148)
Northwest Africans (n = 68)	0.947 (0.148)
Canary Islanders, CBN <sup>a</sup> (n = 15)	0.126 (0.252)
Canary Islanders, CAN <sup>b</sup> (n = 104)	0.174 (0.253)
Canary Islanders <sup>b</sup> , by island:	
El Hierro (n = 7)	0.198 (0.306)
La Palma (n = 7)	0.210 (0.330)
La Gomera (n = 7)	0.425 (0.389)
Tenerife (n = 30)	0.143 (0.253)
Gran Canaria (n = 30)	0.124 (0.181)
Lanzarote (n = 13)	0.164 (0.189)
Fuerteventura (n = 10)	0.216 (0.267)

<sup>a</sup>Samples from the National Spanish DNA Bank;

<sup>b</sup>Samples collected for this study.

doi:10.1371/journal.pone.0018389.t002

**Table 3.** Fraction of markers with significant differences between Canary Islanders and Iberians.

Comparison	Markers with $p \leq 0.05$
Unadjusted Cochran-Armitage trend test	0.29
Adjusted by STRUCTURE estimates:	
93	0.118
69	0.108
46	0.097
23	0.086
Adjusted by PCA <sup>a</sup> scores:	
93	0.097
69	0.053
46	0.065
23	0.065

<sup>a</sup>Using PC1 from the PCA restricted to Iberians, Northwest Africans and Canary Islanders.

doi:10.1371/journal.pone.0018389.t003

of spurious results. This does not imply that African influences are the only source of population stratification in Spanish. Previous studies have indicated that even if both cases and controls are collected from the same European country [38] or when comparing samples across European populations, it might be necessary to control for population stratification [33]. Although this study must be viewed simply as an exercise under particular worst sample settings, our results suggest that, at least, those single nucleotide polymorphisms (SNPs) with large allele frequency differences among Iberians and Northwest Africans might be at risk of being detected as false positives in association studies conducted in Spanish populations. Given the large sample sizes needed to detect the association of genetic variants with modest effects in disease [25], the statistical bias introduced may become more pronounced in real settings [39].

We warn that because we used markers showing large allele frequency differences between populations [30], the reported genetic differences between Spanish and Northwest African populations suffer from ascertainment bias, therefore, not corresponding to the average of the genome. This study also shares the limitations of any other genetic study aiming to estimate ancestry proportions. Among others, ancestry proportions are highly dependent on the samples considered as references or parentals, which are usually derived from a reduced number of individuals from contemporary populations. In this sense, we used a mixture of samples from different locations in Morocco to quantify North African influences in Iberians and as a proxy for the aboriginal population inhabiting the Canary Islands before the conquest. This is a reasonable assumption, given the evidence from the historical records [15,17] and the numerous previous genetic studies [12,16–19,21,40]. Besides, we admit that the use of this sample to represent the North African population constitutes a simplification of the heterogeneous source of African influences that have affected the Spanish populations. It is likely that further sampling of North African regions and the typing of additional markers might allow identifying other influences in Spanish, as has been illustrated recently for African Americans [41].

In conclusion, we have extended the use of EuroAIMs to allow distinguishing the contrasting Northwest African influences existing among the Spanish populations. We have also demonstrated that these differences might increase the risk for false positives in genetic epidemiology studies that can be effectively controlled using a reduced number of EuroAIMs.

## Materials and Methods

### Ethics statement

This study was approved by the Hospital Universitario N.S. de Candelaria and Hospital Universitario Dr. Negrín Ethics Committees. Written informed consent was obtained from all participants involved in the study.

### Samples

DNA samples from individuals of the general Iberian population were obtained from the Spanish National DNA Bank ([www.bancoadn.org](http://www.bancoadn.org)), consisting on 77 samples assigned to different Iberian localities based on the individual's self-reported grandparental origin: Andalusia (n = 15), Murcia (n = 5), Extremadura (n = 5), Castile-La Mancha (n = 4), Valencia (n = 7), Castile and Leon (n = 9), Madrid (n = 1), Catalonia (n = 2), Galicia (n = 4), Cantabria (n = 3), Navarre (n = 4), La Rioja (n = 4), Asturias (n = 1), and mixed Iberian origin (n = 13). The Canary Island population was represented by DNA samples from 104 unrelated healthy donors with at least two generations of ancestors

born in the Canary Islands collected for the study (El Hierro, n = 7; La Palma, n = 7; La Gomera, n = 7; Tenerife, n = 30; Gran Canaria, n = 30; Lanzarote, n = 13; Fuerteventura, n = 10), and 15 additional samples from Gran Canaria obtained from the Spanish National DNA Bank. Additionally, 68 DNA samples from unrelated healthy individuals with at least two generations of Northwest African descent were studied: 25 from Casablanca, 25 from Rabat, and 18 from other regions of Morocco. Details of laboratory procedures can be found in the Text S3.

### Reference population data

To get an initial evidence of ancestry informativeness of EuroAIMs for African and Spanish ancestries, genome-wide data from 29 Mozabite Algerians and 24 French Basques from the HGDP [32] were used as representatives for North African and Spanish populations, although both are well-known population isolates [7,42]. Genotypes from 60 unrelated YRI [43] were used to represent the sub-Saharan African population. To place the study in a European context, we utilized previously published data from 163 Swedish, 57 Polish, 76 English, 119 Italians, 68 Greeks, and 55 Spanish [30], as well as from 60 Utah residents with ancestry from northern and western Europe (CEU) [43].

### Genotyping

Samples were genotyped for EuroAIMs recovering the largest features of the European NNW-SSE axis of differentiation [30], focusing on those 93 SNPs that overlapped across all European samples used in the original study (Table S1). Genotyping was conducted utilizing the iPLEX™ Gold assay on MassARRAY system (Sequenom, San Diego, CA) by the Spanish National Genotyping Center, Santiago de Compostela Node (CeGen, <http://www.cegen.org>). Nine SNPs that gave poor quality data on this platform were finally determined using SNaPshot® Multiplex Kit reactions (Applied Biosystems, Foster City, CA) (Text S3). Genotypes are available from the corresponding author upon request.

### Statistical analysis

**Summary statistics.** Allele counts, exact tests for HWE departures [44], and relevant measures of ancestry informativeness of EuroAIMs were calculated using the SNPInfostats software (available from the corresponding author upon request). For each marker, SNPInfostats was used to estimate the absolute allele frequency difference ( $\delta$ ), the Weir & Cockerham  $F_{ST}$  genetic distance [45], as well as the informativeness of assignment index  $I_n$  [29].

**Population relationships.** EIGENSOFT [24] was used to calculate pairwise population  $F_{ST}$  genetic distances, to assess PCA of individual samples, and to determine the ANOVA statistics for population differences along each of the PCs. Multidimensional scaling was used to represent pairwise population  $F_{ST}$  genetic distances in two dimensions using SPSS ver.15 (SPSS Inc., Chicago, IL).

**Ancestry assessment.** In order to obtain initial estimates of the genetic contribution of populations in Canary Islanders, particularly the influence of sub-Saharan Africans, population-based estimates were calculated using LEADMIX [46] by means of the  $m_R$  least-squares estimator, that ignores sampling and genetic drift in populations [47], and the  $m_W$  maximum likelihood estimator, that allows estimating genetic contributions taking into account the effects of sampling and genetic drift in all populations and the differentiation between parental populations before the admixture event [46].

Individual ancestry estimates were assessed using an unsupervised assignment of individuals to populations by means of



STRUCTURE 2.2 [48]. To infer the number of populations (K), three independent runs with a burn-in length of 50,000 for 200,000 repetitions from K = 1 to K = 7 were performed, using a correlated allele frequency with no-admixture model (as suggested by the software documentation to detect subtle structure) and setting lambda at 1. K was deduced from all runs using the method of Evanno et al. [49], which is based on the rate of change in the estimate of the posterior probability of the data with respect to successive K values. Additional runs with longer iterations were also carried out to check the consistency of the results.

**Ancestry inference with marker subsets.** With the purpose of exploring the ability of reduced EuroAIMs subsets to estimate population membership and ancestry proportions, we selected three subsets containing the most informative 23, 46 and 69 markers. For that, EuroAIMs were ranked based on the average of ranks for  $I_n$  and  $\delta$  values, given that  $F_{ST}$  and  $I_n$  values were highly correlated ( $r^2 = 0.99$ ). STRUCTURE was run for each subset using the same parameters as described above and all correctly identified K = 2. A cut-off membership of 0.70 was used as a criterion for the individual inclusion in the self-reported population. Pearson's correlations between ancestry estimates derived from each subset and those from the 93 EuroAIMs set were computed.

**Allele frequency differences between Canary Islanders and Iberians.** To explore whether allele frequencies of the EuroAIMs differed significantly among the Spanish samples, we treated each individual EuroAIM marker as a candidate locus for association in a mock study where Canary Islanders were taken as "cases" and Iberians were considered "controls". Differences were first tested using the Cochran-Armitage trend test, which is similar to the allele count-test albeit not relying on the assumption of HWE [50]. In order to investigate whether there was a statistical benefit of correcting for population stratification, we then tested the differences adjusting the comparisons for population stratification by means of logistic regression models. These included the mock disease status as the dependent variable and the EuroAIM marker genotypes as well as the PC scores obtained from EIGENSOFT as independent covariates. For this purpose, the genotypes were recoded with pre-specified scores of 0, 1 and 2 as disease risk probabilities based on the number of risk alleles in the genotypes. Alternatively, instead of PC scores, logistic regressions were assessed with the Northwest African ancestry estimates obtained from STRUCTURE as one of the independent covariates. These adjustments were repeated with PC scores and the Northwest African ancestry estimates obtained from the full marker set and subsets of the markers ranking higher for ancestry informativeness. All regression models were done by means of SNPassoc [51].

**Supporting Information**

**Figure S1** STRUCTURE results based on EuroAIMs. This analysis used data from Iberians (IBE), Northwest Africans (NWA) and Yoruba Nigerians (YRI) from HapMap [43] without using

any prior population assignment. The model with best likelihood was K = 3 subpopulations. Each vertical line represents an individual where colors indicate the proportion of the individual's genome derived from each of the two inferred populations. (TIF)

**Figure S2** Quantile-quantile plots of  $p$ -values (as  $-\log_{10}P$ ) for marker allele frequency differences between Spanish populations. Upper panel: adjustments based on STRUCTURE estimates; Lower panel: adjustments based on PC1 scores. Closed circles: trend test statistics; Open circles: statistics adjusted for estimates based on 93 EuroAIMs; dark grey circles: statistics adjusted for estimates based on 69 markers; triangles: statistics adjusted for estimates based on 46 markers; light grey circles: statistics adjusted for estimates based on 23 markers. The discontinuous line indicates the null distribution. (TIF)

**Table S1** Summary statistics for EuroAIMs used in the study. (DOC)

**Table S2** Pairwise population  $F_{ST}$  genetic distances. (DOC)

**Table S3** Functional annotation of EuroAIMs within RefSeq genes. (DOC)

**Text S1** Ancestry informativeness of EuroAIMs in samples from HGDP. (DOC)

**Text S2** Ancestry informativeness of EuroAIMs subsets. (PDF)

**Text S3** Details of laboratory procedures. (DOC)

**Acknowledgments**

We thank David Comas and Jose M. Larruga for providing Northwest African samples and Alkes L. Price for the access to reference genotype data and his help with EIGENSOFT. The authors also want to thank the nursing staff of the Blood Bank at the Hospital Universitario Dr. Negrín Las Palmas de Gran Canaria.

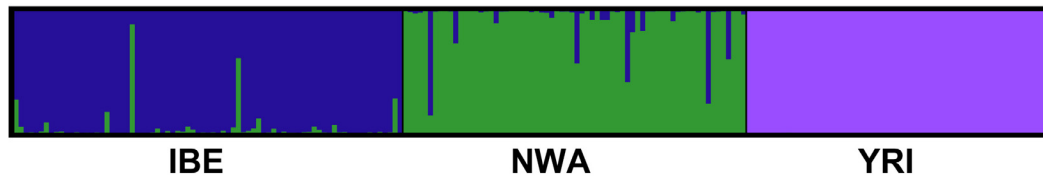
**Author Contributions**

Conceived and designed the experiments: CF. Performed the experiments: AC MP-Y AH. Analyzed the data: MP-Y CF. Contributed reagents/materials/analysis tools: SB AH LG JV. Wrote the paper: MP-Y CF.

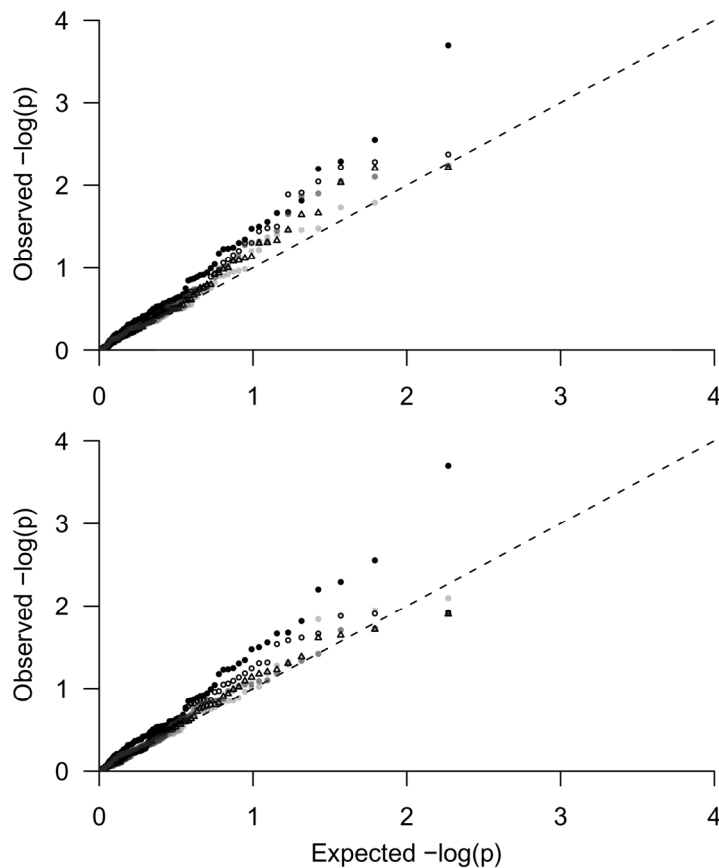
**References**

1. Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton, NJ: Princeton University Press.
2. Barbujani G, Goldstein DB (2004) Africans and Asians abroad: genetic diversity in Europe. *Annu Rev Genomics Hum Genet* 5: 119–150.
3. Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, et al. (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* 290: 1155–1159.
4. Richards M, Macaulay V, Hickey E, Vega E, Sykes B, et al. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 67: 1251–1276.
5. Balesque P, Bowden GR, Adams SM, Leung HY, King TE, et al. (2010) A predominantly neolithic origin for European paternal lineages. *PLoS Biol* 8: e1000285.
6. Flores C, Maca-Meyer N, Gonzalez AM, Oefner PJ, Shen P, et al. (2004) Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. *Eur J Hum Genet* 12: 855–863.
7. Rodriguez-Ezpeleta N, Alvarez-Busto J, Imaz L, Regueiro M, Azcarate MN, et al. (2010) High-density SNP genotyping detects homogeneity of Spanish and French Basques, and confirms their genomic distinctiveness from other European populations. *Hum Genet* 128: 113–117.
8. Gonzalez AM, Brehm A, Perez JA, Maca-Meyer N, Flores C, et al. (2003) Mitochondrial DNA affinities at the Atlantic fringe of Europe. *Am J Phys Anthropol* 120: 391–404.
9. Laayouni H, Calafell F, Bertranpetit J (2010) A genome-wide survey does not show the genetic distinctiveness of Basques. *Hum Genet* 127: 455–458.

Supporting Information



**Figure S1** STRUCTURE results based on EuroAIMs. This analysis used data from Iberians (IBE), Northwest Africans (NWA) and Yoruba Nigerians (YRI) from HapMap [43] without using any prior population assignment. The model with best likelihood was  $K=3$  subpopulations. Each vertical line represents an individual where colors indicate the proportion of the individual's genome derived from each of the two inferred populations.



**Figure S2** Quantile-quantile plots of  $p$ -values (as  $-\log_{10}P$ ) for marker allele frequency differences between Spanish populations. Upper panel: adjustments based on STRUCTURE estimates; Lower panel: adjustments based on PC1 scores. Closed circles: trend test statistics; Open circles: statistics adjusted for estimates based on 93 EuroAIMs; dark grey circles: statistics adjusted for estimates based on 69 markers; triangles: statistics adjusted for estimates based on 46 markers; light grey circles: statistics adjusted for estimates based on 23 markers. The discontinuous line indicates the null distribution.

Table S1. Summary statistics for EuroAIMs used in the study.

rs#	Alleles	Chr.	Position <sup>a</sup>	Call (%)	Ref. allele frequency					Hardy-Weinberg equilibrium p-value <sup>b</sup>					IBE-NWA ancestry measures					IBE-YRI ancestry measures					NWA-YRI ancestry measures				
					Ref.	Cls	IBE	NWA	YRI	Cls	IBE	NWA	YRI	In	δ	F <sub>ST</sub>	In	δ	F <sub>ST</sub>	In	δ	F <sub>ST</sub>	In	δ	F <sub>ST</sub>	In	δ	F <sub>ST</sub>	
rs1157492	C/T	1	211627299	98.5	C	0.40	0.40	0.25	0.45	1.73E-01	3.43E-01	3.30E-01	7.86E-01	1.26E-01	1.49E-01	3.52E-02	9.40E-04	4.29E-02	0.00E+00	2.04E-02	1.92E-01	6.31E-02	2.04E-02	1.92E-01	6.31E-02				
rs11807062	A/C	1	3176394	99.6	C	0.20	0.16	0.38	0.75	2.42E-01	1.18E-01	1.18E-01	4.76E-01	1.80E-02	2.13E-01	9.80E-02	1.86E-01	5.88E-01	5.16E-01	7.34E-02	3.75E-01	2.35E-01	7.34E-02	3.75E-01	2.35E-01				
rs1416467	A/G	1	80470749	99.6	A	0.51	0.51	0.25	0.25	9.96E-02	8.22E-01	1.76E-02	7.14E-01	1.21E-02	1.54E-01	3.35E-02	3.69E-02	1.69E-01	1.19E-01	6.92E-03	1.07E-01	1.06E-02	6.92E-03	1.07E-01	1.06E-02				
rs1890131	C/T	1	164329131	100.0	T	0.24	0.32	0.09	0.03	1.29E-01	1.00E+00	1.00E+00	1.00E+00	4.29E-02	2.30E-01	1.35E-01	8.72E-02	2.93E-01	1.35E-01	9.89E-03	6.32E-02	2.02E-02	9.89E-03	6.32E-02	2.02E-02				
rs2236876	A/G	1	169749143	100.0	A	0.35	0.32	0.43	0.40	6.87E-01	6.12E-01	4.59E-01	7.82E-01	5.54E-03	1.02E-01	8.19E-03	2.93E-03	7.35E-02	0.00E+00	4.14E-04	2.83E-02	0.00E+00	4.14E-04	2.83E-02	0.00E+00				
rs2419063	A/G	1	186988492	99.6	G	0.12	0.05	0.18	0.30	6.63E-01	1.00E+00	4.40E-01	1.00E+00	2.20E-02	1.32E-01	6.98E-02	5.77E-02	2.48E-01	1.37E-01	9.27E-03	1.16E-01	2.07E-02	9.27E-03	1.16E-01	2.07E-02				
rs495347	C/T	1	17909228	97.8	C	0.35	0.27	0.48	0.65	2.22E-01	1.00E+00	6.28E-01	1.40E-01	2.26E-01	7.34E-02	2.05E-01	7.34E-02	2.05E-01	7.34E-02	2.05E-01	1.44E-02	1.68E-01	1.44E-02	1.68E-01	3.95E-02				
rs725974	G/T	1	167344909	93.0	G	0.22	0.18	0.43	0.23	7.75E-01	4.42E-01	4.52E-01	1.00E+00	3.50E-02	2.41E-01	1.18E-01	1.83E-03	4.91E-02	0.00E+00	2.11E-02	1.92E-01	6.47E-02	2.11E-02	1.92E-01	6.47E-02				
rs7552548	A/G	1	82775867	99.6	G	0.36	0.27	0.39	0.49	1.00E+00	1.51E-01	7.99E-01	1.20E-01	8.17E-03	1.20E-01	1.85E-02	2.64E-02	2.22E-01	8.69E-02	5.28E-03	1.02E-01	5.30E-03	5.28E-03	1.02E-01	5.30E-03				
rs10496610	C/G	2	123757692	99.6	G	0.22	0.19	0.27	0.05	7.93E-01	1.00E+00	2.00E-02	1.00E+00	1.89E-02	1.29E-01	5.59E-02	7.28E-02	1.95E-01	1.65E-01	2.35E-02	6.62E-02	4.76E-02	2.35E-02	6.62E-02	4.76E-02				
rs1364394	C/T	2	95617604	98.9	T	0.27	0.29	0.16	0.05	4.91E-01	1.00E+00	6.52E-01	1.00E+00	1.35E-02	1.37E-01	3.79E-02	5.37E-02	2.39E-01	1.55E-01	1.43E-02	1.02E-01	3.57E-02	1.43E-02	1.02E-01	3.57E-02				
rs1448314	A/G	2	223811705	99.6	A	0.10	0.08	0.20	0.25	1.00E+00	1.00E+00	2.76E-01	7.16E-01	1.57E-02	1.21E-01	4.74E-02	2.82E-02	1.72E-01	9.60E-02	1.91E-03	5.15E-02	0.00E+00	1.91E-03	5.15E-02	0.00E+00				
rs1517407	A/G	2	63401982	99.3	G	0.45	0.47	0.36	0.11	4.61E-01	3.55E-01	2.92E-01	1.00E+00	5.89E-03	1.07E-01	9.40E-03	8.26E-02	3.58E-01	2.45E-01	4.58E-02	2.51E-01	1.42E-01	4.58E-02	2.51E-01	1.42E-01				
rs3769005	C/G	2	136437098	98.1	C	0.50	0.44	0.60	0.87	1.35E-01	2.10E-01	2.06E-01	1.00E+00	1.32E-02	1.62E-01	3.77E-02	1.08E-01	4.32E-01	3.20E-01	4.80E-02	2.70E-01	1.54E-01	4.80E-02	2.70E-01	1.54E-01				
rs4832640	C/T	2	19025771	98.9	T	0.29	0.24	0.43	0.71	3.64E-01	2.10E-01	6.22E-01	1.00E+00	1.00E+00	1.86E-01	6.27E-02	1.15E-01	4.69E-01	3.54E-01	4.13E-02	2.83E-01	1.35E-01	4.13E-02	2.83E-01	1.35E-01				
rs6432110	A/T	2	10688616	99.6	T	0.46	0.43	0.39	0.26	4.58E-01	1.00E+00	1.28E-01	3.10E-01	7.81E-04	3.89E-02	0.00E+00	1.62E-02	1.70E-01	4.73E-02	9.90E-03	1.31E-01	2.31E-02	9.90E-03	1.31E-01	2.31E-02				
rs6745653	C/T	2	97748515	99.3	T	0.10	0.14	0.26	0.37	1.00E+00	6.20E-01	3.62E-01	5.63E-01	1.30E-02	1.28E-01	3.75E-02	3.79E-02	2.36E-01	1.30E-01	6.74E-03	1.08E-01	1.05E-02	6.74E-03	1.08E-01	1.05E-02				
rs1879558	C/T	3	155137053	96.7	C	0.24	0.16	0.22	0.19	3.00E-01	6.78E-01	2.88E-01	1.66E-01	3.05E-03	6.15E-02	0.00E+00	2.16E-01	2.86E-02	0.00E+00	8.27E-04	3.30E-02	0.00E+00	8.27E-04	3.30E-02	0.00E+00				
rs2596834	C/T	3	10795067	99.6	C	0.54	0.5	0.43	0.00	1.00E+00	8.20E-01	1.88E-01	1.00E+00	7.02E-03	7.35E-02	0.00E+00	7.02E-03	7.35E-02	0.00E+00	1.77E-04	3.30E-02	0.00E+00	1.77E-04	3.30E-02	0.00E+00				
rs4686497	A/G	3	190124209	100.0	A	0.19	0.14	0.32	0.96	1.00E+00	1.00E+00	1.00E+00	1.00E+00	2.53E-02	1.87E-01	8.32E-02	4.03E-01	8.22E-01	7.99E-01	2.52E-01	6.35E-01	5.92E-01	2.52E-01	6.35E-01	5.92E-01				
rs8227659	G/T	3	22948116	99.3	T	0.39	0.34	0.51	0.45	8.47E-01	4.48E-01	1.00E+00	1.03E-01	1.52E-02	1.72E-01	4.56E-02	6.36E-03	1.10E-01	9.76E-03	1.92E-03	6.19E-02	0.00E+00	1.92E-03	6.19E-02	0.00E+00				
rs9290875	C/T	3	180489335	98.5	C	0.13	0.20	0.20	0.11	4.22E-01	2.28E-01	7.18E-01	5.52E-01	5.60E-04	2.62E-02	0.00E+00	2.28E-03	5.71E-02	0.00E+00	6.52E-03	8.33E-02	8.59E-03	6.52E-03	8.33E-02	8.59E-03				
rs9861816	C/G	3	195649797	99.6	C	0.23	0.22	0.40	0.92	4.31E-01	5.05E-01	3.26E-01	3.95E-01	1.92E-01	1.81E-01	6.08E-02	3.69E-03	6.85E-01	6.31E-01	1.53E-01	5.04E-01	4.23E-01	1.53E-01	5.04E-01	4.23E-01				
rs10518982	G/T	4	98963279	97.4	T	0.33	0.34	0.53	0.84	2.07E-01	7.45E-02	2.24E-01	6.19E-01	1.76E-02	1.86E-01	5.47E-02	1.32E-01	4.92E-01	3.81E-01	5.60E-02	3.07E-01	1.78E-01	5.60E-02	3.07E-01	1.78E-01				
rs1073321	C/T	4	118870455	98.5	T	0.27	0.21	0.19	0.23	1.06E-05	3.07E-02	1.06E-01	5.10E-02	1.88E-04	1.55E-02	0.00E+00	3.37E-04	2.14E-02	0.00E+00	1.03E-03	3.69E-02	0.00E+00	1.03E-03	3.69E-02	0.00E+00				
rs12502036	G/T	4	158829193	98.8	C	0.28	0.23	0.23	0.24	1.00E+00	1.00E+00	7.40E-01	1.00E+00	8.16E-03	1.08E-03	0.00E+00	8.16E-03	1.08E-03	0.00E+00	1.26E-02	1.21E-01	3.12E-02	1.26E-02	1.21E-01	3.12E-02				
rs1373557	C/T	4	167601812	97.8	T	0.26	0.30	0.44	0.18	8.12E-01	5.90E-01	5.15E-02	6.71E-01	1.03E-02	1.39E-01	2.69E-02	1.00E-02	1.21E-01	2.27E-02	4.01E-02	2.59E-01	1.28E-01	2.27E-02	4.01E-02	2.59E-01				
rs17443616	A/G	4	41063093	99.6	A	0.26	0.28	0.26	0.07	1.00E+00	7.78E-01	1.00E+00	1.00E+00	2.06E-04	1.80E-02	0.00E+00	3.88E-02	2.06E-01	1.14E-01	3.36E-02	1.88E-01	1.00E-01	3.36E-02	1.88E-01	1.00E-01				
rs1873195	C/T	4	38713739	98.9	T	0.32	0.28	0.26	0.26	0.07	1.00E+00	2.34E-01	6.11E-01	1.00E+00	9.14E-03	1.28E-01	2.23E-02	2.72E-06	2.03E-03	0.00E+00	9.46E-03	1.28E-01	2.03E-03	0.00E+00	9.46E-03	1.28E-01			
rs2014303	A/G	4	10262125	99.6	A	0.19	0.24	0.28	0.29	1.00E+00	7.56E-01	7.63E-01	1.00E+00	9.97E-04	3.92E-02	0.00E+00	1.65E-03	5.06E-02	0.00E+00	8.11E-05	1.15E-02	0.00E+00	8.11E-05	1.15E-02	0.00E+00				
rs2251432	A/G	4	191063741	99.6	A	0.15	0.24	0.28	0.29	1.00E+00	7.56E-01	7.63E-01	1.00E+00	9.97E-04	3.92E-02	0.00E+00	1.65E-03	5.06E-02	0.00E+00	8.11E-05	1.15E-02	0.00E+00	8.11E-05	1.15E-02	0.00E+00				
rs4555709	C/T	4	53667889	99.6	T	0.09	0.06	0.19	0.25	2.55E-01	1.00E+00	6.95E-01	3.46E-02	2.11E-02	1.33E-01	6.68E-02	3.75E-02	1.92E-01	1.27E-01	2.52E-03	5.88E-02	0.00E+00	2.52E-03	5.88E-02	0.00E+00				
rs9328764	A/G	4	2143685	98.9	A	0.22	0.18	0.48	0.61	1.00E+00	4.42E-01	1.00E+00	5.73E-01	4.99E-02	2.93E-01	1.68E-01	9.82E-02	4.25E-01	3.15E-01	8.74E-03	1.31E-01	1.79E-02	3.15E-01	8.74E-03	1.31E-01				
rs974020	C/T	4	117865302	98.1	C	0.27	0.18	0.22	0.22	4.86E-01	4.72E-01	4.85E-01	4.24E-01	8.10E-04	3.22E-02	0.00E+00	8.98E-04	3.40E-02	0.00E+00	2.29E-06	1.76E-03	0.00E+00	2.29E-06	1.76E-03	0.00E+00				
rs16891982	C/G	5	33987450	95.9	C	0.34	0.15	0.66	1.00	1.36E-01	1.00E+00	2.80E-01	1.00E+00	1.41E-01	5.05E-01	4.15E-01	4.69E-01	8.49E-01	8.28E-01	1.37E-01	3.43E-01	3.15E-01	8.28E-01	1.37E-01	3.43E-01				
rs33706	C/G	5	87540920	99.6	G	0.26	0.29	0.38	0.52	6.31E-01	1.77E-01	8.00E-01	3.09E-01	3.86E-03	8.28E-02	1.53E-02	2.64E-02	2.24E-01	8.69E-02	1.02E-02	1.42E-01	2.46E-02	8.69E-02	1.02E-02	1.42E-01				
rs3822143	A/G	5	11																										

Table S1. (Continued)

rs#	Alleles	Chr.	Position <sup>a</sup>	Call (%)	Ref. allele frequency				Hardy-Weinberg equilibrium $p$ -value <sup>b</sup>				IBE-NWA ancestry measures				IBE-YRI ancestry measures				NWA-YRI ancestry measures			
					Ref.	Cis	IBE	NWA	YRI	Cis	IBE	NWA	YRI	$I_n$	$\delta$	$F_{ST}$	$I_n$	$\delta$	$F_{ST}$	$I_n$	$\delta$	$F_{ST}$	$I_n$	$\delta$
rs1922086	C/T	7	155515702	96.7	T	0.46	0.34	0.58	0.68	1.00E+00	8.02E-01	1.00E+00	1.00E+00	2.88E-02	2.38E-02	9.57E-02	6.05E-02	3.44E-01	1.99E-01	6.04E-03	1.06E-01	7.14E-03		
rs2097884	A/C	7	4125658	99.3	C	0.27	0.31	0.39	0.18	6.41E-01	7.92E-01	6.14E-01	3.53E-01	3.35E-03	7.80E-02	0.00E+00	1.14E-02	1.30E-01	2.80E-02	2.69E-02	2.08E-01	8.33E-02		
rs2219248	C/T	7	113965977	99.6	C	0.28	0.39	0.22	-	3.68E-01	1.00E+00	7.22E-01	-	1.70E-02	1.69E-01	5.14E-02	-	-	-	-	-	-		
rs2367191	A/G	7	141571268	98.1	A	0.41	0.47	0.19	0.10	7.03E-01	1.00E+00	4.32E-01	4.59E-02	2.81E-01	1.50E-01	8.72E-02	3.66E-01	2.55E-01	7.38E-03	8.47E-02	1.14E-02			
rs2905347	A/G	7	22393559	99.6	A	0.48	0.38	0.60	0.69	7.15E-01	6.33E-01	4.56E-01	1.00E+00	2.47E-02	2.21E-01	8.07E-02	4.89E-02	3.09E-01	1.61E-01	4.24E-03	8.80E-02	2.62E-04		
rs10504924	C/T	8	94157437	99.6	T	0.10	0.12	0.12	0.00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	7.05E-07	7.64E-04	0.00E+00	4.23E-02	1.19E-02	9.19E-02	4.28E-02	1.18E-01	9.73E-02		
rs10512122	C/G	9	81715524	96.7	C	0.24	0.15	0.35	0.25	8.02E-01	6.67E-01	5.86E-01	7.42E-01	2.78E-02	2.02E-01	9.28E-02	7.99E-03	1.01E-01	1.74E-02	6.15E-03	1.02E-01	8.14E-03		
rs1408794	A/G	9	12641340	99.6	A	0.48	0.44	0.57	0.94	7.15E-01	8.05E-01	8.05E-01	1.00E+00	8.84E-03	1.33E-01	2.10E-02	1.64E-01	5.01E-01	4.28E-01	1.02E-01	3.68E-01	2.92E-01		
rs2086085	A/C	9	1680761	99.3	C	0.15	0.14	0.16	0.13	1.00E+00	3.44E-01	3.43E-01	2.63E-01	1.40E-04	1.20E-02	0.00E+00	1.38E-04	1.14E-02	0.00E+00	5.52E-04	2.34E-02	0.00E+00		
rs1045873	A/C	10	25177778	95.6	C	0.29	0.24	0.19	0.02	6.46E-01	7.58E-01	6.84E-01	1.00E+00	2.40E-02	5.69E-02	6.50E-02	6.50E-02	2.25E-01	1.69E-01	4.42E-02	1.68E-01	1.20E-01		
rs10509384	A/G	10	78693188	80.4	G	0.20	0.26	0.12	0.04	1.41E-01	1.00E+00	1.94E-01	1.00E+00	1.56E-02	1.38E-01	4.48E-02	5.37E-02	2.20E-01	1.49E-01	1.25E-02	8.27E-02	2.86E-02		
rs10509954	A/G	10	113658378	98.9	A	0.13	0.10	0.13	0.10	3.97E-01	5.00E-01	1.00E+00	4.25E-01	1.11E-03	3.04E-02	0.00E+00	2.07E-05	3.90E-03	0.00E+00	1.43E-03	3.43E-02	0.00E+00		
rs379773	C/G	10	109905342	96.7	G	0.30	0.3	0.42	0.63	1.12E-01	1.00E+00	8.08E-01	1.00E+00	7.75E-03	1.19E-01	1.69E-02	5.53E-02	3.28E-01	1.85E-01	2.21E-02	2.09E-01	6.91E-02		
rs7908825	C/G	10	75224547	100.0	C	0.32	0.29	0.29	0.69	5.28E-01	5.81E-01	1.44E-01	7.51E-01	2.20E-06	1.91E-03	0.00E+00	8.18E-02	3.99E-01	2.64E-01	8.10E-02	3.97E-01	2.60E-01		
rs1560569	A/G	11	8896463	99.3	G	0.34	0.33	0.20	0.15	5.42E-01	1.00E+00	7.14E-01	9.90E-02	1.10E-02	1.30E-01	2.90E-02	2.11E-02	1.74E-02	6.21E-01	1.67E-03	4.40E-02	8.00E+00		
rs2847502	C/T	11	119623708	97.4	T	0.40	0.25	0.38	0.49	4.34E-01	5.48E-01	1.00E+00	3.26E-02	9.41E-03	1.27E-01	2.34E-02	3.06E-02	2.38E-01	1.02E-01	6.20E-03	1.10E-01	1.01E-03		
rs4938377	G/T	11	116801502	99.6	G	0.08	0.08	0.04	0.01	5.42E-01	1.00E+00	1.00E+00	1.00E+00	4.17E-03	4.22E-02	1.99E-03	1.65E-02	6.99E-02	5.53E-02	4.57E-03	2.77E-02	0.00E+00		
rs923031	A/C	11	15777602	99.6	A	0.42	0.44	0.34	0.12	5.70E-01	6.16E-01	5.87E-01	5.52E-01	2.80E-02	2.35E-01	9.23E-02	8.82E-03	1.32E-01	2.02E-02	5.46E-03	1.04E-01	5.88E-03		
rs1003306	A/T	12	2624458	99.6	A	0.33	0.32	0.49	0.69	3.00E-01	1.93E-01	1.00E+00	5.43E-01	1.52E-02	1.74E-01	4.83E-02	7.15E-02	3.73E-01	2.33E-01	2.07E-02	1.99E-01	6.37E-02		
rs1582398	A/G	12	78624019	98.5	G	0.15	0.10	0.35	0.98	1.00E+00	1.00E+00	1.00E+00	1.00E+00	4.63E-02	2.47E-01	1.55E-01	4.88E-01	8.85E-01	8.71E-01	2.73E-01	6.37E-01	6.12E-01		
rs3809125	C/T	12	55130616	98.9	T	0.32	0.37	0.31	0.06	1.00E+00	1.38E-01	1.00E+00	1.00E+00	2.53E-02	6.74E-02	0.00E+00	7.62E-02	3.10E-01	2.18E-01	5.23E-02	2.42E-01	1.56E-01		
rs7965049	A/G	12	7537044	98.5	A	0.24	0.29	0.35	0.94	1.00E+00	2.67E-01	4.19E-01	1.00E+00	1.97E-03	5.85E-02	0.00E+00	2.48E-01	6.44E-01	5.81E-01	2.08E-01	5.86E-01	5.22E-01		
rs998401	A/G	12	53641500	99.6	A	0.08	0.06	0.05	0.07	1.86E-01	1.00E+00	1.00E+00	1.00E+00	4.14E-04	1.35E-02	0.00E+00	1.18E-04	7.79E-03	0.00E+00	9.74E-04	2.13E-02	0.00E+00		
rs1854226	C/T	13	97036243	93.3	T	0.13	0.26	0.10	0.07	6.69E-01	2.12E-01	1.00E+00	2.39E-01	2.26E-02	1.60E-01	6.97E-02	3.24E-02	1.84E-01	9.54E-02	9.52E-04	2.43E-02	0.00E+00		
rs7997100	C/T	13	67988075	99.6	T	0.37	0.38	0.39	0.30	3.36E-01	3.39E-01	4.45E-01	1.00E+00	3.66E-05	8.13E-03	0.00E+00	3.71E-03	8.16E-02	0.00E+00	4.48E-03	8.97E-02	1.12E-03		
rs986642	C/G	13	56944263	99.6	G	0.23	0.29	0.11	0.28	7.13E-02	1.00E+00	1.71E-01	3.40E-01	2.65E-02	1.82E-01	8.33E-02	1.82E-04	1.72E-02	0.00E+00	2.24E-02	1.65E-01	7.07E-02		
rs10463853	A/G	14	72826052	98.1	G	0.36	0.31	0.47	0.20	4.15E-01	1.00E+00	6.33E-01	6.70E-01	1.45E-02	1.65E-01	4.30E-02	7.37E-03	1.05E-01	1.29E-02	4.20E-02	2.71E-01	1.35E-01		
rs10519269	A/C	15	77705433	97.8	C	0.12	0.10	0.23	0.50	1.00E+00	1.00E+00	7.37E-01	7.91E-01	1.49E-02	1.27E-01	4.44E-02	9.92E-02	3.96E-01	3.21E-01	3.98E-02	2.69E-01	1.32E-01		
rs1129038	C/T	15	26030454	98.5	C	0.68	0.62	0.95	-	8.33E-01	3.34E-01	5.93E-03	-	8.60E-02	3.24E-01	2.51E-01	-	-	-	-	-	-		
rs7163907	C/T	15	73632152	98.9	C	0.37	0.29	0.34	0.56	4.98E-02	5.84E-01	1.00E+00	1.00E+00	1.92E-03	5.76E-02	0.00E+00	4.04E-02	2.79E-01	1.36E-01	2.50E-02	2.22E-01	7.93E-02		
rs1107820	C/T	17	41491195	99.6	C	0.26	0.25	0.16	0.00	1.00E+00	1.00E+00	3.58E-01	1.00E+00	5.99E-03	8.92E-02	9.53E-03	9.58E-02	2.50E-01	2.18E-01	5.98E-02	1.62E-01	1.40E-01		
rs1476162	A/G	17	8031808	99.6	G	0.18	0.16	0.31	0.38	5.24E-01	1.94E-01	7.79E-01	9.16E-02	1.61E-02	1.51E-01	4.92E-02	3.28E-02	2.24E-01	1.11E-01	2.95E-03	7.30E-02	0.00E+00		
rs2003092	A/G	17	59516509	96.3	G	0.31	0.34	0.41	0.73	5.10E-01	4.47E-01	6.18E-01	1.00E+00	2.34E-03	6.63E-02	0.00E+00	7.75E-02	3.87E-01	2.48E-01	5.38E-02	3.21E-01	1.75E-01		
rs959260	C/T	17	70881017	97.4	C	0.25	0.19	0.34	0.81	8.02E-01	4.51E-01	5.90E-01	1.00E+00	1.50E-02	1.52E-01	4.49E-02	2.92E-01	7.18E-01	6.70E-01	1.87E-01	5.66E-01	4.90E-01		
rs2418844	A/G	18	26037261	97.8	G	0.49	0.46	0.41	0.59	5.53E-01	8.17E-01	6.18E-01	1.00E+00	1.28E-03	5.01E-02	0.00E+00	8.56E-03	1.31E-01	1.92E-02	1.65E-02	1.81E-01	4.86E-02		
rs4892082	A/C	18	69051672	99.6	A	0.18	0.16	0.40	0.64	5.46E-01	4.07E-01	1.41E-01	2.51E-01	1.62E-02	2.40E-01	1.22E-01	4.72E-01	3.77E-01	2.72E-02	3.32E-01	8.71E-02			
rs523776	C/T	18	7554299	99.3	T	0.30	0.23	0.40	0.48	6.61E-01	2.16E-01	8.04E-01	4.37E-01	3.56E-02	1.63E-01	4.72E-02	3.44E-02	2.50E-01	1.16E-01	3.78E-03	3.00E+00	0.00E+00		
rs959763	G/T	18	56797826	99.3	G	0.38	0.37	0.36	0.48	5.59E-01	1.00E+00	1.00E+00	1.00E+00	5.22E-05	9.84E-03	0.00E+00	6.58E-03	1.13E-01	1.13E-02	7.78E-03	1.23E-01	1.52E-02		
rs103294	C/T	19	59489660	98.1	T	0.17	0.16	0.06	0.05	5.06E-01	4.02E-01	1.00E+00	1.34E-01	1.03E-02	1.03E-01	3.72E-02	1.56E-02	1.08E-01	3.95E-02	6.17E-05	5.16E-03	0.00E+00		
rs202546	C/T	20	1611539	99.6	T	0.61	0.53	0.47	0.34	1.24E-01	3.61E-01	1.00E+00	5.56E-01	1.92E-03	6.19E-02	0.00E+00	1.97E-02	1.96E-01	6.01E-02	9.39E-03	1.34E-01	2.04E-02		
rs477627	A/G	20	47613465	98.1	A	0.18	0.22	0.21	0.08	1.00E+00	1.00E+00	2.99E-01	1.81E-02	1.16E-02	0.00E+00	2.01E-02	1.42E-01	5.95E-02	1.74E-02	1.40E-02	1.30E-01	4.76E-02		
rs7277342	A/C	21	29040845	90.0	A	0.34	0.29	0.48	0.95	1.11E-01	5.90E-02	1.68E-04	3.57E-03	1.81E-02	1.84E-01	5.64E-02	2.59E-01	6.55E-01	5.99E-01	1.51E-01</				

**Table S2.** Pairwise population  $F_{ST}$  genetic distances.

<b>Populations</b>	<b>Canary Is.</b>	<b>Iberians</b>	<b>Northwest Af.</b>	<b>Swedish</b>	<b>Polish</b>	<b>English</b>	<b>Italians</b>	<b>Spanish<sup>a</sup></b>	<b>Greeks</b>
Iberians	0.0040								
Northwest Af.	0.0310	0.0500							
Swedish	0.0480	0.0340	0.1340						
Polish	0.0300	0.0210	0.1060	0.0050					
English	0.0290	0.0170	0.1050	0.0040	0.0040				
Italians	0.0090	0.0140	0.0320	0.0710	0.0450	0.0480			
Spanish <sup>a</sup>	0.0060	0.0010	0.0600	0.0300	0.0150	0.0150	0.0160		
Greeks	0.0100	0.0130	0.0270	0.0720	0.0450	0.0480	0.0020	0.0170	
CEU <sup>b</sup>	0.0260	0.0150	0.0990	0.0010	0.0020	0.0000	0.0460	0.0110	0.0460

<sup>a</sup>From Price et al. [30]; <sup>b</sup>Utah residents with ancestry from northern and western Europe from HapMap [43].

## RESULTADOS

**Table S3** Functional annotation of EuroAIMs within RefSeq genes.

rs#	Entrez ID	Gene Symbol	Gene Description	Function Class
rs1003306	775	<i>CAC1C</i>	Calcium channel, voltage-dependent, L type, alpha 1C subunit	Intron
rs1032143	26230	<i>TIAM2</i>	T-cell lymphoma invasion and metastasis 2	Intron
rs103294	-	-	-	-
rs1045873	56952	<i>PRTFDC1</i>	Phosphoribosyl transferase domain containing 1	UTR
rs10483853	8650	<i>NUMB</i>	Numb homolog (Drosophila)	Intron
rs10484547	-	-	-	-
rs10486207	113263	<i>GLCCI1</i>	Glucocorticoid induced transcript 1	Intron
rs10496610	-	-	-	-
rs10504924	-	-	-	-
rs10508372	-	-	-	-
rs10509384	3778	<i>KCNMA1</i>	Potassium large conductance calcium-activated channel, subfamily M, alpha member 1	Intron
rs10509954	-	-	-	-
rs10512122	-	-	-	-
rs10516982	-	-	-	-
rs10519269	-	-	-	-
rs1073321	-	-	-	-
rs1107820	284058	<i>KIAA1267</i>	Hypothetical protein LOC284058	Intron
rs1129038	8924	<i>HERC2</i>	Hect domain and RLD 2	UTR (miRNA target site)
rs1157492	3776	<i>KCNK2</i>	Potassium channel, subfamily K, member 2	Intron
rs11807062	63976	<i>PRDM16</i>	PR domain containing 16	Intron
rs12502036	-	-	-	-
rs1364394	-	-	-	-
rs1373557	-	-	-	-
rs1408794	-	-	-	-
rs1416467	-	-	-	-
rs1448314	-	-	-	-
rs1476162	54785	<i>C17orf59</i>	Chromosome 17 open reading frame 59	miRNA target site
rs1517407	-	-	-	-
rs153595	-	-	-	-
rs1560569	56672	<i>C11orf17</i>	Chromosome 11 open reading frame 17	Intron
rs1582398	-	-	-	-
rs16891982	51151	<i>SLC45A2</i>	Membrane associated transporter	Coding (F373L)
rs17443616	-	-	-	-
rs17864053	-	-	-	-
rs1854226	-	-	-	-
rs1873195	92689	<i>FAM114A1</i>	Family with sequence similarity 114, member A1	Intron
rs1879558	-	-	-	-
rs1890131	-	-	-	-
rs1922086	-	-	-	-
rs2003092	2081	<i>ERN1</i>	Endoplasmic reticulum to nucleus siglling 1	Intron
rs2014303	116449	<i>MIST</i>	Mast cell immunoreceptor sigl transducer	Intron
rs202546	-	-	-	-
rs2086085	-	-	-	-
rs2097884	-	-	-	-
rs2171209	94120	<i>SYTL3</i>	Syptotagmin-like 3	Intron
rs2187684	-	-	-	-
rs2219248	-	-	-	-
rs2236876	8995	<i>TNFSF18</i>	Tumor necrosis factor (ligand) superfamily, member 18	Intron
rs2251432	-	-	-	-
rs2367191	-	-	-	-
rs2418844	-	-	-	-
rs2419063	339479	<i>FAM5C</i>	Family with sequence similarity 5, member C	Intron
rs2596501	-	-	-	-

**Table S3** (Continued)

rs#	Entrez ID	Gene Symbol	Gene Description	Function Class
rs2596834	-	-	-	-
rs2804756	-	-	-	-
rs2847502	25833	<i>POU2F3</i>	POU domain, class 2, transcription factor 3	Intron
rs2905347	-	-	-	-
rs33706	153396	<i>TMEM161B</i>	Transmembrane protein 161B	Intron
rs3769005	4175	<i>MCM6</i>	MCM6 minichromosome maintenance deficient 6 (MIS5 homolog, <i>S. pombe</i> ) ( <i>S. cerevisiae</i> )	Intron
rs379773	-	-	-	-
rs3809125	4284	<i>MIP</i>	Major intrinsic protein of lens fiber	UTR
rs3822616	9652	<i>KIAA0372</i>	KIAA0372	Intron
rs4555709	152579	<i>SCFD2</i>	Sec1 family domain containing 2	Intron
rs4686497	-	-	-	-
rs477627	5740	<i>PTGIS</i>	Prostaglandin I2 (prostacyclin) synthase	Intron
rs4832640	-	-	-	-
rs4892082	400655	<i>LOC400655</i>	Hypothetical gene LOC400655	Intron
rs4938377	-	-	-	-
rs495347	-	-	-	-
rs523776	-	-	-	-
rs6432110	79954	<i>NOL10</i>	Nucleolar protein 10	Intron
rs6745653	-	-	-	-
rs7108371	-	-	-	-
rs7163907	5780	<i>PTPN9</i>	Protein tyrosine phosphatase, non-receptor type 9	Intron
rs725974	-	-	-	-
rs7277342	-	-	-	-
rs7552548	-	-	-	-
rs756147	-	-	-	-
rs7908825	23053	<i>KIAA0913</i>	KIAA0913	Intron
rs7965049	9332	<i>CD163</i>	CD163 antigen	Intron
rs7997100	-	-	-	-
rs822759	-	-	-	-
rs920590	-	-	-	-
rs923031	-	-	-	-
rs9290675	-	-	-	-
rs9328764	353497	<i>POLN</i>	Polymerase (D directed) nu	Coding (C424R)
rs959260	2885	<i>GRB2</i>	Growth factor receptor-bound protein 2	Intron
rs959763	-	-	-	-
rs969539	23544	<i>SEZ6L</i>	Seizure related 6 homolog (mouse)-like	Intron
rs974020	-	-	-	-
rs9861816	79572	<i>ATP13A3</i>	ATPase type 13A3	Intron
rs986642	-	-	-	-
rs998401	9840	<i>KIAA0748</i>	KIAA0748 gene product	Intron

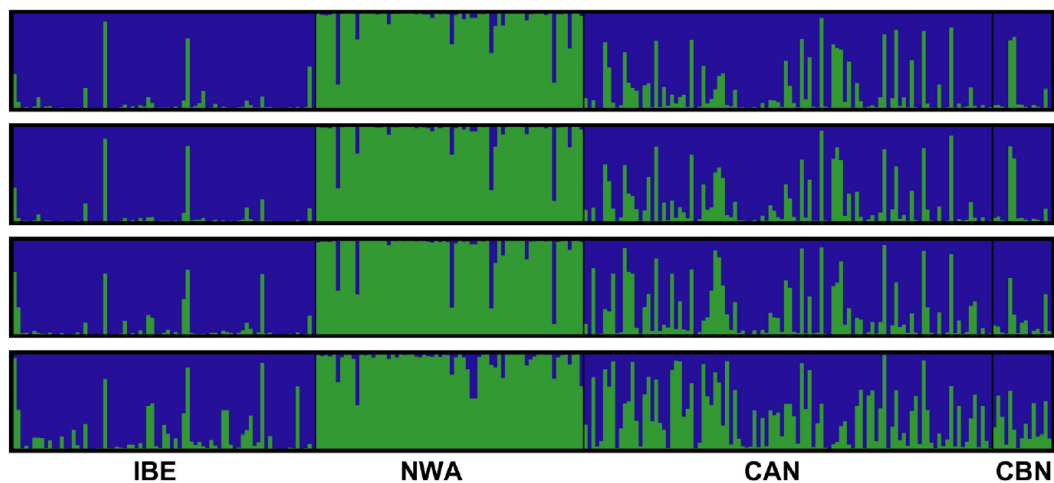
### **Text S1. Ancestry informativeness of EuroAIMs in samples from HGDP**

The initial evidence of ancestry informativeness of EuroAIMs for African and Spanish populations was obtained by comparing data from Mozabite Algerians and French Basques, the closest population proxy to Spanish included in HGDP, obtained using a commercial genome-wide SNP genotyping array [32]. We assumed that EuroAIMs were able to differentiate Algerians and Basques from sub-Saharan Africans given the sharp genetic differences of the latter population [32]. In the comparison of Mozabite Algerians and French Basques, the average  $F_{ST}$  for the 44 EuroAIMs included in the array was in the 85th percentile of the distribution for the  $\approx 642,000$  markers genotyped ( $F_{ST}=0.0669$ ), which is in the range of values allowing to separate NNW and SSE European populations [30]. As expected, this comparison also evidenced the availability of a few thousands AIMs showing large allele frequency differences ( $>0.3$ ) that would be potentially useful to enhance the detection of genetic ancestries for these two populations.



**Text S2. Ancestry informativeness of EuroAIMs subsets.**

Given that EuroAIMs were not ascertained to distinguish the population ancestries under study, we reasoned that subsets of markers would perform similarly in identifying populations and measuring African influences, potentially translating into reductions of genotype costs in future studies. Therefore, we explored the ability of subsets of 23, 46 and 69 markers ranking higher for different measures of ancestry informativeness to estimate individual membership to Iberian and Northwest African populations. For all subsets, STRUCTURE correctly identified the two populations and the individual membership to populations was fairly accurate as shown in the figure below:



**STRUCTURE results based on different subsets of EuroAIMs.** Ancestry inference of Canary Islanders collected for this study (CAN), Canary Islanders available from the National Spanish DNA Bank (CBN), Iberians (IBE) and Northwest Africans (NWA) using STRUCTURE based on 93 EuroAIMs and on the subsets of 69, 46 and 23 markers (from top to bottom). Each vertical line represents an individual where colors indicate the proportion of the individual's genome derived from each of the two inferred populations.

As shown in the table below, individual misclassification rate was lowest when 69 EuroAIMs were used and largest when only 23 of these were employed. Significant differences in individual misclassification were found only between runs with 69 and 23 EuroAIMs (4.1% vs 11.0%, respectively; two-tailed  $\chi^2$  test  $p=0.026$ ).

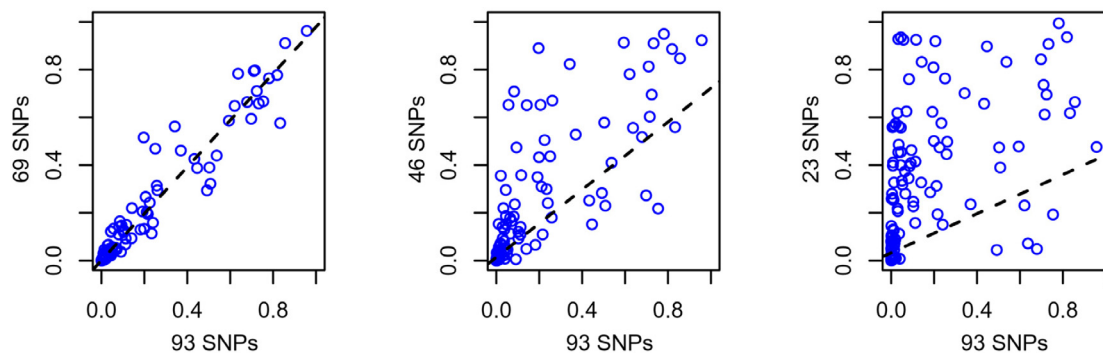
## RESULTADOS

### Sample misclassification and mean Northwest African influence in Canary Islanders with EuroAIMs subsets.

EuroAIMs subset	$F_{ST}^a$	Northwest Africans <sup>b</sup>	Iberians <sup>b</sup>	Northwest Af. influence
93	0.0422	5 (7.4%)	4 (5.2%)	0.168
69	0.0568	3 (4.4%)	3 (3.9%)	0.171
46	0.0805	5 (7.4%)	5 (6.5%)	0.219
23	0.1165	4 (5.9%)	12 (15.6%)	0.333

<sup>a</sup>Average  $F_{ST}$  between Northwest Africans and Iberians; <sup>b</sup>Cut-off for membership at 0.70 to self-reported population.

Since the Northwest African influence in Iberians was barely appreciable, we only evaluated these subsets for the appropriate estimation of ancestries in Canary Islanders. Our results showed that the estimated Northwest African influences biased upwards with decreasing number of markers as a consequence of the reduction of information, being almost two-fold for 23 EuroAIMs relative to those for 93 and 69 EuroAIMs (see the table above). In addition, as shown in the figure below, individual ancestry estimates obtained from the 93 markers were weakly correlated with those from 23 and 46, but highly correlated with those from 69 EuroAIMs, 65 of which would fit into two iPLEX™ reactions and still be sufficient for the appropriate estimation of ancestries ( $r^2=0.845$ ).



**Correlation of estimated Northwest African ancestries in Canary Islanders using subsets of markers.** Using those obtained from 93 EuroAIMs as a standard, correlation values ( $r^2$ ) were 0.930 with those from 69, 0.612 with those from 46, and 0.231 with those from 23.

**Text S3. Details of laboratory procedures**

Genomic DNA from the 104 samples from the Canary Islands was extracted from whole blood using the GFX kit (GE Healthcare, Little Chalfont, UK). All samples used for the study were whole genome amplified from ~10 ng using Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare) following manufacturers' recommendations. In order to verify that the reaction was successful and to quantify the resulting material, a fast SYBR Green I (Molecular Probes, Eugene, OR) real-time-based quantification protocol was established. Briefly, quantifications were performed with a 1/5,000 SYBR Green I final concentration and 1/200 diluted amplified products in 20 µl reactions using iCycler iQ Real-Time detection system (Bio-Rad Laboratories, Hercules, CA), recording the fluorescence at 520 nm under excitation at 490 nm at 25°C. A DNA concentration standard curve (1000, 750, 250, 100, 10 and 0 ng/ml) was prepared on each plate by including serial dilutions of a DNA molecular weight marker XIV (Roche, Basel, Switzerland). All dilutions were prepared fresh using a 5 mM Tris, 0.5 mM EDTA (pH=7.5) solution. DNA concentrations were obtained from the fluorescence intensities (as relative fluorescence units) interpolated on the standard curve.

Data from nine SNPs that gave poor quality data on the iPLEX™ Gold assay were discarded and finally determined utilizing two alternative simultaneous genotyping reactions (containing 7 and 2 SNPs). Briefly, PCR was performed to amplify two different multiplex reactions, using the HotStartTaq® Master Mix Kit (QIAGEN, Valencia, CA) in reactions of a final volume of 2.5 µl using the Veriti™ 96-Well Fast Thermal Cycler (Applied Biosystems, Foster City, CA) and the following thermal conditions: 94°C 15 min, followed by 45 cycles of 94°C 20 sec, 56°C 30 sec, 72°C 1 min, and a final extension step of 72°C 5 min. Reactions were purified by using ExoSAP-IT® (USB Corp., Cleveland, OH) and products used for single base extension reactions using SNaPshot® Multiplex Kit reactions (Applied Biosystems), and fragments resolved on ABI 310 or ABI 3500 (Applied Biosystems).

#### **4.2. Estudio de la relación entre el trasfondo genético poblacional y la susceptibilidad y gravedad del asma y fenotipos relacionados en la población canaria**

En este segundo trabajo se aplicó la metodología desarrollada en el estudio de estratificación poblacional para analizar la relación entre la ascendencia genética poblacional y el riesgo a padecer asma y otros fenotipos relacionados, como posible explicación de su elevada prevalencia en la población canaria comparada con el resto de España. Para ello, los 93 marcadores de ascendencia descritos en el primer artículo se genotiparon en 391 muestras de ADN de individuos asmáticos y 343 controles, todos ellos con dos generaciones de ancestros nacidos en Canarias. De los 93 marcadores, 83 se genotiparon adecuadamente en casos y controles y los datos obtenidos se emplearon en un PCA junto con datos de poblaciones de referencia (norteafricanos y peninsulares). Nuestros resultados mostraron que la ascendencia genética no se asoció con la susceptibilidad al asma ni fenotipos relacionados (función pulmonar, atopia, rinitis, IgE específica...). Sin embargo, no pudo descartarse la presencia de efectos leves en la susceptibilidad al asma o fenotipos relacionados debidos a la ascendencia genética. Por ello, serán necesarios nuevos estudios en los que se analicen un mayor número de muestras y marcadores de ascendencia.

Los resultados de este estudio han sido preparados para su publicación y aparecen a continuación como manuscrito con el título "*Association between genetic ancestry and asthma in the Canary Islands*".

## Association between genetic ancestry and asthma in the Canary Islands

María Pino-Yanes<sup>1,2</sup>, Almudena Corrales<sup>1,2</sup>, José Cumplido<sup>3</sup>, Ruperto González<sup>4</sup>,  
María José Torres-Galván<sup>5</sup>, Orlando Acosta Fernández<sup>6</sup>, Inmaculada Sánchez-Machín<sup>4</sup>,  
Javier Figueroa<sup>7</sup>, Anselmo Sánchez-Palacios<sup>7</sup>, Jesús Villar<sup>1,8,9</sup>, Mariano Hernández<sup>10</sup>,  
Teresa Carrillo<sup>3</sup>, Carlos Flores<sup>1,2</sup>

From:

<sup>1</sup>CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Spain; <sup>2</sup>Research Unit, Hospital Universitario NS de Candelaria, Tenerife, Spain; <sup>3</sup>Allergy Unit, Hospital Universitario Dr. Negrin, Las Palmas de Gran Canaria, Spain; <sup>4</sup>Allergy Unit, Hospital del Tórax, Complejo Hospitalario Universitario NS Candelaria, Tenerife, Spain; <sup>5</sup>Research Unit, Hospital Universitario Dr. Negrin, Las Palmas de Gran Canaria, Spain; <sup>6</sup>Neumology Unit, Hospital Universitario de Canarias, Tenerife, Spain; <sup>7</sup>Allergy Unit, Hospital Universitario Insular de Gran Canaria, Spain; <sup>8</sup>Multidisciplinary Organ Dysfunction Evaluation Research Network (MODERN), Research Unit, Hospital Universitario Dr. Negrin, Las Palmas de Gran Canaria, Spain; <sup>9</sup>Keenan Research Center, St. Michael's Hospital, Toronto, Canada; <sup>10</sup>Genetics Laboratory, Instituto Universitario de Enfermedades Tropicales y Salud Pública de Canarias, Universidad de La Laguna, Tenerife, Spain.

Address for correspondence:

Carlos Flores, PhD  
Unidad de Investigación  
Hospital Universitario N.S. de Candelaria  
Carretera del Rosario s/n  
38010 Santa Cruz de Tenerife  
Phone: (+34) 922-602938  
Fax: (+34) 922-600562  
e-mail: cflores@ull.es

Running title: Genetic ancestry and asthma in the Canary Islands

### Abstract

**Background:** Asthma is a complex respiratory disease characterized by chronic inflammation of airways and frequently associated with atopic symptoms. The population from the Canary Islands, which has resulted from a recent admixture of North African and Iberian populations, shows the highest prevalence of asthma and atopic phenotypes among the Spanish populations (17% compared to an average of 5% in the overall Spanish population). Although environmental particularities would account for the majority of such difference, genetic ancestry might be also involved in the higher prevalence of asthma, as have been demonstrated in other populations showing a recent African admixture.

**Objective:** Here we aimed to explore whether genetic ancestry was associated with asthma or related traits in the Canary Islanders.

**Methods:** A total of 734 DNA samples from the GOA study of unrelated individuals self-reporting at least two generations of ancestors from the Canary Islands (391 asthmatics and 343 controls) were successfully genotyped for 83 ancestry informative markers (AIMs), which allowed to precisely distinguishing between North African and Iberian ancestries.

**Results:** No association was found between genetic ancestry and asthma or asthma related traits after adjusting by demographic variables differing among compared groups. Similarly, none of the individual AIMs were associated with asthma when results were considered in the context of the multiple comparisons performed ( $0.005 \leq p\text{-value} \leq 0.042$ ;  $0.221 \leq q\text{-value} \leq 0.443$ ).

**Conclusion:** Our results suggest that if genetic ancestry is involved in the susceptibility to asthma or asthma related traits among Canary Islanders, its effects would be modest. Larger studies examining more genetic variants would be needed to explore such possibility.

**Word count:** 255 words

**Keywords:** allergy, genetic susceptibility, North Africa, admixture.

## Introduction

Asthma is a complex respiratory disease characterized by chronic inflammation of airways and frequently associated with atopic symptoms [1]. Although environmental factors are involved in its pathogenesis; familiar clustering, twin studies, and genetic studies also support an important genetic contribution to disease predisposition [2,3,4,5].

The prevalence of asthma varies widely around the world (1-18%), and predictions for the coming years indicate that an increase is expected for Western societies [6,7]. In Spain, the prevalence of asthma symptoms has been estimated around 5-6%, with considerable geographic variation among regions (5-12%) [8]. However, the highest prevalence of asthma in adults from Spain has been found in the Canary Islands (17.2%) based on the European Community Respiratory Health Survey (ECHRS) [8,9]. In addition, estimates of asthma prevalence in children aged 6-7 years are two-fold higher in the Canary Islands (18.4%) than the mean for the Spanish population (9.9%) [10]. Similarly, the prevalence of atopy in the Canary Islands is also elevated (40.6%) compared to the mainland Spanish populations [11].

The Canary Islands belongs politically to Spain, in spite of being located at about 1000 km from the closest European point in Iberian Peninsula, and only 100 km off the Northwest coast of Africa. These Islands were inhabited by aboriginal people related to the Berber populations at the time of Spanish occupation in the XV century, as pointed by historical, anthropological, archaeological and linguistic records [12]. Genetic footprints of such Northwest African influence (as a proxy for the aborigines) in Canary Islanders have been demonstrated in contemporary inhabitants using classical [13] and autosomal markers [14,15,16], as well as with mtDNA [17] and the Y-chromosome [18]. Recently, we have estimated North African admixture in Canary Islanders using almost a hundred ancestry informative markers (AIMs), genetic loci showing large allele frequency differences between populations, selected from different regions of the autosomal genome [16]. Intriguingly, North African ancestry averaged  $17\pm 25\%$ , but showed a wide inter-individual variation ranging from 0% to 96%, which was interpreted as a trace of the violent way the Islands were conquered and colonized.

Although prevalence of asthma is generally low within countries from West Africa, individuals with West African ancestry have shown a higher prevalence of asthma and related traits when exposed to a Western style of living compared to European descent populations [19]. This has been explained as a consequence of gene-environment interaction, by which the admixture African descent population would carry risk alleles for asthma and related traits because these alleles would provide them an adaptation to a higher load of parasites in their original populations that are no longer present in the new

## RESULTADOS

---

environment [20]. As a consequence, in a recent admixed population it is expected that genomic segments originated from the ancestral ethnic group with the higher risk will be enriched with risk alleles [21], which will be reflected in a larger proportion of ancestry from the ancestral population in affected individuals compared to the unaffected group. This hypothesis has been confirmed for African American and African Caribbean individuals, in which a higher African admixture has been associated with higher susceptibility for asthma and related traits, and with lower pulmonary function measures [22,23,24,25].

Similar to the West African populations, the prevalence of asthma is generally low in the North African countries [19]; being estimated in 3-4% for clinical diagnosed asthma in Algeria, Morocco and Tunisia [26]. However, a trend for a markedly increase of prevalence has been found in these populations over recent decades, due to the acquisition of a Western lifestyle and continued urban shifts [19]. In fact, in Morocco the highest prevalence of asthma and asthma symptoms, as well as of other allergic diseases, is found in Casablanca, the largest city and the economic capital of Morocco, compared with rural areas [27], which has experienced a significant increase in asthma symptoms (1.25% per year from 1986 to 2005). This increase in prevalence has been accompanied by an increased urbanization and the acquisition of urban life style of living [27]. In line with this data, the genetic expression profile has been found to be profoundly changed by an urban living style compared to a rural living style in the Amazighs, a Northwest African population that is thought to represent a genetically relatively homogeneous [28]. Therefore, a higher risk for asthma and related traits would be expected in populations with North African genetic ancestry with Western living style.

Although the Canary Islands have particular climatic conditions that could promote the emergence of atopic symptoms, it has been suggested that the genetic background of the population might also contribute to the high prevalence of atopic symptoms and asthma [9]. Since no study has explored whether genetic ancestry could partially explain the high prevalence of asthmatic and atopic symptoms in the Canary Islands, here we aimed to study if genetic ancestry, assessed from autosomal AIMs, was associated with asthma or asthma related traits in a case-control sample of Canary Islanders.



---

## Materials and Methods

### ***Ethics statement***

This study was approved by the Ethics Committees at participating hospitals, and written informed consent was obtained from all subjects.

### ***Study subjects***

This study was conducted using a case-control design of 734 DNA samples from unrelated individuals reporting at least two generations of ancestors born in the Canary Islands. See Table 1 for the characteristics of subjects included in the study.

Samples utilized constituted a subset of the Genetics of Asthma study (GOA) in the Spanish population, described elsewhere [29]. Briefly, cases were represented by 391 physician-diagnosed asthmatic patients aged >5 years and fulfilling the Global Initiative for Asthma guidelines for diagnosis and classification of asthma (<http://www.ginasthma.com>). These samples were collected from Respiratory Medicine and Allergy Departments from individuals who reported having at least two generations of ancestors born in the Canary Islands. Data regarding demographic variables (age, sex and smoking habits), age at diagnosis of the disease, severity of illness, family history of allergic diseases, allergic symptoms (rhinitis, atopic dermatitis, food and drug allergy), basal forced expiratory volume in one second (FEV<sub>1</sub>), and the symptomatic treatment were recorded. Atopy was confirmed by a positive prick test (SPT, a wheal with a diameter 3 mm greater than the saline control) and/or elevated specific serum IgE levels (>0.35 UI mL<sup>-1</sup>) to, at least, one of seven common allergens.

Controls consisted of 343 DNA samples from non-asthmatic unrelated adults obtained from blood bank donors who did not report a personal or familiar medical history for allergic or pulmonary diseases, and self-reporting at least two generations of ancestors born in the Canary Islands.

### ***Genotyping***

A total of 93 European AIMs (from now on referred as EuroAIMs) selected for their ability to distinguish between North African and Spanish ancestries were genotyped in cases and controls using iPLEX<sup>™</sup> Gold assays on MassARRAY system, by the Spanish National Genotyping Center, Santiago de Compostela Node (CeGen, <http://www.cegen.org/>) [16]. Briefly, iPLEX<sup>™</sup> assays were scanned by MALDI-TOF mass spectrometry and individual SNP genotype calls were automatically generated using Sequenom TYPER 3.4<sup>™</sup> software. Samples from the Coriell Institute for Medical Research (<http://www.coriell.org>) were included to test allele calling reliability samples of this platform.

### ***Statistical analysis***

Departures from Hardy-Weinberg equilibrium (HWE) were evaluated by separate in cases and controls using an exact test [30], by means of the SNPInfastats software [16]. Given its higher discriminatory power to detect differences among closely related populations, such as North Africans and Iberians [14,31,32], than alternative algorithms based on the assignation of individuals to discrete populations [33,34,35], principal component analysis (PCA) was utilized to derive individual ancestry estimates as scores of the first principal component (PC1) as the major axis covering the Iberian-North African genetic differences [16]. For this purpose, data for the same set of EuroAIMs from individuals self-declaring two generations of ancestors from the Iberian Peninsula (n=889), and Morocco in North Africa (n=68), were jointly analyzed with the case-control sample to serve as reference populations [16,29]. EIGENSOFT [36] was used to assess PCA and to test the association of ancestry estimates with asthma and atopic traits between cases and controls by means of ANOVA statistics.

Then, in order to adjust this association, clinical and demographic variables were included as covariates along with ancestry estimates in regression models utilizing SPSS 15.0 (SPSS Inc., Chicago, IL). Given that genetic ancestry estimates represented a mean value from the genome, we also explored the possibility of having local differences in genetic ancestry that could be detected by using the individual EuroAIMs as proxies. For that, SNPAssoc [37] was employed to test the association of individual EuroAIMs with asthma, adjusting for the genetic ancestry estimates by means of logistic regression models. Finally, in order to limit the expected proportion of false positives incurred in the study when a particular individual EuroAIM association test was called significant and to account for the multiple phenotypes tested for association with genetic ancestry, a False Discovery Rate (FDR) was computed using QVALUE [38] and a threshold  $q$ -value of 0.05 was established to declare significance.

## Results

### *Quality control measures*

Eight SNPs gave poor quality genotype data during genotyping (rs1032143, rs1073321, rs12502036, rs1364394, rs153595, rs1854226, rs2171209 and rs7108371) and were discarded from the study. In addition, two SNPs that deviated significantly from HWE after considering a multiple correction threshold of  $p < 6 \times 10^{-4}$  ( $0.05/85$ ) at least in one study sample (rs7277342 and rs2596501), were also dropped from the analyses. Therefore, further analyses included the remaining 83 EuroAIMs, which had an average completion rate of 98.0% ( $P_{25}$ - $P_{75}$ =97.5-98.6%), for a total of 734 samples, so that none of them was removed from the analyses (mean sample completion rate=98.0%,  $P_{25}$ - $P_{75}$ =98.8-100.0%). As previous findings have showed a high correlation between individual ancestry estimates obtained with the full set of 93 EuroAIMs and with a reduced subset of as few as 65 markers [16], the final number of EuroAIMs considered was still adequate for the purpose of the study. See Table S1 for further details.

### *Association of genetic ancestry with asthma and asthma related traits*

The comparison of genetic ancestries between asthmatic or atopic asthma patients with the control group did not reveal significant differences. Similarly, no association was found between genetic ancestry and a positive result neither for atopy, SPT or rhinitis (Table 2, Figure 1), nor with serum levels of total IgE or with pulmonary function, measured by the percentage of predicted FEV<sub>1</sub> (Table 2, Figure 2). Only among asthmatics patients, genetic ancestry was different between those with a positive result for specific IgE and those with a negative test, with nominal significance ( $p=0.004$ , Table 2, Figure 1), having those individuals with positive specific IgE a mean value of PC1 closer to the Iberian population than to the North Africans. Although this comparison was still significant after considering all traits tested ( $q=0.032$ ), this result was not significant after adjusting by the differences observed in age between the compared groups ( $p=0.056$ ).

### *Association of EuroAIMs with asthma*

Despite no evidence of association was found between genetic ancestry and asthma risk among Canary Islanders, we then determined whether any of the individual EuroAIMs were associated with asthma. Adjusting the association of each marker by ancestry differences between patients and controls, seven EuroAIMs were nominally associated ( $0.005 \leq p\text{-value} \leq 0.042$ ), although none of these associations was considered significant in the context of the multiple comparisons performed ( $0.221 \leq q\text{-value} \leq 0.443$ ) (Table S2).

### Discussion

In this study we have explored for the first time whether genetic ancestry in Canary Islanders was associated with asthma and related traits, as a possible explanation for the high prevalence of asthmatic and atopic symptoms compared with the rest of populations from Spain [9]. However, no association was found between genetic ancestry and asthma or asthma related phenotypes. Similarly, the examination of asthma associations at individual EuroAIMs did not show any asthma risk variant in these regions. These results are in agreement with self-reported information from grandparents of subjects studied within a large cohort study in the Canary Islands, where having two generations of ancestors from the Canary Islands was unrelated to the presence of asthmatic symptoms [9].

Although the Canary Islands have particular climatic conditions, such as high humidity and stable warm temperatures that provide ideal conditions for mites and molds, which favor the emergence of atopic and asthmatic symptoms in the Canary Islands [10], genetic factors are also be involved in this disease. According to the hygiene hypothesis [39], we speculated that genetic African ancestry in the Canary Islanders could provide a higher risk for asthma and atopy because the African ancestral population would have adapted to a pathogen-rich environment that no longer exists in the current urban society in the Canary Islands. In line with this, a recent study has shown that the pathogenic environment (mostly the exposure to helminthes) is the predominant driver of local adaptation, even stronger than climate [40]. In addition, allele frequencies in genes involved in immunity response have been strongly correlated with pathogenic environment [40]. Remarkably, the same study revealed that several genes associated at the genome-wide level with helminth diversity were also associated with susceptibility for asthma and atopy [41]. Under this scenario, the helminth-driven selective pressure is expected to favor individuals carrying alleles that allow a strong Th2 response and, therefore, to promote the transmission and spread of asthma-susceptibility variants [41].

Previous studies have examined the association of genetic ancestry and asthma susceptibility in recently admixed populations, observing a difference of 5-6% in African ancestry between cases and controls in independent African descent populations from the U.S. and the Caribbean [24,25]. Our statistical power was adequate ( $\geq 80\%$ ) to detect differences in North African admixture between cases and controls with effects even lower than the previously observed (up to 3.2% of differences in North African admixture, i.e. a standardized effect  $\geq 0.2$ ) (Figure 3). However, the presence of smaller effects, in case genetic ancestry had a true association with susceptibility to asthma or related traits, cannot be discarded. Similarly, it is likely that the use of a larger number of markers will allow estimating further ancestries among Canary Islanders. In this study, samples from different

locations in Morocco were used as a proxy for the aboriginal population inhabiting the Canary Islands before the conquest. Although this is a reasonable assumption given the evidence from the historical records [13,42] and the numerous previous genetic studies [13,17,18,43,44,45,46], this constitutes a simplification of the heterogeneous African influences that have affected the Canary Islands populations. Therefore, it is likely that further sampling of North African regions and the typing of additional markers might allow identifying other genetic influences in Canary Islanders, which could clarify if a relationship between genetic ancestry and asthma exists in this population.

In this study, we preferred to use PCA for the genetic ancestry estimation instead of model-based ancestry estimation, for example the population assignment admixture provided by STRUCTURE software [47]. This was motivated by a previous study where we observed a good linear correlation between PC1 scores and the STRUCTURE Northwest African estimates except for the values at the ends of the distribution. This is congruent with studies comparing STRUCTURE with alternative algorithms, which have reported biases for its ancestry estimates at the extremes of the distribution [48,49].

In conclusion, no association was found between genetic ancestry and asthma or asthma related traits among Canary Islanders. Thus, waiting for further studies with larger samples sizes and more markers to help resolve additional ancestries of Canary Islanders, the same genetic risks factor for asthma affecting the Spanish populations will be transferable to this population.

## **Acknowledgements**

This work was supported by grants from the Ministry of Science and Innovation of Spain (PI081383) and FUNCIS (27/07), and by a specific agreement between Instituto de Salud Carlos III and Gobierno de Canarias (EMER07/001).

### References

1. Ober C (2005) Perspectives on the past decade of asthma genetics. *J Allergy Clin Immunol* 116: 274-278.
2. Holberg CJ, Halonen M, Wright AL, Martinez FD (1999) Familial aggregation and segregation analysis of eosinophil levels. *Am J Respir Crit Care Med* 160: 1604-1610.
3. Thomsen SF, van der Sluis S, Kyvik KO, Skytthe A, Backer V (2010) Estimates of asthma heritability in a large twin sample. *Clin Exp Allergy* 40: 1054-1061.
4. Ober C, Hoffjan S (2006) Asthma genetics 2006: the long and winding road to gene discovery. *Genes Immun* 7: 95-100.
5. Bouzigon E, Forabosco P, Koppelman GH, Cookson WO, Dizier MH, et al. (2010) Meta-analysis of 20 genome-wide linkage studies evidenced new regions linked to asthma and atopy. *Eur J Hum Genet* 18: 700-706.
6. Braman SS (2006) The global burden of asthma. *Chest* 130: 4S-12S.
7. Pearce N, Ait-Khaled N, Beasley R, Mallo J, Keil U, et al. (2007) Worldwide trends in the prevalence of asthma symptoms: phase III of the International Study of Asthma and Allergies in Childhood (ISAAC). *Thorax* 62: 758-766.
8. The Spanish Group of the European Asthma Study (1995) The European Asthma Study. The prevalence of asthma-related symptoms in 5 Spanish areas. *Med Clin (Barc)* 104: 487-492.
9. Julia-Serda G, Cabrera-Navarro P, Acosta-Fernandez O, Martin-Perez P, Losada-Cabrera P, et al. (2011) High prevalence of asthma and atopy in the Canary Islands, Spain. *Int J Tuberc Lung Dis* 15: 536-541.
10. Sanchez-Lerma B, Morales-Chirivella FJ, Penuelas I, Blanco Guerra C, Mesa Lugo F, et al. (2009) High prevalence of asthma and allergic diseases in children aged 6 and 7 years from the Canary Islands: the International Study of Asthma and Allergies in Childhood. *J Investig Allergol Clin Immunol* 19: 383-390.
11. Julia-Serda G, Cabrera-Navarro P, Acosta-Fernandez O, Martin-Perez P, Losada-Cabrera P, et al. (2011) High prevalence of asthma and atopy in the Canary Islands, Spain. *Int J Tuberc Lung Dis* 15: 536-541.
12. Navarro J (1991) El poblamiento prehistórico. In: Morales-Padrón F, editor. *Historia de Canarias*. Madrid: Prensa Ibérica. pp. 41-60.
13. Flores C, Larruga JM, Gonzalez AM, Hernandez M, Pinto F, et al. (2001) The origin of the Canary Island aborigines and their contribution to the modern population: a molecular genetics perspective. *Curr Anthropol* 42: 749-755.
14. Maca-Meyer N, Villar J, Perez-Mendez L, Cabrera de Leon A, Flores C (2004) A tale of aborigines, conquerors and slaves: Alu insertion polymorphisms and the peopling of Canary Islands. *Ann Hum Genet* 68: 600-605.
15. Fregel R, Maca-Meyer N, Cabrera VM, Gonzalez AM, Larruga JM (2005) Description of a simple multiplex PCR-SSCP method for ABO genotyping and its application to the peopling of the Canary Islands. *Immunogenetics* 57: 572-578.
16. Pino-Yanes M, Corrales A, Basaldúa S, Hernández A, Guerra L, et al. (2011) North African influences and potential bias in case-control association studies in the Spanish population. *PLoS One* 6: e18389.
17. Rando JC, Cabrera VM, Larruga JM, Hernandez M, Gonzalez AM, et al. (1999) Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. *Ann Hum Genet* 63: 413-428.

18. Flores C, Maca-Meyer N, Perez JA, Gonzalez AM, Larruga JM, et al. (2003) A predominant European ancestry of paternal lineages from Canary Islanders. *Ann Hum Genet* 67: 138-152.
19. National Heart, Lung and Blood Institute, National Institute of Health (2004) Global Burden of Asthma (available at [www.ginasthma.org](http://www.ginasthma.org)).
20. Le Souef PN, Candelaria P, Goldblatt J (2006) Evolution and respiratory genetics. *Eur Respir J* 28: 1258-1263.
21. Smith MW, Patterson N, Lautenberger JA, Truelove AL, McDonald GJ, et al. (2004) A high-density admixture map for disease gene discovery in African Americans. *Am J Hum Genet* 74: 1001-1013.
22. Tsai HJ, Kho JY, Shaikh N, Choudhry S, Naqvi M, et al. (2006) Admixture-matched case-control study: a practical approach for genetic association studies in admixed populations. *Hum Genet* 118: 626-639.
23. Kumar R, Seibold MA, Aldrich MC, Williams LK, Reiner AP, et al. (2010) Genetic ancestry in lung-function predictions. *N Engl J Med* 363: 321-330.
24. Flores C, Ma SF, Pino-Yanes M, Wade MS, Pérez-Méndez L, et al. (2011) African ancestry is associated with asthma risk in African Americans. *PLoS One*, in press.
25. Vergara C, Caraballo L, Mercado D, Jimenez S, Rojas W, et al. (2009) African ancestry is associated with risk of asthma and high total serum IgE in a population from the Caribbean Coast of Colombia. *Hum Genet* 125: 565-579.
26. Nafti S, Taright S, El Ftouh M, Yassine N, Benkheder A, et al. (2009) Prevalence of asthma in North Africa: the Asthma Insights and Reality in the Maghreb (AIRMAG) study. *Respir Med* 103 Suppl 2: S2-11.
27. Bouayad Z, Aichane A, Afif A, Benouhoud N, Trombati N, et al. (2006) Prevalence and trend of self-reported asthma and other allergic disease symptoms in Morocco: ISAAC phase I and III. *Int J Tuberc Lung Dis* 10: 371-377.
28. Idaghdour Y, Storey JD, Jadallah SJ, Gibson G (2008) A genome-wide gene expression signature of environmental geography in leukocytes of Moroccan Amazighs. *PLoS Genet* 4: e1000052.
29. Pino-Yanes M, Sánchez-Machín I, Cumplido J, Figueroa J, Torres-Galván MJ, et al. (2011) *IRAK3* gene variants associate with asthma in a replication study in the Spanish population. *J Allergy Clin Immunol*, in press.
30. Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet* 76: 887-893.
31. Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, et al. (2000) Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. *Hum Genet* 107: 312-319.
32. Bosch E, Calafell F, Perez-Lezaun A, Clarimon J, Comas D, et al. (2000) Genetic structure of north-west Africa revealed by STR analysis. *Eur J Hum Genet* 8: 360-366.
33. Price AL, Butler J, Patterson N, Capelli C, Pascali VL, et al. (2008) Discerning the ancestry of European Americans in genetic association studies. *PLoS Genet* 4: e236.
34. Heath SC, Gut IG, Brennan P, McKay JD, Bencko V, et al. (2008) Investigation of the fine structure of European populations with applications to disease association studies. *Eur J Hum Genet* 16: 1413-1429.
35. Li Q, Yu K (2008) Improved correction for population stratification in genome-wide association studies by identifying hidden population structures. *Genet Epidemiol* 32: 215-226.

## RESULTADOS

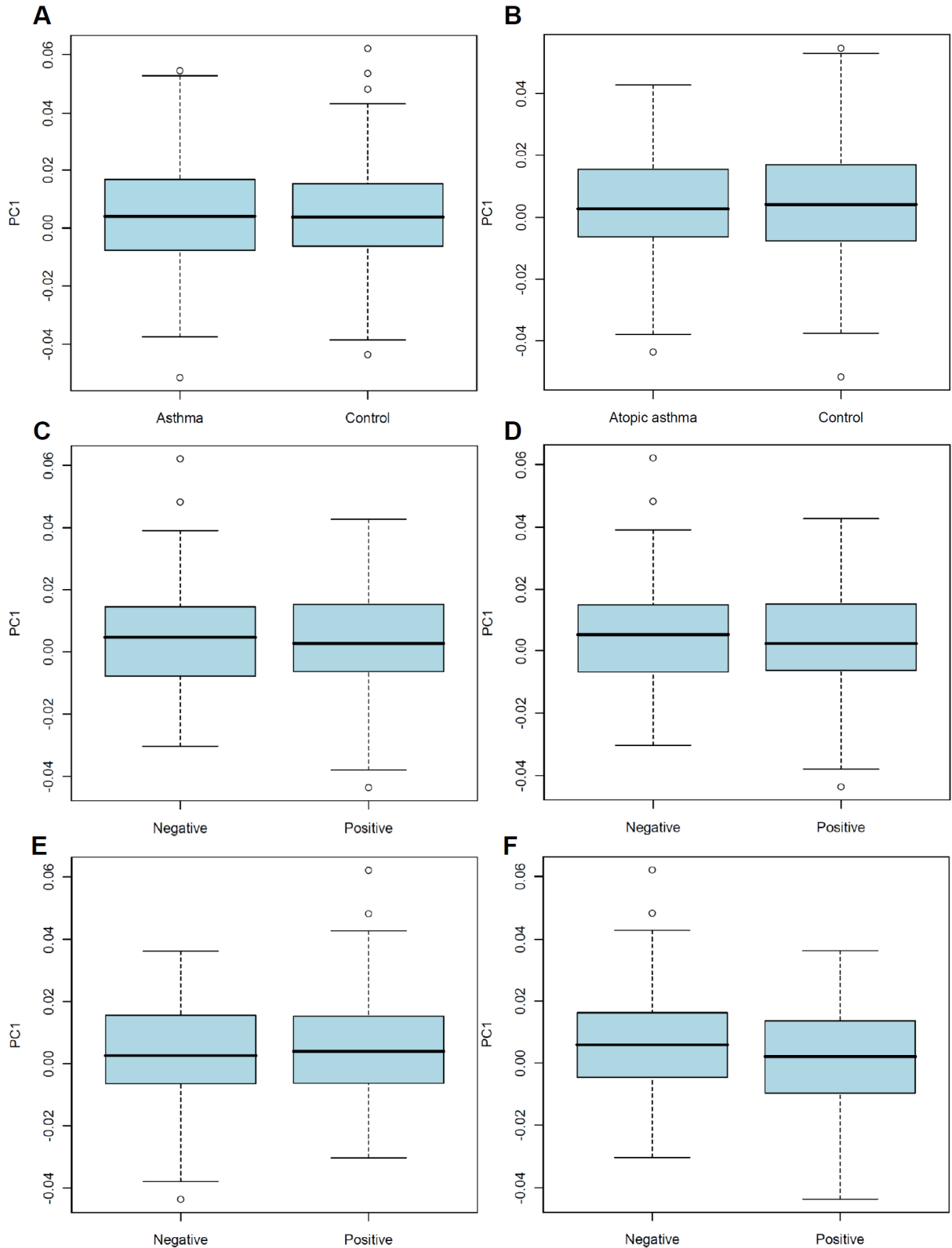
---

36. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38: 904-909.
37. Gonzalez JR, Armengol L, Sole X, Guino E, Mercader JM, et al. (2007) SNPAssoc: an R package to perform whole genome association studies. *Bioinformatics* 23: 644-645.
38. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 100: 9440-9445.
39. Strachan DP (1989) Hay fever, hygiene, and household size. *BMJ* 299: 1259-1260.
40. Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admettla A, Pattini L, et al. (2011) Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genet* 7: e1002355.
41. Fumagalli M, Pozzoli U, Cagliani R, Comi GP, Bresolin N, et al. (2010) The landscape of human genes involved in the immune response to parasitic worms. *BMC Evol Biol* 10: 264.
42. Chejne A (1974) *Muslim Spain. Its History and Culture*. Minneapolis, MN: University of Minnesota Press.
43. Flores C, Maca-Meyer N, Gonzalez AM, Oefner PJ, Shen P, et al. (2004) Reduced genetic structure of the Iberian Peninsula revealed by Y-chromosome analysis: implications for population demography. *Eur J Hum Genet* 12: 855-863.
44. Alonso S, Flores C, Cabrera V, Alonso A, Martin P, et al. (2005) The place of the Basques in the European Y-chromosome diversity landscape. *Eur J Hum Genet* 13: 1293-1302.
45. Adams SM, Bosch E, Balaesque PL, Ballereau SJ, Lee AC, et al. (2008) The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *Am J Hum Genet* 83: 725-736.
46. Fregel R, Gomes V, Gusmao L, Gonzalez AM, Cabrera VM, et al. (2009) Demographic history of Canary Islands male gene-pool: replacement of native lineages by European. *BMC Evol Biol* 9: 181.
47. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
48. Tsai HJ, Choudhry S, Naqvi M, Rodriguez-Cintron W, Burchard EG, et al. (2005) Comparison of three methods to estimate genetic ancestry and control for stratification in genetic association studies among admixed populations. *Hum Genet* 118: 424-433.
49. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19: 1655-1664.



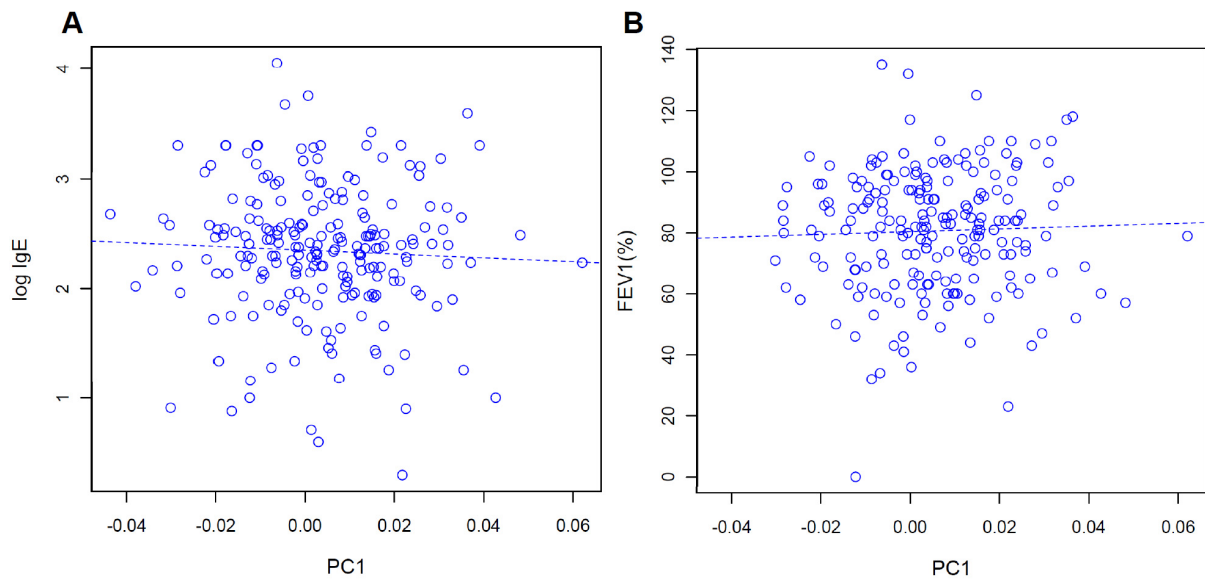
Figures

**Figure 1.** Box plots showing genetic ancestry estimates in compared groups; A) Asthma versus controls, B) Atopic asthma versus control C) Atopy positiveness, D) SPT positiveness, E) Rhinitis positiveness; and F) IgE specific positiveness.

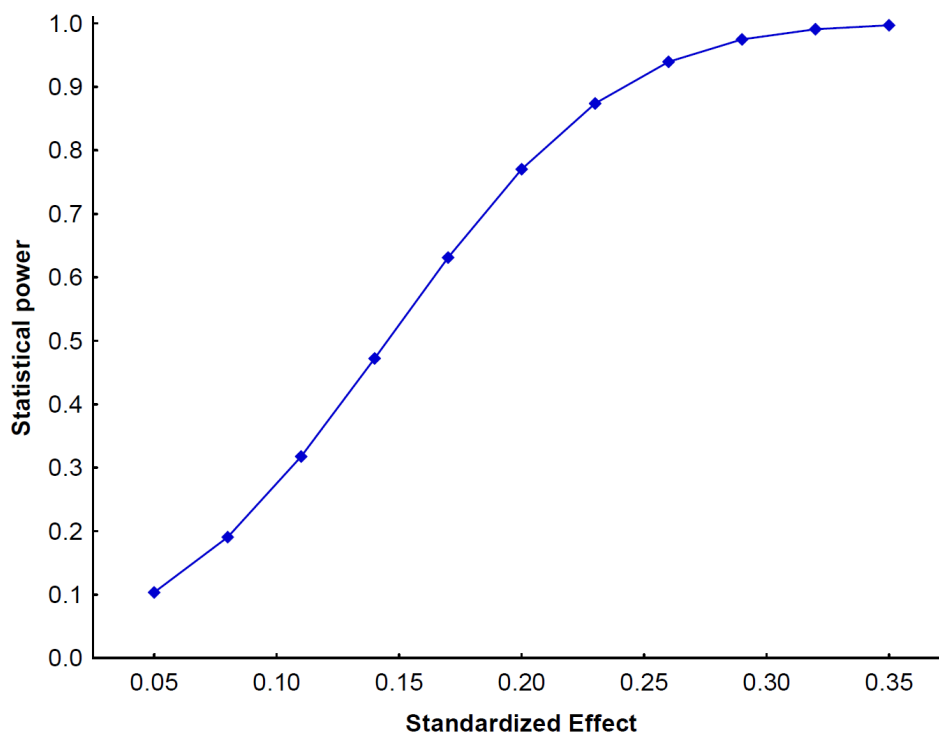


## RESULTADOS

**Figure 2.** Scatter plots showing the correlation between genetic ancestry estimates and clinical measures, A) log specific IgE serum levels and B) FEV1 (%).



**Figure 3.** Statistical power to detect differences in the genetic ancestry proportions by standardized effects for the sample size utilized (391 cases and 343 controls).



## Tables

**Table 1.** Relevant demographic and clinical features of samples.

Variable	Cases (n=391)	Controls (n=343)	p-value
Gender (% of males)	37.4 (387)	62.6 (343)	<0.001 <sup>a</sup>
Age (years); mean, P <sub>25</sub> -P <sub>75</sub> <sup>b</sup>	38, 22-48 (352)	36, 31-45 (343)	0.087 <sup>c</sup>
Smoking habits (% of ever smokers)	20.8 (126)	79.2 (234)	<0.001 <sup>a</sup>
Clinical features:			
Serum IgE levels (kUI mL <sup>-1</sup> ); mean, P <sub>25</sub> -P <sub>75</sub> <sup>b</sup>	499.1, 89.3-425.5 (163)	NA	
Age at diagnosis (years); mean, P <sub>25</sub> -P <sub>75</sub> <sup>b</sup>	24, 10-32 (356)	NA	
FEV1 (% predicted); mean, P <sub>25</sub> -P <sub>75</sub> <sup>b</sup>	81, 67-95 (206)	NA	
Rhinitis (% positive)	44.9 (273)	NA	
Dermatitis (% positive)	4.0 (273)	NA	
SPT (% positive)	60.4 (240)	NA	
Specific IgE (% positive)	42.7 (225)	NA	
Atopy (% positive)	61.7 (256)	NA	
Treatment:			
Inhaled corticoids (%)	86.5 (222)	NA	
Oral corticoids (%)	2.9 (279)	NA	
Long-acting $\beta$ 2-agonists (%)	53.3 (291)	NA	
Short acting $\beta$ 2 agonists (%)	81.5 (205)	NA	

<sup>a</sup> $\chi^2$  test; <sup>b</sup>P<sub>25</sub>, Percentile 25; P<sub>75</sub>, Percentile 75; <sup>c</sup>Mann-Whitney U-test; NA, not available. In parentheses, total number of subjects with available data.

## RESULTADOS

**Table 2.** Association between genetic ancestry and asthmatic and atopic phenotypes.

Clinical variable	Association <i>p</i> -value ( <i>q</i> -value)		Beta
	Unadjusted	Adjusted	
Cases versus controls			
Asthma	0.379 (0.599)	0.667 <sup>a</sup> (0.889)	-3.21
Atopic asthma	0.501 (0.599)	0.861 <sup>a</sup> (0.958)	1.53
Asthmatic patients			
Atopy (positiveness)	0.408 (0.599)	0.473 <sup>b</sup> (0.889)	-5.96
SPT (positiveness)	0.363 (0.599)	0.344 <sup>b</sup> (0.889)	-7.99
Rhinitis (positiveness)	0.137 (0.548)	0.452 <sup>b</sup> (0.889)	6.51
Specific IgE (positiveness)	0.004 (0.032)	0.056 <sup>b</sup> (0.448)	-18.66
Total IgE levels (log scale)	0.524 (0.599)	0.958 <sup>c</sup> (0.958)	-0.13
FEV1 (%)	0.609 (0.609)	0.576 <sup>c</sup> (0.889)	43.95

<sup>a</sup>Logistic regression model including gender and smoking habits as covariates.

<sup>b</sup>Logistic regression model including the age as a covariate.

<sup>c</sup>Linear regression model including the age as a covariate.

## Supplementary material

**Table S1.** Summary and quality control measures for the EuroAIMs successfully genotyped in this study.

**Table S2.** Association of individual EuroAIMs with asthma.

**Table S1.** Summary and quality control measures for the EuroAIMs successfully genotyped in this study.

rs#	Alleles	Chr.	Position <sup>a</sup>	CR (%) <sup>b</sup>	Reference allele			Hardy-Weinberg equilibrium $p$ -value <sup>d</sup>				
					Allele	Frec. IBE	Frec. NWA	Delta <sup>c</sup>	IBE	NWA	Cases	Controls
rs11807062	A/C	1	3176394	97.6	C	0.16	0.38	0.22	0.133	0.118	0.526	0.293
rs495347	C/T	1	17909228	98.6	C	0.31	0.48	0.17	0.874	0.628	0.191	0.437
rs1416467	A/G	1	80470749	98.7	A	0.49	0.35	0.14	0.736	<b>0.018</b>	0.149	0.663
rs7552548	A/G	1	82775867	96.3	G	0.30	0.39	0.09	0.293	0.799	0.224	0.100
rs1890131	C/T	1	164329131	98.3	T	0.32	0.09	0.23	0.352	1.000	0.801	0.112
rs725974	G/T	1	167344909	96.5	G	0.18	0.43	0.25	0.817	0.452	0.486	1.000
rs2236876	A/G	1	169749143	97.1	A	0.30	0.43	0.13	0.523	0.459	0.102	0.807
rs2419063	A/G	1	186988492	97.7	G	0.09	0.18	0.09	0.422	0.440	0.149	0.343
rs1157492	C/T	1	211627299	97.8	C	0.45	0.25	0.20	0.376	0.330	0.515	0.345
rs6432110	A/T	2	10688616	98.0	T	0.44	0.39	0.05	<b>0.029</b>	0.128	0.353	0.428
rs4832640	C/T	2	19025771	97.0	C	0.25	0.43	0.18	0.787	0.622	0.068	0.167
rs1517407	A/G	2	63401982	98.1	G	0.48	0.36	0.12	<b>0.013</b>	0.292	<b>0.040</b>	0.912
rs6745653	C/T	2	97748515	98.8	T	0.11	0.26	0.15	1.000	0.362	1.000	0.820
rs10496610	C/G	2	123757682	99.2	G	0.24	0.07	0.17	0.314	<b>0.020</b>	0.881	<b>0.045</b>
rs3769005	C/G	2	136437098	97.9	C	0.47	0.60	0.13	0.417	0.206	0.305	0.581
rs1448314	A/G	2	223811705	98.1	A	0.08	0.20	0.12	0.504	0.276	0.570	0.119
rs2596834	C/T	3	10795067	95.0	C	0.51	0.43	0.08	0.585	0.138	0.467	<b>0.002</b>
rs822759	G/T	3	22948116	98.4	T	0.32	0.51	0.19	0.534	1.000	<b>0.026</b>	0.175
rs1879558	C/T	3	155137053	98.6	C	0.19	0.22	0.03	0.445	0.288	0.399	0.581
rs9290675	C/T	3	180489335	99.1	C	0.13	0.20	0.07	0.458	0.718	0.167	0.167
rs4686497	A/G	3	190124209	98.2	A	0.16	0.32	0.16	0.615	1.000	0.195	<b>0.018</b>
rs9861816	C/G	3	195649797	98.1	C	0.27	0.40	0.13	0.103	0.326	1.000	0.765
rs9328764	A/G	4	2143685	98.4	A	0.15	0.48	0.33	0.245	1.000	0.731	0.469
rs2014303	A/C	4	10262125	97.5	A	0.23	0.38	0.15	1.000	0.611	0.313	0.297
rs1873195	C/T	4	38713739	98.4	T	0.31	0.26	0.05	1.000	1.000	0.052	0.479
rs17443616	A/G	4	41063093	98.3	A	0.29	0.44	0.15	0.462	0.052	0.231	0.070
rs4555709	C/T	4	53867889	99.4	T	0.07	0.19	0.12	<b>0.032</b>	0.695	0.506	0.469
rs10516982	G/T	4	96963279	98.0	T	0.26	0.53	0.27	0.600	0.224	0.188	0.796
rs974020	C/T	4	117865302	98.4	C	0.31	0.22	0.09	0.050	0.485	0.447	0.779

Table S1. (Continued)

rs#	Alleles	Chr.	Position <sup>a</sup>	CR (%) <sup>b</sup>	Reference allele			Hardy-Weinberg equilibrium p-value <sup>d</sup>				
					Allele	Frec. IBE	Frec. NWA	Delta <sup>c</sup>	IBE	NWA	Cases	Controls
rs1373557	G/T	4	167601812	98.8	T	0.21	0.24	0.03	0.258	<b>0.041</b>	0.057	0.857
rs2251432	A/G	4	191063741	97.8	A	0.21	0.28	0.07	0.609	0.763	0.612	0.393
rs16891982	C/G	5	33987450	99.2	C	0.17	0.66	0.49	0.128	0.280	0.904	0.407
rs33706	C/G	5	87540920	98.5	G	0.27	0.38	0.11	0.447	0.800	0.287	<b>0.029</b>
rs3822616	A/G	5	94828145	98.6	A	0.20	0.34	0.14	<b>0.014</b>	0.102	0.352	0.278
rs2804756	C/T	6	698586	99.1	T	0.08	0.10	0.02	<b>0.007</b>	0.099	0.293	0.568
rs10484547	C/G	6	29560753	99.5	C	0.06	0.15	0.09	0.765	1.000	1.000	1.000
rs2596501	C/T	6	31429190	95.8	C	0.34	0.36	0.02	<b>6.6x10<sup>-5</sup></b>	0.285	<b>0.003</b>	<b>0.008</b>
rs2187684	C/T	6	32872697	97.9	C	0.37	0.33	0.04	0.168	0.174	0.635	0.068
rs756147	A/G	6	141672412	99.6	A	0.05	0.12	0.07	0.075	0.583	1.000	<b>0.035</b>
rs2097884	A/C	7	4125658	97.2	C	0.25	0.39	0.14	0.856	0.614	0.511	0.090
rs10486207	A/G	7	7884020	98.3	G	0.29	0.35	0.06	0.870	1.000	0.439	<b>0.036</b>
rs2905347	A/G	7	22393559	95.8	A	0.43	0.60	0.17	1.000	0.456	0.403	0.305
rs17864053	A/G	7	89801957	98.6	A	0.21	0.25	0.04	0.191	0.518	0.303	0.736
rs2219248	C/T	7	113965977	98.7	C	0.37	0.22	0.15	0.389	0.722	0.202	0.370
rs2367191	A/G	7	141571268	98.2	A	0.42	0.19	0.23	0.781	1.000	0.394	0.363
rs1922086	C/T	7	155515702	97.5	T	0.38	0.58	0.20	0.100	1.000	0.597	0.054
rs920590	C/T	8	19695441	98.0	C	0.35	0.48	0.13	0.303	0.811	0.504	0.216
rs10504924	C/T	8	94157437	97.6	T	0.09	0.12	0.03	0.830	0.215	0.367	0.770
rs2086085	A/C	9	1680761	99.3	C	0.10	0.16	0.06	0.065	0.343	<b>0.015</b>	0.417
rs1408794	A/G	9	12641340	94.6	A	0.46	0.57	0.11	0.248	0.805	<b>0.014</b>	<b>0.002</b>
rs10512122	C/G	9	81715524	98.5	C	0.15	0.35	0.20	0.796	0.586	0.489	0.201
rs10508372	A/G	10	9012024	99.6	A	0.08	0.19	0.11	0.825	0.698	0.757	0.755
rs1045873	A/C	10	25177778	98.4	C	0.31	0.19	0.12	0.814	0.684	0.899	<b>0.005</b>
rs7908825	C/G	10	75224547	97.4	C	0.31	0.29	0.02	0.472	0.144	0.276	0.560
rs10509384	A/G	10	78693188	97.1	G	0.31	0.12	0.19	1.000	0.194	1.000	0.092
rs379773	C/G	10	109905342	95.0	G	0.35	0.42	0.07	0.331	0.806	1.000	<b>0.028</b>
rs10509954	A/G	10	113658378	98.4	A	0.10	0.13	0.03	0.276	1.000	<b>0.011</b>	0.211
rs1560569	A/G	11	8896463	97.3	G	0.35	0.20	0.15	0.882	0.714	0.727	<b>0.032</b>

Table S1. (Continued)

rs#	Alleles	Chr.	Position <sup>a</sup>	CR (%) <sup>b</sup>	Allele	Reference allele			Hardy-Weinberg equilibrium p-value <sup>d</sup>			
						Frec. IBE	Frec. NWA	Delta <sup>c</sup>	IBE	NWA	Cases	Controls
rs923031	A/C	11	15777602	98.2	A	0.48	0.34	0.14	0.280	0.587	0.472	0.498
rs4938377	G/T	11	116801502	98.0	G	0.06	0.04	0.02	0.107	1.000	<b>0.037</b>	0.136
rs2847502	C/T	11	119623708	98.9	T	0.30	0.38	0.08	0.809	1.000	0.273	0.175
rs1003306	A/T	12	2624458	97.9	A	0.32	0.49	0.17	0.282	1.000	0.064	1.000
rs7965049	A/G	12	7537044	97.3	A	0.25	0.35	0.10	0.649	0.419	0.365	0.232
rs998401	A/G	12	53641500	97.9	A	0.08	0.05	0.03	0.054	1.000	1.000	1.000
rs3809125	C/T	12	55130616	98.4	T	0.36	0.31	0.05	0.379	1.000	0.510	0.299
rs1582398	A/G	12	78624019	97.6	G	0.11	0.35	0.24	0.057	1.000	0.536	<b>0.024</b>
rs986642	C/G	13	56944263	98.8	G	0.25	0.11	0.14	0.478	0.171	<b>0.023</b>	0.124
rs7997100	C/T	13	67988075	98.3	T	0.35	0.39	0.04	0.302	0.445	0.910	<b>0.001</b>
rs10483853	A/G	14	72826052	98.5	G	0.22	0.47	0.25	0.145	0.633	0.113	1.000
rs1129038	C/T	15	26030454	98.4	C	0.65	0.95	0.30	0.941	<b>0.006</b>	0.734	1.000
rs7163907	C/T	15	73632152	98.8	C	0.29	0.34	0.05	0.628	1.000	0.219	0.319
rs10519269	A/C	15	77705433	96.3	C	0.10	0.23	0.13	0.247	0.737	0.366	0.452
rs1476162	A/G	17	8031808	98.6	G	0.12	0.31	0.19	0.434	0.779	0.308	1.000
rs1107820	C/T	17	41491195	97.2	C	0.28	0.16	0.12	0.802	0.358	0.157	0.659
rs2003092	A/G	17	59516509	97.8	G	0.26	0.41	0.15	0.484	0.618	0.240	0.526
rs959260	C/T	17	70881017	98.1	C	0.21	0.34	0.13	0.759	0.590	0.655	0.653
rs523776	C/T	18	7554299	98.2	T	0.24	0.40	0.16	0.852	0.804	0.372	0.661
rs2418844	A/G	18	26037261	97.4	G	0.48	0.41	0.07	0.787	0.618	0.682	0.180
rs959763	G/T	18	56797826	98.0	G	0.36	0.36	0.00	0.609	1.000	0.387	0.190
rs4892082	A/C	18	69051672	97.5	A	0.10	0.40	0.30	0.357	0.141	0.681	0.131
rs103294	C/T	19	59489660	98.6	T	0.17	0.06	0.11	0.636	1.000	0.540	0.650
rs202546	C/T	20	1611539	97.0	T	0.57	0.47	0.10	0.240	1.000	1.000	0.497
rs477627	A/G	20	47613465	97.5	A	0.19	0.21	0.02	0.583	1.000	0.093	0.726
rs7277342	A/C	21	29040845	95.0	A	0.32	0.48	0.16	<b>5.4x10<sup>-22</sup></b>	<b>1.7 x10<sup>-4</sup></b>	<b>7.8 x10<sup>-24</sup></b>	<b>1 x10<sup>-30</sup></b>
rs969539	A/G	22	24942341	96.8	A	0.39	0.16	0.23	0.478	0.190	0.218	0.105

<sup>a</sup>According to NCBI Build 35; <sup>b</sup>Completion rate; <sup>c</sup>Allele frequency difference between Iberians and Northwest Africans; <sup>d</sup>Significant p-values in boldface. IBE, Iberians; NWA, Northwest Africans.



**Table S2.** Association of individual EuroAIMs with asthma.

rs	Chr.	Minor allele frequency		Association results <sup>a</sup>	
		Cases	Controls	p-value <sup>b</sup>	q-value
rs11807062	1	0.21	0.20	0.622	0.848
rs495347	1	0.31	0.30	0.642	0.848
rs1416467	1	0.45	0.49	0.091	0.545
rs7552548	1	0.31	0.37	<b>0.014</b>	0.258
rs1890131	1	0.28	0.25	0.240	0.623
rs725974	1	0.24	0.21	0.129	0.560
rs2236876	1	0.33	0.35	0.442	0.841
rs2419063	1	0.12	0.13	0.604	0.848
rs1157492	1	0.38	0.37	0.536	0.841
rs6432110	2	0.46	0.41	<b>0.039</b>	0.443
rs4832640	2	0.27	0.27	0.908	0.852
rs1517407	2	0.47	0.47	0.922	0.852
rs6745653	2	0.13	0.14	0.797	0.848
rs10496610	2	0.22	0.20	0.624	0.848
rs3769005	2	0.48	0.48	0.816	0.848
rs1448314	2	0.10	0.10	0.874	0.852
rs2596834	3	0.47	0.49	0.178	0.593
rs822759	3	0.36	0.33	0.318	0.688
rs1879558	3	0.19	0.18	0.735	0.848
rs9290675	3	0.13	0.14	0.528	0.841
rs4686497	3	0.23	0.19	0.073	0.536
rs9861816	3	0.26	0.24	0.563	0.847
rs9328764	4	0.18	0.18	0.920	0.852
rs2014303	4	0.23	0.20	0.123	0.560
rs1873195	4	0.27	0.26	0.889	0.852
rs17443616	4	0.31	0.35	0.096	0.545
rs4555709	4	0.09	0.08	0.733	0.848
rs10516982	4	0.31	0.31	0.970	0.862
rs974020	4	0.28	0.27	0.675	0.848
rs1373557	4	0.20	0.19	0.481	0.841
rs2251432	4	0.19	0.20	0.536	0.841
rs16891982	5	0.30	0.27	0.201	0.593
rs33706	5	0.31	0.31	0.778	0.848
rs3822616	5	0.21	0.18	0.258	0.634
rs2804756	6	0.08	0.11	0.080	0.536
rs10484547	6	0.09	0.09	0.810	0.848
rs2187684	6	0.31	0.34	0.303	0.688
rs756147	6	0.09	0.06	<b>0.042</b>	0.443
rs2097884	7	0.26	0.26	0.921	0.852
rs10486207	7	0.28	0.27	0.815	0.848
rs2905347	7	0.45	0.47	0.694	0.848
rs17864053	7	0.22	0.20	0.455	0.841
rs2219248	7	0.33	0.31	0.563	0.847

## RESULTADOS

**Table S2.** (Continued)

rs	Chr.	Minor allele frequency		Association results <sup>a</sup>	
		Cases	Controls	p-value <sup>b</sup>	q-value
rs2367191	7	0.39	0.40	0.676	0.848
rs1922086	7	0.41	0.44	0.189	0.593
rs920590	8	0.36	0.42	<b>0.009</b>	0.221
rs10504924	8	0.13	0.10	0.170	0.593
rs2086085	9	0.14	0.11	0.237	0.623
rs1408794	9	0.46	0.47	0.810	0.848
rs10512122	9	0.18	0.22	0.056	0.516
rs10508372	10	0.09	0.10	0.682	0.848
rs1045873	10	0.28	0.27	0.851	0.852
rs7908825	10	0.30	0.37	<b>0.008</b>	0.221
rs10509384	10	0.27	0.27	0.946	0.861
rs379773	10	0.33	0.36	0.315	0.688
rs10509954	10	0.10	0.13	0.193	0.593
rs1560569	11	0.33	0.33	0.906	0.852
rs923031	11	0.46	0.42	0.146	0.567
rs4938377	11	0.08	0.08	0.749	0.848
rs2847502	11	0.30	0.33	0.160	0.590
rs1003306	12	0.38	0.38	0.924	0.852
rs7965049	12	0.28	0.24	0.138	0.565
rs998401	12	0.07	0.06	0.477	0.841
rs3809125	12	0.37	0.39	0.309	0.688
rs1582398	12	0.15	0.18	0.077	0.536
rs986642	13	0.25	0.20	<b>0.019</b>	0.280
rs7997100	13	0.34	0.34	0.963	0.862
rs10483853	14	0.26	0.27	0.667	0.848
rs1129038	15	0.35	0.28	<b>0.005</b>	0.221
rs7163907	15	0.35	0.32	0.245	0.623
rs10519269	15	0.13	0.13	0.741	0.848
rs1476162	17	0.15	0.17	0.237	0.623
rs1107820	17	0.23	0.25	0.497	0.841
rs2003092	17	0.32	0.32	0.809	0.848
rs959260	17	0.22	0.24	0.331	0.688
rs523776	18	0.22	0.25	0.121	0.560
rs2418844	18	0.49	0.48	0.507	0.841
rs959763	18	0.38	0.36	0.468	0.841
rs4892082	18	0.15	0.18	0.105	0.553
rs103294	19	0.15	0.14	0.705	0.848
rs202546	20	0.44	0.42	0.617	0.848
rs477627	20	0.21	0.20	0.456	0.841
rs969539	22	0.36	0.34	0.336	0.688

<sup>a</sup>Significant associations in boldface; <sup>b</sup>Adjusted for genetic ancestry estimates.

### 4.3. Asociación de variantes del gen *IRAK3* con la susceptibilidad al asma en la población española

Dado que el gen *IRAK3* ha sido asociado con el riesgo de asma persistente con comienzo anterior a los 13 años en poblaciones italianas, tras su clonaje posicional (Balaci *et al.*, 2007), y que este resultado no se ha replicado en otros trabajos independientes, nos planteamos realizar un estudio de replicación en la población española.

Para conocer la variación existente en el gen *IRAK3*, a partir de datos de secuenciación del gen en 32 individuos españoles, se seleccionó un conjunto de marcadores que representasen adecuadamente la variación común detectada, y se evaluó la capacidad de los tSNPs seleccionados para inferir la variación en SNPs del gen no genotipados con una tasa de error <5%. Esta evaluación sirvió además para conocer qué parámetros podrían indicar una imputación fiable de SNPs en otros genes para los futuros análisis de asociación. Los resultados de la re-secuenciación y de la evaluación del proceso de imputación han sido publicados como parte de un estudio de asociación de variantes del gen *IRAK3* con la susceptibilidad a la lesión pulmonar aguda (ALI), en un trabajo titulado “Interleukin-1 receptor-associated kinase 3 gene associates with susceptibility to acute lung injury” en la revista *American Journal of Respiratory Cell and Molecular Biology*.

Posteriormente, con los datos de secuenciación en individuos españoles se seleccionó un conjunto de 15 tSNPs para genotiparlos en las 607 muestras de casos GOA y en 1.271 controles, para llevar a cabo el estudio de replicación de la asociación de variantes de este gen con el asma en la población española. En este estudio encontramos asociación de varios SNPs con la susceptibilidad al asma atópico, detectando efectos mayores en individuos con edad de diagnóstico anterior a los 22 años, incluso tras ajustar por las estimas de ascendencia obtenidas con los marcadores de estratificación poblacional. Además, se encontró asociación entre haplotipos de SNPs de la región comprendida desde el intrón 2 al 6 de *IRAK3* con la susceptibilidad al asma y asma atópico. Finalmente, un meta-análisis que incluyó 3.384 muestras, mostró que los efectos fueron consistentes en direccionalidad entre muestras europeas, pero no al incluir muestras japonesas. Estos resultados replicaron, a nivel alélico y haplotípico, y extendieron los hallazgos previos, confirmando la importancia del gen *IRAK3* en la patogénesis del asma. Los resultados de este trabajo han sido publicados en la revista *The Journal of Allergy and Clinical Immunology* bajo el título “*IRAK3* gene variants associate with asthma in a replication study in the Spanish population”.

# Interleukin-1 Receptor–Associated Kinase 3 Gene Associates with Susceptibility to Acute Lung Injury

María Pino-Yanes<sup>1,2</sup>, Shwu-Fan Ma<sup>7</sup>, Xiaoguang Sun<sup>3</sup>, Paula Tejera<sup>4</sup>, Almudena Corrales<sup>1,2</sup>, Jesús Blanco<sup>1,5</sup>, Lina Pérez-Méndez<sup>1,2</sup>, Elena Espinosa<sup>6</sup>, Arturo Muriel<sup>5</sup>, Lluís Blanch<sup>1,8</sup>, Joe G. N. Garcia<sup>3</sup>, Jesús Villar<sup>1,9,10\*</sup>, and Carlos Flores<sup>1,2\*</sup>

<sup>1</sup>Centro de Investigación Biomédica en Red de Enfermedades Respiratorias de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain; <sup>2</sup>Research Unit, and <sup>3</sup>Department of Anesthesia, Hospital Nuestra Señora de Candelaria, Santa Cruz de Tenerife, Spain; <sup>4</sup>Section of Pulmonary and Critical Care, Department of Medicine, University of Chicago, Chicago, Illinois; <sup>5</sup>Institute for Personalized Respiratory Medicine, University of Illinois at Chicago, Chicago, Illinois; <sup>6</sup>Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts; <sup>7</sup>Intensive Care Unit, Hospital Universitario Río Hortega, Valladolid, Spain; <sup>8</sup>Critical Care Center, Corporació Sanitaria Parc Taulí, Sabadell, Barcelona, Spain; <sup>9</sup>Research Unit, Multidisciplinary Organ Dysfunction Evaluation Research Network, Hospital Universitario Dr. Negrin, Las Palmas de Gran Canaria, Spain; and <sup>10</sup>Keenan Research Center, St. Michael's Hospital, Toronto, Ontario, Canada

Sepsis is the most common cause of acute lung injury (ALI), leading to organ dysfunction and death in critically ill patients. Previous studies associated variants of interleukin-1 receptor–associated kinase genes (*IRAKs*) with differential immune responses to pathogens and with outcomes during sepsis, and revealed that increased expression levels of the *IRAK3* gene were correlated with poor outcomes during sepsis. Here we explored whether common variants of the *IRAK3* gene were associated with susceptibility to, and outcomes of, severe sepsis. After our discovery of polymorphism, we genotyped a subset of seven single-nucleotide polymorphisms (SNPs) in 336 population-based control subjects and 214 patients with severe sepsis, collected as part of a prospective study of adults from a Spanish network of intensive care units. Whereas *IRAK3* SNPs were not associated with susceptibility to severe sepsis, rs10506481 showed a significant association with the development of ALI among patients with sepsis ( $P = 0.007$ ). The association remained significant after adjusting for multiple comparisons, population stratification, and clinical variables (odds ratio, 2.50; 95% confidence interval, 1.15–5.47;  $P = 0.021$ ). By imputation, we revealed three additional SNPs independently associated with ALI ( $P < 0.01$ ). One of these (rs1732887) predicted the disruption of a putative human–mouse conserved transcription factor binding site, and demonstrated functional effects *in vitro* ( $P = 0.017$ ). Despite the need for replication in independent studies, our data suggest that common SNPs in the *IRAK3* gene may be determinants of sepsis-induced ALI.

**Keywords:** SNP; polymorphism; case-control; Toll-like receptor; severe sepsis

Sepsis is a devastating clinical syndrome characterized by systemic inflammation, and occurring in the setting of a severe infection (1). It constitutes the most common cause of acute lung injury (ALI) and acute respiratory distress syndrome

## CLINICAL RELEVANCE

Genes encoding key molecules of the Toll-like receptor (TLR) signaling pathway are known to abolish or alter the immune response, and were associated with disease susceptibility and outcomes. Genetic variants in interleukin-1 receptor–associated kinase 1 (*IRAK1*) and *IRAK4* were shown to modify the immune response to pathogens, and to increase the risk for severe complications during sepsis. We show that a functional variant of the *IRAK3* gene, another key element of this signaling pathway, is associated with the development of acute lung injury in patients with severe sepsis.

(ARDS). Severe forms manifest an estimated incidence of 50–100 cases per 100,000 people (2), with an associated mortality in intensive care units (ICUs) at above 40% (3). Clinical and experimental reports indicate that the host response to pathogens is affected by genetic factors (4). In particular, twin studies (5) and inbred-strain murine models (6) support a considerable genetic influence on immune responses to bacterial LPS and a susceptibility to infections. Thus, genetic variations disrupting or modifying innate immune responses may also underlie the diversity of clinical presentations in sepsis, the development of organ dysfunction, and the response to current medical treatments.

Genes encoding key molecules of the Toll-like receptor (TLR) signaling pathway, which is involved in innate immune recognition and cellular activation in response to microbial agents (7), constitute potential determinants of the development and outcomes of sepsis. Both rare and common variants of genes, including *LBP* (8), *CD14* (9), *TLR4* (10), *TLR2* (11–13), *TIRAP* (14), and others (15, 16), are known to abolish or alter the immune response, and are associated with disease susceptibility and outcomes. Genetic variants in interleukin-1 receptor–associated kinase 1 (*IRAK1*) and 4 (*IRAK4*) were shown to modify the immune response to pathogens, and to increase the risk for severe complications during sepsis (17, 18).

The human gene *IRAK3*, located in chromosome 12q14.3, encodes the IRAK-M (also known as IRAK-3) protein, as first described in cells of the monocytic lineage (19), and negatively regulates the Toll/NF- $\kappa$ B signaling pathway. IRAK-3-deficient mice demonstrated a disproportionate response to LPS stimulation, suggesting that IRAK-3 may be implicated in TLR ligand tolerance (20). Congruently, the expression of *IRAK3* in LPS-treated human leukocytes increased early during response (21) and peaked higher and more rapidly when tolerant

(Received in original form July 8, 2010 and in final form January 27, 2011)

This work was supported by grants from the Ministry of Science of Spain (SAF 2004–06833) and Fundación Canaria de Investigación y Salud (53/04), by a specific agreement between the Instituto de Salud Carlos III and the Gobierno de Canarias (EMER07/001) under the National Strategy for Science and Technology 2015 framework, and by grants HL91899, HL58064, and GM07019 from the National Heart Lung Blood Institute of the National Institutes of Health.

\* These authors contributed equally to this work.

Correspondence and requests for reprints should be addressed to Carlos Flores, Ph.D., Unidad de Investigación, Hospital Universitario Nuestra Señora de Candelaria, Carretera del Rosario s/n, 38010 Santa Cruz de Tenerife, Spain. E-mail: cflores@ull.es

This article has an online supplement, which is accessible from this issue's table of contents at [www.atsjournals.org](http://www.atsjournals.org)

Am J Respir Cell Mol Biol Vol 45, pp 740–745, 2011

Originally Published in Press as DOI: 10.1165/rcmb.2010-0292OC on February 4, 2011

Internet address: [www.atsjournals.org](http://www.atsjournals.org)

cells were restimulated with LPS, a situation demonstrated in cells from patients with sepsis (22). Similarly, the marked up-regulation of *IRAK-3* positively correlated with lung injury in septic animals, particularly in Type II epithelial cells and in the interstitial macrophages (23). These findings support a pivotal role of *IRAK-3* in the suppression of downstream signaling pathways required for the inflammatory process and its important role in regulating the generation of protective innate immune responses in the lung *in vivo* (24). In fact, a reduced capacity of peripheral leukocytes from patients with sepsis to produce cytokines, compared with that capacity in healthy control subjects (usually referred to as endotoxin tolerance), is a well-known phenomenon (25), and it may instigate subsequent damage during the course of sepsis. In line with this concept, increased expression levels of the *IRAK3* gene are associated with a decreased capacity to release proinflammatory cytokines from mononuclear cells in patients with sepsis after *ex vivo* stimulation with LPS, and are also associated with poor outcomes (26). Based on this evidence, we explored whether common variants of the *IRAK3* gene were associated with susceptibility to severe sepsis. We also evaluated whether *IRAK3* gene variants were associated with outcomes among patients with severe sepsis, including ALI and mortality.

**MATERIALS AND METHODS**

**Study Population**

This study was approved by the Ethics Committees of the Hospital Universitario Nuestra Señora de Candelaria and the Hospital Universitario Río Hortega. Written, informed consent was obtained from each subject or from appropriate surrogates. The case group of 214 patients, from a Spanish network of ICUs, fulfilled the international criteria for severe sepsis (27, 28), and were followed prospectively for the development of ALI as defined by the American-European Consensus Conference (29). For the purpose of this study, patients with ALI and ARDS were analyzed as a single group of ALI patients. The control group comprised 336 population-based subjects. Table 1. and the online supplement contain a detailed description of samples.

**Discovery of *IRAK3* Gene Polymorphism, Selection of Polymorphisms, and Genotyping**

DNA samples from 32 unrelated, healthy Spanish individuals were sequenced (the primers are listed in Table E1) to search for common variations in the *IRAK3* gene. Seven tagging single-nucleotide polymorphisms (tSNPs) from the *IRAK3* gene were selected and genotyped, using the iPLEX Gold assay in the MassARRAY system (Sequenom, San Diego, CA), by the Santiago de Compostela Node of the Spanish National Genotyping Center (CeGen, <http://www.cegen.org/>). SNaPshot Multiplex Kit reactions (Applied Biosystems, Foster City, CA) were used to verify the association of imputed SNPs. See the online supplement for further details.

**Luciferase Activity Assay**

PupaSuite (<http://pupasuite.bioinfo.cipf.es>) was used to predict the putative functional effects of all associated SNPs (30). The putative *IRAK3* promoter region (spanning -1 to -1698 from the transcription start site), containing the SNP rs1732887, was identified and synthesized with alleles rs1732887\_A (-1464A) and rs1732887\_G (-1464G) by GenScript (Piscataway, NJ), according to the sequence with National Center for Biotechnology Information accession number NC\_000012.10, and was subsequently cloned to the pGL3 basic vector. Human pulmonary artery endothelial cells (Cambrex, Walkersville, MD) were transfected with pGL3 fused constructs and with a plasmid with the renilla luciferase gene as transfection efficacy control. The activity of luciferase was measured 48 hours after transfection with Dual-Luciferase Assay Kits (Promega, Madison, WI), following the manufacturer's recommendations. Values were normalized to the activity of the -1464A reference allele construct. A Student *t* test was

used for comparisons, with two-tailed statistical significance set at *P* < 0.05. See the online supplement for details.

**Statistical Analysis**

Clinical and demographic data were analyzed using SPSS, version 15.0 (SPSS, Inc., Chicago, IL). Genotype associations were tested by means of the Cochran–Armitage test, assuming an additive model. These associations were adjusted for population stratification (31), using previously reported data from 20 independent polymorphisms (32). To control the Type I error rate in multiple comparisons, the false-discovery rate (FDR) was set at 5% (33). For comparison, this rate was also assessed by a permutation procedure (34). Logistic regression with backward elimination was used to adjust for covariates and to calculate effects as odds ratios (ORs) with 95% confidence intervals (CIs). Linear regression was used to test associations with quantitative measures of clinical severity. TUNA (<http://www.stat.uchicago.edu/~wen/tuna>) (35) was used to test the associations of imputed SNPs. EPIDAT 3.0 software (<http://dxsp.sergas.es>) was used to estimate the additive effects of risk alleles from independent SNPs. See the online supplement for details.

**RESULTS**

Sequencing 22.6 kb of the *IRAK3* gene per sample revealed a total of 59 polymorphisms, 58 of which were biallelic (Table E2). Three polymorphisms (one insertion/deletion, one SNP, and the multiallelic locus) did not have an assigned reference in the database for SNPs (dbSNPs), and were considered novel. Most of the polymorphisms revealed by sequencing (81.4%) showed minor allele frequencies (MAFs) of at least 5%. Two coding SNPs were detected, and one of them (rs1152888) predicted a common, nonsynonymous amino-acid change (Val147Ile). A subset of seven tSNPs (Table E2) allowed for the accurate inference of more than 80% of the common variants of the gene (see the online supplement for further details), and was used to genotype case–control samples.

**TABLE 1. DEMOGRAPHIC AND CLINICAL FEATURES OF POPULATION-BASED CONTROL SUBJECTS AND PATIENTS WITH SEVERE SEPSIS**

Variable	Patients (n = 214)	Control Subjects (n = 336)	P Value
Gender (male %)	60.0	48.5	0.012*
Median age (years) (P <sub>25</sub> –P <sub>75</sub> )	67 (54–75)	44 (34–56)	< 0.001†
Previous surgery (%)	72.9	38.8	< 0.001*
Hospitalized > 24 hours (%)	79.2	48.2	< 0.001*
Diabetes (%)	18.4	11.4	0.124*
Hypertension (%)	39.2	22.0	0.004*
Ischemic cardiac disease (%)	9.7	1.6	0.002*
Smoker (%)	30.3	26.1	0.679*
Source of infection (%)			
Pulmonary	35.3		
Extrapulmonary	64.7		
Pathogen (%)			
Gram-negative	23.0		
Gram-positive	15.7		
Fungi	2.1		
Mixed Gram-negative and Gram-positive	8.4		
Negative blood cultures	50.8		
Median APACHE II score (P <sub>25</sub> –P <sub>75</sub> )	24 (18–27)		
ALI/ARDS (%)	86.0		
ICU mortality (%)	45.5		

*Definition of abbreviations:* ALI, acute lung injury; APACHE II, Acute Physiology and Chronic Health Evaluation II; ARDS, acute respiratory distress syndrome; ICU, intensive care unit; P<sub>25</sub>, Percentile 25; P<sub>75</sub>, Percentile 75.

\*  $\chi^2$  test.

† Mann-Whitney U test.

TABLE 2. ASSOCIATION OF *IRAK3* GENE TAGGING SINGLE-NUCLEOTIDE POLYMORPHISMS WITH SEVERE SEPSIS AND ALI

rs/ss Number	Location (Effect)	Position*	MAF, Control Subjects	MAF, Patients with Severe Sepsis	P Value†	MAF, Non-ALI	MAF, ALI	P Value†
ss289570834	Intron 1	64871319	0.30	0.28	0.551	0.37	0.26	0.084
rs11465955	Intron 3	64889666	0.33	0.31	0.489	0.33	0.31	0.763
rs1152888	Exon 5 (Val147Ile)	64891495	0.10	0.08	0.346	0.10	0.07	0.400
rs1624395	Intron 6	64904483	0.40	0.40	0.870	0.45	0.39	0.370
rs1370128	Intron 6	64904905	0.42	0.41	0.652	0.45	0.40	0.440
rs1152912	Intron 8	64920175	0.49	0.50	0.645	0.48	0.50	0.795
rs10506481	3' flanking region	64930378	0.13	0.12	0.543	0.23	0.10	0.007

Definition of abbreviations: ALI, acute lung injury; *IRAK3* gene, interleukin-1 receptor-associated kinase 3 gene; MAF, minor allele frequency.

\* According to the National Center for Biotechnology Information build 36.

† Unadjusted P value.

Quality-control measures for the tSNP genotype data are described in Table E3. We did not observe significant deviations from Hardy-Weinberg equilibrium in cases or control subjects. None of the tSNPs (Table 2) or imputed SNPs (not shown) showed a significant association with susceptibility when comparing patients with severe sepsis and control subjects. Similarly, no significant association was evident for any SNP with ICU mortality among patients with severe sepsis (data not shown). However, the tSNP rs10506481 from the 3' flanking region showed a significant association with the development of ALI among patients with severe sepsis (OR for the T allele, 2.51; 95% CI, 1.29–4.86;  $P = 0.007$ ; Figure 1; genotype counts are listed in Table E4), and the FDR for this result was 4.5%. An alternative method, based on permutations (34), supported the significance of this result in the context of multiple comparisons (permuted  $P = 0.021$ ). Moreover, the association remained significant after applying a conservative adjustment for population stratification ( $P = 0.025$ ). Significance was further

maintained in a regression model adjusting for covariates, with significant differences between patients manifesting severe sepsis with and without ALI (Table 3), including gender, source of infection, and ICU mortality (OR for the T allele, 2.50; 95% CI, 1.15–5.47;  $P = 0.021$ ; Table E5). An *in silico* exploration using PupaSuite (30), which in addition to predicting non-synonymous changes and splicing effects for polymorphisms also scans for allelic differences in binding sites for transcription factors and microRNAs, did not predict any functional effects for rs10506481.

We next used TUNA software to impute and test the associations of 17 SNPs of the gene with ALI. This procedure also predicted three largely correlated SNPs ( $r^2 \geq 0.8$ , MAFs  $\geq 0.25$ ), located on the 5' region of the gene, that were significantly associated with ALI, all three with  $P < 0.01$ : rs1732888 (–1494 T/C), rs1732887 (–1464 A/G), and rs1732886 (c.133 + 550 A/G) (Figure 1 and Table E6). To demonstrate that these results were not spurious, we validated the association by

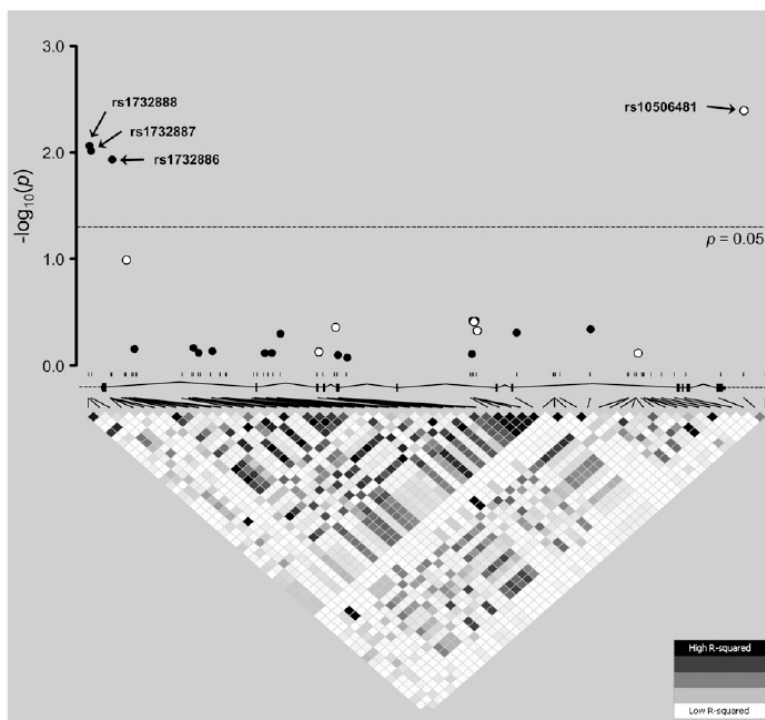


Figure 1. Association P values for interleukin-1 receptor-associated kinase 3 gene (*IRAK3*) single-nucleotide polymorphisms (SNPs) with acute lung injury among patients with severe sepsis. Open and solid circles correspond to P values (in  $-\log_{10}$  scale) resulting from tagging SNPs and imputed SNPs tested for association, respectively. Significantly associated SNPs (i.e., with values above the dotted line corresponding to an unadjusted significance at  $P = 0.05$ ) are indicated. For reference, a linkage disequilibrium (LD) plot of  $r^2$  values across the region, with the approximate location of all 58 biallelic SNPs discovered and a schematic representation of the *IRAK3* gene structure (vertical lines, exons; horizontal lines, introns), is shown below. This plot is based on resequencing data from 32 unrelated, healthy Spanish individuals. Each diamond in the LD plot represents a pairwise SNP comparison with its  $r^2$  value, schematically symbolized by a color gradient ranging from black ( $r^2 = 1$ , corresponding to complete LD) to gray ( $1 > r^2 > 0$ , corresponding to moderate LD) and white ( $r^2 = 0$ , corresponding to an absence of LD).

TABLE 3. RELEVANT DEMOGRAPHIC AND CLINICAL FEATURES OF PATIENTS WITH AND WITHOUT ALI

Variable	ALI (n = 184)	Non-ALI (n = 30)	P Value
Gender, male % (95% CI)	63.8 (56.2–70.5)	33.3 (17.9–52.9)	0.003*
Age, median years (P <sub>25</sub> –P <sub>75</sub> )	67 (54–76)	66.5 (48–74)	0.496†
Clinical features:			
APACHE II, median (P <sub>25</sub> –P <sub>75</sub> )	24 (18–28)	22 (16–24)	0.035†
Source of infection, % (95% CI)			
Pulmonary	40.5 (33.7–48.3)	37.0 (20.6–56.1)	
Extrapulmonary	59.5 (51.8–66.3)	63.0 (43.9–79.5)	<0.001*
ICU mortality, % (95% CI)	51.1 (43.6–58.5)	11.1 (2.6–27.7)	<0.001*

Definition of abbreviations: ALI, acute lung injury; APACHE II, Acute Physiology and Chronic Health Evaluation score II; CI, confidence interval; ICU, intensive care unit; P<sub>25</sub>, Percentile 25; P<sub>75</sub>, Percentile 75.

\*  $\chi^2$  test.

† Mann-Whitney U test.

genotyping two of these SNPs (rs1732886 and rs1732887) in all patients with severe sepsis (genotype counts are listed in Table E4). Borderline significance was obtained for both SNPs, using a regression model to adjust for covariates (OR for the rs1732887 G allele, 2.37; 95% CI, 0.99–5.65;  $P = 0.053$ ; OR for the rs1732886 G allele, 2.57; 95% CI, 1.02–6.10;  $P = 0.033$ ; see Table E5 for further details).

In an overall analysis across all 24 SNPs tested (seven genotyped tSNPs plus 17 imputed SNPs), the FDR was estimated at 7% for all four SNPs with  $P < 0.05$  of an association with ALI (i.e., rs1732888, rs1732887, rs1732886, and rs10506481). However, this value must be regarded with caution, because the underlying linkage disequilibrium (LD) was not taken into account when calculating the FDR.

In a *post hoc* analysis, we tested the association of rs10506481, rs1732887, and rs1732886 with quantitative measures of clinical severity during sepsis, including the number of organs failing in the ICU, the number of days in the ICU, and the Acute Physiology and Chronic Health Evaluation II (APACHE II) score. We did not find rs10506481 to be significantly associated with any of these measures of severity (Table E7). However, borderline associations were evident for rs1732887 and rs1732886 with the number of days in the ICU (Table E7), where the G allele for both SNPs was associated with more days in ICU (not shown). No significant association was evident for any of these SNPs with the number of organs failing in ICU or the APACHE II score.

The SNPs from the 5' and 3' regions of the *IRAK3* gene were poorly correlated (average pairwise  $r^2 < 0.01$ ; Figure 1), indicating the independence of both associations with ALI. We confirmed the independence by reasoning that if risk alleles from both regions were not correlated, the combination of their genotypes would have to show an increased risk for ALI in a dose-dependent manner. Under this assumption, a linear tendency of risk increase resulting from the addition of risk alleles from the two regions would have to be evident. Because of the strong LD among the three SNPs from the 5' flanking region, this trend was investigated only between rs1732887 and rs10506481, assuming a dominant model because of the small sample size (for this reason, we did not test for interaction effects). As expected, a significant trend for the accumulation of risks from alleles at rs1732887 and rs10506481 was evident ( $\chi^2 = 5.24$ ;  $P$  for trend = 0.022).

To examine the potential functionality of any of this 5' region's SNPs, we performed *in silico* analysis using a transcription factor database (TRANSFAC) exploration (30). This predicted that rs1732887 (–1464 A/G) would potentially interrupt a highly

conserved putative transcription factor binding site for the human–murine forkhead box P3 (FOXP3), a protein recently detected not only in regulatory T cells but also in activated T lymphocytes (36). We next used an *in vitro* luciferase activity assay to assess the functional effects of the –1464 A/G alleles of this SNP in gene expression, because it was the only associated SNP with predicted functionality. Constructs included positions –1 to –1698 from the transcription start site of the gene, and varied only at position –1464 with respect to the sequence NC\_000012.10. Luciferase activity, as a proxy for gene expression, was significantly higher for the –1464G construct compared with –1464A construct ( $P = 0.017$ ). The activity of both constructs was significantly higher than the activity of a promoter-less control ( $P \leq 0.01$ ; Figure 2). Because the risk alleles of the other two 5' region SNPs that were associated with ALI were not included in the construct (i.e., the DNA fragment did not contain the rs1732886 locus, and contained only the nonrisk T allele at rs1732888 locus), these results suggest that the –1464 A/G SNP is sufficient to alter the transcriptional control of *IRAK3* gene *in vitro*.

DISCUSSION

To our knowledge, this is the first report examining the role of *IRAK3* gene variants in susceptibility to severe sepsis and sepsis-associated outcomes. We largely characterized common variations across the nonrepetitive sequence of the gene in the population under study, and exhaustively evaluated the tSNP subset selected in the association study for their capacity to impute untyped SNPs. Our findings suggest that common variants of the *IRAK3* gene are associated with the development of ALI during severe sepsis. These findings were confirmed by ruling out effects attributable to the presence of major systematic errors during genotyping, the existence of population stratification, differences in clinical and demographic factors, and chance in the performance of multiple comparisons. Reinforcing these results, the SNPs rs1732887 and rs1732886 also showed borderline association with number of days in the ICU. Finally, we demonstrated that one of the associated variants from the 5' flanking region (rs1732887 and –1464 A/G) modified luciferase activity *in vitro*, suggesting that it may modify the expression of the *IRAK3* gene *in vivo*.

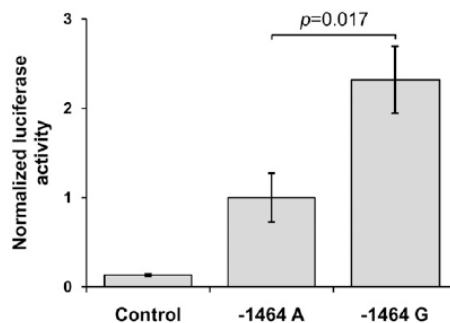


Figure 2. Normalized luciferase activity of the putative *IRAK3* promoter constructs carrying –1464A and –1464G alleles. Control sample is represented by a promoter-less construct, whereas –1464A and –1464G stand for the A and G allele constructs for rs1732887, respectively. Both show higher luciferase activity than the control sample ( $P \leq 0.01$ ). The data represent mean values  $\pm$  SEMs from four independent experiments, normalized to the –1464A construct.

Critical illness in adults is often followed by ALI. Epidemiologic studies indicate that the incidence and outcomes of ALI partly depend on the nature of the precipitating disease and individual susceptibility. Current evidence suggests that variations in genes involved in pulmonary or systemic inflammatory processes are associated with a high risk of developing ALI (16, 37). Therefore, genes/pathways involved in the development of other complex diseases are likely to be relevant in the onset, course, or severity of ALI. The *IRAK3* gene is critical for the early recognition of bacterial products by the Toll-like family of receptors (7, 20), and although it has been poorly studied for associations with susceptibility to other complex diseases and no published reports have associated *IRAK3* polymorphisms with sepsis or ALI (<http://geneticassociationdb.nih.gov>; accessed May 2010), the *IRAK3* gene was recently studied in association with susceptibility to chronic inflammatory conditions such as asthma (38, 39) and inflammatory bowel disease (40). In fact, strong evidence for the role of *IRAK3* variants in susceptibility to asthma was found by Balaci and colleagues (39), who positionally cloned the gene in asthma after using successive linkage and association analyses of samples from families and unrelated case-control subjects from Southern European populations. Notably, rs11465955 and rs1370128, which tagged rs1732887 in our sample, were associated with asthma in samples from Sardinia and Italy (39). This finding may be indicative of a gene region harboring variants with pleiotropic effects in both asthma and ALI, and deserving of further study.

The region where -1464 A/G is located was predicted to be a highly conserved putative human-murine FOXP3 transcription factor binding site that would be disrupted in the presence of the G allele. Because FOXP3 was demonstrated to be a transcriptional repressor (41), patients with severe sepsis and carrying the -1464G allele may manifest increased *IRAK3* gene transcription and higher levels of mRNA expression. This would imply a greater negative regulatory capacity, favoring an imbalance of the inflammatory and anti-inflammatory processes that would enhance the immunosuppression occurring as part of the septic process (25, 42). This hypothesis is not only congruent with our *in vitro* observations of significantly higher luciferase activity for constructs with -1464G (which is the allele at-risk for ALI), but also with the increased levels of the *IRAK3* gene in monocytes from patients with sepsis (22), with the decreased capacity of monocytes to release proinflammatory cytokines after *ex vivo* stimulation with LPS or *Burkholderia* species, and with the positive correlation of *IRAK3* gene expression with poor outcomes in sepsis (26). This evidence testifies to the pivotal role of IRAK-3 in the generation of protective, innate immune responses in the lung (23, 24). However, experimental validation of the binding of FOXP3 to the -1464 A/G region and a demonstration of the functionality of -1464 A/G in its native genomic context (e.g., by gene expression analysis in relevant cell lines or in cells isolated from patients) are necessary.

Although we followed current guidelines for genetic association studies (43) and used a number of approaches to reduce the chance of reporting a false-positive association, such a possibility cannot be excluded. Two major limitations of the study include its lack of a replication sample and its reduced sample size, particularly for detecting associations with secondary outcomes, given the modest effect sizes of common variants (44). In addition, we used population-based subjects as controls to test for the association of *IRAK3* variants with susceptibility to severe sepsis. Control subjects from the general population are not affected by selection bias (i.e., the possibility that health concerns may have influenced participation), and nesting association studies in cohorts of the general population has been productive, leading to many firm genetic associations (45, 46).

However, the use of these cohorts entails some limitations. Most importantly in this study, they did not allow us to control for important environmental factors that might explain the absence of association of *IRAK3* variants with susceptibility to severe sepsis. Moreover, this study was designed to explore the effects of common variants in disease, and not to analyze the effects of rare variants, for which the study was underpowered. Thus, other variants of the gene, neither studied for functionality nor revealed by our resequencing, may also contribute to ALI. Furthermore, although rs10506481 showed a reduced LD with nearby regions, rs1732887 showed moderate-to-strong LD, extending as far as 125 kb upstream from *IRAK3*, where the *LLPH* and *TMBIM4* genes lie (not shown). Thus, whether the observed association of the *IRAK3* gene with ALI is attributable to its biological function or to its LD with nearby genes requires further study.

To the best of our knowledge, we provide the first evidence suggesting that common genetic variants of *IRAK3* may be novel determinants of susceptibility to sepsis-induced ALI. Although our results were adjusted for multiple comparisons, relevant covariates, and population stratification, and although we used functional assays to demonstrate *in vitro* effects of SNPs, we underscore the limitations of our sample size and the need for replication in independent studies. The early tailoring of treatments has been beneficial for outcomes of sepsis in animal models (47). Therefore, an early classification of patients, based on their genetics to predict immunomodulatory status, may offer benefits in more specific therapeutic interventions.

**Author Disclosure:** L.B. received lecture fees from Hamilton, a sponsored grant from Instituto Carlos III, and a patent from the Corporació Sanitària Parc Taulí for a monitoring platform, and he owns stock in Better Care SL. J.V. received lecture fees from GlaxoSmithKline and a sponsored grant from MAQUET (Solna, Sweden). M.P.-Y., S.-F.M., X.S., P.T., A.C., J.B., L.P.-M., E.E., A.M., J.G.N.G., and C.F. do not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript.

**Acknowledgments:** The authors thank Dan Nicolae and Xiaoquan Wen for their advice using TUNA, Santiago Basaldua for help with the statistical analysis, Tobias Felipe for Fortran scripting for handling TUNA outputs, and two anonymous reviewers for their helpful suggestions to improve the manuscript.

## References

- Cohen J. The immunopathogenesis of sepsis. *Nature* 2002;420:885-891.
- Moss M. Epidemiology of sepsis: race, sex, and chronic alcohol abuse. *Clin Infect Dis* 2005;41:S490-S497.
- Blanco J, Muriel-Bombin A, Sagredo V, Taboada F, Gandia F, Tamayo L, Collado J, Garcia-Labattut A, Carriedo D, Villedor M, et al. Grupo de Estudios y Análisis en Cuidados Intensivos: incidence, organ dysfunction and mortality in severe sepsis: a Spanish multi-centre study. *Crit Care* 2008;12:R158.
- Frodsham AJ, Hill AV. Genetics of infectious diseases. *Hum Mol Genet* 2004;13:R187-R194.
- de Craen AJ, Posthuma D, Remarque EJ, van den Biggelaar AH, Westendorp RG, Boomsma DI. Heritability estimates of innate immunity: an extended twin study. *Genes Immun* 2005;6:167-170.
- Aziz RK, Kansal R, Abdeltawab NF, Rowe SL, Su Y, Carrigan D, Nooh MM, Attia RR, Brannen C, Gardner LA, et al. Susceptibility to severe streptococcal sepsis: use of a large set of isogenic mouse lines to study genetic and environmental factors. *Genes Immun* 2007;8:404-415.
- Trinchieri G, Sher A. Cooperation of Toll-like receptor signals in innate immune defence. *Nat Rev Immunol* 2007;7:179-190.
- Flores C, Perez-Mendez L, Maca-Meyer N, Muriel A, Espinosa E, Blanco J, Sanguesa R, Muros M, Garcia JG, Villar J. A common haplotype of the *LBP* gene predisposes to severe sepsis. *Crit Care Med* 2009;37:2759-2766.
- Sutherland AM, Walley KR, Russell JA. Polymorphisms in *CD14*, mannose-binding lectin, and Toll-like receptor-2 are associated with increased prevalence of infection in critically ill adults. *Crit Care Med* 2005;33:638-644.



10. Arbour NC, Lorenz E, Schutte BC, Zabner J, Kline JN, Jones M, Frees K, Watt JL, Schwartz DA. *TLR4* mutations are associated with endotoxin hyporesponsiveness in humans. *Nat Genet* 2000;25:187–191.
11. Bochud PY, Hawn TR, Aderem A. Cutting edge: a Toll-like receptor 2 polymorphism that is associated with lepromatous leprosy is unable to mediate mycobacterial signaling. *J Immunol* 2003;170:3451–3454.
12. Kang TJ, Lee SB, Chae GT. A polymorphism in the Toll-like receptor 2 is associated with IL-12 production from monocyte in lepromatous leprosy. *Cytokine* 2002;20:56–62.
13. Lorenz E, Mira JP, Cornish KL, Arbour NC, Schwartz DA. A novel polymorphism in the Toll-like receptor 2 gene and its potential association with staphylococcal infection. *Infect Immun* 2000;68:6398–6401.
14. Khor CC, Chapman SJ, Vannberg FO, Dunne A, Murphy C, Ling EY, Frodsham AJ, Walley AJ, Kyrieles O, Khan A, et al. A Mal functional variant is associated with protection against invasive pneumococcal disease, bacteremia, malaria and tuberculosis. *Nat Genet* 2007;39:523–528.
15. Clark MF, Baudouin SV. A systematic review of the quality of genetic association studies in human sepsis. *Intensive Care Med* 2006;32:1706–1712.
16. Flores C, Pino-Yanes MM, Villar J. A quality assessment of genetic association studies supporting susceptibility and outcome in acute lung injury. *Crit Care* 2008;12:R130.
17. Picard C, Puel A, Bonnet M, Ku CL, Bustamante J, Yang K, Soudais C, Dupuis S, Feinberg J, Fieschi C, et al. Pyogenic bacterial infections in humans with IRAK-4 deficiency. *Science* 2003;299:2076–2079.
18. Arcaroli J, Silva E, Maloney JP, He Q, Svetkauskaite D, Murphy JR, Abraham E. Variant IRAK-1 haplotype is associated with increased nuclear factor- $\kappa$ B activation and worse outcomes in sepsis. *Am J Respir Crit Care Med* 2006;173:1335–1341.
19. Wesche H, Gao X, Li X, Kirschning CJ, Stark GR, Cao Z. IRAK-M is a novel member of the interleukin-1 receptor-associated kinase (IRAK) family. *J Biol Chem* 1999;274:19403–19410.
20. Kobayashi K, Hernandez LD, Galan JE, Janeway CA Jr, Medzhitov R, Flavell RA. IRAK-M is a negative regulator of Toll-like receptor signaling. *Cell* 2002;110:191–202.
21. Calvano SE, Xiao W, Richards DR, Felciano RM, Baker HV, Cho RJ, Chen RO, Brownstein BH, Cobb JP, Tschoeke SK, et al. Inflammatory and host response to injury: large scale collaborative research program: a network-based analysis of systemic inflammation in humans. *Nature* 2005;437:1032–1037.
22. Escoll P, del Fresno C, Garcia L, Valles G, Lendinez MJ, Arnalich F, Lopez-Collazo E. Rapid up-regulation of IRAK-M expression following a second endotoxin challenge in human monocytes and in monocytes isolated from septic patients. *Biochem Biophys Res Commun* 2003;311:465–472.
23. Villar J, Cabrera N, Casula M, Flores C, Valladares F, Muros M, Slutsky AS, Kacmarek RM. Mechanical ventilation modulates Toll-like receptor signaling pathway in a sepsis-induced lung injury model. *Intensive Care Med* 2010;36:1409–1457.
24. Deng JC, Cheng G, Newstead MW, Zeng X, Kobayashi K, Flavell RA, Standiford TJ. Sepsis-induced suppression of lung innate immunity is mediated by IRAK-M. *J Clin Invest* 2006;116:2532–2542.
25. McCall CE, Grosso-Wilmoth LM, LaRue K, Guzman RN, Cousart SL. Tolerance to endotoxin-induced expression of the interleukin-1 beta gene in blood neutrophils of humans with the sepsis syndrome. *J Clin Invest* 1993;91:853–861.
26. Wiersinga WJ, van't Veer C, van den Pangaart PS, Dondorp AM, Day NP, Peacock SJ, van der Poll T. Immunosuppression associated with interleukin-1 $\alpha$ -associated-kinase-M upregulation predicts mortality in Gram-negative sepsis (melioidosis). *Crit Care Med* 2009;37:569–576.
27. Villar J, Perez-Mendez L, Flores C, Maca-Meyer N, Espinosa E, Muriel A, Sanguesa R, Blanco J, Muros M, Kacmarek RMA. *CXCL2* polymorphism is associated with better outcomes in patients with severe sepsis. *Crit Care Med* 2007;35:2292–2297.
28. Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, Cohen J, Opal SM, Vincent JL, Ramsay G. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Crit Care Med* 2003;31:1250–1256.
29. Bernard GR, Artigas A, Brigham KL, Carlet J, Falke K, Hudson L, Lamy M, Legall JR, Morris A, Spragg R. The American-European Consensus Conference on ARDS: definitions, mechanisms, relevant outcomes, and clinical trial coordination. *Am J Respir Crit Care Med* 1994;149:818–824.
30. Conde L, Vaquerizas JM, Dopazo H, Arbiza L, Reumers J, Rousseau F, Schymkowitz J, Dopazo J. PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes. *Nucleic Acids Res* 2006;34:W621–625.
31. Wang K. Testing for genetic association in the presence of population stratification in genome-wide association studies. *Genet Epidemiol* 2009;33:637–645.
32. Flores C, Maca-Meyer N, Perez-Mendez L, Sanguesa R, Espinosa E, Muriel A, Blanco J, Villar JA. *CXCL2* tandem repeat promoter polymorphism is associated with susceptibility to severe sepsis in the Spanish population. *Genes Immun* 2006;7:141–149.
33. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc, B* 1995;57:289–300.
34. Westfall PH, Young SS. Resampling-based multiple testing: examples and methods for *P*-value adjustment. New York: John Wiley & Sons; 1993.
35. Wen X, Nicolae DL. Association studies for untyped markers with TUNA. *Bioinformatics* 2008;24:435–437.
36. Cassani B, Poliani PL, Moratto D, Sobacchi C, Marrella V, Imperatori L, Vairo D, Plebani A, Giliati S, Vezzoni P, et al. Defect of regulatory T cells in patients with Omenn syndrome. *J Allergy Clin Immunol* 2010;125:209–216.
37. Villar J, Maca-Meyer N, Perez-Mendez L, Flores C. Bench-to bedside review: understanding genetic predisposition to sepsis. *Crit Care* 2004;8:180–189.
38. Nakashima K, Hirota T, Obara K, Shimizu M, Jodo A, Kameda M, Doi S, Fujita K, Shirakawa T, Enomoto T, et al. An association study of asthma and related phenotypes with polymorphisms in negative regulator molecules of the TLR signaling pathway. *J Hum Genet* 2006;51:284–291.
39. Balaci L, Spada MC, Olla N, Sole G, Loddo L, Anedda F, Naitza S, Zuncheddu MA, Maschio A, Altea D, et al. IRAK-M is involved in the pathogenesis of early-onset persistent asthma. *Am J Hum Genet* 2007;80:1103–1114.
40. Weersma RK, Oostenbrug LE, Nolte IM, Van Der Steege G, Oosterom E, Van Dullemen HM, Kleibeuker JH, Dijkstra G. Association of interleukin-1 receptor-associated kinase M (IRAK-M) and inflammatory bowel diseases. *Scand J Gastroenterol* 2007;42:827–833.
41. Schubert LA, Jeffery E, Zhang Y, Ramsdell F, Ziegler SF. Scurfin (FOXP3) acts as a repressor of transcription and regulates T cell activation. *J Biol Chem* 2001;276:37672–37679.
42. Lyn-Kew K, Standiford TJ. Immunosuppression in sepsis. *Curr Pharm Des* 2008;14:1870–1881.
43. Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, von Elm E, Khoury MJ, Cohen B, Davey-Smith G, Grimshaw J, et al. Strengthening the reporting of genetic association studies (STREGA): an extension of the STROBE statement. *Hum Genet* 2009;125:131–151.
44. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science* 2008;322:881–888.
45. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661–678.
46. Moffatt MF, Kabisch M, Liang L, Dixon AL, Strachan D, Heath S, Depner M, von Berg A, Bufe A, Rietschel E, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 2007;448:470–473.
47. Osuchowski MF, Connett J, Welch K, Granger J, Remick DG. Stratification is the key: inflammatory biomarkers accurately direct immunomodulatory therapy in experimental sepsis. *Crit Care Med* 2009;37:1567–1573.

## Online Data Supplement

### Interleukin-1 Receptor-associated Kinase 3 Gene Associates with Susceptibility to Acute Lung Injury

María Pino-Yanes, Shwu-Fan Ma, Xiaoguang Sun, Paula Tejera, Almudena Corrales, Jesús Blanco, Lina Pérez-Méndez, Elena Espinosa, Arturo Muriel, Lluís Blanch, Joe G. N. Garcia, Jesús Villar, Carlos Flores

#### METHODS

##### Study design

Sample characteristics have been reported elsewhere (E1) consisting on unrelated cases and controls of Spanish descent. Cases fulfilled the international criteria for severe sepsis (E2) and basic demographic data, severity of illness scores and clinical information until discharge from the ICU, including source of infection and development of organ dysfunction, were recorded. All patients were followed prospectively for the development of ALI, as defined by the American-European Consensus Conference (E3). The median APACHE II score was 24 (IQR=18-27) and 184 patients (86%) developed ALI/ARDS. The overall ICU mortality was 45.5%. In agreement with previous studies (E4), no pathogens were identified in blood cultures as the causative microorganism for sepsis albeit having an identified site of infection for 50.8% of patients. Controls were randomly drawn from a population-based study of about 7000 samples from unrelated individuals as a representation of the general Spanish population (E5). A health survey consisting on a 2-hour personal interview by trained personnel aiming to characterize occupational and environmental risk factors for common acute and chronic health conditions (supported by personal or familiar medical records) in the population was obtained from all controls. The study sample size was based on *a priori* power calculation for the primary outcome, indicating 80% power to detect a risk of 1.8 at  $p=0.05$

significance level assuming an allele frequency of 30%.

##### *IRAK3* gene polymorphism identification

To study the variation in the *IRAK3* gene, 22.6 kilobases (kb) of non-repetitive gene regions in DNA samples were sequenced in 32 unrelated healthy Spanish individuals. The genomic sequence from the reference assembly (NC\_000012.10) was filtered for interspersed repeats and low complexity DNA sequences using RepeatMasker (<http://www.repeatmasker.org>). Primers for DNA amplification of non-repetitive regions were designed using Primer3 ([http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)). Primer pairs used are listed in Table E1. Sequencing was performed by both strands with BigDye® v3.1 Terminator Cycle Sequencing kit (Applied Biosystems, Foster City, CA, USA) under the manufacturer's recommendations using the amplification primers. Sequencing products were ethanol precipitated and run on an ABI PRISM 310 Genetic Analyzer (Applied Biosystems). Polymorphisms were identified by manual examination of traces. The sequence variants were confirmed by two independent authors and reported following established guidelines (E6).

##### tSNPs selection

In order to efficiently and accurately study common *IRAK3* gene variation in the population, tagging single nucleotide polymorphisms (tSNPs) were

selected using TagIT 3.03 using a SNP-dropping-with-re-sampling method (E7) and tested for their ability to impute untyped variants. Given a genomic region, the allelic frequencies of untyped variants can be imputed using a set of tSNPs with high information content of the variation in the particular region, by means of linear combinations of the tSNPs haplotype frequencies. TUNA software was used to derive such estimates (E8). TUNA also allows estimating the  $M_D$  parameter, representing a multilocus measure of linkage disequilibrium (LD) (E9). This value indicates the amount of information that a set of tSNPs contains for a given untyped SNP, for which a threshold  $\geq 0.7$  has been suggested for accurate predictions to avoid power loss to detect disease associations at untyped SNPs (E8).

In order to infer the association only at those untyped SNPs for which imputation was accurate, we evaluated the accuracy of the selected 7-tSNPs set (average haplotype  $r^2=0.55$ ) in predicting allele frequencies for the remaining 51 untyped biallelic SNPs in the re-sequenced sample, obtaining an error rate from the comparison of observed vs. estimated minor allele frequencies (MAF) from:

$$Error(\%) = \frac{|MAF_{observed} - MAF_{estimated}|}{MAF_{observed}} \times 100$$

To compare the 7-tSNPs set utilized for the study with an optimal tagging alternative (i.e. with an average haplotype  $r^2=1.0$ ), we also selected an additional 15-tSNPs set using the same methodology described. For both sets, the  $M_D$  measures, observed and estimated MAFs, and error values for all untyped SNPs were derived. A high correlation ( $R^2>0.99$ ) between observed and estimated MAFs for the two tSNPs sets was found. Average errors for MAF estimates were similar between both tSNPs sets (6.7% for 15-tSNPs and 7.5% for 7-tSNPs). However, average errors were reduced by more than 20% among the untyped SNPs with  $M_D \geq 0.7$  (5.3% for 15-

tSNPs and 4.6% for 7-tSNPs) and even more (by more than 60%) among the untyped SNPs with  $MAF \geq 10\%$  (1.7% for 15-tSNPs and 2.7% for 7-tSNPs). Thus, using a threshold of  $MAFs \geq 10\%$  rather than the  $M_D \geq 0.7$ , reduced MAF error estimates at least by two-fold. This is supported by recent observations across diverse populations indicating that genotype imputation errors produced the greatest power reduction for SNPs with  $MAFs \leq 10\%$  (E10). To reduce the MAF error rate as low as possible, we considered the overlapping SNPs of both thresholds for the association study (with  $MAF \geq 10\%$  and  $M_D \geq 0.7$ ). MAF error estimates derived from these two tSNPs sets were minimal (0.7% for 15-tSNPs and 2.6% for 7-tSNPs), still allowing for the accurate imputation of more than 80% of SNPs with  $MAF \geq 10\%$  (27 SNPs for 15-tSNPs and 24 SNPs for 7-tSNPs). Worth noting, when a set of 6 tSNPs was selected to provide an average haplotype  $r^2=1.0$  for the Utah residents with ancestry from Northern and Western Europe (CEU) from HapMap Phase II (E11) and was tested against our re-sequencing sample, MAF error estimates were five-fold larger on average irrespective of MAF or  $M_D$  values (not shown).

Taken together, these data support that the utilized set of 7 tSNPs allowed accurate imputing of a similar number of untyped SNPs in the *IRAK3* gene as a set of 15 tSNPs with maximum theoretical coverage (haplotype  $r^2=1.0$ ). In addition, we adopted a conservative approach by imputing and studying the association of those 17 untyped SNPs with  $MAF \geq 10\%$  and  $M_D \geq 0.7$  to reduce errors to the minimum.

### Genotyping quality control

Genotyping was blind to case and control status. Approximately 9% of the samples were genotyped in duplicate to monitor genotyping quality. Samples used for polymorphism identification were also genotyped with this platform to test alle-

calling reliability. Genotypes were assigned using all data simultaneously. Among all 550 samples analyzed in this study, completion rate and the average confidence of automatic genotype calls were over 90% for all tSNPs assessed. Among duplicated samples, no discrepancies were found with the original genotypes for any of the SNPs, giving an estimated overall discordance rate of 0.0%. Only 3 alleles out of the 268 genotypes obtained by both sequencing and iPLEX Gold assays were discordant (1.11%; 95% CI=0.23-3.23%). Taken together, these data are indicative of accurate genotyping methods.

### Statistical analysis

Departures from Hardy-Weinberg equilibrium (HWE) were assessed using the Pearson's goodness-of-fit  $\chi^2$ -test.

Logistic regression model was constructed by backwards elimination, starting with a model including all the variables showing significant differences between septic patients with ALI and septic patients without ALI from Table 3, and maintaining in the model the variables showing  $p$ -values  $\leq 0.1$ . The final model excluded APACHE II, and considered as covariates the SNP of interest, gender (male), source of infection (extra-pulmonary) and ICU mortality.

To adjust for the multiple comparisons, a permutation procedure was also performed. The permuted  $p$ -value was calculated by testing the association of 7 tSNPs over permuted samples, obtained from 1000 swappings of the ALI and non-ALI labels, and saving best chi-square value overall loci for each single sample. This resulted in an empirical distribution of association values considering the multiple tests done.  $P$ -values adjusted for the multiple tests being performed were then obtained by comparing the observed chi-square of each tSNP in this distribution (E12). Because the permutation scheme preserves the LD structure, this provides a less stringent correction for multiple

testing in comparison to the commonly used Bonferroni correction, which assumes that all tests are independent.

Information contained in the tSNPs set was used to impute untyped SNPs in ALI and non-ALI patients and to infer the association of these SNPs by means of TUNA ver. 1.1 (E8). For this aim, PHASE 2.1 (E13) was used to estimate the haplotypes from sequencing data that were employed by TUNA software as a reference for the underlying LD across all gene variants in the population. To maintain an extra level of confidence in the imputation of untyped variation, only common well-tagged variants (i.e. with multilocus LD values  $\geq 0.7$ ) (E8) with  $MAF \geq 10\%$  were considered for this analysis. The justification for this decision is explained above.

The patterns of LD, in terms of  $r^2$  values, were explored using Haploview 3.32 (E14) and SNAP (E15). The latter was applied to plot LD across the 250 kb region centered on *IRAK3* gene for the HapMap CEU sample.

### Luciferase activity assay

PupaSuite was used to predict the putative functionality of associated SNPs, including effects on microRNA binding sites, transcription factor binding sites, splicing sites, exonic splicing enhancers, and exonic splicing silencers, using the default parameters of the on-line tool (E16). The putative *IRAK3* promoter region was identified by Cap-Analysis of Gene Expression (CAGE) and Genomatix database (<http://www.genomitix.de>). Human pulmonary artery endothelial cells (HPAEC) (Cambrex, Walkersville, MD, USA), which are known to express *IRAK3* (E17), were cultured as previously described (E18) and transfected with FuGENE 6 Transfection Reagent (Roche Applied Science, Indianapolis, IN) in Opti-MEM medium for 2 h. A plasmid with renilla luciferase gene (phRL-TK) was co-transfected with pGL3 fused constructs as transfection efficacy control. Cells were

harvested 48 h post-transfection and lysed in Passive Lysis Buffer. Luciferase activity was measured by Dual-Luciferase Assay Kits (Promega). The relative construct activities were expressed as the ratio of firefly luciferase in pGL3 to renilla luciferase in pRL-TK (RLU) following the Dual-Glo luciferase reporter assay protocols recommended by the manufacturer. Four independent

transfections and duplicate luciferase assays were performed for each condition, along with pGL3-basic vector used as a control. Values were normalized to the activity of the construct containing the reference allele rs1732887\_A (-1464A). A Student *t*-test was used for comparisons with a two-tailed statistical significance set at  $p < 0.05$ .

## REFERENCES

- E1. Villar J, Perez-Mendez L, Flores C, Maca-Meyer N, Espinosa E, Muriel A, Sanguesa R, Blanco J, Muros M, Kacmarek RM. A *CXCL2* polymorphism is associated with better outcomes in patients with severe sepsis. *Crit Care Med* 2007;35:2292-2297.
- E2. Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, Cohen J, Opal SM, Vincent JL, Ramsay G. 2001 SCCM/ESICM/ACCP/ATS/SIS international sepsis definitions conference. *Crit Care Med* 2003;31:1250-1256.
- E3. Bernard GR, Artigas A, Brigham KL, Carlet J, Falke K, Hudson L, Lamy M, Legall JR, Morris A, Spragg R. The American-European consensus conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination. *Am J Respir Crit Care Med* 1994;149:818-824.
- E4. Baudouin SV, Saunders D, Tiangyou W, Elson JL, Poynter J, Pyle A, Keers S, Turnbull DM, Howell N, Chinnery PF. Mitochondrial DNA and survival after sepsis: A prospective study. *Lancet* 2005;366:2118-2121.
- E5. Cabrera de Leon A, Rodriguez Perez Mdel C, Almeida Gonzalez D, Dominguez Coello S, Aguirre Jaime A, Brito Diaz B, Gonzalez Hernandez A, Perez Mendez LI. [presentation of the "CDC de Canarias" Cohort: Objectives, design and preliminary results]. *Rev Esp Salud Publica* 2008;82:519-534.
- E6. Antonarakis SE. Recommendations for a nomenclature system for human gene mutations. Nomenclature working group. *Hum Mutat* 1998;11:1-3.
- E7. Ahmadi KR, Weale ME, Xue ZY, Soranzo N, Yarnall DP, Briley JD, Maruyama Y, Kobayashi M, Wood NW, Spurr NK, Burns DK, Roses AD, Saunders AM, Goldstein DB. A single-nucleotide polymorphism tagging set for human drug metabolism and transport. *Nat Genet* 2005;37:84-89.
- E8. Wen X, Nicolae DL. Association studies for untyped markers with tuna. *Bioinformatics* 2008;24:435-437.
- E9. Nicolae DL. Quantifying the amount of missing information in genetic association studies. *Genet Epidemiol* 2006;30:703-717.
- E10. Huang L, Wang C, Rosenberg NA. The relationship between imputation error and statistical power in genetic association studies in diverse populations. *Am J Hum Genet* 2009;85:692-698.
- E11. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Wayne MM,

Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallée C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;449:851-861.

E12. Westfall PH, Young SS. Resampling-based multiple testing: examples and methods for p-value adjustment. NY: John Wiley & Sons; 1993.

E13. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001;68:978-989.

E14. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263-265.

E15. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008;24:2938-2939.

E16. Conde L, Vaquerizas JM, Dopazo H, Arbiza L, Reumers J, Rousseau F, Schymkowitz J, Dopazo J. PupaSuite: Finding functional single nucleotide polymorphisms for large-scale genotyping purposes. *Nucleic Acids Res* 2006;34:W621-625.

E17. Karoly ED, Li Z, Dailey LA, Hyseni X, Huang YC. Up-regulation of tissue factor in human pulmonary artery endothelial cells after ultrafine particle exposure. *Environ Health Perspect* 2007;115:535-540.

E18. Garcia JG, Liu F, Verin AD, Birukova A, Dechert MA, Gerthoffer WT, Bamberg JR, English D. Sphingosine 1-phosphate promotes endothelial cell barrier integrity by EDG-dependent cytoskeletal rearrangement. *J Clin Invest* 2001;108:689-701.

## RESULTADOS

### SUPPLEMENTARY RESULTS

**Table E1.** Primers used for polymorphism discovery in *IRAK3* gene

Fragment	Forward primer	Reverse primer	Size*
1	ATCGAAATTCTTGCCACAC	GAAAAGACACCAAATCAGCGTA	538
2	AGTTCGACTCTGCTTGTC	CTGCATCCAACCCAGGAA	531
3	CAGCGTGAAAGCATCTTCTG	GGCCTTCTGCTAGTCTGGT	562
4	GCCTGCTGGAAGTGTGAGT	CCCCATTACCAACATTAAAC	530
5	AGGACCACCCAAATGAGAA	ATGACCACCTGCACAATGAA	566
6	TTCATTGTGCAGGTGGTCAT	AAGAAGCTGGACCCTCAACC	550
7	CCAGTGTGTGCAGGATAAGG	GGTGTGAGCTGCTCCAAATA	550
8	CCTTGGTGTGAGGGTGAAT	TGCAGAAATCACGTCTCCAT	543
9	CTGAAAGGGGACACGTAAGG	TCCTGTGACTTGGCTGAGAA	533
10	GCTACTATTGCCACCCGATG	CGTTCCTGCTACCCTTTGAA	522
11	CCTAAACTTTCCCTTTGCTGAA	GCATGAAGACAGGCCTCTAAC	558
12	TGGCATTGGATAGTCTTTGC	TGCCTCTGTGTCTCTCCAG	539
13	AACGAAAAATTGGGCATGG	CCAATGGGAGACTTCAAAGG	551
14	CATTAGCAGAAGGCAGGTGA	CACAAGAAAAATGTGAAGCTGGA	519
15	AGGGACTTCCAGCTCACA	TGGAACAAACAGAACAGGTGA	563
16	TGGGTCTGTGTCATGCCTTA	TTTCACTTGCTCAGTCAACTTCA	555
17	TTCCCTTTGTCTTTTGCACAT	TTGAAACATCAACTCAATCAATTACA	528
18	GCCACCGTGGACAGATTTAT	GGGCTGTATGGGCACTAAAA	531
19	CGTAATGTTGCATGGGAATG	TGCAAAACACCAACCAAAAA	515
20	TTGGTTGGTGTTCATGT	TTAGCCAATGGAGCCTTAACA	591
21	TTAAAATAACTTCAATTTGATTGACAC	TTGGCTATGTCCTTCAGAACC	553
22	AAGGCTTTGGTTCTTCAACA	TAGAGAAAGGCCAGGAACC	568
23	CTCTGTGGAATGGTGGAAC	CTGGGGATGGAAAGCATTTA	548
24	TAGTGCTAGCCTGCCCTGAT	TTGCTATCTTTTATGTGCTTCTGTG	518
25	GGGGCTAATGGATTGTTGAA	GGAAAGGAGGAAATGGTGAAC	558
26	GGCTATTCTGGGGTACCAT	CCTCGAGAAAGGAGGGAAAA	523
27	CCTGGCTTTTGGCTTAACCT	CTGGTTGGTTCTCCTGCAAC	579
28	ATCCCCATCCCACCTTACT	TTGATGCCAACTCACTGGAG	531
29	TCACCCACCTTGCTATCTA	TGGGATACTCAAAGTATGATTTCT	600
30	TGGGGACAGAGGGTCACTAA	GGCTAAAGATGGCTCACAGC	559
31	AAGAGGAGGCCAGTTCTTCC	ACAGGCAGGGACACATTGAT	389
32	CAGGCATGTGTCAGGGTATG	TCCATTGTTCACTCCTCCAA	648
33	CCGCGTCCTCCGATATTATT	ATTCACACTTTGGGCACCAG	544
34	TCCCTCCTAGCCCTTAGTGA	TTGCATTCCCTGTAGGCTTT	534
35	CCACCCTAATCCCCAAC	AGGCAGGTCTGAGAGACACA	596
36	AAGGTAATGGACATAGGATGGA	CCAGAGATAAAGAGGACCCAAC	600
37	AGCCTGCAGTTCTAGGAGA	CTGAATCCCCCAAACAAATG	550
38	AGGGGATGGCCTCTCTGTTA	GGCCATCAACAAGATGTGAA	545
39	CCAAAGTTGAAGGGCAGAAG	GGTGGCAAGGGAATTGTAAA	545
40	GGAATTGCATGCCTTTCTGT	AGCGCAGAGCCACTAACTTG	564
41	GCTAACAGGAAATCAATGAGTCAA	GCTAACAGGAAATCAATGAGTCAA	545
42	TTGACTCATTGATTTCTGTAGC	AGATTTCCATGGTTCTGTGG	548
43	GGTGGCAAATCTGAGGCTTA	AGAAGAAGTGGGCATGTTGC	560
44	CCTTCTGATGAAGGCCTGAG	AATTGAGCGTGGATTGGTC	558
45	TCCAAATTTTCTGGGATGA	CAAGCAGGCTAGTCAGGGATT	578
46	GGCCATCTCTTCCAGAGCCTA	GTCTCCCAACTGCTTACCG	532
47	AGTCCCTCCATCAGGTCCTT	AGCAAAGTGAGGTCCACCAT	589

\*Amplification product in base pairs.

**Table E2.** Summary of *IRAK3* gene re-sequencing in population controls (tSNPs are in bold and underlined)

#	Position*	Location (effect)	rs/ss#	5' <sup>†</sup>	3' <sup>†</sup>	MAF <sup>‡</sup>	Val <sup>§</sup>
1	64867853	-1494T>C	rs1732888	TGCCAGTTTT	AGATACATTT	0.28	
2	64867883	-1464A>G	rs1732887	ATTTTCGCTG	CAAAACCATT	0.25	
3	64868152	-1195C>T	rs2701653	TTCAAAACAA	AATTTTACTT	0.11	
4	64870002	c.133+523T>G	rs11465939	TTTACCCAAT	TCGTTGAATG	0.02	
5	64870029	c.133+550A>G	rs1732886	TATGCTAAAG	GTGGCTTTTT	0.28	
6	64870071	c.133+592G>A	rs11465940	CTTCAGAACT	TATTTTATGC	0.02	††
<u>7</u>	64871319	c.133+1841_1853delGGAAGTGGTCAGT	ss289570834	CAAAAATTGA	AATTATCAA	0.30	§§
8	64871396	c.133+1917C>G	rs1625441	GCTGCCTGGG	TAGTTGCTAC	0.50	
9	64872005	c.133+2526T>A	rs2701654	GCAGCCTATA	GATATGACCA	0.06	
10	64872107	c.133+2628C>T	rs2576183	ACTCCTAGAG	TGCTGATTCA	0.07	**
11	64872409	c.133+2931T>C	rs56156535	TTATGGCTTT	AGTGAAAGCA	0.39	**
12	64872503	c.133+3024C>T	rs11176084	GGAGACAAC	ATCTGATTAC	0.03	
13	64876748	c.134-7010T>C	rs1168772	ATATATTAAC	AGTAAGTGTT	0.06	
14	64877754	c.134-6004A>G	rs1168774	CTTAGATGTG	TGCAAACCTCA	0.44	
15	64877849	c.134-5909G>C	rs4762088	CTGAATCTCA	AAATGTGCTT	0.37	
16	64878292	c.134-5466C>T	rs12823414	GTCTGTCAAT	AGCAAAAATA	0.06	
17	64878509	c.134-5249C>A	rs17767286	CCTAGTTAAA	CTTACCCAC	0.36	
18	64879508	c.134-4250A>G	rs17767298	CTTGTTTTCA	TCTGAAACTG	0.34	
19	64880383	c.134-3375T>C	rs1168757	GCAAACCTCTC	TCCATGTAGT	0.08	
20	64880401	c.134-3357C>G	rs1168758	AGTGACTCAG	ATTTCAGGCT	0.06	
21	64880943	c.134-2815T>C	rs1185253	CTGCAACCTC	GCCTCCTGGG	0.08	††
22	64881027	c.134-2731T>G	rs1732875	CTGGCTAATT	TTGTATTTTT	0.06	††
23	64883686	c.134-72T>A	rs1185630	AATAAAACAT	AGGTACACAA	0.09	
24	64883968	c.316+28C>T	rs1882200	TAATGTGGCT	TTAATCTGTA	0.28	
25	64884556	c.316+616A>G	rs1617632	ATTTAGAACA	CATGATCAAT	0.02	
26	64884954	c.316+1014T>C	rs2289134	TATTAGTTC	AATAGACTTC	0.34	
27	64885411	c.316+1471T>C	rs1732877	AAGACTTTAC	GCCCTAGTAC	0.39	
28	64886149	c.316+2209delT	rs11465951	TTTCAATCT	CCATTTTATA	0.08	
<u>29</u>	64889666	c.381+99C>T	rs11465955	CTGGATTTGC	GGCACTTTTT	0.31	
30	64890247	c.435A>G (p.Lys145Lys)	rs56001649	ATAATGAAAA	GGTATGAAAA	0.03	
31	64891273	c.437-220A>T	rs2293657	GGTAAAAATC	CAATGCTGCT	0.31	
<u>32</u>	64891495	c.439G>A (p.Val147Ile)	rs1152888	TTCCTAAGGA	TACTGCTTAA	0.10	
33	64892495	c.588+851C>T	rs1821777	CACACACAAA	ACAGAAGAGC	0.31	
34	64904334	c.654-2436C>T	rs55715236	CCTGAATCAT	TTCTCAGATG	0.30	**
35	64904449	c.654-2321G>T	rs1623665	CTAATTTAAG	GATAAGATGA	0.39	
<u>36</u>	64904483	c.654-2287G>A	rs1624395	ACAGAGGAGGA	AATGATAACT	0.38	
37	64904584	c.654-2186G>A	rs1370127	AATATTAAGG	AATTACCTTA	0.38	
<u>38</u>	64904905	c.654-1865C>T	rs1370128	AACAAGGCAC	CAGAACATTT	0.39	
39	64908730	c.887+313delA	rs10716217	GATTTTTTTT	CTTACAATG	0.39	
40	64912239	c.887+3822G>T	rs17102235	GCTGACAATT	TTGCCTTTCT	0.02	
41	64912278	c.887+3861A>G	ss289570833	AAGGACTTCA	TGTAACCCAT	0.02	§§
42	64912424	c.887+4007G>A	rs726302	TCCCTCCAC	GCCATTGAAC	0.02	
43	64912865	c.887+4448A>G	rs17102243	TAATTGGCTG	GAAAAGGACA	0.02	
44	64915771	c.887+7354T>A	rs1152908	AGCGTATAGT	TATAATTATA	0.38	



## RESULTADOS

---

45	64919287	c.888-5246T>C	rs1152911	AACAGTCTTT ATACAAGGGA	0.08	
46	64919431	c.888-5102G>A	rs17826057	CAAAAGACTA AAAAAGCTGAG	0.07	
47	64919952	c.888-4581T>C	rs11176095	TGTCTTAAAA GTCTGGAAAA	0.08	
<b>48</b>	64920175	c.888-4358G>A	rs1152912	TGGCCAGCCT TATCCTTTAA	0.42	
49	64920682	c.888-3851T>A	rs1152913	CAGACAGCAC TGTGTGCTGT	0.44	
50	64920772	c.888-3761T>C	rs1152914	CCCTCTTCCC AGGTCCTTCT	0.05	**
51	64920868	c.888-3652TA(6_8)	ss289570835	TATACCCACA AAATATATCG	0.05	§§
52	64920925	c.888-3608A>C	rs1152915	GTATATATAT TCTCTCTCTT	0.26	**
53	64921571	c.888-2962A>T	rs1152916	TTAAGGTTGA GCAGGGGAAA	0.45	
54	64922569	c.888-1964C>T	rs1152918	TGGGGGTGTG GAGGCCAGG	0.05	
55	64923752	c.888-781C>T	rs3782348	CTGAGGCTTA ATGGCGAAGG	0.05	
56	64924916	c.1087-79A>G	rs11176097	AACATATATC TAGTCATGGT	0.03	
57	64928242	c.*21A>G	rs7135413	ATAAAGAAAA AGCAAGTATT	0.03	
<b>58</b>	64930378	*2157T>C	rs10506481	GTTAATCATA TCACATTAGA	0.09	
59	64932466	*4245G>T	rs7972963	GTCATTIGTT ATTCCACAGA	0.08	

\*According to NCBI build 36; †Flanking sequences; ‡Minor allele frequency; §Validation status as October 2009: frequency available, ||; 1000 genomes Project, \*\*; unknown or by 2-hit as in dbSNP build 129, ††; described here for the first time, §§.

**Table E3.** Summary of quality control measures for genotyped *IRAK3* gene tSNPs

rs/ss#	Location (effect)	Position*	CR† (%)	CAC‡ (%)	HWE§ controls	HWE§ cases
ss289570834	Intron 1	64871319	95	98	0.295	0.295
rs11465955	Intron 3	64889666	91	98	0.295	0.295
rs1152888	Exon 5 (Val147Ile)	64891495	94	99	0.355	0.303
rs1624395	Intron 6	64904483	95	98	0.295	0.295
rs1370128	Intron 6	64904905	95	97	0.295	0.295
rs1152912	Intron 8	64920175	95	94	0.063	0.295
rs10506481	3' flanking region	64930378	94	91	0.295	0.063

\*According to NCBI build 36; †Completion rate; ‡Average confidence of automatic calls; §Hardy-Weinberg equilibrium *p*-values.

**Table E4.** Genotype counts of genotyped SNPs associated with ALI.

rs#	Ref. allele/Risk allele	ALI			non-ALI		
		1/1	1/2	2/2	1/1	1/2	2/2
rs1732887	A/G	103	63	12	21	9	0
rs1732886	A/G	102	63	13	21	9	0
rs10506481	C/T	2	28	139	4	6	20

**Table E5.** Final regression models used to adjust *IRAK3* SNP associations with ALI.

	<i>p</i> -value	OR	95% CI	
			Lower	Upper
<b>(A)</b> Regression model for rs10506481 (T as risk)				
rs10506481	<b>0.021</b>	2.50	1.15	5.47
Gender	0.073	2.38	0.92	6.17
ICU mortality	<b>0.002</b>	8.46	2.25	31.81
Source of infection	<b>0.011</b>	0.07	0.10	0.54
<b>(B)</b> Regression model for rs1732887 (G as risk)				
rs1732887	0.053	2.37	0.99	5.65
Gender	0.074	2.37	0.92	60.09
ICU mortality	<b>0.001</b>	8.48	2.34	30.76
Source of infection	<b>0.017</b>	0.08	0.01	0.63
<b>(C)</b> Regression model for rs1732886 (G as risk)				
rs1732886	<b>0.033</b>	2.57	1.02	6.10
Gender	0.063	2.46	0.95	6.38
ICU mortality	<b>0.001</b>	8.82	2.42	32.11
Source of infection	<b>0.016</b>	0.08	0.01	0.62

Significant *p*-values in bold.

## RESULTADOS

**Table E6.** Association of imputed SNPs of *IRAK3* gene with ALI.

rs#	Position*	MAF <sup>†</sup>		p-value
		non-ALI	ALI	
rs1732888	64867853	0.141	0.269	<b>0.009</b>
rs1732887	64867883	0.138	0.265	<b>0.010</b>
rs1732886	64870029	0.165	0.299	<b>0.012</b>
rs56156535	64872409	0.410	0.389	0.713
rs4762088	64877849	0.373	0.396	0.703
rs17767286	64878509	0.378	0.359	0.772
rs17767298	64879508	0.362	0.343	0.759
rs1882200	64883968	0.331	0.314	0.785
rs2289134	64884954	0.377	0.363	0.791
rs1732877	64885411	0.460	0.416	0.520
rs2293657	64891273	0.325	0.312	0.827
rs1821777	64892495	0.338	0.327	0.868
rs55715236	64904334	0.325	0.312	0.812
rs1623665	64904449	0.462	0.406	0.387
rs1370127	64904584	0.448	0.390	0.387
rs10716217	64908730	0.448	0.400	0.507
rs1152908	64915771	0.428	0.385	0.470

\*According to NCBI build 36; <sup>†</sup>Minor allele frequency. Significant p-values in bold.

**Table E7.** Association of *IRAK3* SNPs with quantitative measures of clinical severity.

rs#	Location (effect)	Position*	Days in ICU <sup>‡</sup>	Organs failing <sup>‡</sup>	Apache II <sup>‡</sup>
rs1732887	-1464A>G	64867883	0.063	0.102	0.771
rs1732886	c.133+550A>G	64870029	0.051	0.107	0.751
rs10506481	3' flanking region	64930378	0.388	0.374	0.702

\*According to NCBI build 36; <sup>‡</sup>p-value from a linear regression model.

## Letter to the Editor

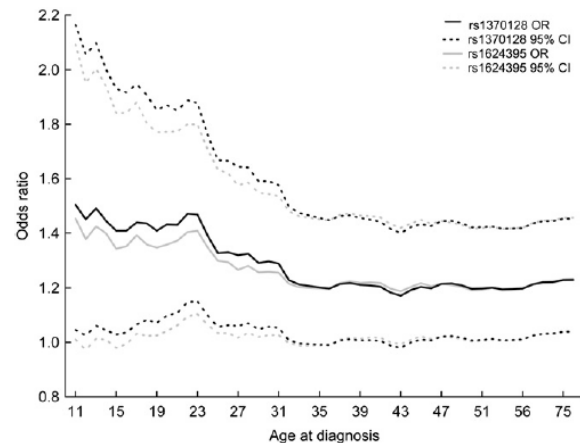
**IL-1 receptor-associated kinase 3 gene (*IRAK3*) variants associate with asthma in a replication study in the Spanish population***To the Editor:*

Asthma is a chronic inflammatory disease associated with genetic and environmental factors. Although the 12q13-24 chromosome region had previously shown linkage to asthma, the IL-1 receptor-associated kinase 3 gene (*IRAK3*) has recently emerged as a susceptibility candidate for asthma as a result of positional cloning of persistent asthma with age of onset less than 13 years in Italian samples.<sup>1</sup> *IRAK3* encodes a protein that negatively regulates Toll-like receptor signaling pathways involved in innate host defense and in the control of adaptive immune responses. However, single nucleotide polymorphisms (SNPs) near *IRAK3* have not reached the stringent significance thresholds in any of the genome-wide association studies (GWASs) performed in asthmatic patients.<sup>2</sup> Here we aimed to replicate the association of *IRAK3* genetic variants with susceptibility to asthma in a case-control study with unrelated Spanish subjects.

For this purpose, DNA samples from 607 patients aged 5 years or older with physician-diagnosed asthma and fulfilling the Global Initiative for Asthma guidelines ([www.ginasthma.com](http://www.ginasthma.com)) were collected from respiratory medicine and allergy departments as part of the Genetics of Asthma (GOA) study in the Spanish population. Samples from 1271 nonasthmatic patients without a personal or familial history of allergic or pulmonary diseases were obtained from the Spanish National DNA Biobank ([www.bancoadn.org](http://www.bancoadn.org)), and these were used as control samples. Further sample details can be found in the Methods section and Table E1 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org).

Fifteen tagging SNPs (tSNPs) were selected and genotyped, as described elsewhere,<sup>3</sup> to comprehensively survey common gene variation. All of them passed our quality control filters and were considered for the analyses (see the Methods section and Table E2 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org)). Additionally, 83 European ancestry informative markers (EuroAIMs; see Table E3 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org)) were genotyped and used for principal component analysis with EIGENSOFT to derive the scores for the first principal component (PC1).<sup>4</sup> Association of individual tSNPs was tested with the Cochran-Armitage trend test, and then a logistic regression analysis was used to calculate odds ratios (ORs) with 95% CIs and to adjust for population stratification by using the PC1 scores. Additionally, 15 untyped SNPs of the gene were imputed and tested for association. Haplotype associations were also tested to replicate previous findings.<sup>1</sup> A false discovery rate (FDR) threshold of 5% was established to limit the expected proportion of false-positive results incurred in the study when a particular subject's SNP test result was considered significant. Further method details are described in the Methods section in this article's Online Repository.

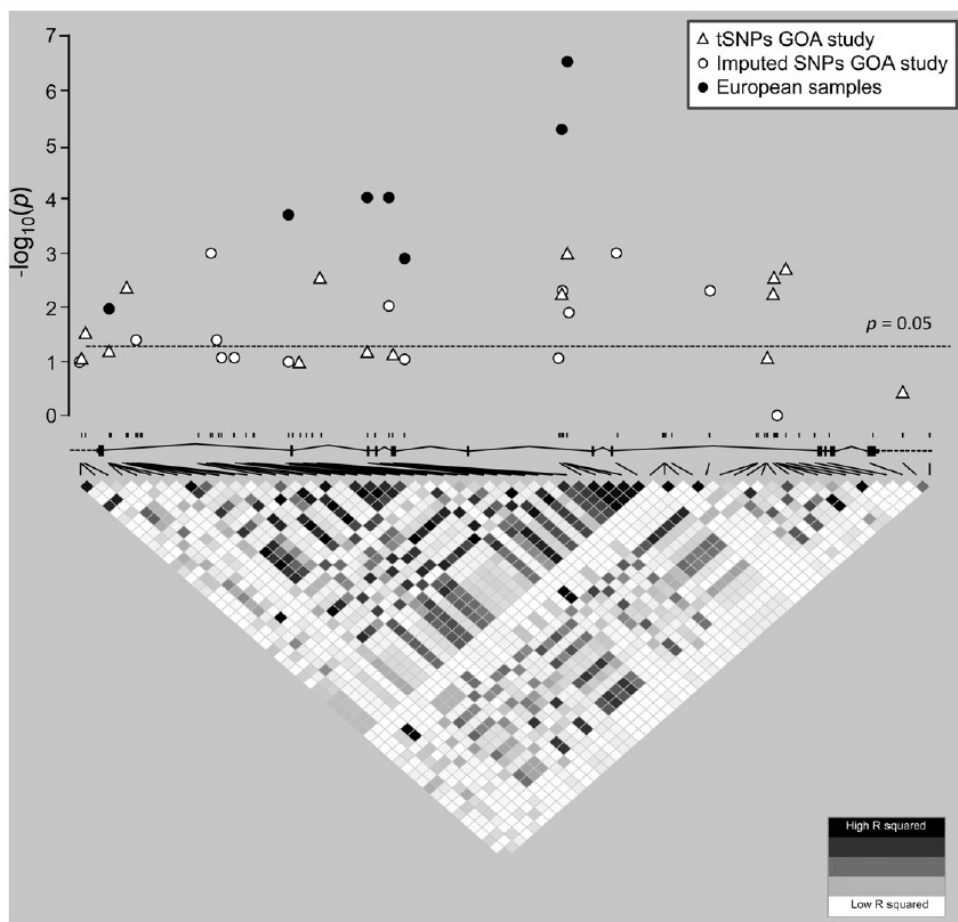
For asthma, only 1 SNP (rs1168774) was nominally associated but was not considered significant in the context of the multiple tests performed ( $P = .038$ , FDR = 42%, see Table E4 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org)). However, 4



**FIG 1.** ORs and their 95% CIs for the SNPs rs1624395 and rs1370128 in the sequential addition of atopic asthma cases according to age at diagnosis.

tSNPs showed nominal association with atopic asthma (see Table E4 for unadjusted results), 3 of them remaining significant after adjusting for population stratification: rs1732877 (OR, 1.24; 95% CI, 1.04-1.46;  $P = .014$ ), rs1624395 (OR, 1.22; 95% CI, 1.03-1.45;  $P = .019$ ), and rs1370128 (OR, 1.22; 95% CI, 1.03-1.44;  $P = .019$ ). Testing of untyped variants revealed 8 additional SNPs with nominal association with atopic asthma ( $.014 \leq P$  value  $\leq .048$ , see Table E4). All 11 SNPs were considered significantly associated with atopic asthma (FDR = 2%) and showed moderate-to-strong linkage disequilibrium (LD;  $0.48 \leq r^2 \leq 1.0$ ). Three of them replicated previous findings (rs1624395, rs1370128, and rs2293657).<sup>1</sup>

A previous study suggested that association of *IRAK3* with asthma varied with the age of onset.<sup>1</sup> Analyzing the tSNPs replicating previous findings (rs1370128 and rs1624395), we observed that the effects were largest for 22 years of age at diagnosis (Fig 1) both for rs1370128 (OR, 1.46; 95% CI, 1.14-1.87;  $P = .001$ ; permuted  $P = .021$ ) and rs1624395 (OR, 1.40; 95% CI, 1.09-1.79;  $P = .007$ ; permuted  $P = .043$ ). Thereafter, considering only those atopic asthma cases diagnosed at 22 years of age or less, we observed nominal associations for 8 tSNPs, all showing larger effects than when considering all atopic asthma cases (see Table E4). After adjusting for population stratification, all these associations remained significant: rs2701653 (OR, 1.51; 95% CI, 1.04-2.19;  $P = .030$ ), rs78503618 (OR, 1.57; 95% CI, 1.15-2.16;  $P = .005$ ), rs1732877 (OR, 1.43; 95% CI, 1.11-1.85;  $P = .007$ ), rs1624395 (OR, 1.37; 95% CI, 1.07-1.76;  $P = .013$ ), rs1370128 (OR, 1.44; 95% CI, 1.12-1.84;  $P = .004$ ), rs1152912 (OR, 1.36; 95% CI, 1.06-1.76;  $P = .018$ ), rs1152913 (OR, 1.48; 95% CI, 1.14-1.91;  $P = .003$ ), and rs1152916 (OR, 1.44; 95% CI, 1.11-1.86;  $P = .007$ ). On the contrary, none of the tSNPs was significantly associated with atopic asthma diagnosed at age greater than 22 years ( $.228 \leq P$  value  $\leq .971$ , not shown), suggesting that associations found with atopic asthma were likely to be due to the cases with an age at diagnosis of 22 years or less.



**FIG 2.** Association  $P$  values (in log scale) of *IRAK3* gene SNPs with atopic asthma diagnosed at age 22 years or less and in a meta-analysis across Europeans. An LD plot of  $r^2$  values based on resequencing data is shown for the 58 SNPs of *IRAK3*, as well as their relative position in the gene. Each diamond of the LD plot represents the pairwise  $r^2$  correlation between SNPs.

Testing of untyped variants revealed 8 additional SNPs nominally associated with atopic asthma diagnosed at age 22 years or less ( $.001 \leq P \text{ value} \leq .040$ , Fig 2 and see Table E4). All 16 SNPs nominally associated with atopic asthma diagnosed at age 22 years or less were considered significant ( $FDR \leq 2\%$ ), and most of them were in moderate-to-strong LD. The exception were rs2701653 and rs78503618 from the 5' region, which showed weak LD in pairwise comparisons between them ( $r^2 = .001$ ) and with the rest of the SNPs of the gene ( $0.001 \leq r^2 \leq .319$ ). However, only the inclusion of the SNP rs2701653 as a covariate in logistic regression left a significant association for the 2 SNPs that replicated previous findings ( $P = .012$  and  $P = .007$  for rs1624395 and rs1370128, respectively), indicating the presence of independent associations.

Haplotype testing of a previously associated 3-SNP region (rs11465955, rs1624395 and rs1370128)<sup>1</sup> replicated the association of the protective CGC haplotype ( $.001 \leq P \leq .032$ , see Table E5 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org)). However, the association of the TAT risk haplotype<sup>1</sup> was not replicated ( $P \geq .068$ ), although it showed a significant effect in logistic

regression (OR, 1.39; 95% CI, 1.07-1.81;  $P = .013$ ). Nevertheless, the CAT haplotype was detected as a risk haplotype for all asthma phenotypes ( $P \leq .023$ ,  $1.37 \leq OR \leq 1.75$ ; see Table E5). A further exploration suggested that the 3 SNPs included in haplotypes had almost perfect LD ( $D'$ ) with 2 other nearby SNPs (rs2289134 and rs1732877), revealing similar effects when extending the haplotypes to intron 2 of the *IRAK3* gene (see Table E5).

Because our results suggested that atopic asthma cases diagnosed at age 22 years or less accounted for the discovered associations, we explored whether these associations had directionally consistent effects with previous findings by performing a meta-analysis with data from candidate-gene case-control association studies of unrelated patients with childhood and early-onset asthma (see the Methods section in this article's Online Repository for further details).<sup>1,5</sup> This analysis included 3384 samples (736 cases and 2648 control subjects) and compared 7, 3, and 5 SNPs overlapping between our samples and Sardinian, Italian, and Japanese subjects, respectively. This joint analysis only evidenced allele and effect-wise replication among Europeans (Fig 2 and see Table E6 in this article's Online Repository

at [www.jacionline.org](http://www.jacionline.org)), with 1 SNP associated with suggestive significance at a genome-wide level (rs1370128; OR, 1.54; 95% CI, 1.31-1.82;  $P = 2.9 \times 10^{-7}$ ). Congruent with these results, a study of 39 candidate genes in non-Hispanic white nuclear families with asthma<sup>6</sup> reported a nominal association for 2 *IRAK3* SNPs: rs1821777, which shows an  $r^2$  value of 1 with rs2289134 associated into haplotypes of 5 SNPs in our study, and rs1152912, which is also associated in our study. However, a large-scale asthma GWAS among Europeans reported a  $P$  value of .07 for 1 *IRAK3* SNP not tested in this study (rs17826057) in a meta-analysis combining cases with childhood-onset, later-onset, severe, and occupational asthma, showing nonsignificant results also when the analysis was restricted to asthma cases with onset at less than 16 years.<sup>7</sup> Similarly, *IRAK3* was not associated with childhood asthma among Japanese or Mexican nuclear families in which most cases were atopic (90%).<sup>5,8</sup> Although more studies need to be accomplished, the nontransferability of the *IRAK3* association with asthma between studies could be due to differences in LD between populations, genetic and environmental interactions, genetic heterogeneity, or the inadequate gene coverage of previous studies (see Table E7 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org) for gene coverage in Europeans in commercial arrays).<sup>8</sup> Thus, despite the fact that SNPs from the *IRAK3* gene are not among the GWAS hits for asthma to date,<sup>2</sup> our results support the importance of this gene in patients with atopic asthma.

In this study we performed the first independent replication of *IRAK3* association with asthma,<sup>1</sup> examining extensively the most common variation of the gene. Our results replicated the association of *IRAK3* at the SNP and haplotype levels and evidenced that other variants of the gene had independent effects in atopic asthma. These results were robust to population stratification and to the differences in age and sex between cases and control subjects (not shown). Strikingly, the association of haplotypes was not restricted to atopic asthma diagnosed at age 22 years or less but also was found among all asthma cases. However, the TAT risk haplotype was not replicated, which could be motivated by our limited power to detect association on the basis of previous findings<sup>1</sup> (<60% for risks  $\geq 1.2$  at a 2-sided  $P$  value of .05 and a variant frequency of 0.35 to 0.40) or the potential presence of private risk variants in Sardinians. Moreover, we could not distinguish the causal variant because all but a few associated SNPs were in moderate-to-strong LD. Despite the fact that the tested trait was not exactly the same as that analyzed among Italian samples,<sup>1</sup> because the persistency of asthma symptoms was not recorded among GOA samples, the fact that we replicated the same SNPs and the protective haplotype supports that these results are unlikely to be spurious.

In conclusion, our results confirm the importance of the *IRAK3* gene in asthma pathogenesis in European populations, supporting the hypothesis of shared risk factors among complex diseases<sup>3</sup> and bolstering the relevance of the hygiene hypothesis in asthma development.<sup>9</sup> Further research is needed to discern whether this association is due to rare gene variants,<sup>10</sup> to the biological function of *IRAK3*, or to LD with nearby genes, as well as its relevance in non-European populations.

We thank Carole Ober, Deborah Meyers, and the 4 reviewers who helped improve the article and Tobías Felipe for Fortran scripting.

María Pino-Yanes, MSc<sup>a,b</sup>  
Inmaculada Sánchez-Machín, MD<sup>c</sup>  
José Cumplido, MD<sup>d</sup>  
Javier Figueroa, MD<sup>e</sup>  
María José Torres-Galván, PhD<sup>f</sup>  
Ruperto González, MD<sup>g</sup>  
Almudena Corrales, CLT<sup>h,b</sup>  
Orlando Acosta-Fernández, MD<sup>g</sup>  
José Carlos García-Robaina, MD<sup>f</sup>  
Teresa Carrillo, MD<sup>d</sup>  
Anselmo Sánchez-Palacios, MD<sup>e</sup>  
Jesús Villar, MD, PhD<sup>a,h,i</sup>  
Mariano Hernández, PhD<sup>d</sup>  
Carlos Flores, PhD<sup>a,b</sup>

From <sup>a</sup>CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain; <sup>b</sup>the Research Unit, Hospital Universitario NS de Candelaria, Tenerife, Spain; <sup>c</sup>the Allergy Unit, Hospital Universitario NS de Candelaria, Tenerife, Spain; <sup>d</sup>the Allergy Unit, Hospital Universitario Dr. Negrin, Las Palmas de Gran Canaria, Spain; <sup>e</sup>the Allergy Unit, Hospital Universitario Insular de Gran Canaria, Spain; <sup>f</sup>the Research Unit, Hospital Universitario Dr. Negrin, Las Palmas de Gran Canaria, Spain; <sup>g</sup>the Neumology Unit, Hospital Universitario de Canarias, Tenerife, Spain; <sup>h</sup>the Multidisciplinary Organ Dysfunction Evaluation Research Network (MODERN), Research Unit, Hospital Universitario Dr. Negrin, Las Palmas de Gran Canaria, Spain; <sup>i</sup>the Keenan Research Center, St Michael's Hospital, Toronto, Ontario, Canada; and <sup>l</sup>the Genetics Laboratory, Instituto Universitario de Enfermedades Tropicales y Salud Pública de Canarias, Universidad de La Laguna, Tenerife, Spain. E-mail: cflores@ull.es.

Supported by grants from the Ministry of Science and Innovation of Spain (PI081383) and FUNCIS (27/07) and by a specific agreement between Instituto de Salud Carlos III and Gobierno de Canarias (EMER07/001) under the ENCYT 2015 framework.

Disclosure of potential conflict of interest: The authors declare that they have no relevant conflicts of interest.

#### REFERENCES

- Balaci L, Spada MC, Olla N, Sole G, Loddo L, Anedda F, et al. *IRAK-M* is involved in the pathogenesis of early-onset persistent asthma. *Am J Hum Genet* 2007;80:1103-14.
- Ober C, Yao TC. The genetics of asthma and allergic disease: a 21st century perspective. *Immunol Rev* 2011;242:10-30.
- Pino-Yanes M, Ma SF, Sun X, Tejera P, Corrales A, Blanco J, et al. Interleukin-1 receptor-associated kinase 3 gene associates with susceptibility to acute lung injury. *Am J Respir Cell Mol Biol* 2011;45:740-5.
- Pino-Yanes M, Corrales A, Basaldúa S, Hernández A, Guerra L, Villar J, et al. North African influences and potential bias in case-control association studies in the Spanish population. *PLoS One* 2011;6:e18389.
- Nakashima K, Hirota T, Obara K, Shimizu M, Jodo A, Kameda M, et al. An association study of asthma and related phenotypes with polymorphisms in negative regulator molecules of the TLR signaling pathway. *J Hum Genet* 2006;51:284-91.
- Rogers AJ, Raby BA, Lasky-Su JA, Murphy A, Lazarus R, Klanderman BJ, et al. Assessing the reproducibility of asthma candidate gene associations, using genome-wide data. *Am J Respir Crit Care Med* 2009;179:1084-90.
- Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 2010;363:1211-21.
- Wu H, Romieu I, Shi M, Hancock DB, Li H, Sienra-Monge JJ, et al. Evaluation of candidate genes in a genome-wide association study of childhood asthma in Mexicans. *J Allergy Clin Immunol* 2010;125:321-7, e13.
- Strachan DP. Hay fever, hygiene, and household size. *BMJ* 1989;299:1259-60.
- Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. *PLoS Biol* 2010;8:e1000294.

doi:10.1016/j.jaci.2011.10.001

## METHODS

### Ethics statement

This study was approved by the ethics committees at participating hospitals, and written informed consent was obtained from all subjects.

### Study subjects

This study was conducted by using a case-control design of DNA samples from unrelated subjects reporting at least 2 generations of Spanish descent. Case samples that fulfilled the Global Initiative for Asthma guidelines for the diagnosis and classification of asthma ([www.ginasthma.com](http://www.ginasthma.com)) were collected from respiratory medicine and allergy departments as part of the GOA study in the Spanish population. This sample included 607 patients with physician-diagnosed asthma with active specific medication for asthmatic symptoms and age of 5 years or older. We recorded basic demographic data (age, sex, and smoking habits), age at diagnosis of the disease, severity of the illness, family history of allergic disease, allergic symptoms (rhinitis, atopic dermatitis, and food and drug allergy), basal FEV<sub>1</sub>, and symptomatic treatment. Atopy was confirmed by a positive skin prick test response (a wheal with a diameter 3 mm greater than that elicited by the saline control) to one of 7 common allergens, including dust mite, epithelium, pollen, fungi, food, latex, and others, or by specific serum IgE levels of greater than 0.35 IU/mL. Allergens evaluated for specific IgE levels were dust mite (*Dermatophagoides pteronyssinus*, *Dermatophagoides farinae*, *Glycyphagus* species, *Blomia tropicalis*, *Acarus siro*, *Tyrophagus putrescentiae*, *Lepidoglyphus destructor*, and *Euroglyphus maynei*), epithelium (*Felis domesticus*, *Canis familiaris*, *Equus caballus*, and *Oryctolagus cuniculus*), pollen (*Olea europaea*, *Salsola* species, *Lolium perenne*, *Artemisia vulgaris*, *Parietaria* species, *Platanus* species, *Chenopodium* species, *Plantago* species, *Rumex* species, and *Cupressus* species), fungi (*Penicillium notatum*, *Alternaria alternata*, *Aspergillus fumigatus*, and *Cladosporium herbarum*), food (cow's milk, hen's egg, peanut, soybean, wheat, fish, shrimp, crabs, lobster, clams, oysters, mussels, banana, chestnut, and kiwi), latex, and cockroach (*Blattella germanica*).

A total of 1271 DNA samples from subjects self-reporting at least 2 generations of ancestors born in Spain and without personal or familial history of allergic or pulmonary diseases were used as control samples. These were obtained from the Spanish National DNA Biobank ([www.bancoadn.org](http://www.bancoadn.org)), a member of the Public Population Project in Genomics Consortium ([www.p3gobservatory.org](http://www.p3gobservatory.org)). Briefly, samples were collected from branches of the National Blood Service from unrelated subjects residing in Spain. After signing an informed consent form, donors self-declared general health status, physical activity, transportation and nutrition habits, employment and qualification, demographics, tobacco smoke, alcohol consumption, genealogical information, residence and mother tongue, and personal and familiar history of diseases, including infectious, cancerous, blood and circulatory, endocrine, mental and behavioral, respiratory (including asthma symptoms), immunologic (including allergies), bone, congenital, skin and digestive, as well as eye and hearing disorders. No information from medical records was incorporated or revised, and no medical testing was performed on donors.

### Selection of tSNPs

tSNPs were selected by using TagIT 3.03<sup>E1</sup> with published resequencing data from 22.6 kb in 32 unrelated healthy Spanish subjects and forcing the inclusion of previously associated SNPs<sup>E2,E3</sup> and all potential functional variants, as predicted by PupaSuite<sup>E4</sup> or previously established by us.<sup>E3</sup> A final set of 15 tSNPs was selected (Table E2) and tested for performance by using an SNP-dropping-with-resampling method,<sup>E1</sup> satisfying a haplotype  $r^2$  value of 1.0 in the resequencing data.

### Genotyping of tSNPs

DNA was extracted from blood (GFX kit; GE Healthcare, Little Chalfont, United Kingdom) or saliva (Oragene DNA; DNA Genotek, Inc, Kanata, Ontario, Canada) and used for whole-genome amplification with the Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare). Dilutions of this product were quantified by using an SYBR Green I-based DNA quantification

method (Molecular Probes, Eugene, Ore) and genotyped in 2 reactions containing 7 and 8 tSNPs by means of the SNaPshot Multiplex kit (Applied Biosystems, Foster City, Calif), as described elsewhere.<sup>E3</sup> Electrophoresis was performed on an ABI 310 Genetic Analyzer (Applied Biosystems), and genotypes were automatically assigned by using GeneMapper v3.2 (Applied Biosystems). Genotyping was blind to disease status. Approximately 6% of the samples were genotyped in duplicate to monitor genotyping quality, yielding an estimated overall discordance rate among duplicated samples of 0.07% (95% CI, 0.02% to 0.19%). The genotyping completion rate was greater than 97.7% across all 15 tSNPs (see Table E2).

### Assessment of population stratification

Ninety-three EuroAIMs were genotyped in case and control samples to reduce the risk for false-positive results, allowing correcting for major population stratification effects among European and Spanish populations.<sup>E5,E6</sup> Genotyping of EuroAIMs was conducted by using the iPLEX Gold assay on the MassARRAY system, and genotype calls were automatically generated by using all data from the study simultaneously with TYPER 3.4 software (Sequenom, San Diego, Calif) by the Spanish National Genotyping Center, Santiago de Compostela Node (CeGen, [www.cegen.org](http://www.cegen.org)). Of 93 EuroAIMs, 83 markers were successfully genotyped and considered for the final analysis (see Table E3). Our previous studies have demonstrated that a subset of as few as 65 of these EuroAIMs effectively controlled false-positive results in Spanish case-control association studies.<sup>E6</sup> On the basis of these markers, principal component analysis was used to derive the scores for the PC1 as ancestry estimates in cases and control subjects. This analysis was performed with EIGENSOFT,<sup>E7</sup> as described elsewhere.<sup>E6</sup>

### Statistical analysis

Clinical and demographic data were analyzed by using  $\chi^2$  and Mann-Whitney  $U$  tests with SPSS 15.0 software (SPSS, Inc, Chicago, Ill). Departures from Hardy-Weinberg equilibrium were evaluated separately for cases and control subjects by using an exact test.<sup>E8</sup> Because none of the tSNPs deviated significantly from Hardy-Weinberg equilibrium in the control group (threshold for significance after considering the multiple comparisons performed was a  $P$  value of .003), all tSNPs were retained for the association analyses. Further details can be found in Table E2.

Individual tSNP associations were tested by using the Cochran-Armitage trend test, assuming an additive model, and the empiric  $P$  value was obtained by using 1000 permutations with a custom script for STATISTICA (StatSoft, Inc, Tulsa, Okla).<sup>E9</sup> Logistic regression analysis was used to calculate allelic effects as ORs with 95% CIs, to adjust for population stratification by using the PC1 scores as a covariate, and to test the independence of associations between SNPs. These analyses were performed with SPSS 15.0 software (SPSS, Inc).

Because a previous study evidenced an increase of allelic effects among cases with onset at less than 13 years, we explored whether the allelic effects varied with the age at diagnosis to estimate the trait cutoff at which the allelic effects were largest. For that, a previously described algorithm was used.<sup>E10</sup> Briefly, individual SNP associations were computed as above but repeatedly for each group of cases resulting from the sequential incorporation of atopic asthmatic patients for every increase in 1 year in the age of diagnosis (starting from the age of 11 years, for which the study accumulated >50 cases). The empiric  $P$  value, adjusted for the multiple tests done, was then obtained from the comparison of the smallest nominal  $P$  value from all tests done with  $P$  values calculated from permuting genotypes 1000 times among the whole sample (by reshuffling labels of cases and controls) as the proportion of permutation replicates that yielded  $P$  values smaller than that observed in the actual dataset.

Untyped SNPs with multilocus LD values of .7 or greater<sup>E11</sup> or with a minor allele frequency (MAF) of 10% or greater were imputed from tSNPs and tested for association by means of TUNA version 1.1<sup>E11</sup> to limit the imputation error,<sup>E3</sup> allowing us to analyze 15 additional SNPs of the gene. Haplotypes from the resequencing data were estimated with PHASE 2.1<sup>E12</sup> to be used as a reference for the underlying LD across the gene. An FDR was assessed with QVALUE<sup>E13</sup> considering all 30 SNPs tested (15 tSNPs and 15 imputed

SNPs) to judge the SNP associations in the context of the multiple comparisons performed. An omnibus test overall haplotype frequency distribution was first obtained with WHAP to determine whether there were haplotype associations.<sup>E14</sup> Then the statistical significance for the association of specific haplotypes was assessed by using logistic regression with HAPSTAT,<sup>E15</sup> and ORs and 95% CIs were obtained by using THESIAS,<sup>E16</sup> with the latter using the most common haplotype as a reference.

Under the principle that if SNPs had consistent effects (ie, allele and effect wise) among different studies, a joint analysis would improve the results obtained from the studies by using a separate analysis, a meta-analysis of association for the SNPs overlapping across our study and the previous ones<sup>E2,E17</sup> was performed. For this purpose, fixed and random effects meta-analyses were examined with the R library rmeta version 2.16.<sup>E18</sup> The patterns of LD, as  $r^2$  and  $D'$  values, were explored with Haploview 3.32.<sup>E19</sup>

Finally, an evaluation of the array coverage of SNPs from the *IRAK3* gene with MAFs of 5% or greater in the GWAS performed in samples of European ancestry to date was assessed.<sup>E20-E26</sup> For that, the information of SNPs contained in the arrays was downloaded from the UCSC Genome Bioinformatics Web page (<http://genome.ucsc.edu>) for the region delimited by our resequencing survey. Gene coverage was then assessed as the proportion of SNPs that were captured by the array both directly and indirectly (with a multimarker  $r^2 \geq 0.8$ ) with respect to the total number of SNPs with MAFs of 5% or greater in Utah residents with ancestry from northern and western Europe from Hap-Map release 28 and in our resequencing data.

## REFERENCES

- E1. Ahmadi KR, Weale ME, Xue ZY, Soranzo N, Yarnall DP, Briley JD, et al. A single-nucleotide polymorphism tagging set for human drug metabolism and transport. *Nat Genet* 2005;37:84-9.
- E2. Balaci L, Spada MC, Olla N, Sole G, Loddio L, Anedda F, et al. IRAK-M is involved in the pathogenesis of early-onset persistent asthma. *Am J Hum Genet* 2007;80:1103-14.
- E3. Pino-Yanes M, Ma SF, Sun X, Tejera P, Corrales A, Blanco J, et al. Interleukin-1 receptor-associated kinase 3 gene associates with susceptibility to acute lung injury. *Am J Respir Cell Mol Biol* 2011;45:740-5.
- E4. Conde L, Vaquerizas JM, Dopazo H, Arbiza L, Reumers J, Rousseau F, et al. PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes. *Nucleic Acids Res* 2006;34:W621-5.
- E5. Price AL, Butler J, Patterson N, Capelli C, Pascali VL, Scarnicci F, et al. Discerning the ancestry of European Americans in genetic association studies. *PLoS Genet* 2008;4:e236.
- E6. Pino-Yanes M, Corrales A, Basaldúa S, Hernández A, Guerra L, Villar J, et al. North African influences and potential bias in case-control association studies in the Spanish population. *PLoS One* 2011;6:e18389.
- E7. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904-9.
- E8. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet* 2005;76:887-93.
- E9. Sun X, Ma SF, Wade MS, Flores C, Pino-Yanes M, Moitra J, et al. Functional variants of sphingosine-1-phosphate receptor 1 gene associate with asthma susceptibility. *J Allergy Clin Immunol* 2010;126:241-9, e1-3.
- E10. Macgregor S, Craddock N, Holmans PA. Use of phenotypic covariates in association analysis by sequential addition of cases. *Eur J Hum Genet* 2006;14:529-34.
- E11. Wen X, Nicolae DL. Association studies for untyped markers with TUNA. *Bioinformatics* 2008;24:435-7.
- E12. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001;68:978-89.
- E13. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 2003;100:9440-5.
- E14. Purcell S, Daly MJ, Sham PC. WHAP: haplotype-based association analysis. *Bioinformatics* 2007;23:255-6.
- E15. Lin DY, Hu Y, Huang BE. Simple and efficient analysis of disease association with missing genotype data. *Am J Hum Genet* 2008;82:444-52.
- E16. Tregouet DA, Escolano S, Tiret L, Mallet A, Golmard JL. A new algorithm for haplotype-based association analysis: the Stochastic-EM algorithm. *Ann Hum Genet* 2004;68:165-77.
- E17. Nakashima K, Hirota T, Obara K, Shimizu M, Jodo A, Kameda M, et al. An association study of asthma and related phenotypes with polymorphisms in negative regulator molecules of the TLR signaling pathway. *J Hum Genet* 2006;51:284-91.
- E18. R Development Core Team. R: a language and environment for statistical computing. Vienna (Austria): R Development Core Team; 2008.
- E19. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263-5.
- E20. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, et al. Genetic variants regulating *ORMDL3* expression contribute to the risk of childhood asthma. *Nature* 2007;448:470-3.
- E21. Himes BE, Hunninghake GM, Baurley JW, Rafaels NM, Sleiman P, Strachan DP, et al. Genome-wide association analysis identifies *PDE4D* as an asthma-susceptibility gene. *Am J Hum Genet* 2009;84:581-93.
- E22. Sleiman PM, Flory J, Imielinski M, Bradfield JP, Annaiah K, Willis-Owen SA, et al. Variants of *DENND1B* associated with asthma in children. *N Engl J Med* 2010;362:36-44.
- E23. Li X, Howard TD, Zheng SL, Haselkorn T, Peters SP, Meyers DA, et al. Genome-wide association study of asthma identifies *RAD50-IL13* and *HLA-DR/DQ* regions. *J Allergy Clin Immunol* 2010;125:328-35, e11.
- E24. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 2010;363:1211-21.
- E25. Ferreira MA, McRae AF, Medland SE, Nyholt DR, Gordon SD, Wright MJ, et al. Association between *ORMDL3*, *IL13* and a deletion on chromosome 17q21 with asthma risk in Australia. *Eur J Hum Genet* 2011;19:458-64.
- E26. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet* 2011;43:887-92.



TABLE E1. Relevant demographic and clinical features of samples

Variable	Cases (n = 607)	Control subjects (n = 1271)	P value
Sex (% males)	23.3 (567)	59.1 (1270)	<.001*
Age (y), mean, P <sub>25</sub> -P <sub>75</sub>	37, 25-44 (569)	40, 32-48 (1270)	<.001†
Smoking habits (% ever smokers)	27.3 (344)	48.1 (1122)	<.001*
Clinical features			
Serum IgE levels (IU/mL), mean, P <sub>25</sub> -P <sub>75</sub>	478.1, 89.3-394.5 (259)	NA	
Age at diagnosis (y), mean, P <sub>25</sub> -P <sub>75</sub>	28, 14-39 (570)	NA	
FEV <sub>1</sub> (% predicted), mean, P <sub>25</sub> -P <sub>75</sub>	82, 68-95 (231)	NA	
Rhinitis (% positive)	74.2 (485)	NA	
Dermatitis (% positive)	20.2 (480)	NA	
SPT response (% positive)	76.6 (445)	NA	
Specific IgE (% positive)	51.2 (287)	NA	
Atopy (% positive)	76.7 (464)	NA	
Treatment			
Inhaled corticoids (%)	85.6 (250)	NA	
Oral corticoids (%)	4.6 (305)	NA	
Long-acting β <sub>2</sub> -agonists (%)	53.0 (317)	NA	
Short-acting β <sub>2</sub> -agonists (%)	81.8 (231)	NA	
Body mass index (kg m <sup>-2</sup> ), mean, P <sub>25</sub> -P <sub>75</sub>	24.6, 21.1-27.5 (227)	NA	

Numbers in parentheses indicate the total number of subjects with available data.

NA, Not available; P<sub>25</sub>, percentile 25; P<sub>75</sub>, percentile 75.

\*χ<sup>2</sup> Test.

†Mann-Whitney U test.

ARTICLE IN PRESS

J ALLERGY CLIN IMMUNOL  
VOLUME ■■■, NUMBER ■■

LETTER TO THE EDITOR 3.e4

TABLE E2. Information, CRs, and HWE *P* values for *IRAK3* tSNPs

rs No.	Position*	SNP (effect)	CR (%)	HWE <i>P</i> value			
				Control subjects	Asthmatic patients	Atopic asthmatic patients	Diagnosed at $\leq 22$ y
rs1732887	64867883	-1464A>G	98.7	<b>.004</b>	.579	.779	.404
rs2701653	64868152	-1195C>T	98.1	.312	.276	.570	.754
rs1732886	64870029	c.133+550A>G	98.1	<b>.014</b>	.912	.888	.579
rs78503618	64871319	c.133+1841_1853delGGAAGTGGTCAGT	98.7	.278	.534	.368	.201
rs2289134	64884954	c.316+1014T>C	98.3	.307	.654	.909	1.000
rs1732877	64885411	c.316+1471T>C	97.7	.642	.616	.157	.515
rs11465955	64889666	c.381+99C>T	98.2	.250	.721	.909	.907
rs1152888	64891495	c.439G>A (p.Val147Ile)	98.0	.434	.139	.766	.289
rs1624395	64904483	c.654-2287G>A	98.9	.593	1.000	.519	.513
rs1370128	64904905	c.654-1865C>T	98.5	.522	.457	.164	.589
rs11176095	64919952	c.888-4581T>C	98.7	<b>.005</b>	.168	.421	<b>.034</b>
rs1152912	64920175	c.888-4358G>A	98.3	.280	.808	.245	.750
rs1152913	64920682	c.888-3851T>A	98.2	.865	.060	<b>.031</b>	.139
rs1152916	64921571	c.888-2962A>T	98.9	1.000	.056	.152	.072
rs10506481	64930378	*2157T>C	98.3	.114	<b>.011</b>	.218	<b>.003</b>

Nominally significant associations are shown in boldface.

CR, Completion rate; HWE, hardy-Weinberg equilibrium.

\*According to National Center for Biotechnology Information build 36.3.

TABLE E3. Summary statistics for EuroAIMs successfully genotyped in this study

rs No.	Alleles	Chromosome	Position*	CR (%)	MAF		HWE P values	
					Control subjects	Asthmatic patients	Control subjects	Asthmatic patients
rs1157492	C/T	1	211627299	97.9	0.430	0.397	.908	.549
rs11807062	A/C	1	3176394	97.5	0.171	0.179	.070	.089
rs1416467	A/G	1	80470749	98.5	0.491	0.469	.910	.162
rs1890131	C/T	1	164329131	98.3	0.303	0.293	.893	.694
rs2236876	A/G	1	169749143	97.2	0.315	0.325	.427	.347
rs2419063	A/G	1	186988492	97.8	0.105	0.122	1.000	.702
rs495347	C/T	1	17909228	98.6	0.304	0.300	.592	.285
rs725974	G/T	1	167344909	96.7	0.187	0.215	1.000	.275
rs7552548	A/G	1	82775867	96.5	0.320	0.310	.030	.177
rs10496610	C/G	2	123757682	99.2	0.230	0.222	.749	1.000
rs1448314	A/G	2	223811705	98.1	0.087	0.089	.101	.307
rs1517407	A/G	2	63401982	98.2	0.474	0.469	.046	.323
rs3769005	C/G	2	136437098	98.2	0.482	0.523	.255	.805
rs4832640	C/T	2	19025771	97.1	0.255	0.272	.200	.214
rs6432110	A/T	2	10688616	98.1	0.431	0.458	.017	.620
rs6745653	C/T	2	97748515	98.8	0.116	0.124	.679	.709
rs1879558	C/T	3	155137053	98.7	0.188	0.176	.307	.122
rs2596834	C/T	3	10795067	95.4	0.495	0.470	.013	.561
rs4686497	A/G	3	190124209	98.3	0.168	0.205	.042	.209
rs822759	G/T	3	22948116	98.6	0.324	0.356	.846	.214
rs9290675	C/T	3	180489335	99.1	0.131	0.120	.135	.562
rs9861816	C/G	3	195649797	98.2	0.259	0.240	.235	.501
rs10516982	G/T	4	96963279	98.1	0.272	0.289	.566	.135
rs1373557	G/T	4	167601812	98.9	0.202	0.199	.332	.201
rs17443616	A/G	4	41063093	98.5	0.306	0.311	.640	.182
rs1873195	C/T	4	38713739	98.5	0.301	0.267	.788	.021
rs2014303	A/C	4	10262125	97.8	0.221	0.224	.869	.289
rs2251432	A/G	4	191063741	97.7	0.208	0.190	.195	.503
rs4555709	C/T	4	53867889	99.2	0.073	0.074	.035	.559
rs9328764	A/G	4	2143685	98.6	0.160	0.174	.599	.398
rs974020	C/T	4	117865302	98.6	0.296	0.270	.153	.470
rs16891982	C/G	5	33987450	99.3	0.201	0.253	.429	.452
rs33706	C/G	5	87540920	98.6	0.284	0.305	.036	.290
rs3822616	A/G	5	94828145	98.4	0.192	0.183	.143	.338
rs10484547	C/G	6	29560753	99.4	0.071	0.083	.828	.592
rs2187684	C/T	6	32872697	98.0	0.360	0.341	.902	.273
rs2804756	C/T	6	698586	99.1	0.086	0.079	.151	.253
rs756147	A/G	6	141672412	99.5	0.055	0.077	.009	.773
rs10486207	A/G	7	7884020	98.4	0.282	0.273	.184	.917
rs17864053	A/G	7	89801957	98.6	0.209	0.210	.229	.806
rs1922086	C/T	7	155515702	97.7	0.394	0.414	.006	.312
rs2097884	A/C	7	4125658	97.4	0.251	0.257	.172	.747
rs2219248	C/T	7	113965977	98.7	0.355	0.350	1.000	.059
rs2367191	A/G	7	141571268	98.3	0.413	0.405	.412	.671
rs2905347	A/G	7	22393559	96.4	0.435	0.449	.640	.677
rs10504924	C/T	8	94157437	97.7	0.092	0.123	.864	.704
rs920590	C/T	8	19695441	98.2	0.372	0.366	.114	.479
rs10512122	C/G	9	81715524	98.6	0.175	0.165	.170	.458
rs1408794	A/G	9	12641340	94.8	0.477	0.523	.011	.004
rs2086085	A/C	9	1680761	99.1	0.106	0.120	.455	.019
rs1045873	A/C	10	25177778	98.6	0.301	0.281	.105	.268
rs10508372	A/G	10	9012024	99.6	0.087	0.092	.723	.216
rs10509384	A/G	10	78693188	97.3	0.298	0.284	.276	.611
rs10509954	A/G	10	113658378	98.4	0.109	0.115	.143	.105
rs379773	C/G	10	109905342	95.4	0.353	0.354	.704	.470
rs7908825	C/G	10	75224547	97.4	0.324	0.312	1.000	.104
rs1560569	A/G	11	8896463	97.2	0.344	0.340	.344	.781
rs2847502	C/T	11	119623708	98.9	0.306	0.304	.206	.847
rs4938377	G/T	11	116801502	98.1	0.063	0.076	1.000	.141
rs923031	A/C	11	15777602	98.3	0.468	0.449	.138	.741
rs1003306	A/T	12	2624458	98.0	0.339	0.372	.447	.005

(Continued)

ARTICLE IN PRESS

J ALLERGY CLIN IMMUNOL  
VOLUME ■■■, NUMBER ■■

LETTER TO THE EDITOR 3.e6

TABLE E3. (Continued)

rs No.	Alleles	Chromosome	Position*	CR (%)	MAF		HWE <i>P</i> values	
					Control subjects	Asthmatic patients	Control subjects	Asthmatic patients
rs1582398	A/G	12	78624019	97.5	0.129	0.140	.377	.732
rs3809125	C/T	12	55130616	98.3	0.367	0.369	.112	.724
rs7965049	A/G	12	7537044	97.5	0.245	0.253	.357	.227
rs998401	A/G	12	53641500	98.0	0.074	0.074	.058	1.000
rs7997100	C/T	13	67988075	98.4	0.349	0.348	.010	.719
rs986642	C/G	13	56944263	98.8	0.239	0.268	.160	.009
rs10483853	A/G	14	72826052	98.7	0.234	0.257	.344	.236
rs10519269	A/C	15	77705433	96.5	0.106	0.130	.095	.586
rs1129038	C/T	15	26030454	98.4	0.331	0.341	.898	.927
rs7163907	C/T	15	73632152	98.9	0.303	0.328	.254	.164
rs1107820	C/T	17	41491195	97.4	0.274	0.242	1.000	.120
rs1476162	A/G	17	8031808	98.6	0.135	0.146	.544	1.000
rs2003092	A/G	17	59516509	97.7	0.274	0.314	1.000	.337
rs959260	C/T	17	70881017	98.3	0.218	0.215	.934	.716
rs2418844	A/G	18	26037261	97.8	0.489	0.503	.253	.806
rs4892082	A/C	18	69051672	97.5	0.123	0.131	1.000	.721
rs523776	C/T	18	7554299	98.4	0.242	0.220	1.000	.722
rs959763	G/T	18	56797826	98.0	0.363	0.386	.854	.796
rs103294	C/T	19	59489660	98.7	0.162	0.156	.835	.643
rs202546	C/T	20	1611539	97.3	0.429	0.438	.449	.675
rs477627	A/G	20	47613465	97.6	0.193	0.217	.647	.904
rs969539	A/G	22	24942341	97.2	0.374	0.362	.626	.211

CR, Completion rate; HWE, Hardy-Weinberg equilibrium.

\*According to National Center for Biotechnology Information build 35.

TABLE E4. Association summary, including unadjusted effects and P values, of IRAK3 SNPs with asthma and atopic asthma

rs No.	Position*	SNP † (effect)	Risk allele	Minor allele frequency				OR (95% CI) ‡				P values §		
				Control subjects	Asthmatic patients	Atopic asthmatic patients	Diagnosed at ≤22 y	Asthmatic patients	Atopic asthmatic patients	Diagnosed at ≤22 y	Asthmatic patients	Atopic asthmatic patients	Diagnosed at ≤22 y	
rs1732888	64867853	-1494T>C	T	0.273	0.258	0.269	0.228					.344	.819	.098
rs1732887	64867883	-1464A>G	A	0.258	0.243	0.254	0.210	1.09 (0.93-1.27)	1.02 (0.85-1.23)	1.28 (0.96-1.72)		.309	.819	.082
rs2701653	64868152	-1195C>T	T	0.094	0.104	0.106	0.134	1.14 (0.91-1.42)	1.15 (0.88-1.51)	1.49 (1.03-2.14)		.307	.355	<b>.034</b>
rs1732886	64870029	c.133+550A>G	A	0.256	0.242	0.247	0.205	1.08 (0.92-1.25)	1.04 (0.87-1.27)	1.32 (0.98-1.75)		.368	.610	.061
rs78503618	64871319	c.133+1841_1853delG	ins	0.267	0.269	0.231	0.192	.99 (0.85-1.16)	1.22 (1.00-1.49)	1.54 (1.12-2.13)		.866	.057	<b>.004</b>
		GAAAGTGG												
		TCAGT												
rs56156535	64872409	c.133+2931T>C	C	0.402	0.422	0.439	0.463					.181	<b>.046</b>	<b>.040</b>
rs1168774	64877754	c.134-600A>G	G	0.429	0.458	0.472	0.483					<b>.038</b>	<b>.014</b>	<b>.001</b>
rs4762088	64877849	c.134-5909G>C	C	0.386	0.407	0.425	0.449					.167	<b>.048</b>	<b>.039</b>
rs17767286	64878509	c.134-5249C>A	A	0.379	0.394	0.411	0.433					.325	.111	.071
rs17767298	64879508	c.134-4250A>G	G	0.362	0.377	0.393	0.413					.316	.108	.072
rs1882200	64883968	c.316+28C>T	T	0.331	0.347	0.365	0.384					.343	.084	.105
rs2289134	64884954	c.316+1014T>C	C	0.334	0.347	0.365	0.385	1.06 (0.91-1.22)	1.14 (0.96-1.36)	1.24 (0.96-1.59)		.410	.117	.095
rs1732877	64885411	c.316+1471T>C	C	0.417	0.445	0.470	0.488	1.13 (0.98-1.29)	1.24 (1.04-1.46)	1.46 (1.13-1.89)		.113	<b>.008</b>	<b>.003</b>
rs11465955	64889666	c.381+99C>T	T	0.334	0.350	0.368	0.392	1.07 (0.93-1.24)	1.16 (0.97-1.37)	1.28 (0.99-1.64)		.293	.082	.058
rs2293657	64891273	c.437-220A>T	T	0.286	0.310	0.327	0.348					.067	<b>.018</b>	<b>.010</b>
rs1152888	64891495	c.439G>A	A	0.080	0.091	0.100	0.112	1.17 (0.92-1.49)	1.29 (0.97-1.72)	1.46 (0.98-2.17)		.270	.085	.070
		(p. Val147Ile)												
rs1821777	64892495	c.588+851C>T	T	0.343	0.350	0.369	0.390					.677	.177	.088
rs5715236	64904334	c.654-2436C>T	T	0.329	0.338	0.359	0.383					.589	.158	.068
rs1623665	64904449	c.654-2321G>T	T	0.415	0.442	0.466	0.489					.124	<b>.023</b>	<b>.005</b>
rs1624395	64904483	c.654-2287G>A	A	0.392	0.419	0.442	0.475	1.12 (0.98-1.29)	1.23 (1.04-1.45)	1.40 (1.09-1.79)		.101	<b>.013</b>	<b>.007</b>
rs1370127	64904584	c.654-2186G>A	A	0.392	0.419	0.442	0.475					.106	<b>.022</b>	<b>.012</b>
rs1370128	64904905	c.654-1865C>T	T	0.415	0.443	0.466	0.489	1.12 (0.98-1.29)	1.22 (1.04-1.44)	1.46 (1.14-1.87)		.107	<b>.011</b>	<b>.001</b>
rs10716217	64908730	c.887+313delA	delA	0.415	0.442	0.466	0.489					.125	<b>.017</b>	<b>.001</b>
rs1152908	64915771	c.887+7354T>A	A	0.399	0.423	0.444	0.484					.126	<b>.015</b>	<b>.005</b>
rs11176095	64919952	c.888-4581T>C	C	0.062	0.066	0.074	0.088	1.10 (0.84-1.44)	1.24 (0.90-1.70)	1.48 (0.96-2.28)		.618	.235	.086
rs1152912	64920175	c.888-4358G>A	A	0.479	0.483	0.499	0.561	1.02 (0.88-1.17)	1.10 (0.93-1.30)	1.40 (1.09-1.81)		.856	.273	<b>.006</b>
rs1152913	64920682	c.888-3851T>A	A	0.496	0.492	0.463	0.401	1.05 (0.92-1.20)	1.18 (1.00-1.39)	1.51 (1.17-1.94)		.515	.057	<b>.003</b>
rs1152915	64920925	c.888-3608A>C	C	0.304	0.312	0.318	0.304					.605	.428	.986
rs1152916	64921571	c.888-2962A>T	A	0.443	0.425	0.402	0.353	1.08 (0.94-1.23)	1.18 (1.00-1.40)	1.46 (1.13-1.89)		.310	<b>.043</b>	<b>.002</b>
rs10506481	64930378	*2157T>C	T	0.123	0.120	0.124	0.104	1.03 (0.84-1.27)	1.00 (0.78-1.28)	1.20 (0.81-1.79)		.767	.986	.362

rsSNPs are underlined. Nominally significant associations are shown in boldface.  
 \*According to National Center for Biotechnology Information (NCBI) build 36.3.  
 †According to the sequence with NCBI accession no. NC\_000012.10.  
 ‡Calculated only for the genotyped SNPs.  
 §Obtained from 1000 permutations.

TABLE E5. Analysis of haplotypes of 3 or 5 SNPs

Haplotype*	Frequencies					P value		OR (95% CI); P value†	
	Control subjects	Asthmatic patients	Atopic asthmatic patients	Diagnosed at ≤22 y	Asthmatic patients	Atopic asthmatic patients	Diagnosed at ≤22 y	Asthmatic patients	Atopic asthmatic patients
CGC	0.582	0.545	0.525	0.481	<b>.032</b>	<b>.006</b>	<b>.001</b>	Reference	Reference
TAT	.329	.340	.361	.383	.494	.113	.068	1.09 (0.94-1.27); .267	1.18 (0.98-1.41); .077
CAT	0.062	0.082	0.084	0.091	<b>.029</b>	<b>.043</b>	.066	1.37 (1.05-1.80); <b>.020</b>	1.46 (1.05-2.02); <b>.023</b>
CGT	0.022	0.022	0.027	0.036	.984	.957	.154	1.05 (0.65-1.68); .854	0.96 (0.52-1.80); .909
Omnibus					.071	<b>.035</b>	<b>.010</b>	Reference	Reference
TTCCG	0.578	0.542	0.522	0.477	.077	<b>.011</b>	<b>.003</b>	Reference	Reference
CCTAT	0.327	0.336	0.358	0.376	.417	.118	.093	1.08 (0.93-1.26); .288	1.15 (0.97-1.39); .111
TCCAT	0.060	0.082	0.085	0.093	<b>.016</b>	<b>.019</b>	<b>.035</b>	1.37 (1.05-1.80); <b>.022</b>	1.43 (1.03-1.98); <b>.032</b>
TCCGT	0.021	0.021	0.022	0.035	.973	.941	.199	1.04 (0.64-1.70); .877	1.02 (0.55-1.88); .958
Omnibus					.084	<b>.035</b>	<b>.019</b>	Reference	Reference

Nominally significant associations are shown in boldface. SNPs included in the haplotypes studied by Balaci et al.<sup>22</sup> are underlined.

\*Haplotypes are defined by 3 rSNPs (rs11465955, rs1624395, and rs1370128) or 5 SNPs SNPs (rs2289134, rs1732877, rs11465955, rs1624395, and rs1370128).

†P value from regression logistic model by using the most common haplotype as reference.

TABLE E6. Meta-analysis of /RAK3 SNPs in unrelated case-control samples

Rs no.	Risk allele	Minor allele frequency												Association tests			
		Spanish			Sardinian			Italian			Japanese			Individual studies		Meta-analysis	
		Control subjects (n = 1271)	Cases† (n = 139)	P value	Control subjects (n = 460)	Cases (n = 139)	P value	Control subjects (n = 278)	Cases (n = 67)	P value	Control subjects (n = 639)	Cases (n = 391)	P value	OR (95% CI)*	P value	OR (95% CI)*	P value
rs1732887	A	0.26	0.21	.082	0.10	0.09	.082	1.28 (0.96-1.72)	.082	1.06 (0.78-1.43)	.725	1.05 (0.87-1.25)	.620	1.18 (0.95-1.46)	.1420	1.11 (0.95-1.31)	.1791
rs2701653	T	0.09	0.13	.034	0.41	0.43	.034	1.32 (0.98-1.75)	.061	1.49 (1.03-2.14)	.034	1.32 (0.98-1.75)	.061	1.36 (1.08-1.72)	.0105	1.42 (1.18-1.71)	.0002
rs1732886	A	0.26	0.21	.061	0.16	0.16	.106	1.25 (0.97-1.61)	.105	1.65 (1.26-2.17)	.001	1.48 (1.01-2.18)	.029	1.42 (1.18-1.71)	.0002	1.42 (1.2-1.68)	.0001
rs1882200	T	0.33	0.38	.046	0.34	0.46	.046	1.28 (0.99-1.64)	.058	1.56 (1.18-2.04)	.001	1.07 (0.82-1.39)	.180	1.07 (0.82-1.39)	.0001	1.31 (1.13-1.53)	.0004
rs11465955	T	0.33	0.39	.046	0.34	0.46	.046	1.32 (1.01-1.71)	.010	1.65 (1.26-2.17)	.001	0.96 (0.81-1.15)	.839	1.03 (0.87-1.21)	.3719		
rs2293657	T	0.29	0.35	.046	0.34	0.46	.046	1.46 (0.98-2.17)	.07	1.22 (0.94-1.57)	.088	1.54 (1.17-2.02)	.001	1.36 (1.13-1.63)	.0013		
rs1152888	A	0.08	0.11	.046	0.40	0.40	.046	1.40 (1.09-1.79)	.007	1.46 (1.11-1.91)	.002	1.68 (1.15-2.46)	.004	1.47 (1.25-1.74)			
rs1821777	T	0.34	0.39	.046	0.36	0.46	.046	1.46 (1.14-1.87)	.001	1.53 (1.16-2.00)	.001	1.76 (1.21-2.58)	.002	1.54 (1.31-1.82)			
rs1624395	A	0.39	0.48	.053	0.40	0.53	.053	1.46 (1.14-1.87)	.001	1.53 (1.16-2.00)	.001	1.76 (1.21-2.58)	.002	1.54 (1.31-1.82)			
rs1370128	T	0.42	0.49	.041	0.41	0.55	.041	1.46 (1.14-1.87)	.001	1.53 (1.16-2.00)	.001	1.76 (1.21-2.58)	.002	1.54 (1.31-1.82)			
rs10716217	delA	0.42	0.49	.001	0.54	0.55	.001	1.50 (1.17-1.92)	.001	1.04 (0.87-1.24)	.685	1.17 (1.02-1.36)	.0295				

\*From an additive model.  
 †Atopic asthma diagnosed at 22 years or younger.  
 ‡From Table E4.  
 §Only childhood cases from the Japanese study were considered in this analysis.

## ARTICLE IN PRESS

J ALLERGY CLIN IMMUNOL  
VOLUME ■■■, NUMBER ■■■

LETTER TO THE EDITOR 3.e10

**TABLE E7.** Coverage of SNPs from the *IRAK3* gene with MAFs of 5% or greater on commercial arrays used in asthma GWAS in samples of European ancestry\*

SNP array	Coverage (%)	
	HapMap CEU	Resequencing
Affymetrix 500K	63	29
Affymetrix Genome-Wide Human SNP Array 6.0	93	71
Illumina HumanHap300†	77	19
Illumina HumanHap550‡	83	35
Illumina HumanHap650Y	83	42
Illumina 1M-Duo	97	65

CEU, Ancestry from northern and western Europe.

\*Includes studies described in Moffatt et al, 2007<sup>E20</sup>; Himes et al, 2009<sup>E21</sup>; Sleiman et al, 2009<sup>E22</sup>; Li et al, 2010<sup>E23</sup>; Moffatt et al, 2010<sup>E24</sup>; Ferreira et al, 2011<sup>E25</sup>; and Torgerson et al, 2011.<sup>E26</sup>

†Including the Illumina HumanCNV370-Duo as it was built on the Illumina HumanHap300 BeadChip (317,000 SNPs) with an additional 52,000 markers aimed at detecting copy number variations.

‡Including the Illumina Human610 because it is similar to the Illumina HumanHap550 with the addition of 60,000 markers to target regions with copy number variations.



#### 4.4. Asociación de variantes de ocho de los genes candidatos más firmemente replicados en asma en la población española

En este último artículo se presentan los resultados del estudio de asociación de ocho de los genes candidatos más robustos de susceptibilidad al asma. En este trabajo, un total de 607 muestras de casos GOA y 1.271 controles fueron genotipadas para 75 tSNPs seleccionados de modo que asegurasen una cobertura adecuada de los genes *ADAM33*, *ADRB2*, *CD14*, *MS4A2*, *IL13*, *IL4*, *IL4R*, y *LTA-TNF*. Estos tSNPs permitieron además imputar 119 variantes en estos genes para su análisis en asociación con asma. Los resultados mostraron la replicación de muchos de los polimorfismos descritos en la bibliografía, pertenecientes a los genes *IL13-IL4*, *LTA-TNF*, *MS4A2*, *IL4R* y *ADAM33*, a la vez que ampliaron los hallazgos de asociación a nuevas variantes génicas. Asimismo, en este trabajo se muestra la importancia de tener en cuenta la edad de comienzo del asma en los estudios de asociación para mejorar la replicación tanto en términos de número de SNPs asociados, como de sus efectos en la enfermedad. Finalmente, se encontró que la acumulación de variantes de riesgo en varios genes se relaciona con la susceptibilidad al asma y con un comienzo más temprano de la enfermedad.

Este artículo ha sido preparado para su publicación y se presenta a continuación en forma de manuscrito bajo el título “Assessing the validity of asthma associations for the ‘elite’ non-*HLA* candidate genes and age-at-onset effects”.

---

## Assessing the validity of asthma associations for the 'elite' non-*HLA* candidate genes and age-at-onset effects

María Pino-Yanes<sup>1,2</sup>, Almudena Corrales<sup>1,2</sup>, José Cumplido<sup>3</sup>, Paloma Poza<sup>4</sup>, Inmaculada Sánchez-Machín<sup>4</sup>, Anselmo Sánchez-Palacios<sup>5</sup>, Javier Figueroa<sup>5</sup>, Orlando Acosta Fernández<sup>6</sup>, María José Torres-Galván<sup>7</sup>, José Carlos García-Robaina<sup>4</sup>, Mariano Hernández<sup>8</sup>, Jesús Villar<sup>1,7,9</sup>, Teresa Carrillo<sup>3</sup>, Carlos Flores<sup>1,2#</sup>

### From:

<sup>1</sup>*CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Spain;* <sup>2</sup>*Research Unit, Hospital Universitario NS de Candelaria, Tenerife, Spain;* <sup>3</sup>*Allergy Unit, Hospital Universitario Dr. Negrin, Gran Canaria, Spain;* <sup>4</sup>*Allergy Unit, Hospital Universitario NS de Candelaria, Tenerife, Spain;* <sup>5</sup>*Allergy Unit, Hospital Universitario Insular de Gran Canaria, Gran Canaria, Spain;* <sup>6</sup>*Neumology Unit, Hospital Universitario de Canarias, Tenerife, Spain;* <sup>7</sup>*Research Unit, Hospital Universitario Dr. Negrin, Gran Canaria, Spain;* <sup>8</sup>*Genetics Laboratory, Instituto Universitario de Enfermedades Tropicales y Salud Pública de Canarias, Universidad de La Laguna, Tenerife, Spain;* <sup>9</sup>*Keenan Research Center, St. Michael's Hospital, Toronto, Canada.*

#Address for correspondence:

Carlos Flores, PhD  
Unidad de Investigación  
Hospital Universitario N.S. de Candelaria  
Carretera del Rosario s/n  
38010 Santa Cruz de Tenerife  
Phone: (+34) 922-602938  
Fax: (+34) 922-600562  
e-mail: cflores@ull.es

**Running title:** Age at diagnosis effects in eight asthma candidate genes

### Abstract

Before the advent of genome-wide association studies (GWAS), a massive number of independent candidate gene associations suggested that *ADAM33*, *ADRB2*, *CD14*, *MS4A2*, *IL13*, *IL4*, *IL4R*, and *TNF* constituted the most replicated non-*HLA* genes with asthma and related traits. Except for the *IL13-IL4* region, none of these genes have been found in close proximity of significant hits among the published GWAS for asthma or related traits to date. Here we aimed to assess the reproducibility of associations with asthma and atopic asthma in an independent sample, and to test if replications improved considering the effect of the age at onset of the disease. We systematically evaluated common single nucleotide polymorphisms (SNPs) of these 8 genes (75 by genotyping and 119 by imputation, including previously associated variants) in a sample of 1865 unrelated Spanish participants (606 asthmatics and 1259 controls). We found nominal associations with asthma or atopic asthma for 38 SNPs in the *IL13-IL4* and *TNF* regions, *IL4R*, *ADAM33*, and *MS4A2*. Replicated associations demonstrated varying effects with the age-at-diagnosis, allowing selecting an age cutoff for each SNP that showed the largest effects. A re-analysis of the associations based on these cutoffs provided further evidences of replication, in terms of the number of SNPs and/or increases in effect sizes, particularly for *MS4A2*, *IL4R* and *ADAM33*. In addition, the accumulation of risk variants was strongly associated with asthma susceptibility, providing an OR up to 1.39 (95% CI=1.22-1.60,  $p=1.67E-06$ ) for each risk allele, an OR up to 3.45 (95% CI=1.62-7.35) for carriers of  $\geq 6$  risk alleles, and was inversely associated with the age-at-diagnosis of asthma ( $p=3.0E-04$ ). Taken together, our results support the important role of these genes in disease susceptibility and bolster the hypothesis that an earlier onset of asthma correlates with the accumulation of risk variants.

**Word count:** 293

**Keywords:** allergy, genetic susceptibility, association study, polymorphism

## Introduction

Asthma (OMIM #600807) is a complex respiratory disease characterized by chronic inflammation of airways and frequently associated with atopy, high eosinophil count, pulmonary obstruction and bronchial hyper-responsiveness against a diversity of stimulus [1]. Its prevalence varies widely around the world (1-18%), and predictions of increasing are expected for Western societies in the coming years [2]. Familiar clustering [3], twin studies [4], and genetic studies [5,6] support an important genetic component of the disease, with an estimated heritability of 60% [7].

Before the advent of genome-wide association studies (GWAS) [8], a number of independent candidate-gene association studies were performed aiming to find susceptibility genes for asthma and related traits, resulting in almost a thousand publications so far [9]. However, given the conflicting results accumulated by these studies, replication of associations in independent study populations is required in order to assess the consistency of findings; the greater number of independent replications, the higher confidence to support that a particular gene is an asthma susceptibility gene [10]. Considering the gene as the unit of replication and using a broad definition for asthma, Ober & Hoffjan [5] summarized elegantly the accumulated evidence for candidate-gene association studies from the literature. This yielded a ranking of genes based on the number of studies for which a positive association between any polymorphism and any asthma trait was found [5]. As a result, eight non-*HLA* genes were put forward among the most replicated (in >10 populations) and, therefore, suggested as firm candidates for asthma susceptibility [5]. Four of these genes were located in the linked region for asthma on chromosome 5q: interleukin (IL) 4 (*IL4*), *IL13*, *CD14* and the  $\beta$ 2-adrenergic receptor (*ADRB2*); one gene was located in the linked region 6p21: the tumor necrosis factor (*TNF*); another was the first positionally cloned asthma gene, ADAM metallopeptidase domain 33 (*ADAM33*); another was the gene encoding the  $\alpha$  chain of the IL-4 and IL-13 receptors (*IL4R*); and finally, the gene encoding the IgE Fc receptor beta-subunit (*MS4A2*, alias *FCER1B*). However, most of them have not been systematically analyzed by surveying more than few variants (for example, by using tagging SNPs [tSNPs]) in additional powered and well-characterized study populations. Note that the majority of previous studies of these genes were assessed with limited sample sizes averaging  $\approx$ 200 individuals for each study group [5,11].

To date, despite the fact that more than ten GWAS of asthma have been published discovering novel susceptibility disease *loci*, none of these firm candidates have been replicated at genome wide significance, nor have been found in close proximity of GWAS hits, except for the *IL13-IL4* region [8,12,13,14,15,16,17,18,19,20,21]. Therefore, the question of whether GWAS have failed to evidence them because of study limitations or the

## RESULTADOS

---

necessity of improved phenotype definitions [22], or alternatively, if the associations at these genes only represent false positives, still remains to be solved. In this respect, asthma is clinically recognized as an amalgam of several distinct phenotypes [23], which contributes to blur the complex genetic architecture underlying asthma susceptibility. Among these phenotypes, the age-at-onset of asthma could differentiate asthmatic groups, so that genetic variants might inconsistently associate with childhood and later-onset disease [17]. However, few studies have investigated the varying effects of particular genetic risk factors by the age-at-onset of asthma, demonstrating the strongest effects for earlier onset and almost no effects for later onset [17,24,25]. Motivated by this evidence, here we aimed to assess the reliability of asthma associations for the eight most replicated non-*HLA* asthma genes existing before the availability of GWAS, by means of a systematic assessment in a sufficiently large study population. In addition, we explored whether effects of risk alleles varied with the age-at-onset of the disease and if this fact allowed improving the replication of risk variants of these genes. Finally, we also ascertained the extent to which the independently associated risk variants allowed distinguishing the disease status.

---

## Materials and Methods

### ***Ethics statement***

This study was approved by the Ethics Committees at participating hospitals, and written informed consent was obtained from all subjects.

### ***Study subjects***

This study was conducted using a case-control design of 1878 DNA samples from unrelated individuals, all reporting at least two generations of Spanish descent. Sample details have been described elsewhere [24]. In brief, cases included 607 asthmatic patients aged >5 years and diagnosed by physicians following the Global Initiative for Asthma (GINA) guidelines for diagnosis and classification of asthma (<http://www.ginasthma.com>). These samples were collected and characterized for allergic and asthmatic symptoms in Respiratory Medicine and Allergy Departments, as part of the Genetics of Asthma study (GOA) in the Spanish population. Control group consisted of 1271 DNA samples from adults self-reporting no personal or familiar medical history of allergic or pulmonary diseases. These were collected from branches of the National Blood Service from unrelated individuals residing in Spain (visit <http://www.bancoadn.org> for further information). Further sample details can be found in Table S1.

### ***Selection of tagging SNPs***

Tagging SNPs (tSNPs) were selected by means of TagIT 3.03 [26], using available re-sequencing data from European samples from different sources (Tables 1 and S2). It is worth noting that re-sequencing data for the *MS4A2* gene was unavailable from any public database when the study initiated. Thus, for this gene, tSNP selection was done using data from 60 unrelated samples from Utah residents with ancestry from northern and western Europe (CEU) from HapMap phase 2 [27], and tagging performance was verified against re-sequencing data from 96 type 1 diabetes individuals of white European ancestry (v0.1, accessed on 19<sup>th</sup> March 2008) [28]. The *IL13* and *IL4* genes, which lie in close proximity, were considered as a single region for tSNP selection and further analyses. Similarly, given the strong linkage disequilibrium (LD) between *LTA* and *TNF* genes [29], common variants of the *LTA* gene were also tagged and jointly analyzed with *TNF* for association. In all cases, CEU genotypes were also used to complement the set of tSNPs selected from all genes to ensure an adequate gene coverage, not only for the initial source of variation, but also for HapMap phase 2 data. During tSNP selection, we forced the inclusion in the final tSNPs set of previously associated SNPs and all potential functional variants as predicted by PupaSuite [30]. This resulted in a final list of 82 tSNPs that were further tested for tagging performance

## RESULTADOS

---

using a SNP-dropping-with-resampling method [26], satisfying a haplotype  $r^2 \geq 0.85$  in the reference genotype data. See Table S2 for further details.

### **Assessment of population stratification**

To reduce the risk for false positives due to major population stratification effects, a total of 83 European ancestry informative markers (termed EuroAIMs) were determined in case and control samples, allowing to correct for major differences in Spanish populations [31]. A principal component analysis (PCA) based on these genetic markers was utilized to derive the ancestry estimates in cases and controls as scores of the first principal component (PC1). This analysis was performed by means of EIGENSOFT [32]. A full list of EuroAIMs utilized and the genotyping procedures have been detailed elsewhere [24,31].

### **Genotyping**

DNA was extracted from blood (GFX kit, GE Healthcare, Little Chalfont, UK) or saliva (Oragene™ DNA, DNA Genotek Inc., Ontario, Canada) and was used for whole genome amplification with Illustra *GenomiPhi* V2 DNA Amplification Kit (GE Healthcare). Dilutions of these products were used in a SYBR Green I (Molecular Probes, Eugene, OR)-based DNA quantification method, as described elsewhere [33].

Genotyping was conducted using the iPLEX® Gold assay on MassARRAY® system (Sequenom Inc., San Diego, CA) by the Spanish National Genotyping Center, Santiago de Compostela Node (CeGen, <http://www.cegen.org>). Briefly, iPLEX® assays were scanned by MALDI-TOF mass spectrometry and individual SNP genotype calls were automatically generated using Sequenom TYPER 3.4® software (Sequenom Inc.). Samples from the Coriell Institute for Medical Research (<http://www.coriell.org>) were included on each SpectroCHIP® (Sequenom Inc.) to test allele calling reliability samples of this platform. The SNPs that gave poor quality data on this platform were finally determined at the Hospital Universitario N. S. de Candelaria using either SNaPshot® Multiplex kit reactions (Applied Biosystems, Foster City, CA) or KASPar SNP Genotyping System assays (KBiosciences, Hertfordshire, UK). For that, we alternatively utilized an ABI 3500 Genetic Analyzer (Applied Biosystems), automatically assigning genotypes using GeneMapper v3.2 (Applied Biosystems), or a 7500 Fast Real-Time PCR System (Applied Biosystems), with automated calls generated using the 7500 software 2.0.1 based on discriminating plots with 95% confidence, respectively. Genotyping was blind to the disease status and approximately 6% of the samples were genotyped in duplicate to monitor genotyping quality.

Contrary to the dbSNPs database, that defined the SNP rs570269 in *ADAM33* gene as a C/G change, this SNP was detected as a triallelic *locus* during genotyping with iPLEX® Gold assays. This result was confirmed by sequencing of individuals showing the new T

allele (present in 0.4% of our controls). For that purpose, sequencing was performed by both strands with BigDye® v3.1 Terminator Cycle Sequencing kit (Applied Biosystems) under the manufacturer's recommendations and using the amplification primers. Sequencing products were ethanol precipitated and run on an ABI 3500 Genetic Analyzer. However, for the purposes of association, no distinction was made between T and C alleles during the statistical analysis.

### **Statistical analysis**

Clinical and demographical data were analyzed by means of the  $\chi^2$ -test and the Mann-Whitney U-test using SPSS 15.0 (SPSS Inc., Chicago, IL). Departures from Hardy-Weinberg equilibrium (HWE) were evaluated separately for cases and controls using an exact test [34] by means of a custom script for STATISTICA (StatSoft Inc., Tulsa, OK) [35]. However, as deviations in cases have been considered a symptom of disease association [36,37,38], only those tSNPs deviating significantly from HWE in the control group were filtered out from further analyses (threshold for significance was  $p$ -value=7.0E-04 after considering the multiple comparisons performed). Individual tSNP associations were tested under an additive model by means of regression analysis with SNPAssoc [39], allowing to adjust associations for population stratification, by incorporating the PC1 scores as a covariate in regression models, and to estimate allele effects as odds ratios (ORs) with 95% confidence intervals (CIs). The independence of tSNP associations within each gene was then tested by means conditional regression analyses using SPSS 15.0. Additionally, untyped SNPs were imputed from tSNPs and tested for association by means of TUNA ver. 1.1 [40]. For this aim, re-sequencing and/or CEU data were used to reconstruct haplotypes with PHASE 2.1 [41] to be used as a reference for the underlying LD across the gene. Based on previous studies [33], we chose to test the association of those imputed SNPs showing  $MAF \geq 10\%$ , and  $M_r \geq 0.7$  [42], which added an extra level of confidence on the association results, albeit at the cost of a trivial reduction in the number of common SNPs tested for association.

In order to test if replication improved considering age-at-onset-varying effects, we implemented a sequential addition (SA) of cases for the genotyped SNPs that were previously associated and showed a nominal and independent association at  $p \leq 0.05$  in this study (irrespective of the allele or direction of effect), to obtain the age cutoff at which the allele effects were largest [43]. For that, age-at-diagnosis was utilized as a proxy for the age-at-onset of the disease, which was not recorded for most patients. Briefly, in this process, individual SNP associations were computed iteratively using a Cochran-Armitage trend test assuming an additive model for each group of cases resulting from the sequential incorporation of asthmatic individuals for every increase in 1 year in the age of diagnosis.



## RESULTADOS

---

The empirical  $p$ -value adjusted for the multiple comparisons was then obtained as the proportion of permutation replicates, calculated from reshuffling labels of cases and controls 1000 times among the whole sample, which yielded  $p$ -values smaller than the observed in the actual data set. We next used this age-at-diagnosis cutoff to select a sub-sample of cases for which associations were tested again as done before, both for tSNPs and for imputed SNPs. Thus, the age-at-diagnosis cutoff (and therefore the selected sub-sample of cases) was identical for all tests of a given gene. LD patterns and regional association results were represented using LocusZoom 1.1 [44].

To judge the SNP associations in the context of the multiple comparisons performed, a false discovery rate (FDR) was assessed considering the gene as the unit of replication using QVALUE [45]. A FDR threshold of 5% was established to limit the expected proportion of false positives incurred in the study when a particular individual SNP test was called significant, considering all SNPs (genotyped and imputed) and the three group comparisons assessed per gene.

Finally, in order to determine the risk for asthma by the accumulation of risk alleles from independent SNPs at different genes, a genetic risk score (GRS) was calculated by summing risk alleles contained in each SNP genotype (i.e. 0, 1, or 2). For this, SNPs were equally weighted and a trend test was assessed to evaluate the linearity between the number of risk alleles and the risk of asthma. In order to verify if these estimates were robust to the presence of population stratification, we also assessed the GRS for asthma development by means of regression models incorporating the PC1 scores as a covariate. Based on this sum score, we also calculated the Receiver Operating Characteristic curve (ROC), and its area under the curve (AUC) as the indicator of disease prediction. Nagelkerke's  $R^2$  was calculated to measure the explained variance. All these analyses were assessed using SPSS 15.0.

## Results

A total of 13 samples (1 case and 12 controls) were excluded from the analyses because of genotype quality (completion rate <90%). Out of the initial set of 82 tSNPs, 6 were found monomorphic (rs5744440, rs35684381, rs597040, rs8124875, rs614971 and rs17548816) by using both iPLEX<sup>®</sup> and one alternative genotyping method. Only one tSNP (rs12361312 at *MS4A2*) deviated significantly from HWE expectations in the control group and was discarded for the association analyses (Table S2). Therefore, a total of 75 tSNPs, which maintained an adequate coverage for all genes ( $r^2 \geq 0.85$ ), and 119 imputed SNPs were used for association studies in 1865 samples (606 cases and 1259 controls). The mean completion rate among the 75 tSNPs was 98.5% ( $P_{25}-P_{75}=98.7-100.0\%$ ), and the estimated overall genotype discordance rate among duplicated samples was 0.30% (95% CI=0.08%-1.09%).

Association testing revealed a total of 16 tSNPs, for which an adjustment for population stratification was possible, and 21 imputed SNPs which were associated with either asthma or atopic asthma at nominal significance (Table S3). Asthma associations were distributed in *LTA-TNF*, *MS4A2*, *ADAM33*, and *IL4R* ( $0.002 \leq p\text{-value} \leq 0.047$ ), while associations with atopic asthma distributed in *IL13-IL4*, *LTA-TNF*, *MS4A2*, and *IL4R* ( $1.0E-04 \leq p\text{-value} \leq 0.049$ ). Among these, 18 (47%) constituted SNP-level replications of a positive finding in at least one previous study sample regardless of the allele or the direction of the effect, 14 of them were found associated with asthma and 17 with atopic asthma, being distributed among *IL13-IL4*, *LTA-TNF*, *MS4A2*, *ADAM33*, and *IL4R* genes. Conditional logistic regressions including the tSNPs that constituted SNP-level replications of a positive finding within each gene, consequently precluding the analysis of *ADAM33*, evidenced independent associations with asthma or atopic asthma for rs1800925 (-1112 C/T) in *IL13-IL4* (only for atopic asthma), rs2071590 in *LTA-TNF* (only for asthma), rs569108 (Gly237Glu) in *MS4A2*, and rs1805015 (Ser478Pro) in *IL4R* (both for asthma and atopic asthma) (data not shown). None of the SNPs from *CD14* and *ADRB2* genes showed a nominal significant association with either asthma or atopic asthma.

With the premise that the age-at-onset might increase the power to detect association by assisting the enrichment of the case sample for patients with larger allele effects [24,25], we used SA of asthma patients to estimate the age-at-diagnosis cutoff maximizing allele effects. This was initially done only for the 4 tSNPs that replicated previous findings and showed independent associations on each gene. SA did not show any age-at-diagnosis cutoff that significantly maximized the association with atopic asthma for rs1800925 in *IL13-IL4* (lowest  $p\text{-perm}=0.063$ ). On the contrary, SA revealed allele effects peaking at different age-at-diagnosis of the disease: rs2071590 in *LTA-TNF* showing a maximum at 24 years ( $p\text{-}$

## RESULTADOS

---

$p_{perm}=0.003$ ), rs569108 in *MS4A2* showing a maximum at 36 years ( $p_{perm}=0.009$ ), and rs1805015 in *IL4R* showing a maximum at 38 years ( $p_{perm}=0.002$ ) (Figure 1). Testing associations on the sub-sample of cases with the age-at-diagnosis of asthma before the maxima determined by SA, revealed 4 additional SNPs associated at nominal significance (Table S3), all of them being tSNPs of *IL4R* but none constituted SNP-level replications of previous studies. However, this evaluation evidenced increases in allele effect sizes along with the improvement of association  $p$ -values, particularly for most of the SNP-level replications of *LTA-TNF* and *IL4R* genes, and the only one that remained associated in *MS4A2* (Table S3, and Figure 2).

Given the improvements of using SA of cases to determine age-at-diagnosis cutoffs before testing the associations, we speculated whether SA would be capable to increase the replication in *ADAM33*, a gene for which SNP-level replications are difficult to be found in the literature [5,46]. Applying SA to rs2787095, the tSNP showing the lowest nominal significant and independent association with asthma in *ADAM33*, revealed allele effects peaking at an age-at-diagnosis of the disease at 13 years ( $p_{perm}=0.002$ ) (Figure 3). Testing associations on the sub-sample of cases determined by this maximum of age-at-diagnosis of asthma revealed 9 additional SNPs associated at nominal significance ( $4.9E-05 \leq p \leq 0.020$ ), 5 of these constituting SNP-level replications of a positive finding in at least one previous study (Table S3, and Figure 3). In addition, as rs2787095 was weakly correlated with all those significant SNPs that were genotyped ( $0.032 \leq r^2 \leq 0.275$ ), evidence for more than one asthma-associated variant within *ADAM33* was obtained. The inclusion of rs2787095 and the alternative SNP showing the strongest evidence for association (rs628977;  $r^2=0.17$  with rs2787095) as covariates in a logistic regression still left a significant association for both SNPs ( $OR_{rs628977}=1.42$ , 95% CI=1.09-1.86,  $p=0.009$ ;  $OR_{rs2787095}=1.86$ , 95% CI=1.22-2.06,  $p=0.001$ ). Therefore, the advantages of taking into account the age-at-onset varying effects for replication studies in asthma were clearly evidenced in *ADAM33*, as we would have missed >55% SNP-level replications of this gene.

After all tests done for asthma, atopic asthma, and asthma with an age-at-diagnosis before the cutoff maximizing allele effects, a total of 51 SNPs were associated at nominal significance. That is 26% of all SNPs tested, 23 of them (45%) constituting SNP-level replications of a positive finding in at least one previous study sample, most of them being also significant in the context of the multiple tests performed for each gene (46 SNPs with an  $FDR \leq 5\%$ ). The exceptions were: the 4 nominally associated SNPs with atopic asthma from *IL13-IL4* (the minimum FDR was 48% for rs1295685, which constituted a SNP-level replication of previous studies); and rs2583477 in *MS4A2* (minimum FDR of 43%). The top SNPs per gene were: rs2891057 and rs1805015 in *IL4R* (both with  $FDR=0.01\%$ ), rs628977

and rs2787095 in *ADAM33* (both with FDR=0.2%), rs569108 in *MS4A2* (FDR=0.6%), and rs2071590 in *LTA-TNF* (2.5%).

Finally, the associated tSNPs that showed independent effects with asthma or atopic asthma for each gene were used for risk prediction in the same samples by assessing a GRS. For asthma, GRS included four SNPs: rs2071590 from the *LTA-TNF* region, rs2787095 from *ADAM33*, rs569108 from *MS4A2*, and rs1805015 from *IL4R* gene. The cumulative addition of each risk allele resulted in a significant addition of risk for asthma susceptibility (OR=1.23, 95% CI=1.12-1.34 for the addition of each risk allele,  $p=7.3E-07$  [OR=1.22, 95% CI=1.11-1.33,  $p=1.5E-05$ , after adjusting for population stratification]), explaining up to an OR=3.45 (95% CI=1.62-7.35) for carriers of  $\geq 6$  risk alleles (see the case-control frequency change in Figure S1), an AUC of 0.56 (95% CI=0.53-0.59,  $p=3.3E-05$ ) (Figure S2), and accounting for 1.7% of genetic variance. For comparison, similarly built GRSs from 5 random sets of 4 alternative tSNPs showed an average AUC for asthma of 0.52 (95% CI=0.49-0.55).

As for asthma, a GRS of atopic asthma was calculated considering: rs1800925 from *IL13-IL4* region, rs1800610 from the *LTA-TNF* region, rs569108 from *MS4A2*, and rs1805015 from *IL4R*. In this case, the cumulative addition of each risk allele resulted in a significant addition of risk for atopic asthma susceptibility (OR=1.39, 95% CI=1.22-1.60,  $p=1.7E-07$  [OR=1.40, 95% CI=1.22-1.60,  $p=1.7E-06$ , after adjusting for population stratification]), adding up to an OR=2.47 (95% CI=1.44-4.21) for carriers of  $\geq 6$  risk alleles (Figure S1), an AUC of 0.58 (95% CI=0.55-0.62,  $p=2.9E-06$ ) (Figure S2), and explaining a genetic variance of 2.3%. However, the similarly built GRSs from 5 random sets of 4 alternative tSNPs showed an average AUC for atopic asthma of 0.52 (95% CI=0.49-0.56), although where not significant. Moreover, among patients, the GRS was inversely associated with the age-at-diagnosis of asthma ( $p=3.0E-04$ ) (Figure 4), although not the GRS of atopic asthma ( $p=0.106$ ). In agreement, the differences on the average count of risk alleles against controls (5.61) were larger when compared to that of asthma cases diagnosed by a pre-specified age of 23 years (6.08, Mann-Whitney U-test  $p=1.6E-07$ ), than when compared to that of asthma cases diagnosed at all ages (5.89, Mann-Whitney U-test  $p=1.7E-05$ ).

### Discussion

In this study, we have comprehensively analyzed the association of 194 common variants, including previously associated SNPs, of eight candidate genes with asthma and atopic asthma in a reasonably well-sized case-control Spanish sample and found nominal associations in 5 of these genes (*IL13-IL4*, *IL4R*, *ADAM33*, *MS4A2*, and *LTA-TNF*). While SNPs from *IL4R*, *ADAM33*, *MS4A2*, and *LTA-TNF* genes showed association with asthma and atopic asthma, *IL13-IL4* was found associated only with atopic asthma. However, not a single SNP from *ADRB2* and *CD14* genes was found associated. It is well known that the age-at-onset of asthma is associated with different phenotypic characteristics [47], and it has recently evidenced that age-varying genetic associations can cause non-replication and, consequently, lead to missing important genetic associations [48]. Therefore, we also re-evaluated the association of these genes by restricting the analysis to case subjects with an age-at-diagnosis of asthma before a cutoff that maximized allele effects of replicated variants. This allowed us to verify that replication improved for certain genes, such as *ADAM33*, as recently supported for other firm candidates [25,49,50], and also to gain insight in the genetic complexity of asthma associations at these *loci*.

Moving towards the application of personalized medicine requires first to identify firm susceptibility genes underlying disease risk. For that, replication in independent populations consisting of large and well-characterized samples is essential for establishing the credibility of a genotype-phenotype association, regardless of whether the first evidence of association was provided by a GWAS study or a candidate gene survey [51]. Besides, replication efforts allow testing the generality of findings in other populations, and discovering novel genetic *loci* contributing to phenotypic trait variability [52]. Particularly, testing the associations in populations of recent African ancestry will likely improve the detection of new risk variants [53], as they may offer the opportunity to refine the signal or to allocate the causal variants [54]. Our study line up with these considerations, as it was performed in a population with sizeable African genetic influences [31], and with a sample size representing a substantially larger population of cases (>97%) than the vast majority of prior published studies of these genes that utilized an unrelated case-control design. This sample size provided >80% power to detect a minimum risk of 1.45 for a risk frequency of 20% with a two-sided  $p=0.05$  significance level. Despite that, we acknowledge that risk effects of this range are in the upper bound of those expected for common variants in complex traits [55], which may have contributed to our failure to replicate the association of *ADRB2* and *CD14* genes. Alternatively, our failure to replicate associations at these genes may possibly be attributed to: i) Our impossibility to test their association with more relevant traits or patient sub-samples, as *ADRB2* has been related to asthma drug responses [56,57] and *CD14* has

shown association once the sample has been stratified by environmental exposures [58]; ii) The lack of a true association with asthma susceptibility, as has been suggested for particularly relevant variants by meta-analyses [56]. Whichever one is correct, a recently published and similarly powered replication study showed positive and negative replication results fully congruent for all genes overlapping with ours [59].

Among the final list of 51 SNPs found associated at nominal significance, 45% constituted SNP-level replications [5,11], of which 90% would be considered significant in the context of the multiple tests done. For some of them, we found evidence of slight but significant increases of effect sizes among cases diagnosed before certain ages, so that some effects were observed to increase within asthmatic group with early diagnosis (SNPs at *LTA-TNF* and *ADAM33* genes), while for others (*IL4R* and *MS4A2*) the effects were also observed after including cases with a latter diagnosis. Intriguingly, many of the effects peaked in the range of age at diagnosis between 20 and 45 years, coincidental with the age range with the maximum expression of the disease [60,61]. Whether a true biological mechanism is underlying these and previous observations [17,25,49] will be worth noting to be considered in further studies, because of its potential impact in the design of independent association studies [48], as well as in the research of improved asthma treatment and prevention.

Our percentage of replication at gene level is in the upper bound of those reported by previous large and recent replication studies, where a range of 9-26% of the genes analyzed showed replication with an asthma trait [11,54,59,62,63,64,65]. This might have been motivated by the different stringency of criteria for selecting the candidates for the analyses. Note that our study focused on candidate genes showing at least one SNP associated in >10 study populations, while previous studies required associations in fewer independent study populations, location of candidates under linkage peaks or evidences from studies with knockout animal models. In addition to that, many of these studies have utilized available arrays for genome wide genotyping, where common variants of many key asthma candidate genes were insufficiently covered. For example, Michel et al. [59] indicated that only 37% of the previously associated SNPs from 14 candidate genes were captured by the array utilized by the same authors on the first GWAS for asthma [8] and, surprisingly, not a single SNP from key asthma genes such as *ADAM33*, *IL4* and *CD14* was contained in their array. Only after extending the study by further genotyping (and by imputation) on the same samples of their GWAS, Michel *et al.* were able to consistently replicate many of the biological candidates that were missing from the GWAS [59]. We confirmed that the coverage of published GWAS for asthma to date in European populations has been inadequate for *ADAM33* (<30%) and for *LTA-TNF* region (<62%) if the HapMap phase 2 data is used as a reference (Table S4). If array comparisons were made against the 1000 Genomes Project

## RESULTADOS

---

(1KGP) sequencing data [66], the coverage would be even lower. Note, however, that the estimated coverage of genes was calculated under a best-case scenario, as these were implicitly derived for HapMap CEU data and the same data was used to inform the SNP contents of the array. In addition, we have assumed that the 100% of SNPs contained in the array were successfully genotyped, and that all individuals from the study were determined for the same amount of SNPs (not incurring in power loss because of uncertainty in SNP during the imputation process or even SNP exclusion during a joint analysis of samples genotyped with different arrays). Effects similar to those related to the age-of-onset of asthma, which have been only exceptionally explored [17], could have also contributed to their absence from the significant hits obtained by the published GWAS for asthma or related traits to date. Therefore, albeit most of them have been missed by published GWAS for asthma, the available evidences support that *IL13-IL4*, *IL4R*, *ADAM33*, *MS4A2*, and *LTA-TNF* genes replicate consistently among different studies. Additional support to their genuine role in asthma susceptibility comes from the evidence from their derived multi-SNP GRS scores that, despite showing marginal disease predictability, demonstrated similar AUC values and associated *p*-values to those revealed by the top GWAS hits [17,18,67], indicating that they are effectively different from a random guess (i.e. from an AUC=0.5).

In conclusion, we replicated the association of common variants in five biological candidate genes of asthma giving independent support for their role as risk factors for asthma and atopic asthma. Moreover, our results evidenced the genetic complexity at some of these susceptibility *loci* and the importance of considering age-at-onset effects for the evaluation of genetic risk factors in asthma, both for replication assessment and for the detection of further susceptibility genes. Finally, the observation that the accumulated risk and the explained proportion of genetic risk for asthma increased among patients with earlier age of diagnosis supports the notion that asthma patients with earlier onset of the disease are enriched for the genetic disease [15].

## **Acknowledgements**

The authors want to thank Tobías Felipe for programming, and María Torres (CeGen, Santiago de Compostela Node) for her exceptional technical support with iPLEX<sup>®</sup> Gold assays. In addition, we would like to express our gratitude to the Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory (JDRF/WT DIL, Cambridge Institute for Medical Research) for granting the access to the *MS4A2* re-sequencing data, and acknowledge that those who carried out the original analysis and collection of the data bear no responsibility for the analysis or interpretations of this study. This work was supported by grants from the Ministry of Science and Innovation of Spain (PI081383) and FUNCIS (27/07), and by a specific agreement between Instituto de Salud Carlos III and Gobierno de Canarias (EMER07/001).

## **Conflict of Interest Statement**

The authors declare having no conflicts of interest.



### References

1. Ober C (2005) Perspectives on the past decade of asthma genetics. *J Allergy Clin Immunol* 116: 274-278.
2. Braman SS (2006) The global burden of asthma. *Chest* 130: 4S-12S.
3. Holberg CJ, Halonen M, Wright AL, Martinez FD (1999) Familial aggregation and segregation analysis of eosinophil levels. *Am J Respir Crit Care Med* 160: 1604-1610.
4. Thomsen SF, van der Sluis S, Kyvik KO, Skytthe A, Backer V (2010) Estimates of asthma heritability in a large twin sample. *Clin Exp Allergy* 40: 1054-1061.
5. Ober C, Hoffjan S (2006) Asthma genetics 2006: the long and winding road to gene discovery. *Genes Immun* 7: 95-100.
6. Bouzigon E, Forabosco P, Koppelman GH, Cookson WO, Dizier MH, et al. (2010) Meta-analysis of 20 genome-wide linkage studies evidenced new regions linked to asthma and atopy. *Eur J Hum Genet* 18: 700-706.
7. Duffy DL, Martin NG, Battistutta D, Hopper JL, Mathews JD (1990) Genetics of asthma and hay fever in Australian twins. *Am Rev Respir Dis* 142: 1351-1358.
8. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, et al. (2007) Genetic variants regulating *ORMDL3* expression contribute to the risk of childhood asthma. *Nature* 448: 470-473.
9. Swarr DT, Hakonarson H (2010) Unraveling the complex genetic underpinnings of asthma and allergic disorders. *Curr Opin Allergy Clin Immunol* 10: 434-442.
10. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, et al. (2007) Replicating genotype-phenotype associations. *Nature* 447: 655-660.
11. Daley D, Lemire M, Akhbir L, Chan-Yeung M, He JQ, et al. (2009) Analyses of associations with asthma in four asthma population samples from Canada and Australia. *Hum Genet* 125: 445-459.
12. Hancock DB, Romieu I, Shi M, Sienna-Monge JJ, Wu H, et al. (2009) Genome-wide association study implicates chromosome 9q21.31 as a susceptibility locus for asthma in Mexican children. *PLoS Genet* 5: e1000623.
13. Himes BE, Hunninghake GM, Baurley JW, Rafaels NM, Sleiman P, et al. (2009) Genome-wide association analysis identifies *PDE4D* as an asthma-susceptibility gene. *Am J Hum Genet* 84: 581-593.
14. Mathias RA, Grant AV, Rafaels N, Hand T, Gao L, et al. (2010) A genome-wide association study on African-ancestry populations for asthma. *J Allergy Clin Immunol* 125: 336-346 e334.
15. Sleiman PM, Flory J, Imielinski M, Bradfield JP, Annaiah K, et al. (2010) Variants of *DENND1B* associated with asthma in children. *N Engl J Med* 362: 36-44.
16. Li X, Howard TD, Zheng SL, Haselkorn T, Peters SP, et al. (2010) Genome-wide association study of asthma identifies *RAD50-IL13* and *HLA-DR/DQ* regions. *J Allergy Clin Immunol* 125: 328-335 e311.
17. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, et al. (2010) A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 363: 1211-1221.

18. Ferreira MA, McRae AF, Medland SE, Nyholt DR, Gordon SD, et al. (2011) Association between *ORMDL3*, *IL1RL1* and a deletion on chromosome 17q21 with asthma risk in Australia. *Eur J Hum Genet* 19: 458-464.
19. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, et al. (2011) Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet* 43: 887-892.
20. Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, et al. (2011) Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat Genet* 43: 893-896.
21. Ferreira MA, Matheson MC, Duffy DL, Marks GB, Hui J, et al. (2011) Identification of *IL6R* and chromosome 11q13.5 as risk loci for asthma. *Lancet* 378: 1006-1014.
22. Haldar P, Pavord ID, Shaw DE, Berry MA, Thomas M, et al. (2008) Cluster analysis and clinical asthma phenotypes. *Am J Respir Crit Care Med* 178: 218-224.
23. Henderson J, Granell R, Sterne J (2009) The search for new asthma phenotypes. *Arch Dis Child* 94: 333-336.
24. Pino-Yanes M, Sánchez-Machín I, Cumplido J, Figueroa J, Torres-Galván MJ, et al. (2011) *IRAK3* gene variants associate with asthma in a replication study in the Spanish population. *J Allergy Clin Immunol*, in press.
25. Halapi E, Gudbjartsson DF, Jonsdottir GM, Bjornsdottir US, Thorleifsson G, et al. (2010) A sequence variant on 17q21 is associated with age at onset and severity of asthma. *Eur J Hum Genet* 18: 902-908.
26. Ahmadi KR, Weale ME, Xue ZY, Soranzo N, Yarnall DP, et al. (2005) A single-nucleotide polymorphism tagging set for human drug metabolism and transport. *Nat Genet* 37: 84-89.
27. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449: 851-861.
28. Smyth DJ, Howson JM, Payne F, Maier LM, Bailey R, et al. (2006) Analysis of polymorphisms in 16 genes in type 1 diabetes that have been associated with other immune-mediated diseases. *BMC Med Genet* 7: 20.
29. Belfer I, Buzas B, Hipp H, Dean M, Evans C, et al. (2004) Haplotype structure of inflammatory cytokines genes (*IL1B*, *IL6* and *TNF/LTA*) in US Caucasians and African Americans. *Genes Immun* 5: 505-512.
30. Conde L, Vaquerizas JM, Dopazo H, Arbiza L, Reumers J, et al. (2006) PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes. *Nucleic Acids Res* 34: W621-625.
31. Pino-Yanes M, Corrales A, Basaldúa S, Hernández A, Guerra L, et al. (2011) North African influences and potential bias in case-control association studies in the Spanish population. *PLoS One* 6: e18389.
32. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38: 904-909.
33. Pino-Yanes M, Ma SF, Sun X, Tejera P, Corrales A, et al. (2011) Interleukin-1 receptor-associated kinase 3 gene associates with susceptibility to acute lung injury. *Am J Respir Cell Mol Biol* 45: 740-745.
34. Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet* 76: 887-893.

## RESULTADOS

---

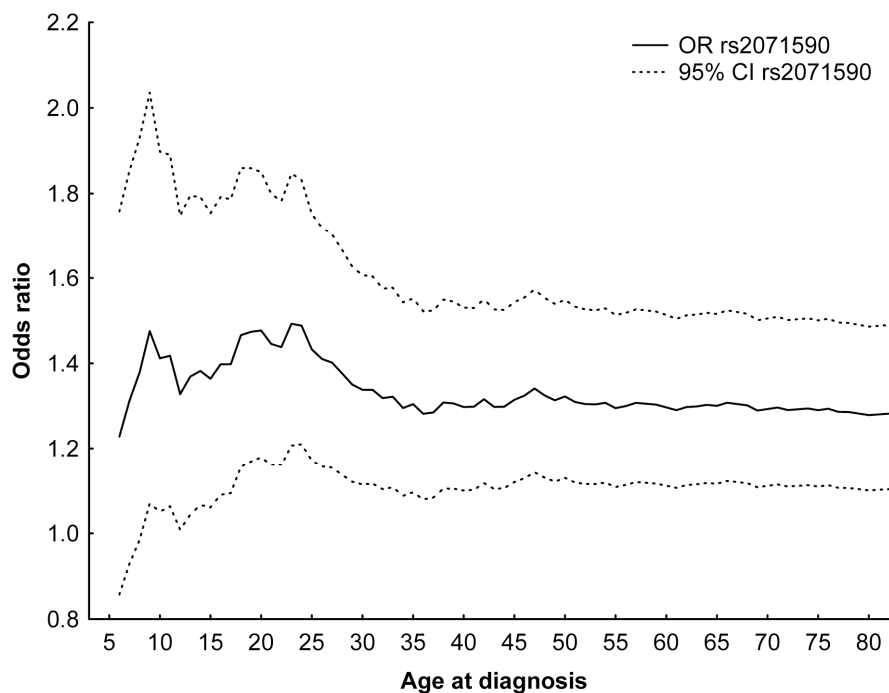
35. Sun X, Ma SF, Wade MS, Flores C, Pino-Yanes M, et al. (2010) Functional variants of sphingosine-1-phosphate receptor 1 gene associate with asthma susceptibility. *J Allergy Clin Immunol* 126: 241-249, 249.e1-3.
36. Chakraborty R, Zhong Y (1994) Statistical power of an exact test of Hardy-Weinberg proportions of genotypic data at a multiallelic locus. *Hum Hered* 44: 1-9.
37. Salanti G, Amountza G, Ntzani EE, Ioannidis JP (2005) Hardy-Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power. *Eur J Hum Genet* 13: 840-848.
38. Balding DJ (2006) A tutorial on statistical methods for population association studies. *Nat Rev Genet* 7: 781-791.
39. Gonzalez JR, Armengol L, Sole X, Guino E, Mercader JM, et al. (2007) SNPAssoc: an R package to perform whole genome association studies. *Bioinformatics* 23: 644-645.
40. Wen X, Nicolae DL (2008) Association studies for untyped markers with TUNA. *Bioinformatics* 24: 435-437.
41. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68: 978-989.
42. Nicolae DL (2006) Testing untyped alleles (TUNA)-applications to genome-wide association studies. *Genet Epidemiol* 30: 718-727.
43. Macgregor S, Craddock N, Holmans PA (2006) Use of phenotypic covariates in association analysis by sequential addition of cases. *Eur J Hum Genet* 14: 529-534.
44. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, et al. (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26: 2336-2337.
45. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 100: 9440-9445.
46. Vercelli D (2008) Discovering susceptibility genes for asthma and allergy. *Nat Rev Immunol* 8: 169-182.
47. Gelfand EW (2008) Is asthma in childhood different from asthma in adults? Why do we need special approaches to asthma in children? *Allergy Asthma Proc* 29: 99-102.
48. Lasky-Su J, Lyon HN, Emilsson V, Heid IM, Molony C, et al. (2008) On the replication of genetic associations: timing can be everything! *Am J Hum Genet* 82: 849-858.
49. Bouzigon E, Corda E, Aschard H, Dizier MH, Boland A, et al. (2008) Effect of 17q21 variants and smoking exposure in early-onset asthma. *N Engl J Med* 359: 1985-1994.
50. Castro-Giner F, de Cid R, Gonzalez JR, Jarvis D, Heinrich J, et al. (2010) Positionally cloned genes and age-specific effects in asthma and atopy: an international population-based cohort study (ECRHS). *Thorax* 65: 124-131.
51. Nicolae DL, Ober C (2009) (Too) great expectations: the challenges in replicating asthma disease genes. *Am J Respir Crit Care Med* 179: 1078-1079.
52. Pulit SL, Voight BF, de Bakker PI (2010) Multiethnic genetic association studies improve power for locus discovery. *PLoS One* 5: e12600.
53. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461: 747-753.

54. Baye TM, Butsch Kovacic M, Biagini Myers JM, Martin LJ, Lindsey M, et al. (2011) Differences in candidate gene association between European ancestry and African American asthmatic children. *PLoS One* 6: e16522.
55. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, et al. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 106: 9362-9367.
56. Contopoulos-Ioannidis DG, Manoli EN, Ioannidis JP (2005) Meta-analysis of the association of beta2-adrenergic receptor polymorphisms with asthma phenotypes. *J Allergy Clin Immunol* 115: 963-972.
57. Finkelstein Y, Bournissen FG, Hutson JR, Shannon M (2009) Polymorphism of the *ADRB2* gene and response to inhaled beta- agonists in children with asthma: a meta-analysis. *J Asthma* 46: 900-905.
58. Simpson A, John SL, Jury F, Niven R, Woodcock A, et al. (2006) Endotoxin exposure, *CD14*, and allergic disease: an interaction between genes and the environment. *Am J Respir Crit Care Med* 174: 386-392.
59. Michel S, Liang L, Depner M, Klopp N, Ruether A, et al. (2010) Unifying candidate gene and GWAS approaches in asthma. *PLoS One* 5: e13894.
60. Barbee RA, Kaltenborn W, Lebowitz MD, Burrows B (1987) Longitudinal changes in allergen skin test reactivity in a community population sample. *J Allergy Clin Immunol* 79: 16-24.
61. Boulet LP, Turcotte H, Laprise C, Lavertu C, Bedard PM, et al. (1997) Comparative degree and type of sensitization to common indoor and outdoor allergens in subjects with allergic rhinitis and/or asthma. *Clin Exp Allergy* 27: 52-59.
62. Ferreira MA, Zhao ZZ, Thomsen SF, James M, Evans DM, et al. (2009) Association and interaction analyses of eight genes under asthma linkage peaks. *Allergy* 64: 1623-1628.
63. Rogers AJ, Raby BA, Lasky-Su JA, Murphy A, Lazarus R, et al. (2009) Assessing the reproducibility of asthma candidate gene associations, using genome-wide data. *Am J Respir Crit Care Med* 179: 1084-1090.
64. Wu H, Romieu I, Shi M, Hancock DB, Li H, et al. (2010) Evaluation of candidate genes in a genome-wide association study of childhood asthma in Mexicans. *J Allergy Clin Immunol* 125: 321-327 e313.
65. Galanter JM, Torgerson D, Gignoux CR, Sen S, Roth LA, et al. (2011) Cosmopolitan and ethnic-specific replication of genetic risk factors for asthma in 2 Latino populations. *J Allergy Clin Immunol* 128: 37-43 e12.
66. Consortium TGP (2010) A map of human genome variation from population-scale sequencing. *Nature* 467: 1061-1073.
67. Xu M, Tantisira KG, Wu A, Litonjua AA, Chu JH, et al. (2011) Genome wide association study to predict severe asthma exacerbations in children using random forests classifiers. *BMC Med Genet* 12: 90.

Figure legends

Figure 1. ORs and its 95% CIs obtained as a result of the sequential addition of cases according to the age-at-diagnosis of asthma for replicated SNPs from A) *LTA-TNF* (rs2071590), B) *MS4A2* (rs569108), and *IL4R* (rs1805015) [Continued on page 161].

A)



B)

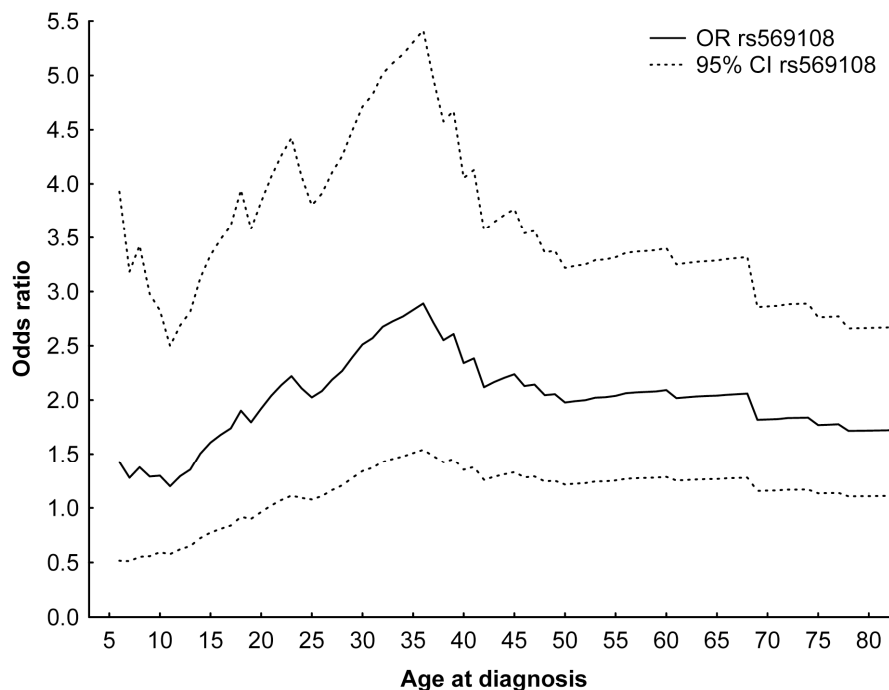
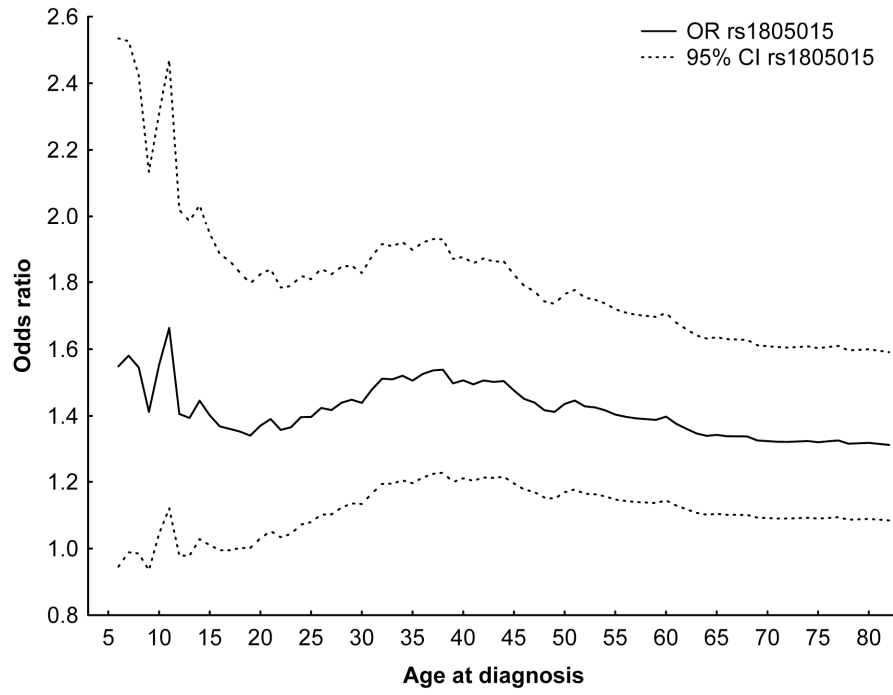


Figure 1. (Continued)

C)



## RESULTADOS

**Figure 2.** *P*-values of association by chromosome position with A) atopic asthma (*IL13-IL4*), and asthma with age-at-diagnosis of the disease before the cutoff of B)  $\leq 23$  years for *LTA-TNF* C)  $\leq 36$  years for *MS4A2* and, D)  $\leq 38$  years for *IL4R*. *P*-values are expressed in  $-\log_{10}$  scale. The SNP number shown on the plot denotes the result for the most significant SNP for each gene and the results for the remaining were color coded to reflect their LD with this SNP based on pairwise  $r^2$  values from the 1KGP. Estimated recombination rates (from 1KGP) were also plotted in cyan to reflect the local LD structure [Continued on pages 163 and 164].

A)

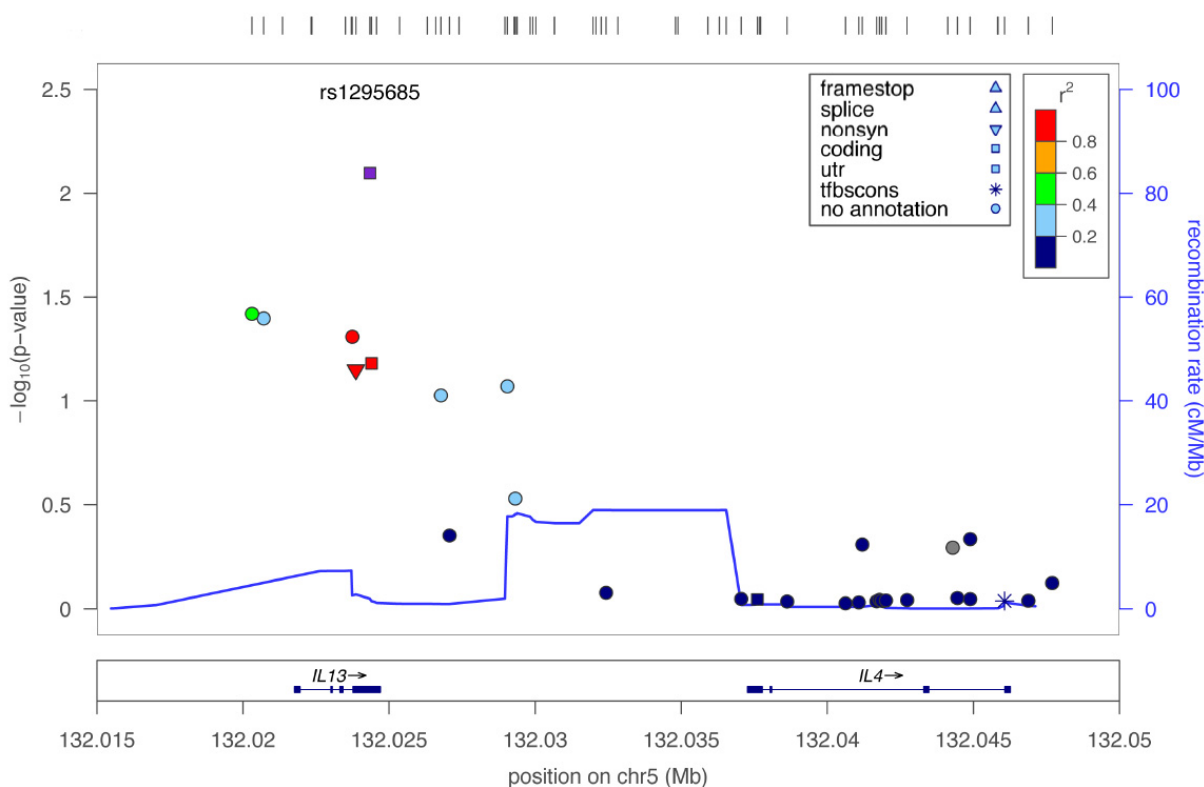
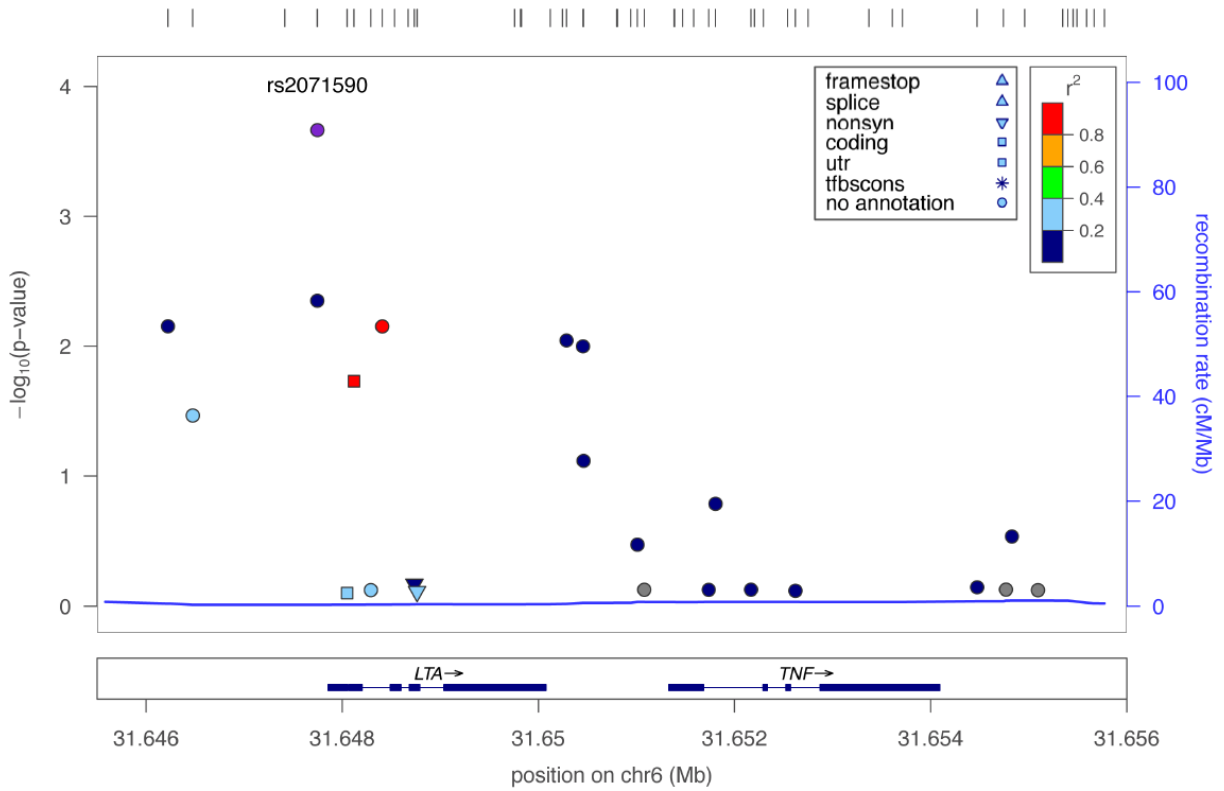
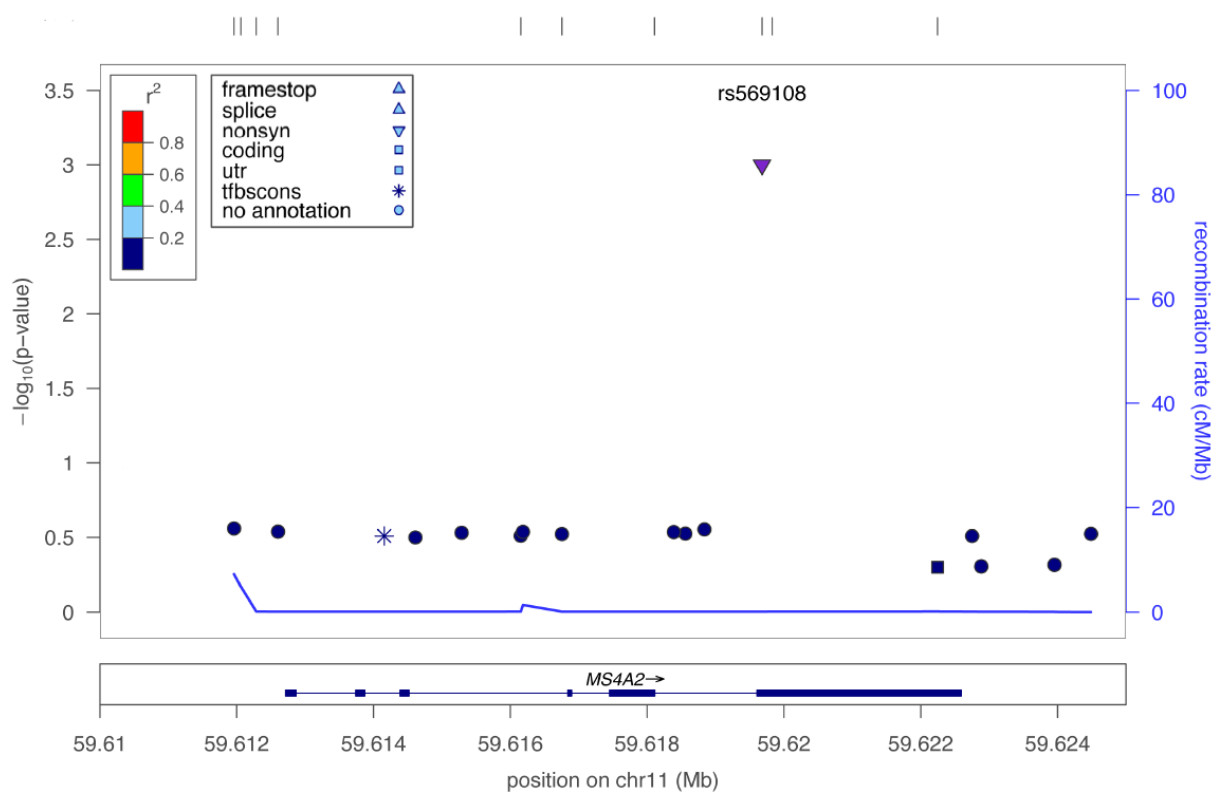


Figure 2. (Continued)

B)



C)

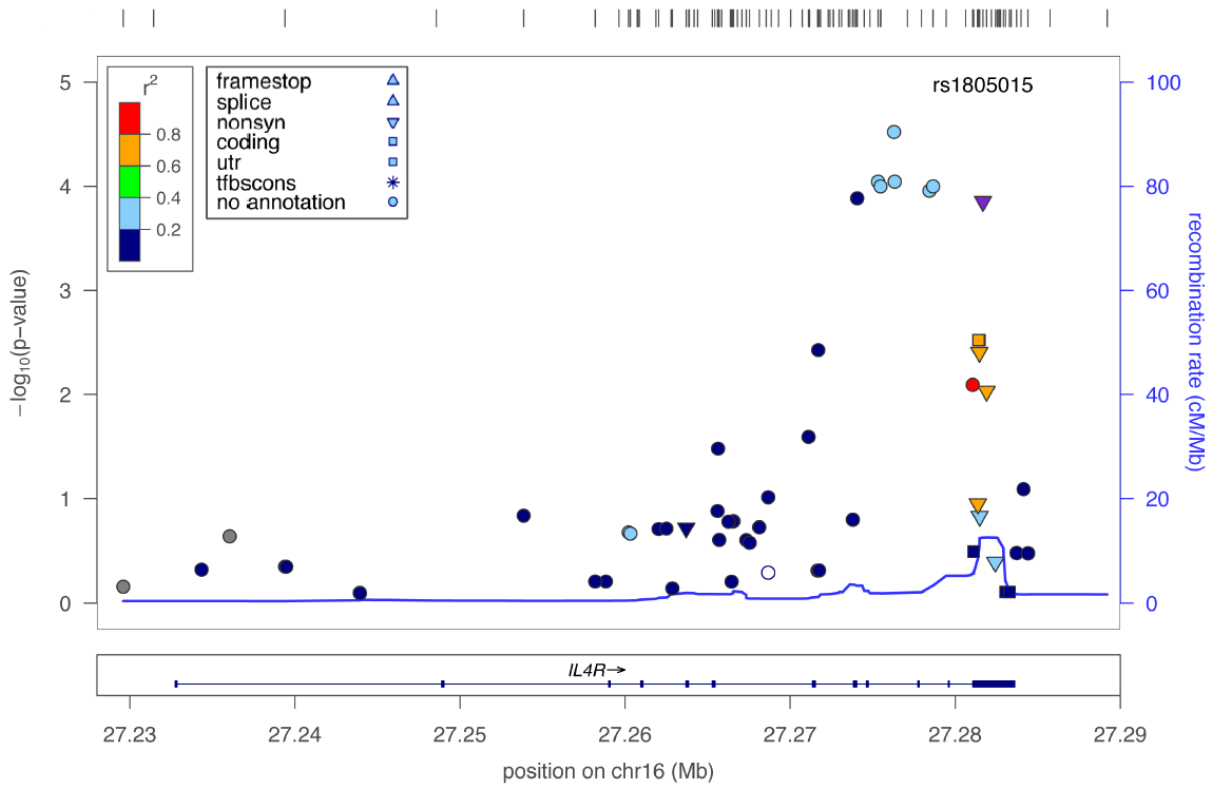




# RESULTADOS

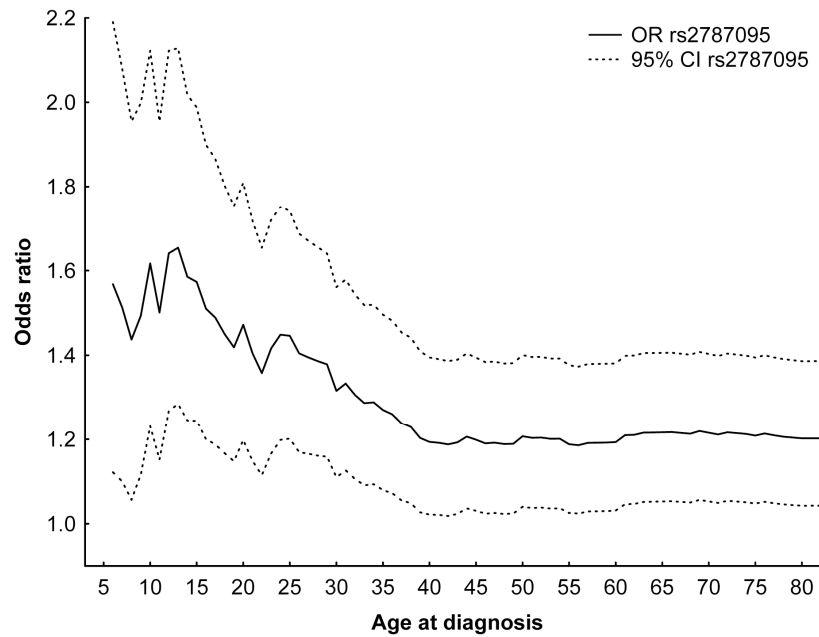
Figure 2. (Continued)

D)

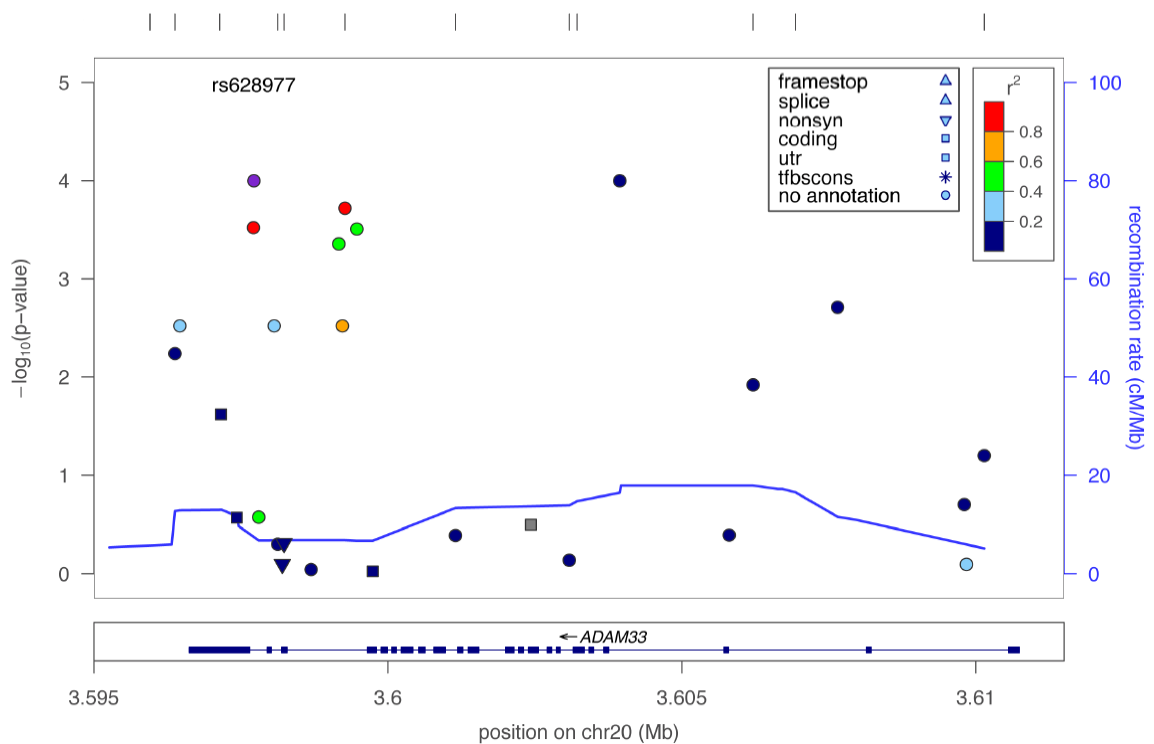


**Figure 3.** Association results for *ADAM33*. A) ORs and its 95% CIs obtained from the sequential addition of cases according to the age-at-diagnosis of the disease for the SNP rs2787095. B) *P*-values of association with asthma diagnosed  $\leq 13$  years (in  $-\log_{10}$  scale) by chromosome position. The SNP number shown on the plot (rs628977) denotes the result for the most significant SNP and the results for the remaining were color coded to reflect their LD with this SNP based on pairwise  $r^2$  values from the 1KGP. Estimated recombination rates (from 1KGP) were also plotted in cyan to reflect the local LD structure.

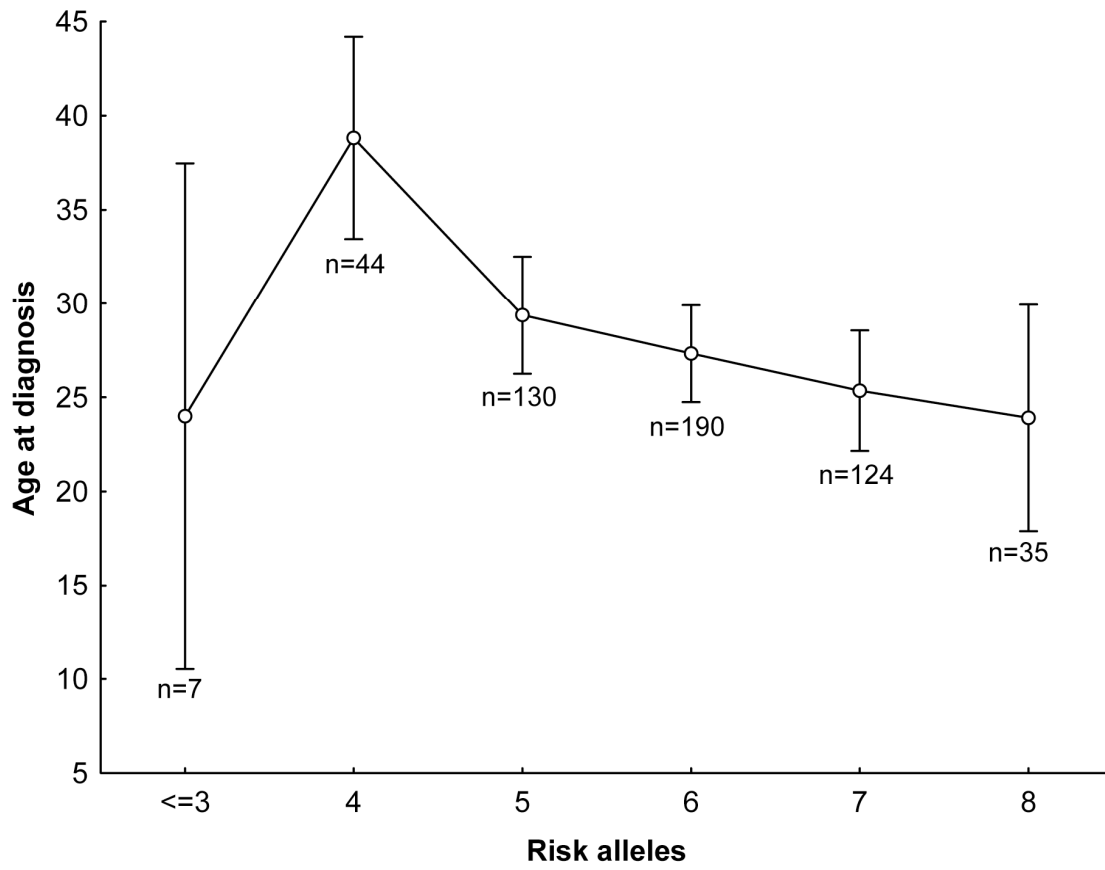
A)



B)



**Figure 4.** Accumulation of risk alleles (GRS) within asthmatics by age-at-diagnosis of the disease (mean±SE)



Tables

Table 1. Summary information used for the selection of tagging SNPs (tSNPs) on the candidate genes.

Gene	Chr. (Mb)	Size (kb)	Covered region (kb)	Data sources	Selected tSNPs	Monomorphic	Final tSNPs	Final haplotype $r^2$
<i>IL13-IL4</i>	5q31.1 (132.0)	12	29.0	SeattleSNPs <sup>a</sup>	10	0	10	1.00
<i>CD14</i>	5q31.1 (140.0)	2	7.0	Innate Immunity <sup>b</sup>	6	1	5	1.00
<i>ADRB2</i>	5q31 (148.2)	2	9.5	SeattleSNPs <sup>a</sup>	8	1	7	1.00
<i>TNF-LTA</i>	6p21.3 (31.5)	6	9.3	SeattleSNPs <sup>a</sup>	11	0	11	0.85
<i>MS4A2</i>	11q13 (59.9)	10	15.3	HapMap/T1D <sup>c</sup>	7	0	7	0.97
<i>IL4R</i>	16p12.1 (27.3)	51	56.0	SeattleSNPs <sup>a</sup>	21	0	21	0.92
<i>ADAM33</i>	20p13 (3.6)	14	15.2	EGP <sup>d</sup>	19	4	15	1.00
Total		97	141.3		82	6	76	

<sup>a</sup>The National Heart Lung and Blood Institute's (NHLBI) Programs for Genomic Applications (PGA) SeattleSNPs database (<http://pga.gs.washington.edu>).

<sup>b</sup>The Innate Immunity NHLBI Program for PGA (<https://regepi.bwh.harvard.edu/IIPGA2>).

<sup>c</sup>HapMap phase 2 (<http://hapmap.ncbi.nlm.nih.gov>) and data from 96 type 1 diabetes individuals [28].

<sup>d</sup>The NIEHS Environmental Genome Project (<http://egp.gs.washington.edu>).

### Supplementary material

**Table S1.** Relevant demographic and clinical features of GOA samples.

**Table S2.** Information, completion rates and Hardy-Weinberg equilibrium (HWE)  $p$ -values for the tSNPs.

**Table S3.** Association summary of genotyped and imputed SNPs with asthma and atopic asthma for all candidate genes analyzed. Association summary of SNPs with asthma diagnosed before the cutoff demonstrating the largest effects is also shown for *IL13-IL4*, *LTA-TNF*, *CD14*, *MS4A2*, *ADRB2*, *ADAM33*, and *IL4R*.

**Table S4.** Coverage of candidate genes on commercial arrays used in asthma GWAS in samples of European ancestry.

**Figure S1.** Frequency distributions of the accumulated risk alleles for asthma and atopic asthma patients, and controls.

**Figure S2.** Receiver operating characteristic (ROC) curves derived from the multi-SNP GRS scores of associated risk alleles to distinguish asthmatics and atopic asthmatics from controls.

**Table S1.** Relevant demographic and clinical features of GOA samples.

Variable	Cases (n = 607)	Controls (n = 1271)	p-value
Gender (male %)	23.3 (567)	59.1 (1270)	<0.001 <sup>a</sup>
Age, median years (P <sub>25</sub> -P <sub>75</sub> ) <sup>b</sup>	25-44 (569)	32-48 (1270)	<0.001 <sup>c</sup>
Smoking habits (ever smokers, %)	27.3 (344)	48.1 (1122)	<0.001 <sup>a</sup>
Clinical features:			
log <sub>10</sub> Serum IgE levels (P <sub>25</sub> -P <sub>75</sub> ) <sup>b</sup>	1.95-2.60 (259)	NA	
Age at diagnosis (P <sub>25</sub> -P <sub>75</sub> ) <sup>b</sup>	14-39 (570)	NA	
FEV1 (% predicted, P <sub>25</sub> -P <sub>75</sub> ) <sup>b</sup>	68-95 (231)	NA	
Rhinitis (%)	74.2 (485)	NA	
Dermatitis (%)	20.2 (480)	NA	
SPT (% positive)	76.9 (445)	NA	
Specific IgE (% positive)	51.6 (287)	NA	
Atopy (%)	76.7 (464)	NA	
Treatment:			
Inhaled corticoids (%)	85.6 (250)	NA	
Oral corticoids (%)	4.6 (305)	NA	
Long-acting β2-agonists (%)	53.0 (317)	NA	
Short acting β2 agonists (%)	81.8 (231)	NA	
Body mass index kg/m <sup>2</sup> (P <sub>25</sub> -P <sub>75</sub> ) <sup>b</sup>	21.1-27.5 (227)	NA	

<sup>a</sup>χ<sup>2</sup> test; <sup>b</sup>P<sub>25</sub>, Percentile 25; P<sub>75</sub>, Percentile 75; <sup>c</sup>Mann-Whitney U-test; NA, not available.

Numbers in parentheses indicates total with non-missing data for each characteristic.

## RESULTADOS

**Table S2.** Information, completion rates and Hardy-Weinberg equilibrium (HWE)  $p$ -values for the tSNPs.

Gene	rs#	Position <sup>a</sup>	Effects and other designations	CR <sup>b</sup> (%)	HWE Controls	HWE Cases
<i>IL13-IL4</i>	rs1800925	132020708	-1112 C/T	97	0.858	0.283
	rs20541	132023863	Gln130Arg	99	<b>0.031</b>	0.637
	rs1295683	132026775		96	0.121	0.607
	rs2243297	132027070		96	0.408	1.000
	rs2243208	132029050		99	1.000	0.428
	rs2243211	132029321		99	0.836	0.394
	rs2243302	132032430		97	0.711	0.704
	rs2070874	132037609	Prom+33	91	<b>0.027</b>	<b>0.026</b>
	rs2227284	132040624	SNP+3017	99	0.667	<b>0.006</b>
	rs2243281	132044294		97	0.341	0.789
<i>LTA-TNF</i>	rs915654	31646476		96	<b>0.010</b>	0.863
	rs2844482	31647746	-753 G/A	98	0.602	1.000
	rs2071590	31647747		97	<b>0.045</b>	0.852
	rs2239704	31648120		96	0.518	0.793
	rs909253	31648292	NcoI Intron 1, 252 A/G	98	0.721	0.684
	rs746868	31648408		99	0.643	0.664
	rs2229092	31648736		99	0.826	1.000
	rs1041981	31648763		94	0.609	<b>0.026</b>
	rs1799724	31650461	-857 C/T	99	0.672	0.117
	rs1800629	31651010	-308 G/A	98	0.612	0.185
rs1800610	31651806		99	1.000	0.056	
<i>CD14</i>	rs2915863	139994561	-1601 intronic C/T	92	<b>0.002</b>	0.223
	rs2569191	139994087		97	0.059	0.621
	rs2569190	139993100	-159 C/T	98	0.231	1.000
	rs4914	139991652	1309 synonymous coding C/G	98	0.214	0.682
	rs2563298	139991499	1462 C/A	93	0.497	0.741
<i>MS4A2</i>	rs12361312	59609172		98	<b>9.9E-06</b>	0.358
	rs513986	59609429		99	0.728	0.799
	rs2583477	59610431		97	0.605	0.679
	rs573790	59611961		98	0.249	0.653
	rs1441586	59612604	-109 C/T	98	0.425	0.680
	rs569108	59619680	Gly237Glu	99	0.433	<b>0.042</b>
	rs2847655	59622247	RsaIex7	99	0.272	0.616
<i>ADRB2</i>	rs12654778	148185934	-654 A/G	98	0.051	0.547
	rs11959427	148186221		94	0.285	0.857
	rs1042713	148186633	Arg16Gly	98	0.129	0.800
	rs1042718	148187110		97	0.126	<b>0.030</b>
	rs1042719	148187640		97	0.319	0.314
	rs1042720	148187826		98	1.000	0.785
	rs4705271	148192184		99	0.091	0.379
<i>ADAM33</i>	rs554743	3610142		98	0.884	0.102
	rs2853213	3609840		98	0.660	0.195
	rs17548962	3609803		98	1.000	1.000
	rs570269 <sup>c</sup>	3607647		98	0.109	0.793

Table S2. (Continued)

Gene	rs#	Position <sup>a</sup>	Effects and other designations	CR <sup>b</sup> (%)	HWE Controls	HWE Cases
	rs2853210	3606211		97	0.815	0.910
	rs597165	3605804		98	1.000	0.602
	rs2787095	3603943		99	0.449	0.068
	rs511898	3603085	216_F_+1 Intron	99	0.622	0.663
	rs528557	3599742	216_S_2 ORF (Gly717Gly )	97	0.883	0.262
	rs574174	3598694	216_ST_+7 Intron	98	0.924	0.433
	rs2280091	3598234	216_T_1 ORF (Thr764Met)	98	0.320	0.314
	rs2280090	3598205	216_T_2 ORF (Ser774Pro)	99	0.152	0.254
	rs678881	3597803		98	1.000	0.917
	rs628977	3597721		98	0.433	1.000
	rs512625	3596378		98	0.623	0.848
<i>IL4R</i>	rs2057768	27229596	-3223 C/T	98	0.423	1.000
	rs1110470	27243928		98	0.954	0.115
	rs2283563	27253855		98	0.519	0.693
	rs3024537	27260320		99	0.382	0.626
	rs1805010	27263704	Ile50Val	97	0.730	0.184
	rs2301807	27265599		97	0.698	1.000
	rs3024585	27267345		98	0.122	0.624
	rs3024586	27267542		99	<b>0.027</b>	0.572
	rs2283561	27268668		95	0.859	0.800
	rs3024607	27271112		99	0.359	0.759
	rs3024613	27271754		98	0.955	0.461
	rs3024632	27273797		99	0.563	0.239
	rs3024676	27281059		99	0.165	0.175
	rs2234897	27281113		100	1.000	0.224
	rs1805011	27281373	Glu375Ala	99	0.185	0.866
	rs1805012	27281465	Cys406Arg	99	0.156	0.173
	rs1805013	27281481	Ser411Leu	99	0.140	1.000
	rs1805015	27281681	Ser478Pro	98	0.359	<b>0.039</b>
	rs1801275	27281901	Gln551Arg	98	<b>0.026</b>	0.104
	rs1805016	27282428	Ser727Ala	98	0.457	0.715
	rs1049631	27283043		98	<b>0.040</b>	<b>0.032</b>

<sup>a</sup>According to NCBI build 36.3; <sup>b</sup>Completion rate; <sup>c</sup>No distinction was made between C and the new T allele for this analysis. Nominal significant deviations in bold.



Table S3. Association summary of SNPs with asthma, atopic asthma and asthma with age-at-diagnosis of the disease before the cutoff demonstrating the largest effects at replicated SNPs.

Gene	rs#	Position <sup>a</sup>	Risk allele	Minor allele	Minor allele frequency			OR (95% CI) <sup>b</sup>			p-value	
					Controls	Asthma	Atopic asthma	Asthma	Atopic asthma	Asthma before age-at-diagnosis cutoff		Asthma before age-at-diagnosis cutoff
IL13-IL4	rs1811457	132020308	C	C	0.199	0.221	0.236				0.117	0.038 <sup>c</sup>
	rs1900925	132020708	T	T	0.199	0.221	0.236	1.16 (0.97-1.37)	1.24 (1.01-1.52)		0.100	0.037 <sup>c</sup>
	rs12956886	132023742	T	T	0.165	0.179	0.201				0.301	0.049 <sup>c</sup>
	rs20541	132023863	A	A	0.146	0.152	0.174				0.462	0.066
	rs12956885	132024344	A	A	0.183	0.198	0.222	1.07 (0.89-1.30)	1.23 (0.99-1.53)		0.222	0.008
	rs848	132024399	A	A	0.158	0.165	0.186				0.501	0.066
	rs12956883	132026775	A	A	0.081	0.088	0.101	1.13 (0.88-1.46)	1.29 (0.96-1.72)		0.400	0.109
	rs2243297	132027070	T	A	0.037	0.034	0.044	1.08 (0.72-1.59)	0.83 (0.53-1.28)		0.733	0.398
	rs2243208	132029050	G	G	0.077	0.086	0.098	1.12 (0.87-1.44)	1.29 (0.97-1.73)		0.367	0.082
	rs2243211	132029321	A	A	0.074	0.078	0.086	1.08 (0.83-1.39)	1.19 (0.88-1.60)		0.572	0.263
	rs2243302	132032430	G	A	0.134	0.124	0.130	1.09 (0.88-1.33)	1.02 (0.79-1.32)		0.460	0.874
	rs2243250	132037053	T	T	0.132	0.134	0.133				0.874	0.894
	rs2070874	132037609	T	T	0.132	0.133	0.133	1.02 (0.83-1.26)	1.02 (0.79-1.30)		0.816	0.883
	rs734244	132038625	T	T	0.132	0.134	0.133				0.857	0.920
	rs2227284	132040624	G	T	0.272	0.266	0.273				0.771	0.932
	rs2227282	132041078	C	G	0.272	0.266	0.273	1.02 (0.88-1.19)	0.99 (0.83-1.19)		0.709	0.929
	rs2243263	132041198	C	G	0.131	0.118	0.122				0.229	0.491
	rs2243266	132041688	A	A	0.132	0.134	0.133				0.885	0.918
	rs2243267	132041785	C	C	0.132	0.134	0.133				0.901	0.903
	rs2243268	132041862	C	C	0.132	0.134	0.133				0.882	0.911
rs2243269	132041913	AA	AA	0.132	0.134	0.133				0.879	0.916	
rs2243270	132042008	G	A	0.165	0.164	0.167				0.939	0.909	
rs2243274	132042731	A	G	0.165	0.164	0.167	1.16 (0.92-1.49)	1.10 (0.83-1.47)		0.940	0.905	
rs2243281	132044294	T	C	0.099	0.085	0.090				0.218	0.514	
rs2243282	132044453	A	A	0.132	0.134	0.133				0.889	0.885	
rs2243284	132044891	A	G	0.165	0.164	0.167				0.945	0.896	
rs2243285	132044892	T	G	0.131	0.118	0.122				0.222	0.462	
rs2243290	132046068	A	A	0.132	0.134	0.133				0.870	0.916	
rs2243291	132046882	C	G	0.165	0.164	0.167				0.949	0.912	
rs2243293	132047698	G	A	0.259	0.249	0.254				0.508	0.750	
CD14	rs2563298	139991499	A	A	0.258	0.263	0.239	1.02 (0.87-1.20)	0.90 (0.73-1.10)		0.774	0.322
	rs4914	139991652	C	G	0.116	0.113	0.105	1.05 (0.85-1.32)	1.12 (0.85-1.47)		0.670	0.411
	rs2569190	139993100	G	G	0.482	0.494	0.487	1.08 (0.94-1.24)	1.01 (0.86-1.20)		0.276	0.883
	rs2569191	139994087	T	T	0.484	0.496	0.488	1.07 (0.93-1.23)	1.01 (0.86-1.19)		0.355	0.890
	rs2915863	139994561	T	C	0.460	0.444	0.460	1.04 (0.90-1.20)	0.99 (0.84-1.18)		0.571	0.918
	rs3138076	139994797	C	C	0.223	0.240	0.245				0.269	0.227
	rs2569192	139995392	C	C	0.258	0.263	0.239				0.772	0.268
	rs2569193	139995679	A	A	0.258	0.263	0.239				0.749	0.301
	rs2563310	139995925	G	G	0.482	0.494	0.487				0.184	0.818
	rs3138074	139996116	T	T	0.223	0.240	0.245				0.259	0.233
	rs5744441	139997031	A	A	0.223	0.240	0.245				0.265	0.234
	rs4912717	139998419	A	A	0.258	0.263	0.239				0.749	0.304
	rs4912718	139998562	T	T	0.482	0.494	0.487				0.188	0.809

Table S3. (Continued)

Gene	rs#	Position <sup>a</sup>	Risk allele	Minor allele frequency				OR (95% CI) <sup>b</sup>			p-value	
				Minor allele	Controls	Asthma	Atopic asthma	Asthma before age-at-diagnosis cutoff	Asthma	Atopic asthma	Asthma before age-at-diagnosis cutoff	Asthma
ADRB2	rs33910799	148182683	dG	dG	0.418	0.418	0.399				0.995	0.393
	rs11746634	148182861	C	G	0.342	0.340	0.348				0.920	0.761
	rs11168067	148182994	A	G	0.361	0.359	0.367				0.913	0.749
	rs11957351	148183297	T	C	0.361	0.359	0.367				0.913	0.776
	rs11948371	148183301	T	A	0.361	0.359	0.367				0.910	0.761
	rs11960649	148183337	A	C	0.361	0.359	0.367				0.914	0.775
	rs1432622	148183955	T	C	0.361	0.359	0.367				0.900	0.729
	rs1432623	148184201	C	T	0.361	0.359	0.367				0.926	0.756
	rs17778257	148184770	T	T	0.418	0.418	0.399				0.993	0.376
	rs2400706	148185057	T	C	0.235	0.225	0.232				0.462	0.870
	rs2895795	148185159	A	T	0.235	0.225	0.232				0.482	0.886
	rs2400707	148185245	A	G	0.361	0.359	0.367				0.908	0.751
	rs2053044	148185565	A	G	0.361	0.359	0.367				0.919	0.752
	rs12654778	148185934	G	A	0.399	0.388	0.371	1.04 (0.91-1.20)	1.12 (0.94-1.33)		0.565	1.184
	rs11168070	148186120	G	C	0.361	0.359	0.367				0.920	0.740
	rs11959427	148186221	C	C	0.361	0.359	0.367	1.00 (0.87-1.16)	1.04 (0.87-1.24)		0.979	0.706
	rs1042711	148186541	C	T	0.344	0.342	0.350				0.855	0.765
	rs1801704	148186568	C	T	0.363	0.360	0.369				0.893	0.755
	rs1042713	148186633	G	A	0.416	0.414	0.393	1.01 (0.88-1.16)	1.11 (0.93-1.32)		0.872	0.250
	rs1042714	148186666	G	C	0.361	0.359	0.367				0.919	0.771
rs1042717	148186839	A	G	0.218	0.216	0.223				0.860	0.759	
rs1042718	148187110	A	A	0.184	0.187	0.193	1.03 (0.87-1.23)	1.06 (0.86-1.31)		0.703	0.552	
rs1042719	148187640	G	C	0.313	0.284	0.291	1.15 (0.98-1.33)	1.11 (0.93-1.33)		0.086	0.265	
rs1042720	148187826	G	A	0.362	0.347	0.353	1.06 (0.92-1.23)	1.04 (0.88-1.25)		0.418	0.657	
rs4705271	148192184	A	A	0.199	0.203	0.208	1.02 (0.86-1.21)	1.05 (0.86-1.28)		0.793	0.636	
LTA-TNF	rs2009658	31646223	G	G	0.163	0.195	0.198			0.218	0.018	0.032
	rs915654	31646476	A	A	0.350	0.391	0.387	1.18 (1.02-1.35)	1.16 (0.98-1.37)	1.23 (1.01-1.49)	0.025	0.085
	rs2844482	31647746	T	T	0.163	0.196	0.198	1.24 (1.03-1.48)	1.26 (1.02-1.56)	1.42 (1.12-1.81)	0.020 <sup>c</sup>	0.033 <sup>c</sup>
	rs2071590	31647747	G	A	0.378	0.327	0.324	1.20 (1.04-1.39)	1.25 (1.04-1.49)	1.41 (1.15-1.75)	0.013 <sup>c</sup>	0.013 <sup>c</sup>
	rs1800683	31648050	A	A	0.274	0.279	0.268			0.280	0.740	0.780
	rs2239704	31648120	C	A	0.425	0.385	0.376	1.15 (1.00-1.33)	1.22 (1.02-1.45)	1.23 (1.00-1.52)	0.055	0.029
	rs909253	31648292	G	G	0.274	0.279	0.268	1.00 (0.86-1.17)	0.96 (0.80-1.17)	1.00 (0.80-1.24)	0.978	0.694
	rs746868	31648408	G	C	0.422	0.379	0.373	1.16 (1.01-1.35)	1.22 (1.02-1.45)	1.27 (1.03-1.56)	0.037	0.025
	rs2229092	31648736	A	C	0.068	0.060	0.058	1.15 (0.86-1.54)	1.18 (0.83-1.67)	1.16 (0.77-1.75)	0.360	0.354
	rs1041981	31648763	A	A	0.270	0.278	0.266	1.02 (0.87-1.19)	0.97 (0.80-1.18)	1.00 (0.81-1.25)	0.841	0.774
	rs1799964	31650287	C	C	0.213	0.243	0.246			0.264	0.014 <sup>c</sup>	0.041 <sup>c</sup>
	rs1800630	31650455	A	A	0.163	0.196	0.198			0.218	0.014 <sup>c</sup>	0.048 <sup>c</sup>
	rs1799724	31650461	C	T	0.112	0.088	0.079	1.32 (1.04-1.67)	1.47 (1.09-1.96)	1.35 (0.95-1.89)	0.021 <sup>c</sup>	0.009 <sup>c</sup>
	rs1800629	31651010	A	A	0.130	0.145	0.154	1.11 (0.91-1.36)	1.21 (0.95-1.53)	1.09 (0.83-1.45)	0.321	0.123
	rs361525	31651080	A	G	0.099	0.095	0.103			0.104	0.737	0.775
	rs3093861	31651737	A	G	0.099	0.095	0.103			0.104	0.722	0.777
	rs1800610	31651806	G	A	0.111	0.084	0.080	1.35 (1.06-1.72)	1.43 (1.06-1.92)	1.25 (0.89-1.75)	0.014	0.016
	rs3093662	31652168	G	A	0.099	0.095	0.103			0.104	0.716	0.785

Table S3. (Continued)

Gene	rs#	Position <sup>a</sup>	Risk allele	Minor allele	Minor allele frequency				OR (95% CI) <sup>b</sup>				P-value			
					Asthma before age-at-diagnosis cutoff		Asthma before age-at-diagnosis cutoff		Asthma before age-at-diagnosis cutoff		Asthma before age-at-diagnosis cutoff		Asthma	Atopic asthma	Asthma	Atopic asthma
					Controls	Asthma	Atopic asthma	age-at-diagnosis cutoff	Controls	Asthma	Atopic asthma	age-at-diagnosis cutoff				
LTA-TNF	rs3093664	31652621	G	A	0.099	0.095	0.103	0.104	1.11 (0.96-1.28)	1.12 (0.94-1.33)	1.08 (0.92-1.27)	0.138	0.194	0.758	0.758	
	rs3093668	31654474	C	G	0.099	0.095	0.103	0.104	1.16 (1.01-1.33)	1.15 (0.97-1.37)	1.15 (0.98-1.37)	0.038	0.102	0.713	0.713	
	rs3093726	31654768	C	T	0.099	0.095	0.103	0.104	1.08 (0.93-1.25)	1.04 (0.87-1.24)	1.09 (0.92-1.29)	0.306	0.648	0.742	0.742	
	rs1800628	31654829	A	A	0.111	0.124	0.132	0.127	1.03 (0.90-1.19)	0.97 (0.82-1.15)	1.09 (0.93-1.28)	0.636	0.722	0.291	0.291	
	rs3093727	31655094	A	G	0.099	0.095	0.103	0.104				0.734	0.792	0.750	0.750	
	rs513986	59609429	C	T	0.430	0.405	0.403	0.403	1.11 (0.96-1.28)	1.12 (0.94-1.33)	1.08 (0.92-1.27)	0.138	0.194	0.361	0.361	
	rs253477	59610431	T	C	0.481	0.447	0.447	0.442	1.16 (1.01-1.33)	1.15 (0.97-1.37)	1.15 (0.98-1.37)	0.038	0.102	0.081	0.081	
	rs573790	59611961	T	T	0.333	0.351	0.342	0.351	1.08 (0.93-1.25)	1.04 (0.87-1.24)	1.09 (0.92-1.29)	0.306	0.648	0.320	0.320	
	rs1441586	59612604	C	T	0.466	0.457	0.473	0.450	1.03 (0.90-1.19)	0.97 (0.82-1.15)	1.09 (0.93-1.28)	0.636	0.722	0.290	0.290	
	rs2583476	59614157	A	G	0.422	0.414	0.429	0.402				0.634	0.736	0.310	0.310	
rs2847663	59614612	G	C	0.429	0.422	0.436	0.409				0.637	0.738	0.316	0.316		
rs556917	59615288	A	T	0.466	0.457	0.473	0.444				0.603	0.756	0.294	0.294		
rs2847666	59616152	G	A	0.429	0.422	0.436	0.409				0.642	0.708	0.308	0.308		
rs2847667	59616185	T	C	0.451	0.443	0.458	0.430				0.643	0.716	0.290	0.290		
rs502581	59616754	G	T	0.466	0.457	0.473	0.444				0.636	0.722	0.300	0.300		
rs2583471	59618390	A	G	0.429	0.422	0.436	0.409				0.641	0.750	0.291	0.291		
rs2070970	59618559	T	C	0.429	0.422	0.436	0.409				0.671	0.717	0.298	0.298		
rs2847668	59618837	T	A	0.429	0.422	0.436	0.409				0.618	0.730	0.279	0.279		
rs569108	59619680	A	G	0.039	0.024	0.023	0.018	1.67 (1.10-2.56)	1.75 (1.03-2.94)	2.94 (1.56-5.56)	0.012 <sup>c</sup>	0.027 <sup>c</sup>	1.0E-04 <sup>c</sup>	1.0E-04 <sup>c</sup>		
rs2847655	59622247	T	C	0.434	0.428	0.441	0.424	1.02 (0.88-1.16)	0.96 (0.81-1.14)	1.05 (0.89-1.23)	0.818	0.685	0.639	0.639		
rs502419	59622751	G	A	0.466	0.457	0.473	0.444				0.628	0.722	0.309	0.309		
rs2855017	59622885	C	T	0.434	0.428	0.441	0.420				0.774	0.707	0.493	0.493		
rs17528859	59623955	C	T	0.434	0.428	0.441	0.420				0.751	0.725	0.481	0.481		
rs574704	59624489	A	G	0.466	0.457	0.473	0.444				0.658	0.760	0.299	0.299		
rs2057768	27229596	T	T	0.306	0.309	0.313	0.313	1.01 (0.87-1.18)	1.04 (0.86-1.24)	1.03 (0.87-1.23)	0.869	0.710	0.700	0.700		
rs8052962	27234343	T	T	0.400	0.415	0.423	0.414				0.362	0.273	0.475	0.475		
rs8044444	27236044	T	C	0.155	0.152	0.142	0.140				0.795	0.371	0.229	0.229		
rs6498011	27239395	A	A	0.401	0.417	0.424	0.415				0.373	0.295	0.445	0.445		
rs6498012	27239475	C	C	0.401	0.417	0.424	0.415				0.349	0.250	0.446	0.446		
rs1110470	27243928	G	A	0.453	0.442	0.451	0.458	1.04 (0.91-1.20)	1.01 (0.85-1.20)	0.99 (0.84-1.16)	0.505	0.863	0.878	0.878		
rs2040789	27243938	G	A	0.457	0.450	0.457	0.461				0.669	0.978	0.801	0.801		
rs2283563	27253855	C	T	0.328	0.297	0.292	0.300	1.18 (1.01-1.37)	1.19 (0.99-1.43)	1.15 (0.97-1.37)	0.041	0.059	0.103	0.103		
rs3024530	27258188	G	G	0.478	0.495	0.496	0.489				0.144	0.255	0.618	0.618		
rs2057767	27258839	G	G	0.478	0.495	0.496	0.489				0.139	0.265	0.618	0.618		
rs3024536	27260214	T	C	0.153	0.148	0.137	0.135				0.684	0.303	0.210	0.210		
rs3024537	27260320	G	A	0.153	0.148	0.137	0.135	1.02 (0.84-1.23)	1.12 (0.88-1.43)	1.12 (0.89-1.41)	0.855	0.355	0.314	0.314		
rs3024548	27262032	G	C	0.458	0.447	0.429	0.431				0.549	0.208	0.195	0.195		
rs3024552	27262505	C	G	0.458	0.447	0.429	0.431				0.560	0.193	0.193	0.193		
rs3024556	27262863	A	G	0.270	0.263	0.257	0.264				0.603	0.440	0.720	0.720		
rs1805010	27263704	A	G	0.458	0.447	0.429	0.431	1.05 (0.92-1.20)	1.12 (0.95-1.33)	1.11 (0.95-1.32)	0.478	0.180	0.176	0.176		
rs3024560	27264168	G	G	0.354	0.386	0.396	0.354				0.055	0.028	0.035	0.035		
rs2301807	27265599	C	A	0.038	0.026	0.029	0.023	1.49 (0.98-2.27)	1.37 (0.83-2.22)	1.67 (1.01-2.78)	0.053	0.208	0.036	0.036		
rs3024575	27265633	T	T	0.370	0.402	0.417	0.412				0.055	0.016	0.033	0.033		

Table S3. (Continued)

Gene	rs#	Position <sup>a</sup>	Risk allele	Minor allele	Minor allele frequency				OR (95% CI) <sup>b</sup>				P-value	
					Controls	Asthma	Atopic asthma	Asthma before age-at-diagnosis cutoff	Asthma	Atopic asthma	Asthma before age-at-diagnosis cutoff	Asthma	Atopic asthma	Asthma before age-at-diagnosis cutoff
IL4R	rs3024577	27265704	G	G	0.473	0.496	0.499	0.498	0.200	0.196	0.200	0.196	0.196	0.248
	rs2283562	27266264	C	C	0.461	0.490	0.494	0.488	0.095	0.106	0.095	0.106	0.106	0.166
	rs2239348	27266449	A	G	0.440	0.421	0.426	0.429	0.284	0.542	0.284	0.542	0.542	0.620
	rs2239347	27266522	C	C	0.461	0.490	0.494	0.488	0.064	0.101	0.064	0.101	0.101	0.165
	rs2239346	27266562	T	T	0.461	0.490	0.494	0.488	0.069	0.101	0.069	0.101	0.101	0.164
	rs3024585	27267345	A	A	0.473	0.496	0.499	0.498	1.09 (0.95-1.25)	1.12 (0.94-1.32)	1.10 (0.94-1.28)	1.10 (0.94-1.28)	1.10 (0.94-1.28)	0.243
	rs3024586	27267542	G	A	0.044	0.037	0.049	0.035	1.16 (0.83-1.67)	0.89 (0.61-1.30)	1.25 (0.83-1.89)	0.384	0.550	0.268
	rs3024589	27268128	G	A	0.499	0.478	0.469	0.475	0.187	0.111	0.187	0.111	0.111	0.187
	rs3024590	27268129	C	T	0.499	0.478	0.469	0.475	0.195	0.097	0.195	0.097	0.097	0.188
	rs2283561	27268668	C	C	0.424	0.455	0.462	0.459	1.13 (0.98-1.30)	1.17 (0.98-1.39)	1.14 (0.97-1.34)	0.095	0.080	0.117
	rs2283561	27268668	C	C	0.422	0.481	0.481	0.485	0.301	0.409	0.301	0.409	0.409	0.508
	rs3024601	27270352	C	C	0.423	0.441	0.438	0.438	0.239	0.387	0.239	0.387	0.387	0.394
	rs3024607	27271112	G	A	0.084	0.071	0.068	0.060	1.16 (0.89-1.52)	1.25 (0.90-1.72)	1.39 (1.01-1.92)	0.257	0.179	0.039
	rs3024611	27271701	A	C	0.117	0.093	0.085	0.079	0.003	0.002	0.003	0.002	0.002	5.00E-05
	rs3024613	27271754	T	C	0.500	0.482	0.486	0.485	1.06 (0.93-1.22)	1.05 (0.89-1.25)	1.05 (0.89-1.23)	0.397	0.538	0.553
	rs3024632	27273797	T	C	0.079	0.074	0.069	0.064	1.03 (0.79-1.35)	1.14 (0.83-1.59)	1.20 (0.88-1.67)	0.820	0.430	0.237
	rs3024634	27274063	G	A	0.152	0.123	0.113	0.106	0.003	0.002	0.003	0.002	0.002	1.30E-04
	rs3024644	27275327	C	G	0.152	0.123	0.113	0.106	0.004	1.00E-04	0.004	1.00E-04	1.00E-04	9.00E-05
	rs3024647	27275473	G	A	0.152	0.123	0.113	0.106	0.005	0.001	0.005	0.001	0.001	1.00E-04
	rs2891057	27276291	A	T	0.152	0.123	0.113	0.106	0.005	0.001	0.005	0.001	0.001	3.00E-05
	rs2891058	27276334	G	A	0.152	0.123	0.113	0.106	0.002	0.004	0.002	0.004	0.004	9.00E-05
	rs3024659	27278437	G/G	G/G	0.152	0.123	0.113	0.106	0.006	0.002	0.006	0.002	0.002	1.10E-04
	rs3024660	27278659	T	C	0.152	0.123	0.113	0.106	0.010	0.001	0.010	0.001	0.001	1.00E-04
	rs3024676	27281059	C	A	0.206	0.186	0.173	0.164	1.15 (0.97-1.37)	1.23 (1.00-1.54)	1.32 (1.08-1.61)	0.121	0.054	0.007
	rs2234897	27281113	C	C	0.019	0.021	0.017	0.022	1.11 (0.68-1.82)	0.91 (0.47-1.73)	1.16 (0.67-2.01)	0.667	0.767	0.599
	rs1805011	27281373	A	C	0.151	0.140	0.131	0.128	1.11 (0.91-1.35)	1.18 (0.93-1.49)	1.22 (0.96-1.54)	0.306	0.195	0.096
	rs2234898	27281416	T	G	0.126	0.100	0.094	0.089	0.025 <sup>c</sup>	0.017 <sup>c</sup>	0.025 <sup>c</sup>	0.017 <sup>c</sup>	0.017 <sup>c</sup>	0.003 <sup>c</sup>
	rs1805012	27281465	T	C	0.126	0.100	0.094	0.089	1.30 (1.04-1.61)	1.37 (1.04-1.82)	1.47 (1.14-1.92)	0.018 <sup>c</sup>	0.019 <sup>c</sup>	0.003 <sup>c</sup>
	rs2234900	27281473	C	T	0.126	0.100	0.094	0.089	0.023 <sup>c</sup>	0.023 <sup>c</sup>	0.023 <sup>c</sup>	0.023 <sup>c</sup>	0.023 <sup>c</sup>	0.003 <sup>c</sup>
	rs1805013	27281481	C	T	0.040	0.032	0.028	0.025	1.28 (0.88-1.89)	1.43 (0.88-2.33)	1.56 (0.98-2.56)	0.188	0.131	0.050
	rs1805015	27281681	T	T	0.191	0.154	0.141	0.133	1.32 (1.09-1.56)	1.43 (1.12-1.79)	1.54 (1.23-1.92)	0.004 <sup>c</sup>	0.002 <sup>c</sup>	1.0E-04 <sup>c</sup>
	rs1801275	27281901	A	G	0.235	0.207	0.197	0.190	1.18 (1.00-1.39)	1.23 (1.01-1.52)	1.30 (1.08-1.59)	0.044 <sup>c</sup>	0.034 <sup>c</sup>	0.007 <sup>c</sup>
	rs1805016	27282428	T	G	0.062	0.060	0.056	0.051	1.04 (0.78-1.39)	1.11 (0.78-1.59)	1.19 (0.84-1.69)	0.822	0.582	0.314
	rs1049631	27283043	A	A	0.468	0.475	0.478	0.475	1.02 (0.89-1.16)	1.03 (0.88-1.22)	1.02 (0.87-1.19)	0.771	0.663	0.800
	rs8832	27283288	A	A	0.468	0.475	0.478	0.475	0.730	0.648	0.730	0.648	0.648	0.782
	rs1029489	27283718	G	G	0.367	0.380	0.389	0.388	0.472	0.311	0.472	0.311	0.311	0.330
	rs2382721	27284138	C	C	0.309	0.328	0.339	0.343	0.256	0.140	0.256	0.140	0.140	0.081
	rs3024685	27284411	C	C	0.367	0.380	0.389	0.388	0.451	0.309	0.451	0.309	0.309	0.332
ADAM33	rs512625	3596378	A	A	0.281	0.313	0.309	0.363	1.17 (1.00-1.36)	1.14 (0.95-1.37)	1.46 (1.12-1.91)	0.047	0.147	0.006
	rs2787093	3596462	C	C	0.104	0.125	0.116	0.184	0.058	0.394	0.058	0.394	0.394	0.003
	rs2787094	3597161	C	C	0.206	0.243	0.241	0.278	0.024 <sup>c</sup>	0.079	0.024 <sup>c</sup>	0.079	0.079	0.024 <sup>c</sup>
	rs677044	3597431	G	A	0.257	0.266	0.263	0.284	0.923	0.691	0.923	0.691	0.691	0.270
	rs628965	3597713	A	A	0.383	0.401	0.400	0.498	0.305	0.408	0.305	0.408	0.408	3.0E-04 <sup>c</sup>

Table S3. (Continued)

Gene	rs#	Position <sup>a</sup>	Risk allele	Minor allele	Minor allele frequency				OR (95% CI) <sup>b</sup>				p-value	
					Controls	Asthma	Atopic asthma	Asthma before age-at-diagnosis cutoff	Asthma	Atopic asthma	Asthma before age-at-diagnosis cutoff	Asthma	Atopic asthma	Asthma before age-at-diagnosis cutoff
<u>ADAM33</u>	<u>rs628977</u>	3597721	T	T	0.383	0.400	0.397	0.496	1.09 (0.95-1.26)	1.07 (0.90-1.27)	1.71 (1.32-2.22)	0.231	0.455	<b>4.9E-05<sup>c</sup></b>
	<u>rs678881</u>	3597803	G	C	0.271	0.270	0.279	0.304	1.00 (0.85-1.16)	0.96 (0.79-1.16)	0.84 (0.64-1.11)	0.966	0.664	0.217
	<u>rs630712</u>	3598066	C	C	0.104	0.125	0.116	0.184				0.071	0.361	<b>0.003</b>
	<u>rs2280089</u>	3598127	A	A	0.132	0.142	0.141	0.147				0.459	0.567	0.499
	<u>rs2280090</u>	3598205	A	A	0.124	0.124	0.113	0.130	0.98 (0.79-1.20)	0.90 (0.69-1.16)	1.02 (0.70-1.48)	0.816	0.390	0.915
	<u>rs2280091</u>	3598234	G	G	0.132	0.142	0.141	0.147	1.07 (0.87-1.30)	1.07 (0.84-1.35)	1.11 (0.77-1.58)	0.547	0.616	0.583
	<u>rs574174</u>	3598694	T	T	0.182	0.193	0.203	0.179	1.07 (0.89-1.28)	1.15 (0.93-1.41)	0.97 (0.70-1.35)	0.443	0.203	0.850
	<u>rs597980</u>	3599165	A	G	0.441	0.411	0.428	0.330				0.082	0.555	<b>4.4E-04<sup>c</sup></b>
	<u>rs44707</u>	3599226	G	T	0.486	0.478	0.468	0.412				0.069	<b>0.041<sup>c</sup></b>	<b>0.003<sup>c</sup></b>
	<u>rs598418</u>	3599269	G	G	0.383	0.401	0.400	0.498				0.315	0.411	<b>1.9E-04<sup>c</sup></b>
	<u>rs2853209</u>	3599472	A	T	0.411	0.397	0.405	0.325				0.286	0.702	<b>3.1E-04<sup>c</sup></b>
	<u>rs528557</u>	3599742	G	G	0.263	0.279	0.278	0.261	1.07 (0.91-1.25)	1.07 (0.89-1.29)	0.96 (0.72-1.27)	0.411	0.476	0.757
	<u>rs3918395</u>	3601149	A	A	0.298	0.337	0.350	0.322				<b>0.019<sup>c</sup></b>	<b>0.004<sup>c</sup></b>	0.407
	<u>rs2271511</u>	3602433	T	T	0.325	0.364	0.378	0.354				<b>0.014<sup>c</sup></b>	<b>0.013<sup>c</sup></b>	0.318
	<u>rs511898</u>	3603085	T	T	0.359	0.375	0.368	0.370				0.480	0.730	0.898
	<u>rs2787095</u>	3603943	C	C	0.416	0.458	0.432	0.459	1.05 (0.91-1.22)	1.03 (0.86-1.23)	1.02 (0.78-1.33)	0.480	0.730	0.898
	<u>rs597165</u>	3605804	A	A	0.136	0.136	0.145	0.154	1.17 (1.02-1.35)	1.06 (0.89-1.26)	1.63 (1.26-2.11)	<b>0.027</b>	0.485	<b>1.90E-04</b>
	<u>rs2853210</u>	3606211	C	T	0.240	0.237	0.228	0.312	0.99 (0.81-1.22)	1.07 (0.84-1.36)	1.14 (0.80-1.63)	0.956	0.551	0.476
	<u>rs570269<sup>d</sup></u>	3607647	G	G	0.205	0.234	0.225	0.290	1.03 (0.88-1.22)	1.08 (0.88-1.32)	0.71 (0.54-0.94)	0.720	0.494	<b>0.020</b>
	<u>rs1754896<sup>e</sup></u>	3609803	T	T	0.033	0.040	0.038	0.051	1.15 (0.98-1.36)	1.11 (0.91-1.36)	1.54 (1.16-2.04)	0.086	0.287	<b>0.004</b>
	<u>rs2853213</u>	3609840	C	G	0.345	0.333	0.336	0.337	1.22 (0.84-1.76)	1.15 (0.74-1.80)	1.52 (0.83-2.75)	0.289	0.534	0.189
	<u>rs554743</u>	3610142	C	C	0.267	0.276	0.276	0.320	1.08 (0.93-1.25)	1.05 (0.88-1.25)	1.06 (0.81-1.39)	0.319	0.580	0.647

tSNPs are underlined. Nominal significant associations in bold.

<sup>a</sup>According to NCBI build 36.3; <sup>b</sup>Calculated only for the genotyped tSNPs; <sup>c</sup>Replication of previous studies; <sup>d</sup>No distinction was made between C and the new T allele for this analysis.

**Table S4.** Coverage of candidate genes on commercial arrays used in asthma GWAS in samples of European ancestry.<sup>a</sup>

Gene	SNP Array	HapMap CEU				1KGP CEU			
		SNPs MAF>5 %	SNPs in array	# tagged SNPs <sup>b</sup>	Coverage (%)	SNPs MAF>5 %	SNPs in array	# tagged SNPs <sup>b</sup>	Coverage (%)
<i>IL13-IL4</i>	Affymetrix 500K	34	11	30	88	47	7	36	77
	Affymetrix Genome-Wide Human SNP Array 6.0	34	11	30	88	47	10	37	79
	Illumina HumanHap 300 <sup>c</sup>	34	5	14	41	47	4	17	36
	Illumina HumanHap 550 <sup>d</sup>	34	7	24	71	47	7	32	68
	Illumina HumanHap 650	34	12	28	82	47	9	32	68
<i>CD14</i>	Illumina HumanHap 1M-Duo	34	33	34	100	47	40	40	85
	Affymetrix 500K	6	2	5	83	15	2	13	87
	Affymetrix Genome-Wide Human SNP Array 6.0	6	3	5	83	15	3	13	87
	Illumina HumanHap 300 <sup>c</sup>	6	0	0	0	15	0	0	0
	Illumina HumanHap 550 <sup>d</sup> /650	6	1	3	50	15	1	4	27
<i>ADRB2</i>	Illumina HumanHap 1M-Duo	6	2	3	50	15	2	4	27
	Affymetrix 500K, Affymetrix Genome-Wide Human SNP Array 6.0	10	2	7	70	27	18	17	63
	Illumina HumanHap 300 <sup>c</sup>	10	3	7	70	27	3	4	15
	Illumina HumanHap 550 <sup>d</sup>	10	4	7	70	27	5	21	78
	Illumina HumanHap 650	10	5	9	90	27	7	23	85
<i>LTA-TNF</i>	Illumina HumanHap 1M-Duo	10	10	10	100	27	18	25	93
	Affymetrix 500K, Affymetrix Genome-Wide Human SNP Array 6.0	13	1	1	8	25	1	1	4
	Illumina HumanHap 300 <sup>c</sup> /550 <sup>d</sup> /650	13	5	8	62	25	5	14	56
	Illumina HumanHap 1M-Duo	13	13	13	100	25	20	20	80
	Affymetrix 500K	21	2	20	95	34	2	24	71
<i>MS442</i>	Affymetrix Genome-Wide Human SNP Array 6.0	21	3	20	95	34	3	24	71
	Illumina HumanHap 300 <sup>c</sup>	21	0	0	0	34	0	0	0
	Illumina HumanHap 550 <sup>d</sup> /650	21	3	20	95	34	3	24	71
	Illumina HumanHap 1M	21	9	21	100	34	8	31	91
	Affymetrix 500K	55	11	19	35	143	10	56	39
<i>IL4R</i>	Affymetrix Genome-Wide Human SNP Array 6.0	55	20	39	71	143	24	110	77
	Illumina HumanHap 300 <sup>c</sup>	55	12	35	64	143	12	86	60
	Illumina HumanHap 550 <sup>d</sup>	55	20	51	93	143	18	115	80
	Illumina HumanHap 650	55	26	53	96	143	22	128	90
	Illumina HumanHap 1M-Duo	55	47	54	98	143	81	138	97

Table S4. (Continued)

Gene	SNP Array	HapMap CEU					1KGP CEU						
		SNPs MAF>5 %	SNPs in array	# tagged SNPs <sup>b</sup>	Coverage (%)	SNPs MAF>5%	SNPs in array	# tagged SNPs <sup>b</sup>	Coverage (%)	SNPs MAF>5%	SNPs in array	# tagged SNPs <sup>b</sup>	Coverage (%)
ADAM33	Affymetrix 500K Affymetrix Genome-Wide Human SNP Array 6.0	17	2	4	24	50	2	4	8				
	Illumina HumanHap 300 <sup>c</sup>	17	3	3	18	50	3	3	6				
	Illumina HumanHap 550 <sup>d</sup> /650	17	5	5	29	50	5	9	18				
	Illumina HumanHap 11M-Duo	17	11	13	76	50	11	14	28				

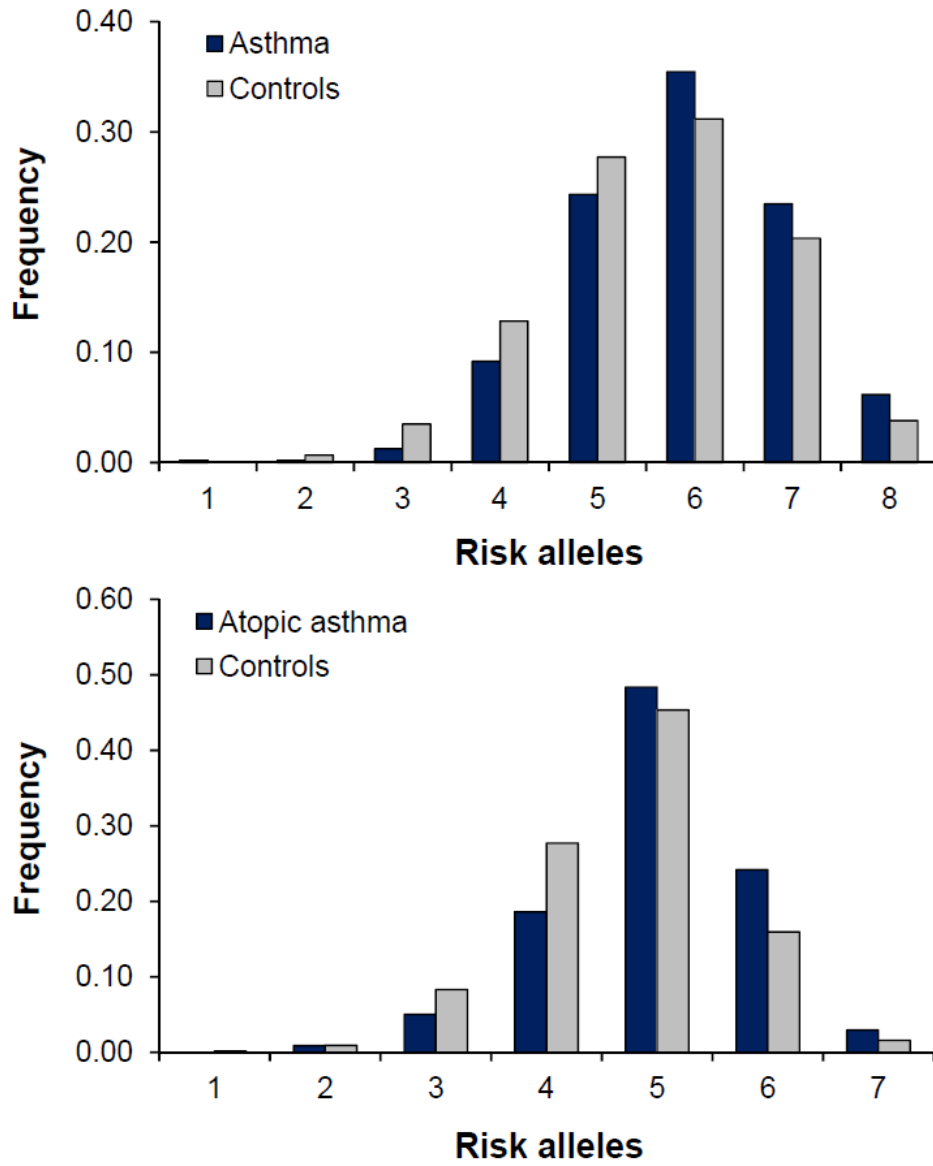
<sup>a</sup>Includes studies described in Moffatt et al. 2007, Himes et al. 2009, Sleiman et al. 2009, Li et al. 2010, Moffatt et al. 2010, Ferreira et al. 2011, Torgerson et al. 2011.

<sup>b</sup>SNPs captured by the array both directly and indirectly (with a multi-marker  $r^2 \geq 0.8$ )

<sup>c</sup>Including the Illumina HumanCNV370-Duo as it was built upon the Illumina HumanHap300 BeadChip (317K SNPs) with additional 52K markers aimed at detecting CNVs.

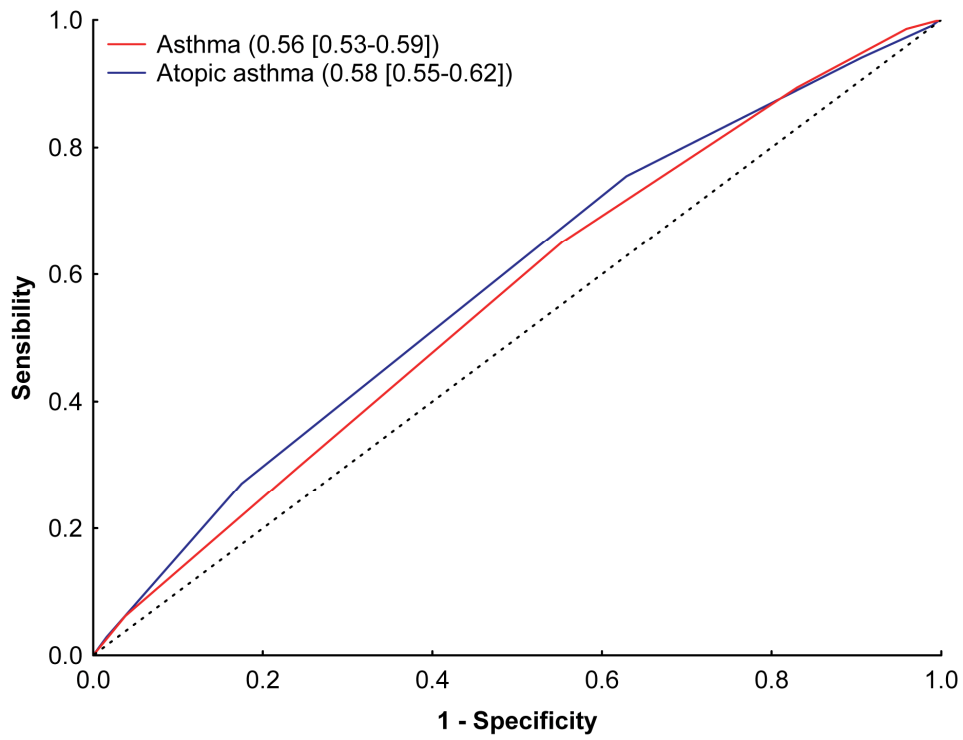
<sup>d</sup>Including the Illumina Human610 as it is similar to the Illumina HumanHap550 with the addition of 60K markers to target regions with CNVs.

**Figure S1.** Frequency distributions of the accumulated risk alleles for asthma and atopic asthma patients, and controls.





**Figure S2.** Receiver operating characteristic (ROC) curves derived from the multi-SNP GRS scores of associated risk alleles to distinguish asthmatics and atopic asthmatics from controls.



## ***5. DISCUSIÓN***

---

## 5. DISCUSIÓN

### 5.1. Estructura genética de la población española

En esta tesis doctoral se ha evaluado por primera vez la posible existencia de estructura poblacional en España debida a la presencia de un componente genético norteafricano desigualmente distribuido entre regiones. Para ello, en primer lugar se demostró que unos pocos EuroAIMs, marcadores de ascendencia seleccionados en un estudio anterior para detectar estructura poblacional en Europa (Price *et al.*, 2008), también resultaron útiles para distinguir entre poblaciones norteafricanas y europeas. Tras esto, los EuroAIMs permitieron estimar que la influencia norteafricana es menor en la población peninsular que en la población canaria (5% frente a 17%, respectivamente). Además, en la población canaria se observó una gran variación entre individuos para esas estimas, oscilando desde un 0% hasta un 96% de mezcla norteafricana. Asimismo, se demostró que esta distribución desigual de mezcla norteafricana en la población española podría sesgar los estudios de asociación de tipo caso-control realizados con individuos sin relación de parentesco, y que estos efectos podrían minimizarse usando un conjunto reducido de EuroAIMs.

Este estudio reveló la existencia de un flujo genético bidireccional entre poblaciones norteafricanas y peninsulares, algo que ya había sido descrito en estudios previos (Rando *et al.*, 1998; Bosch *et al.*, 2001). Sin embargo, mientras que en esta tesis doctoral se estimó un 5% de influencia norteafricana en la población peninsular, los estudios anteriores encontraron un valor de influencia norteafricana en la población peninsular del 8-10%, empleando marcadores de herencia uniparental (Flores *et al.*, 2004; Adams *et al.*, 2008; Capelli *et al.*, 2009). Nuestros resultados, basados en marcadores autosómicos, mostraron que esta mezcla podría variar desde tan sólo un  $2,1 \pm 4,6\%$  en las poblaciones del este peninsular hasta un  $9,0 \pm 28,8\%$  en las poblaciones del oeste peninsular, si bien estas diferencias no llegaron a ser estadísticamente significativas ( $p=0,156$ ), dado el número limitado de muestras correspondientes a cada región. En línea con nuestros resultados, un estudio previo realizado con marcadores del cromosoma Y también encontró diferencias geográficas en la influencia norteafricana entre poblaciones del este y oeste de la Península Ibérica (Adams *et al.*, 2008), mientras que otro estudio detectó igualmente la presencia de subestructura poblacional entre distintas poblaciones de la Península (Gayán *et al.*, 2010). Sin embargo, serán necesarios más estudios en los que se incluyan mayor número de muestras y de marcadores genéticos para confirmar la existencia de tales patrones geográficos de mezcla norteafricana diferencial en las poblaciones peninsulares.

A pesar de que la conquista española durante el siglo XV se realizó de una manera agresiva, y que desde entonces han existido en Canarias numerosos procesos migratorios, el elevado porcentaje de mezcla norteafricana encontrado en la población canaria actual apoya el hecho de que aún existe una huella genética de la población aborigen. Curiosamente, para las muestras procedentes de La Gomera se obtuvo la estima más elevada de mezcla norteafricana, aunque sin mostrar diferencias estadísticamente significativas con las de otras islas. Sin embargo, dado que nuestro estudio incluyó un número reducido de individuos de cada isla, con tan sólo 7 representantes de La Gomera, estos resultados deben interpretarse con cautela, si bien solapan con los hallazgos proporcionados por estudios realizados con ADN mitocondrial (Rando *et al.*, 1999).

Un hallazgo interesante de esta tesis es el no haber detectado contribución genética subsahariana en la población canaria actual. Tal como se mostró en el apartado de resultados, esta ascendencia se pudo distinguir de manera adecuada de las poblaciones norteafricana y peninsular con los marcadores autosómicos empleados. Además, el hecho de no haber detectado influencia subsahariana en la población canaria no puede explicarse por la utilización de muestras de individuos nigerianos de etnia yoruba como única población de referencia, puesto que se comprobó los EuroAIMs son capaces de distinguir otras poblaciones subsaharianas disponibles en el HGDP (bantú, aka, mandinga, mbuti y san) (resultados no mostrados). Por tanto, confirmada la capacidad resolutive de los marcadores autosómicos empleados, el resultado obtenido contrasta con las evidencias previas acerca de la existencia de influencia subsahariana en la población canaria, sugerida por la bien documentada introducción de población subsahariana en Canarias como resultado del tráfico de esclavos posterior a la conquista (Flores *et al.*, 2001), y confirmada por los datos proporcionados por el estudio del ADN mitocondrial (Rando *et al.*, 1999) y de la porción no recombinante del cromosoma Y (Flores *et al.*, 2003). Sin embargo, estos resultados pueden explicarse por el distinto modo de herencia de estos marcadores, y por la compleja historia de la población canaria, de manera que los distintos marcadores proporcionan puntos de vista complementarios (Flores *et al.*, 2001; Fregel *et al.*, 2009). En definitiva, encontramos pocas evidencias que apoyen la existencia de una contribución subsahariana en la población canaria distinta de la que provendría del sustrato genético de los aborígenes canarios. En este sentido, varios estudios genéticos han demostrado la existencia de un sustrato subsahariano en restos arqueológicos de aborígenes canarios y en las poblaciones actuales del Sahara occidental (Fregel *et al.*, 2009; Rando *et al.*, 1998; Bosch *et al.*, 2001).

## 5.2. Implicaciones de la estructura genética de la población española en los estudios de asociación de casos-control

En vista de las estimas dispares de mezcla norteafricana encontradas entre poblaciones peninsulares y canarias, anticipamos que existiría mayor probabilidad de obtener falsos positivos en los estudios de asociación realizados con muestras de individuos españoles sin relación familiar, simplemente debido a las diferencias poblacionales (Marchini *et al.*, 2004). Esta premisa se confirmó al demostrar que el número de asociaciones que se detectaron como falsos positivos al comparar muestras de canarios con individuos peninsulares resultó ser más del doble de lo esperado por azar cuando no se realizó ningún ajuste de estratificación poblacional. Sin embargo, en esta tesis doctoral demostramos que en los estudios de la población española que incluyan muestras de individuos canarios, los efectos debidos a la estratificación poblacional podrían corregirse de manera eficiente usando PCA. En contraste, los ajustes basados en STRUCTURE no resultaron adecuados.

La distinta eficiencia en la corrección de la estratificación poblacional proporcionada por el programa STRUCTURE y por el PCA no es sorprendente. Como se comentó anteriormente, el PCA presenta un mayor poder discriminatorio para detectar diferencias entre poblaciones estrechamente relacionadas (Heath *et al.*, 2008; Li & Yu, 2008). De hecho, algunos estudios en los que se han comparado las estimas de ascendencia obtenidas con STRUCTURE frente a los resultados obtenidos con otros algoritmos, han mostrado que el STRUCTURE presenta sesgos en la estimación de los extremos de la distribución (Tsai *et al.*, 2005; Alexander *et al.*, 2009). Del mismo modo, en nuestro estudio observamos una buena correlación lineal entre el PC1 y las estimas de ascendencia norteafricanas obtenidas mediante el programa STRUCTURE, salvo en los extremos de la distribución, tendiendo a infra-estimar los valores en el rango bajo de ascendencia y a sobreestimarlos en el rango alto (resultados no mostrados). Lo expuesto anteriormente, así como el interés de utilizar un modelo libre de asunciones poblacionales subyacentes, nos llevó a escoger al PCA para los ajustes de estratificación poblacional y para estudiar la relación entre ascendencia y asma.

Es importante destacar que los resultados obtenidos no implican que las influencias norteafricanas que detectamos sean la única fuente de estructura poblacional en España. Aún así, nuestros resultados sí que señalan la importancia de controlar el efecto de las diferencias de ascendencia norteafricana en la comparación de individuos peninsulares y canarios dentro de un mismo estudio de asociación. Aunque la aproximación empleada puede verse como un simple ejercicio en el que se asume el peor escenario, nuestros resultados sugieren que al menos cuando los SNPs analizados en relación con la enfermedad presenten grandes diferencias en frecuencia alélica entre peninsulares y

norteafricanos, los estudios de asociación sufrirán un exceso de falsos positivos. Más aún, dado los tamaños muestrales requeridos para detectar la asociación de variantes genéticas con efectos leves en enfermedad (Chanock *et al.*, 2007), los sesgos estadísticos introducidos por la estratificación poblacional podrían incluso ser más pronunciados en los estudios reales (Pritchard & Rosenberg, 1999). Además, esta situación en la que todos los individuos enfermos proceden de una misma región, como por ejemplo Canarias, y los controles se muestreen en el resto del país será con el tiempo cada vez más habitual. Esto se debe a la existencia del Banco Nacional de ADN ([www.bancoadn.org](http://www.bancoadn.org)) o las fuentes de datos genómicos como la base de datos dbGaP (<http://www.ncbi.nlm.nih.gov/gap>), cuyo propósito es precisamente proporcionar muestras de individuos sanos para facilitar y abaratar los estudios de asociación. Precisamente, este fue el caso del estudio de asociación realizado en esta tesis, en el que los individuos enfermos fueron recogidos en su mayoría en Canarias (64%), mientras que los controles procedieron principalmente de bancos de sangre de centros peninsulares (71%). Sin embargo, incluso aquellos estudios que sólo incluyan muestras de casos y controles de la población canaria estarían expuestos a los sesgos originados por la estratificación poblacional, dadas las grandes diferencias de ascendencia estimadas entre los individuos de esta población.

Cabe destacar que al haber usado marcadores genéticos que muestran grandes diferencias alélicas entre poblaciones (Price *et al.*, 2008), las diferencias genéticas entre peninsulares y norteafricanos no representan la media del genoma al completo, por lo que podrían estar sesgadas. Además, nuestro estudio presenta las mismas limitaciones que cualquier estudio genético en el que se estimen proporciones de ascendencia en una población, ya que éstas dependen en gran medida de las poblaciones que se consideren como parentales o poblaciones de referencia. En este sentido, en esta tesis doctoral se emplearon muestras de distintas localidades de Marruecos para cuantificar la mezcla norteafricana en la Península y como referencia para las poblaciones aborígenes canarias, lo cual es razonable dadas las evidencias procedentes de registros históricos (Chejne, 1974; Flores *et al.*, 2001) y los estudios genéticos previos (Rando *et al.*, 1999; Flores *et al.*, 2001; Flores *et al.*, 2003; Maca-Meyer *et al.* 2004; Alonso *et al.*, 2005; Adams *et al.*, 2008; Fregel *et al.*, 2009). Sin embargo, el uso de esta muestra como representativa del norte de África no es más que una simplificación de las influencias africanas heterogéneas que han afectado a las poblaciones españolas a lo largo de su historia. Posiblemente, una mayor caracterización genética de las poblaciones norteafricanas y la utilización de un mayor número de marcadores genéticos permitan identificar en un futuro otras influencias genéticas en la población española, tal y como ha sido mostrado recientemente para las poblaciones afroamericanas (Tishkoff *et al.*, 2009).

### 5.3. Análisis de la relación entre la ascendencia norteafricana y el riesgo de asma en Canarias

Los resultados del primer artículo de esta tesis nos han permitido validar una herramienta con la que estimar adecuadamente la contribución genética norteafricana en la población canaria. Con ello, se exploró por primera vez la posible relación entre la ascendencia genética norteafricana y el riesgo a padecer asma y atopia en la población canaria, en busca de una explicación a la mayor prevalencia de fenotipos asmáticos en las Islas en comparación con la media española (The Spanish Group of the European Asthma Study, 1995; Juliá-Serdá *et al.*, 2005; Sánchez-Lerma *et al.*, 2009).

Dado que en el capítulo anterior habíamos demostrado que la reducción de marcadores de ascendencia hasta 69 EuroAIMs producía resultados equiparables al análisis del conjunto completo de los 93 marcadores, y teniendo en cuenta que disponíamos de un total de 83 marcadores que habían superado los controles de calidad de genotipado, estos últimos se consideraron adecuados para su utilización en la evaluación de la influencia genética norteafricana en la población canaria. En primer lugar, un PCA con estos 83 EuroAIMs permitió evaluar la asociación entre la ascendencia genética con el riesgo a padecer asma y distintos fenotipos atópicos, no encontrándose relación alguna al ajustar por variables clínicas, esto es al incluir en los modelos de asociación aquellas variables que diferían entre los grupos comparados. Del mismo modo, ninguno de los 83 EuroAIMs mostró individualmente asociación con asma tras realizar ajustes por las múltiples comparaciones realizadas.

El único trabajo previo en el que se había realizado una aproximación al estudio de la relación entre la ascendencia y el asma en la población canaria, se basó en la comparación entre individuos que declararon dos generaciones de ancestros nacidos en Canarias y los que declararon tener ancestros de otros lugares (Juliá-Serdá *et al.*, 2005). Los autores tampoco encontraron relación alguna entre tener ancestros canarios y el riesgo a sufrir asma.

Basado en la estima de la media de mezcla genética norteafricana en la población canaria ( $17\pm 25\%$ ) y en el número de casos y controles analizados, estimamos que nuestro estudio tendría un poder estadístico adecuado para detectar efectos de hasta un 21% de efecto estandarizado (esto es una diferencia de mezcla norteafricana de un 3,2%; 20,2% en casos y 17% en controles). Este valor de efecto sería incluso menor que el observado en poblaciones afroamericanas y afro-caribeñas, en las que se han encontrado diferencias del 5-6% en ascendencia africana entre asmáticos y controles (Vergara *et al.*, 2009; Flores *et al.*, 2011). Por tanto, llegado a este punto, no es posible afirmar que no exista relación entre

la ascendencia genética y el riesgo al asma o fenotipos atópicos en la población canaria. Tan sólo se podría concluir que, de existir tal relación, sus efectos serían moderados o leves y menores que los descritos en las poblaciones afroamericanas y afro-caribeñas, lo que se traduciría en pequeñas diferencias de mezcla norteafricana (<3,2%). No obstante, serán necesarios nuevos estudios con tamaños muestrales mayores y un mayor número de marcadores genéticos para explorar tal posibilidad.

### 5.4. Asociación de genes candidatos con el asma en la población española

La determinación de las variantes genéticas de susceptibilidad al asma es el primer paso necesario para la comprensión de las causas de esta enfermedad y su aplicación clínica (Hamburg & Collins, 2010). Uno de los desafíos de la medicina personalizada es identificar adecuadamente a los individuos antes de que desarrollen los síntomas de enfermedad, lo que permitiría el establecimiento de medidas preventivas, además de cambiar la actitud terapéutica para hacerla más específica y económica (Kraft, 2011). Para ello, primero es necesario establecer los genes de susceptibilidad firmemente asociados con la enfermedad, siendo la replicación en estudios independientes con un gran número de muestras de individuos y una caracterización fenotípica adecuada, la única herramienta disponible para evaluar la validez de las asociaciones detectadas en un estudio (Chanock *et al.*, 2007; Nicolae & Ober, 2009).

En este trabajo se han analizado aproximadamente 1.950 muestras de individuos de población española para más de 220 polimorfismos, correspondientes a nueve regiones génicas candidatas, en asociación con la susceptibilidad al asma y al asma atópico. Entre los genes candidatos estudiados, se encontró asociación de los genes *IL13-IL4*, *LTA-TNF*, *IRAK3*, *MS4A2*, *IL4R* y *ADAM33* con el desarrollo de fenotipos asmáticos, tanto de variantes asociadas en estudios previos, como de polimorfismos hasta ahora no citados en trabajos anteriores. Por el contrario, otras asociaciones previamente descritas para los genes *ADRB2* y *CD14* no fueron validadas en nuestro estudio.

Por otra parte, nuestros resultados han puesto de manifiesto el efecto de la edad de comienzo de la enfermedad en la asociación de algunos de los genes, mostrando mayores efectos al incluir en los análisis sólo aquellos asmáticos con una edad de diagnóstico menor que un umbral estimado. Más aún, la acumulación de variantes de riesgo mostró una fuerte asociación con el desarrollo de asma y asma atópico y una relación inversa con la edad de comienzo de la enfermedad.

La mayoría de estudios de asociación de asma se han centrado en poblaciones con ascendencia europea, principalmente del norte, pese a que se ha sugerido la importancia de realizarlos en todas las poblaciones mundiales para así descubrir los *loci* genéticos que



contribuyen en la variabilidad fenotípica (Pulit *et al.*, 2010). Asimismo, la realización de estudios en poblaciones con mezcla de ascendencia africana reciente se ha señalado como un factor que probablemente mejoraría el descubrimiento de nuevas variantes de riesgo, ya que ello permite refinar la señal de asociación o incluso descubrir las variantes causales (Manolio *et al.*, 2009). Por ello, la presencia de mezcla genética norteafricana en la población española, confirmada en esta tesis, justifica el interés que tiene la realización de estudios de asociación en esta población.

#### 5.4.1. Replicación de la asociación de *IRAK3* con el asma

En esta tesis se realizó el primer estudio de replicación independiente de la asociación de variantes génicas de *IRAK3* con el asma, encontrándose los mismos SNPs y alelos de riesgo y el mismo haplotipo protector que en el estudio que permitió su clonaje posicional en poblaciones italianas (Balaci *et al.*, 2007). Los resultados obtenidos ampliaron las evidencias anteriores al encontrar dos señales de asociación independientes dentro del gen y al extender la región asociada por medio de haplotipos. En este último caso, se demostró incluso que es posible detectar la asociación de *IRAK3* considerando el conjunto total de asmáticos, sin distinguir entre atópicos y no atópicos o casos con comienzo temprano.

Es interesante destacar que, como parte de esta tesis, *IRAK3* también se asoció por primera vez con la susceptibilidad al ALI producido como consecuencia de la sepsis. De confirmarse en estudios independientes, este gen se sumaría a la lista de genes que participan en la interacción con patógenos con efectos pleiotrópicos en la susceptibilidad al asma y a la sepsis o sus complicaciones, como lo son *CD14* (Gu *et al.*, 2008; Little *et al.*, 2009) y *NOD2* (Weidinger *et al.*, 2005; Henckaerts *et al.*, 2009). Estos resultados apoyan la hipótesis de la existencia de factores de riesgo genéticos compartidos entre distintas enfermedades complejas y refuerza la importancia de la hipótesis de higiene en el desarrollo del asma (Strachan, 1989).

La realización de un meta-análisis de los estudios de tipo caso-control que han analizado la asociación de variantes de *IRAK3* con el asma, demostró la presencia de efectos consistentes sólo entre las muestras de ascendencia europea, coincidente con lo observado en estudios previos (Rogers *et al.*, 2009; Wu *et al.*, 2010), pero no al incluir muestras japonesas en el análisis (Nakashima *et al.*, 2006). Por tanto, futuros estudios deberán analizar la importancia de *IRAK3* en la susceptibilidad al asma en otras poblaciones no europeas.

### 5.4.2. Replicación de la asociación de genes candidatos de susceptibilidad al asma más robustamente replicados

En el análisis de asociación de los 8 genes más robustamente replicados en la bibliografía, se encontró asociación de 37 SNPs a nivel de significación nominal, 18 de los cuales constituyeron replicaciones de estudios anteriores con distintos fenotipos de asma (Ober & Hoffjan, 2006; Daley *et al.*, 2009).

Este no es el primer estudio que intenta validar la asociación de genes candidatos previamente asociados empleando un mayor número de muestras y cubriendo de manera adecuada el gen de interés. El estudio de Ferreira *et al.* (2009) se centró en validar la asociación de 8 genes candidatos situados en regiones cromosómicas que habían mostrado ligamiento al asma y considerando el gen como unidad de replicación (Neale & Sham, 2004), lo que les permitió encontrar asociación entre polimorfismos del gen *MS4A2* y niveles séricos de IgE. Por su parte, Michel *et al.* (2010) analizaron 6 de los genes estudiados en esta tesis y sus resultados fueron congruentes con los nuestros: *IL13*, *IL4R*, *TNF* y *ADAM33* replicaron las asociaciones descritas en estudios anteriores mientras que *CD14* e *IL4* no lo hicieron. Otra estrategia de estudio de la replicación de los genes asociados con el asma ha sido el empleo de datos de GWAS para la evaluación específica de genes candidatos (Daley *et al.*, 2009; Rogers *et al.*, 2009; Wu *et al.*, 2010; Galanter *et al.*, 2011). En algunos casos, como veremos más adelante, sus resultados solapan con los obtenidos en la presente tesis.

En el gen *IL4R*, las asociaciones más significativas se centraron en SNPs localizados entre los exones 10 al 12, codificantes de los dominios intracelulares del receptor (Kruse *et al.*, 1999a). Además, muchos de estos SNPs se correspondieron con cambios de aminoácido que podrían afectar a la unión de moléculas efectoras de la vía de señalización de este receptor y a su fosforilación (Hershey *et al.*, 1997). En concreto, dos de los cambios de aminoácido asociados con asma en esta tesis (Ser478Pro [rs1805015] y Gln551Arg [rs1801275]) han mostrado tener relación con la fosforilación de las moléculas que se unen al dominio intracelular de la *IL4R* en ensayos *in vivo* cuando los alelos señalados como protectores en esta tesis doctoral están presentes (Kruse *et al.*, 1999b). En individuos portadores de los dos alelos protectores, prolina y arginina, se ha comprobado que se produce mayor fosforilación de las moléculas de señalización IRS-1 e IRS-2, debidos a cambios en la carga del receptor y en su conformación, conduciendo a una menor unión de sus sustratos que se encargan de continuar con la vía de señalización del receptor corriente abajo (Hershey *et al.*, 1997). Esta menor señalización conduciría finalmente a una menor activación de las células B, desencadenando una menor producción de IgE (Kruse *et al.*, 1999b). Precisamente, un GWAS reciente con niveles de IgE ha mostrado que los alelos protectores de este estudio se asocian con una menor concentración sérica de IgE, con una

significación cercana a los umbrales genómicos ( $p \leq 7 \times 10^{-6}$ ) (Granada *et al.*, 2011). Sin embargo, es sorprendente que los alelos que en nuestro estudio resultaron ser protectores hayan mostrado el mismo efecto sólo en unos pocos estudios (Genuneit *et al.*, 2009; Rogers *et al.*, 2009). Por el contrario, estos alelos protectores han revelado efectos totalmente contrarios en otros estudios (Michel *et al.*, 2010), incluso en un meta-análisis del polimorfismo Gln551Arg que reunió datos de 13 estudios independientes (Loza & Chang, 2007). Precisamente, en ese meta-análisis, un estudio realizado en la población española fue descartado por mostrar efectos protectores como los mostrados en la presente tesis (Isidoro-García *et al.*, 2005a; Loza & Chang, 2007). Es más, al analizar conjuntamente los resultados de asma atópico de dicho estudio y el nuestro, los efectos de este SNP resultaron consistentes, dado que mejoraron la significación obtenida en cada estudio por separado. Este no es el primer gen para el cual se encuentra asociación de un mismo polimorfismo pero con efectos contrarios de los alelos asociados en diferentes poblaciones (Ober & Hoffjan, 2006; Vercelli, 2008). Este fenómeno, denominado *flip-flop*, se ha observado incluso en algunos GWAS, en los que determinados polimorfismos resultaron ser de riesgo o protección dependiendo de la población analizada (Sleiman *et al.*, 2010). Por lo tanto, la asociación específica encontrada para la *IL4R* en la población española podría estar causada por frecuencias alélicas propias de esta población, distintos patrones de LD, y/o por exposiciones ambientales diferentes entre poblaciones. Esta observación refuerza la importancia de estudiar distintas poblaciones para obtener un catálogo de genes de importancia en cada una de ellas (Ober & Hoffjan, 2006; Manolio *et al.*, 2009), y la necesidad de no basar los esfuerzos de replicación únicamente en el estudio de SNPs individuales (Neale & Sham, 2004).

En la región génica *IL13-IL4*, los SNPs asociados pertenecieron al gen *IL13*, precisamente el único de los genes candidatos analizados que ha mostrado significación genómica en un GWAS de asma (Moffatt *et al.*, 2010). Concretamente, los 3 SNPs que replicamos en este estudio (rs1881457, rs1800925, rs1295686) se correspondieron con tres de los SNPs más significativos de un GWAS de IgE (Granada *et al.*, 2011). Además, el SNP rs1295686 se asoció también en un estudio a gran escala de genes candidatos y un GWAS de asma (Daley *et al.*, 2009; Moffatt *et al.*, 2010).

Por su parte, el único polimorfismo del gen *MS4A2* asociado en esta tesis doctoral, el SNP rs569108, replicó resultados de estudios previos (Shirakawa *et al.*, 1996; Green *et al.*, 1998; Cui *et al.*, 2003; Kim *et al.*, 2007; Wang *et al.*, 2009). Sin embargo, los estudios anteriores no habían analizado todas las variantes comunes de este gen con las que dilucidar si existían otras variantes asociadas con asma. En la presente tesis doctoral se mostró que este SNP no presentó LD con ningún otro analizado, por lo que este SNP podría

## DISCUSIÓN

---

ser causal de la asociación observada en el gen *MS4A2*. Este SNP predice la presencia de un cambio de aminoácido no sinónimo Glu237Gly en una región génica altamente conservada (Kim *et al.*, 2007) y representaría un cambio de un aminoácido cargado negativamente (glutámico) por otro aminoácido polar no cargado (glicina). Este cambio de carga supondría la alteración de la hidrofobicidad en el extremo C-terminal del receptor FCER1B en una posición adyacente a un motivo de activación del inmunoreceptor por tirosina (ITAM). El dominio ITAM permite la unión de tirosinas quinasas para producir fosforilaciones, que a su vez provocan la activación celular, transmitiéndose una señal intracelular de calcio esencial para la liberación de vesículas con mediadores inflamatorios (Cookson *et al.*, 1989; Scharenberg & Kinet, 1994; Lin *et al.*, 1996). Estos mediadores son responsables, en última instancia, de la hiper-respuesta de la vía aérea (Kim *et al.*, 2007). Por tanto, el cambio de aminoácido podría alterar esta unión y en consecuencia el nivel de señalización del FCER1B (Cookson *et al.*, 1989). Se ha demostrado que la liberación de histamina por parte de basófilos de pacientes asmáticos estimulados con IgE resultó ser mayor en aquellos individuos portadores del alelo Glu237, que en los que presentan el alelo Gly237 (Kim *et al.*, 2002), que resultó protector en esta tesis doctoral. Otros estudios han mostrado asociación de otros polimorfismos de este gen (Genuneit *et al.*, 2009; Galanter *et al.*, 2011), e incluso al comparar dos poblaciones distintas dentro del mismo estudio (Galanter *et al.*, 2011). Estos resultados podrían estar motivados porque en las poblaciones analizadas en esos estudios los SNPs asociados presenten LD con el SNP rs569108, o que este SNP no sea realmente el causal de la asociación encontrada.

En cuanto a la región génica *LTA-TNF*, dos de los SNPs asociados, correspondientes a la región promotora del *TNF* (-863 C/A [rs1800630] y -1031 C/T [rs1799964]), han mostrado modular los niveles de expresión de IgE y TNF en suero (Sharma *et al.*, 2006). De ellos, el SNP -863 C/A mostró mediante ensayos de movilidad electroforética una unión específica al factor de transcripción NF- $\kappa$ B (Sharma *et al.*, 2006). Además, un haplotipo formado por la combinación de dos SNPs, rs2239704 (LTA+80) y rs746868 (LTA+368), que previamente había mostrado efectos funcionales en la producción de LTA (Knight *et al.*, 2004), mostró asociación significativa con asma ( $p=0,007$ ). Concretamente, el haplotipo AC resultó ser menos frecuente en asmáticos (38,4%) que en controles (42,0%), OR=0,86; 95% CI=0,74-1,00  $p=0,039$ . Precisamente, este haplotipo protector mostró en estudios anteriores una menor accesibilidad a la RNA polimerasa II, una menor abundancia del ARN transcrito para la *LTA* y, en definitiva, una menor producción de la proteína LTA (Knight *et al.*, 2004; Knight *et al.*, 2003).

### 5.4.3. Efecto de la edad de comienzo del asma

Se ha apuntado que la heterogeneidad fenotípica del asma en cuanto a la edad de comienzo, presencia de atopia, función pulmonar y respuesta a tratamiento no deberían ignorarse en los estudios de asociación genéticos (Meyers, 2010). Entre estos factores, se conoce que la edad de comienzo determina distintas características de la enfermedad (Gelfand, 2008), y ya había sido sugerida como causante de la falta de replicación de asociaciones genéticas importantes en la enfermedad (Lasky-Su *et al.*, 2008). Por ello, en este estudio se re-evaluaron las asociaciones en función de la edad de diagnóstico, obteniéndose un umbral de edad de comienzo que maximizase los efectos encontrados. El re-análisis de la asociación considerando aquellos asmáticos con una edad de comienzo anterior al umbral determinado, mejoró la replicación de ciertos genes, tal y como había sido observado en trabajos previos para otros genes (Bouzigon *et al.*, 2008; Halapi *et al.*, 2010; Castro-Giner *et al.*, 2010), lo cual refuerza la idea de la complejidad de las asociaciones con el asma. En concreto, la asociación de los genes *IL4R*, *LTA-TNF*, *MS4A2* e *IRAK3*, mejoró principalmente en cuanto a los efectos. Sin embargo, el cambio más drástico se observó en el caso de *ADAM33*, para el cual aumentó considerablemente el número de SNPs replicados. Precisamente este gen ha mostrado una enorme heterogeneidad en los SNPs asociados en distintas poblaciones (Ober & Hoffjan, 2006; Vercelli, 2008). Por tanto, el efecto de la consideración de la edad de comienzo en los estudios podría explicar, al menos en parte, los problemas de replicación observados para éstos y otros genes.

Dado que la edad de comienzo no se registró en nuestro estudio para un gran número de asmáticos, pero sí la de diagnóstico, la variable empleada en los análisis de adición secuencial fue esta última. Por ello, es probable que los efectos se relacionen con puntos de corte de la edad de diagnóstico del asma que se traduzcan en edades más tempranas en términos de comienzo, sin poder obtener una estima precisa de la misma. No obstante, es sorprendente que en muchos casos la edad que maximizó los efectos se encuentre en el rango de los 20 a los 45 años, lo cual coincide con la edad de máxima expresión de la enfermedad (Barbee & Murphy, 1998). Con independencia del establecimiento de un valor exacto de edad de comienzo, nuestros resultados sugieren el interés de tener en cuenta este efecto en el diseño y análisis de futuros estudios de asociación.

### 5.4.4. Cobertura de la variación de los genes candidatos en las plataformas de genotipado en los GWAS de poblaciones europeas

Dado que los genes asociados en nuestro estudio no han sido asociados en los GWAS de asma, salvo por el caso de *IL13* (Moffatt *et al.*, 2010), en esta tesis doctoral nos planteamos si la cobertura de las variantes de los genes analizados estaría adecuadamente representada en las plataformas de genotipado empleadas en dichos estudios, centrándonos en las poblaciones europeas. En adelante, la cobertura se referirá a las variantes comunes del gen ( $MAF \geq 5\%$ ), entendida como aquellas que pudiesen ser analizadas en asociación con la enfermedad por haber sido directamente genotipadas o por la posibilidad de imputarlas. Un estudio previo ya había señalado la existencia de una mala cobertura para algunos de los SNPs asociados en genes de importancia en el asma, como *ADAM33*, *IL4* y *CD14* (Michel *et al.*, 2010), en el *array* de genotipado empleado en el primer GWAS de asma (Moffatt *et al.*, 2007). Sin embargo, ningún estudio anterior había realizado una evaluación sistemática de la cobertura de todos los *arrays* empleados en los GWAS de asma en poblaciones europeas (Moffatt *et al.*, 2007; Hancock *et al.*, 2009; Himes *et al.*, 2009; Mathias *et al.*, 2010; Sleiman *et al.*, 2010; Li *et al.*, 2010a; Moffatt *et al.*, 2010; Ferreira *et al.*, 2011b; Ferreira *et al.*, 2011a; Torgerson *et al.*, 2011; Hirota *et al.*, 2011).

Nuestra evaluación señaló que para los genes analizados en el presente trabajo, las plataformas de genotipado empleadas en los GWAS de asma proporcionarían una cobertura que podría parecer adecuada sobre datos del HapMap, en los cuales se basó el diseño de la mayoría de *arrays*. Sin embargo, al examinar su cobertura teórica sobre datos del 1KGP, ésta resultó más limitada. Aún asumiendo de manera optimista un CR del 100%, la cobertura teórica resultó sub-óptima para todos los genes asociados en este estudio, salvo para el gen *IL4R*, siendo totalmente inadecuada para el gen *ADAM33* (<30%). Para casi todos los *loci* asociados en este estudio, el único *array* que teóricamente proporcionaría una cobertura teórica mayor (Illumina HumanHap 1M-Duo), sólo se ha empleado en una parte de las muestras de un GWAS (Torgerson *et al.*, 2011). Precisamente al no emplear el mismo *array* para el genotipado de todas las muestras, en dicho estudio el análisis de asociación se ciñó a las variantes solapantes entre los distintos *arrays* utilizados, tras imputar las 2,5 variantes millones del genoma presentes en los datos de la fase II del HapMap (Torgerson *et al.*, 2011). En el caso de *IRAK3*, pese a que dicho *array* mostró una cobertura teórica del 71%, tras pasar los controles de calidad del genotipado y de la imputación, el solape de los datos de los dos *arrays* sólo permitió la evaluación de la asociación de dos variantes de *IRAK3*, y ninguna de ellas fue el SNP rs1370128, que mostró una significación cercana al umbral genómico en nuestro estudio (Dra. Carole Ober, University of Chicago, comunicación personal). Esto supone que el único GWAS de asma que ha empleado el *array* con la

cobertura teórica más adecuada, mostró una cobertura sobre las variantes comunes de *IRAK3* de tan sólo un 4,2%. Consecuentemente, es de esperar que para algunos de los genes asociados en este estudio la cobertura alcanzada sea también inferior a la estimada.

Dado que los GWAS podrían estar perdiendo poder de detección de genes con gran importancia, dada su cobertura limitada para algunas regiones génicas y los estrictos umbrales de significación establecidos en este tipo de estudios, nuestros resultados confirman la importancia que siguen desempeñando los estudios de asociación en genes candidatos, en los que se examine exhaustivamente la variación de los genes analizados.

#### **5.4.5. Fortalezas y limitaciones del estudio de asociación**

Esta tesis representa el mayor estudio de asociación entre variantes comunes en genes candidatos y el asma llevado a cabo en la población española, al realizarse en una muestra siete veces mayor que las que se han empleado como media en esta población y analizar un gran número de variantes genéticas. Es más, incluso si lo comparamos con los estudios que han examinado los mismos genes candidatos a nivel mundial, el número de muestras analizado ha sido mayor que el utilizado en el 97% de los mismos, siendo casi cinco veces mayor que el tamaño medio empleado en ellos. Asimismo, el número total de variantes génicas examinadas en los previos estudios no ha superado los 100 SNPs considerándolos todos en conjunto (Daley *et al.*, 2009; Ober & Hoffjan, 2006). Aún así, el poder estadístico alcanzado en la presente tesis fue óptimo para la detección de ORs situados en el límite superior de la estimación de los efectos esperados para la asociación de variantes comunes en enfermedades complejas (Hindorff *et al.*, 2009), por lo que variantes de riesgo con efectos menores podrían haber pasado desapercibidas.

En nuestro estudio se descartó la existencia de errores sistemáticos durante el genotipado, mediante varios tipos de controles de calidad, así como los posibles efectos de confusión debidos a la presencia de estratificación poblacional.

Para el análisis de asociación se realizó una evaluación exhaustiva de la asociación de variantes génicas de los genes candidatos cubriendo casi por completo la variación común presente en estos genes, al analizar tSNPs seleccionados en base a datos de re-secuenciación y no del HapMap, en contraposición con la mayoría de estudios en los que sólo se han analizado polimorfismos situados en regiones promotoras, polimorfismos que predicen cambios de aminoácidos, etc (Ober & Hoffjan, 2006; Daley *et al.*, 2009). La selección de tSNPs permitió también la realización de imputación de variantes no genotipadas, en cuyo proceso se minimizó la posibilidad de cometer errores. Para ello, en el artículo de asociación de *IRAK3* con ALI pudimos determinar qué variantes podrían ser imputadas de manera fiable, al haber empleado datos de re-secuenciación de la población

española para la comprobación de la exactitud de la imputación de las frecuencias alélicas. Nuestros resultados mostraron que no sólo la capacidad informativa de los tSNPs, medida como correlación multilocus  $M_D$  (Nicolae, 2006), es importante en la imputación, sino que más aún lo es la frecuencia alélica de las variantes imputadas (Huang *et al.*, 2009). Al menos con nuestros datos, las variantes con  $MAF \geq 10\%$  resultaron imputadas con errores menores que los cometidos al considerar las variantes con  $M_D \geq 0,7$ . Además, aplicando ambos umbrales simultáneamente los errores en la imputación de polimorfismos se redujeron drásticamente, lo que nos llevo a adoptar este criterio para los estudios de asociación del resto de genes. Es interesante destacar que en el caso de la asociación de *IRAK3* con ALI, la asociación predicha mediante imputación de una variante del promotor del gen fue confirmada empíricamente tras su genotipado, revelando una variante funcional que no hubiese sido detectada de no ser por el proceso de imputación.

Como ya se ha comentado anteriormente, pese a que *ADRB2* y *CD14* han sido señalados como genes robustamente asociados con el asma (Ober & Hoffjan, 2006; Vercelli, 2008), en el presente estudio su asociación con asma no fue confirmada. Si bien este resultado pudo deberse a limitaciones en el tamaño muestral, otros factores también podrían haber influido en la falta de replicación de la asociación. En primer lugar, el fenotipo analizado podría tener gran importancia, ya que en este estudio no se registraron las variables que se han visto relacionadas con la asociación de estos dos genes, como la respuesta al tratamiento (Lee *et al.*, 2004; Contopoulos-loannidis *et al.*, 2005; Finkelstein *et al.*, 2009), y la presencia de asma nocturno (Contopoulos-loannidis *et al.*, 2005) en el caso de *ADRB2*, o variables relacionadas con las exposiciones ambientales en el caso de la asociación de *CD14* (Simpson *et al.*, 2006), que podrían alterar los patrones de metilación de este gen (Munthe-Kaas *et al.*, 2010). Otro de los fenotipos que ha mostrado asociación con el gen *ADRB2* es el asma grave (Contopoulos-loannidis *et al.*, 2005). Si bien la gravedad de cada paciente se registró en nuestro estudio siguiendo las guías GINA, dado que sólo 89 del total de asmáticos de la muestra fueron clasificados como asmáticos graves, el poder estadístico proporcionado sería muy limitado (<40%), lo que llevó a no evaluar la asociación de polimorfismos con este fenotipo. Sin embargo, una exploración de la correlación entre la gravedad según las categorías del GINA y la edad de comienzo indicó que estas dos variables se encuentran asociadas significativamente ( $p=1.84 \times 10^{-7}$ ), de manera que una edad más temprana de comienzo se correlacionó con una mayor gravedad de la enfermedad.

En segundo lugar, podría ser que la asociación de los polimorfismos de los genes *ADRB2* y *CD14* con asma esté regida por un modelo genético distinto del modelo considerado en este estudio (Zhao & Bracken, 2011). En nuestro trabajo empleamos un



modelo aditivo de herencia, que se ha sugerido que subyace a la mayoría de asociaciones con enfermedades complejas, y proporciona un mayor poder estadístico para detectar asociación, aunque podría no ser tan adecuado si el modelo de herencia es recesivo (Balding, 2006). Una tercera alternativa podría ser que estos genes sólo muestren asociación en un grupo poblacional concreto, como por ejemplo en individuos asiáticos, como se ha propuesto en un meta-análisis (Zhang *et al.*, 2011). Por último, quizá los genes *ADRB2* y *CD14* no tengan una implicación real en la susceptibilidad al asma, como han sugerido otros meta-análisis (Contopoulos-Ioannidis *et al.*, 2005; Kedda *et al.*, 2005; Nishimura *et al.*, 2006; Li *et al.*, 2010b).

Otras limitaciones de este estudio de asociación merecen ser mencionadas. En primer lugar, dada la presencia de múltiples SNPs asociados en cada gen, que en muchos casos mostraron un LD moderado-alto, la variante causal de la asociación encontrada en cada gen no pudo ser determinada. Esto podría ser especialmente importante en el caso del *locus LTA-TNF*, localizado en la región 6p21 a una distancia de 1 megabase del *locus HLA-DRB1*. Sin embargo, la exploración de este efecto se descartó por medio del genotipado de variantes dentro de este segundo *locus*, observándose la independencia de la asociación entre estas dos regiones (resultados no mostrados), como ya había sido encontrado en el caso de la asociación de estos dos *loci* con lepra (Alcais *et al.*, 2007). También podría ocurrir que fueran variantes de baja frecuencia las causantes de las asociaciones encontradas, quizás por su correlación con las variantes comunes mediante un fenómeno denominado “asociaciones sinténicas” (Dickson *et al.*, 2010), para las que este estudio presentó un poder estadístico limitado. Alternativamente, la causalidad podría deberse a otro tipo de variación más compleja, por ejemplo los CNVs, que podrían subyacer a las asociaciones detectadas mediante SNPs por su correlación con ellos (Manolio *et al.*, 2009).

En este estudio, la selección de controles no se realizó asegurando un emparejamiento de cada caso con un control en términos de edad y sexo, cuya influencia en los estudios de asociación es motivo de debate (Zeggini & Morris, 2011). En su lugar, se le dio más importancia a que los controles perteneciesen a la población española y que presentasen una baja probabilidad de presentar asma o enfermedades relacionadas y que tampoco tuviesen antecedentes familiares. Esto condujo a que los casos y controles presentasen diferencias significativas para las variables edad y sexo. Sin embargo, es muy poco probable que los resultados presentados sean espurios, dado que representan repeticiones de asociaciones anteriores. Asimismo, y aunque no se ha especificado en los artículos surgidos de esta tesis doctoral, muchos de los análisis fueron repetidos estratificando por género y por cuartiles de edad, encontrándose resultados consistentes entre los grupos comparados, al mantenerse las frecuencias alélicas constantes entre

grupos. Esto es congruente con la observación de que en presencia de HWE los genotipos de un determinado *locus* son independientes de otros factores (Clayton & McKeigue, 2001). Por ello, y puesto que el hecho de ajustar por variables innecesarias podría disminuir el poder de detección de asociación (Cox & McCullagh, 1982), en este estudio optamos por ajustar los resultados sólo por la presencia de estratificación poblacional, que sí se ha demostrado ser un factor de confusión en la detección de asociación (Amos, 2007; Zondervan & Cardon, 2007). De hecho, esta estrategia es una práctica habitual en los GWAS, en los que muchas veces se emplean como controles individuos tomados de la población que ya han sido genotipados para otros estudios distintos, sin conocer datos de edad o sexo e incluso sin disponer de una caracterización clínica (Li *et al.*, 2010a; Ferreira *et al.*, 2011a), y se da más importancia a que cada caso esté genéticamente emparejado con un control (Guan *et al.*, 2009).

### **5.4.6. Capacidad predictiva de las variantes analizadas y varianza de la enfermedad explicada por la acumulación de variantes de riesgo**

Dado que los efectos encontrados en la asociación de variantes genéticas en enfermedades complejas han mostrado ser muy leves, la capacidad predictiva de cada uno de ellos considerados por separado sería insignificante (Holtzman & Marteau, 2000; Vineis *et al.*, 2001). Por ello, se ha propuesto la elaboración de perfiles genéticos de riesgo definidos por distintas variantes genéticas asociadas en genes independientes, que han mostrado utilidad en estudios de simulación y en enfermedades como la degeneración macular (Janssens *et al.*, 2006; Maller *et al.*, 2006).

En el presente trabajo se mostró que cuanto mayor fuese el número de variantes de riesgo acumuladas en cuatro genes de susceptibilidad, mayor sería el riesgo a padecer asma y asma atópico y más temprano ocurriría el diagnóstico de la enfermedad. A pesar de los sorprendentes niveles de significación obtenidos en estas asociaciones, el poder discriminatorio resultó limitado, con AUCs entre 0,56 y 0,61, dependiendo del fenotipo considerado. Aunque no aparece comentado en los artículos, al incluir el gen *IRAK3* en el GRS de asma atópico, el poder discriminatorio alcanzó un valor de AUC de 0,63. Estos resultados fueron sorprendentemente altos si se consideran en el contexto de los estudios previos que han elaborado puntuaciones de riesgo genético a partir de datos de GWAS de asma, en los que el AUC obtenido ha sido de 0,57-0,58 (Moffatt *et al.*, 2010; Ferreira *et al.*, 2011a), independientemente del número de SNPs considerados. Sin embargo, es de esperar que el AUC proporcionado por el GRS de este estudio sea menor en una muestra independiente, tal y como se ha observado en estudios anteriores (Eleftherohorinou *et al.*, 2009).

Dado que en todos los casos la clasificación de los individuos mejoró respecto a lo esperado por azar, estos datos sugieren que las asociaciones encontradas en este estudio representan asociaciones genuinas similares a las reveladas mediante los GWAS. Sin embargo, dado que la varianza de la enfermedad explicada por los genes asociados en esta tesis no superó un 4,5%, incluso en el mejor modelo que incluyó a *IRAK3*, muchos factores genéticos quedan aún por determinarse.

Para otras enfermedades sí que se han encontrado modelos acumulativos de variantes de riesgo en genes candidatos con un mayor poder discriminatorio, como por ejemplo para la diabetes tipo I (AUC=0,78) o la enfermedad de Crohn (AUC=0,77) (Evans *et al.*, 2009). No obstante, hay que tener en cuenta que la capacidad discriminativa teórica máxima que puede proporcionar un conjunto de variantes genéticas comunes depende de la heredabilidad de la enfermedad (Janssens *et al.*, 2006). Así, cuanto mayor sea la heredabilidad de una enfermedad, mayor valor predictivo podrán mostrar las variantes genéticas (Janssens *et al.*, 2006). De esta manera, una enfermedad con una heredabilidad del 50% tendría un máximo teórico de AUC=0,8 (Evans *et al.*, 2009). Algo mayor sería el AUC máximo teórico para el asma, con una heredabilidad estimada en un 60% (Duffy *et al.*, 1990). Esto indica que será necesario tener en cuenta otros factores, como la edad, sexo o exposiciones ambientales, para llevar a cabo una predicción adecuada de la presencia de enfermedad. En este sentido, un GWAS en el que no sólo se incluyeron variantes genéticas, sino que también se consideraron conjuntamente factores clínicos y características demográficas, mostró valores predictivos que variaron desde AUC=0,57, al incluir los 10 SNPs con asociación más significativa, hasta AUC=0,66 al considerar los 160 SNPs más significativos (Xu *et al.*, 2011).

En definitiva, dada la baja heredabilidad explicada hasta el momento por los genes asociados con el asma y su baja capacidad discriminativa entre asmáticos y controles, serán necesarios más estudios en esta disciplina para disponer de una predicción de la enfermedad basada en variantes genéticas (Meyers, 2010).

## ***6. CONCLUSIONES***

---

## 6. CONCLUSIONES

1. Al menos 23 EuroAIMs, seleccionados por su capacidad para distinguir poblaciones del noroeste y sureste de Europa, fueron útiles para diferenciar ascendencia norteafricana y europea en individuos españoles.
2. El porcentaje de mezcla norteafricana estimado en peninsulares fue menor (5%) que el observado en canarios (17%), en los que además se observó una gran variación interindividual (rango de 0-96%). Estos datos apoyan la existencia de un rastro genético de la población aborígen en la población canaria actual. Sin embargo, la mezcla subsahariana fue indetectable en esta población.
3. Se confirmó que las diferencias genéticas de la población canaria con respecto a la peninsular podrían generar falsos positivos en los estudios de asociación caso-control que se realicen en la población española. No obstante, este efecto fue corregido adecuadamente mediante estimas de ascendencia proporcionadas por el conjunto reducido de EuroAIMs propuesto en este trabajo.
4. No se encontró asociación entre la ascendencia genética y el asma o fenotipos relacionados en la población canaria. A falta de estudios que confirmen esta observación, los resultados de este trabajo sugieren que de existir una influencia de la ascendencia, sus efectos en el riesgo serían leves.
5. La asociación de *IRAK3* mostró efectos pleiotrópicos en el desarrollo de fenotipos asmáticos y de la lesión pulmonar aguda derivada de la sepsis, como lo han hecho anteriormente otros genes involucrados en la interacción patógeno-hospedador. Asimismo, se replicó la asociación del gen *IRAK3* con asma atópico en la población española, además de ampliarse a otras variantes nunca antes asociadas.
6. De las 194 variantes comunes estudiadas en los genes de susceptibilidad al asma más robustamente asociados, un 24% mostró asociación con alguno de los fenotipos asmáticos analizados, y de ellas, el 45% constituyeron replicaciones de estudios anteriores. Las variantes asociadas pertenecieron a los genes *IL4R*, *ADAM33*, *MS4A2*, y a las regiones *LTA-TNF* e *IL13-IL4*.
7. Se demostró que la edad de diagnóstico del asma es un factor importante a considerar en los análisis de asociación. Esto se confirmó particularmente para el gen *ADAM33*, para el que mejoró la capacidad de replicación de resultados previos, y permitió la asociación de nuevas variantes genéticas.
8. La carga genética de predisposición al asma se relacionó con un diagnóstico más temprano de la enfermedad.

## CONCLUSIONES

---

9. Pese a no mostrar valor diagnóstico adecuado, la acumulación de variantes de riesgo presentes en los distintos genes replicados mostró una fuerte asociación con fenotipos asmáticos, proporcionando hasta 3,5 veces más riesgo en individuos portadores de 6 o más alelos de susceptibilidad.

## ***7. BIBLIOGRAFÍA***

---

## 7. BIBLIOGRAFÍA

- Adams, S.M., Bosch, E., Balaesque, P.L., Ballereau, S.J., Lee, A.C., Arroyo, E., López-Parra, A.M., Aler, M., Grifo, M.S., Brión, M., Carracedo, A., Lavinha, J., Martínez-Jarreta, B., Quintana-Murci, L., Picornell, A., Ramón, M., Skorecki, K., Behar, D.M., Calafell, F. & Jobling, M.A. (2008) The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *Am J Hum Genet*, **83**, 725-36.
- Aggarwal, B.B. (2003) Signalling pathways of the TNF superfamily: a double-edged sword. *Nat Rev Immunol*, **3**, 745-56.
- Ahmadi, K.R., Weale, M.E., Xue, Z.Y., Soranzo, N., Yarnall, D.P., Briley, J.D., Maruyama, Y., Kobayashi, M., Wood, N.W., Spurr, N.K., Burns, D.K., Roses, A.D., Saunders, A.M. & Goldstein, D.B. (2005) A single-nucleotide polymorphism tagging set for human drug metabolism and transport. *Nat Genet*, **37**, 84-9.
- Alcais, A., Alter, A., Antoni, G., Orlova, M., Nguyen, V.T., Singh, M., Vanderborght, P.R., Katoch, K., Mira, M.T., Vu, H.T., Ngyuen, T.H., Nguyen, N.B., Moraes, M., Mehra, N., Schurr, E. & Abel, L. (2007) Stepwise replication identifies a low-producing lymphotoxin-alpha allele as a major risk factor for early-onset leprosy. *Nat Genet*, **39**, 517-22.
- Alexander, D.H., Novembre, J. & Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*, **19**, 1655-64.
- Allen, M., Heinzmann, A., Noguchi, E., Abecasis, G., Broxholme, J., Ponting, C.P., Bhattacharyya, S., Tinsley, J., Zhang, Y., Holt, R., Jones, E.Y., Lench, N., Carey, A., Jones, H., Dickens, N.J., Dimon, C., Nicholls, R., Baker, C., Xue, L., Townsend, E., Kabesch, M., Weiland, S.K., Carr, D., Von Mutius, E., Adcock, I.M., Barnes, P.J., Lathrop, G.M., Edwards, M., Moffatt, M.F. & Cookson, W.O. (2003) Positional cloning of a novel gene influencing asthma from chromosome 2q14. *Nat Genet*, **35**, 258-63.
- Alonso, S., Flores, C., Cabrera, V., Alonso, A., Martín, P., Albarrán, C., Izagirre, N., De La Rua, C. & García, O. (2005) The place of the Basques in the European Y-chromosome diversity landscape. *Eur J Hum Genet*, **13**, 1293-302.
- Altshuler, D., Daly, M.J. & Lander, E.S. (2008) Genetic mapping in human disease. *Science*, **322**, 881-8.
- Amigo, J., Salas A., Phillips C. (2011) ENGINES: exploring single nucleotide variation in entire human genomes. *BMC Bioinformatics*, **12**:105
- Amos, C.I. (2007) Successful design and conduct of genome-wide association studies. *Hum Mol Genet*, **16** (SPEC) No. 2, R220-5.
- Antonarakis, S.E. (1998) Recommendations for a nomenclature system for human gene mutations. Nomenclature Working Group. *Hum Mutat*, **11**, 1-3.
- Arnedo-Pena, A., García-Marcos, L., García Hernández, G., Aguinagua Ontoso, I., González Díaz, C., Morales Suárez-Varela, M., Dominguez Aurrecochea, B., Busquets Monge, R., Blanco Quirós, A., Batlles Garrido, J., Miner Kanflanka, I. & López-Silvarrey Varela, A. (2005) Time trends and geographical variations in the prevalence of symptoms of allergic rhinitis in 6-7-year-old children from eight areas of Spain according to the ISAAC. *An Pediatr (Barc)*, **62**, 229-36.
- Arriba-Méndez, S., Sanz, C., Isidoro-García, M., Davild, I., Laffond, E., Horeno, E., Avila, C. & Lorente, F. (2006) 927T>C polymorphism of the cysteinyl-leukotriene type-1 receptor (CYSLTR1) gene in children with asthma and atopic dermatitis. *Pediatr Allergy Immunol*, **17**, 323-8.



## BIBLIOGRAFÍA

---

- Arriba-Méndez, S., Sanz, C., Isidoro-García, M., Pascual, M., Avila, C., Dávila, I. & Lorente, F. (2008) Analysis of 927T > C CYSLTR1 and -444A > C LTC4S polymorphisms in children with asthma. *Allergol Immunopathol (Madr)*, **36**, 259-63.
- Auton, A., Bryc, K., Boyko, A.R., Lohmueller, K.E., Novembre, J., Reynolds, A., Indap, A., Wright, M.H., Degenhardt, J.D., Gutenkunst, R.N., King, K.S., Nelson, M.R. & Bustamante, C.D. (2009) Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Res*, **19**, 795-803.
- Balaci, L., Spada, M.C., Olla, N., Sole, G., Loddo, L., Anedda, F., Naitza, S., Zuncheddu, M.A., Maschio, A., Altea, D., Uda, M., Pilia, S., Sanna, S., Masala, M., Crisponi, L., Fattori, M., Devoto, M., Doratiotto, S., Rassu, S., Mereu, S., Giua, E., Cadeddu, N.G., Atzeni, R., Pelosi, U., Corrias, A., Perra, R., Torrazza, P.L., Pirina, P., Ginesu, F., Marcias, S., Schintu, M.G., Del Giacco, G.S., Manconi, P.E., Malerba, G., Bisognin, A., Trabetti, E., Boner, A., Pescolderung, L., Pignatti, P.F., Schlessinger, D., Cao, A. & Pilia, G. (2007) IRAK-M is involved in the pathogenesis of early-onset persistent asthma. *Am J Hum Genet*, **80**, 1103-14.
- Balaresque, P., Bowden, G.R., Adams, S.M., Leung, H.Y., King, T.E., Rosser, Z.H., Goodwin, J., Moisan, J.P., Richard, C., Millward, A., Demaine, A.G., Barbujani, G., Previdere, C., Wilson, I.J., Tyler-Smith, C. & Jobling, M.A. (2010) A predominantly neolithic origin for European paternal lineages. *PLoS Biol*, **8**, e1000285.
- Balding, D.J. (2006) A tutorial on statistical methods for population association studies. *Nat Rev Genet*, **7**, 781-91.
- Baldini, M., Lohman, I.C., Halonen, M., Erickson, R.P., Holt, P.G. & Martínez, F.D. (1999) A Polymorphism\* in the 5' flanking region of the CD14 gene is associated with circulating soluble CD14 levels and with total serum immunoglobulin E. *Am J Respir Cell Mol Biol*, **20**, 976-83.
- Barbee, R.A. & Murphy, S. (1998) The natural history of asthma. *J Allergy Clin Immunol*, **102**, S65-72.
- Barbujani, G. & Goldstein, D.B. (2004) Africans and Asians abroad: genetic diversity in Europe. *Annu Rev Genomics Hum Genet*, **5**, 119-50.
- Barnholtz-Sloan, J.S., Chakraborty, R., Sellers, T.A. & Schwartz, A.G. (2005) Examining population stratification via individual ancestry estimates versus self-reported race. *Cancer Epidemiol Biomarkers Prev*, **14**, 1545-51.
- Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263-5.
- Belfer, I., Buzas, B., Hipp, H., Dean, M., Evans, C., Lorincz, I., Max, M.B. & Goldman, D. (2004) Haplotype structure of inflammatory cytokines genes (IL1B, IL6 and TNF/LTA) in US Caucasians and African Americans. *Genes Immun*, **5**, 505-12.
- Black, R.A., Rauch, C.T., Kozlosky, C.J., Peschon, J.J., Slack, J.L., Wolfson, M.F., Castner, B.J., Stocking, K.L., Reddy, P., Srinivasan, S., Nelson, N., Boiani, N., Schooley, K.A., Gerhart, M., Davis, R., Fitzner, J.N., Johnson, R.S., Paxton, R.J., March, C.J. & Cerretti, D.P. (1997) A metalloproteinase disintegrin that releases tumour-necrosis factor-alpha from cells. *Nature*, **385**, 729-33.
- Boronat, M., Varillas, V.F., Saavedra, P., Suárez, V., Bosch, E., Carrillo, A. & Novoa, F.J. (2006) Diabetes mellitus and impaired glucose regulation in the Canary Islands (Spain): prevalence and associated factors in the adult population of Telde, Gran Canaria. *Diabet Med*, **23**, 148-55.
- Bosch, E., Calafell, F., Comas, D., Oefner, P.J., Underhill, P.A. & Bertranpetit, J. (2001) High-resolution analysis of human Y-chromosome variation shows a sharp

- discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. *Am J Hum Genet*, **68**, 1019-29.
- Bossé, Y. & Hudson, T.J. (2007) Toward a comprehensive set of asthma susceptibility genes. *Annu Rev Med*, **58**, 171-84.
- Bouzigon, E., Corda, E., Aschard, H., Dizier, M.H., Boland, A., Bousquet, J., Chateigner, N., Gormand, F., Just, J., Le Moual, N., Scheinmann, P., Siroux, V., Vervloet, D., Zelenika, D., Pin, I., Kauffmann, F., Lathrop, M. & Demenais, F. (2008) Effect of 17q21 variants and smoking exposure in early-onset asthma. *N Engl J Med*, **359**, 1985-94.
- Bouzigon, E., Forabosco, P., Koppelman, G.H., Cookson, W.O., Dizier, M.H., Duffy, D.L., Evans, D.M., Ferreira, M.A., Kere, J., Laitinen, T., Malerba, G., Meyers, D.A., Moffatt, M., Martín, N.G., Ng, M.Y., Pignatti, P.F., Wjst, M., Kauffmann, F., Demenais, F. & Lewis, C.M. (2010) Meta-analysis of 20 genome-wide linkage studies evidenced new regions linked to asthma and atopy. *Eur J Hum Genet*, **18**, 700-6.
- Braman, S.S. (2006) The global burden of asthma. *Chest*, **130**, 4S-12S.
- Cantero-Recasens, G., González, J.R., Fandos, C., Duran-Tauleria, E., Smit, L.A., Kauffmann, F., Anto, J.M. & Valverde, M.A. (2010) Loss of function of transient receptor potential vanilloid 1 (TRPV1) genetic variant is associated with lower risk of active childhood asthma. *J Biol Chem*, **285**, 27532-5.
- Capelli, C., Onofri, V., Brisighelli, F., Boschi, I., Scarnicci, F., Masullo, M., Ferri, G., Tofanelli, S., Tagliabracci, A., Gusmao, L., Amorim, A., Gatto, F., Kirin, M., Merlitti, D., Brión, M., Vereá, A.B., Romano, V., Cali, F. & Pascali, V. (2009) Moors and Saracens in Europe: estimating the medieval North African male legacy in southern Europe. *Eur J Hum Genet*, **17**, 848-52.
- Cárdaba, B., Moffatt, M.F., Fernández, E., Jurado, A., Rojo, M., García, M., Ansotegui, I.J., Cortegano, I., Arrieta, I., Etxenagusia, M.A., Del Pozo, V., Urraca, J., Aceituno, E., Gallardo, S., Palomino, P., Cookson, W. & Lahoz, C. (2001) Allergy to dermatophagoides in a group of Spanish gypsies: genetic restrictions. *Int Arch Allergy Immunol*, **125**, 297-306.
- Carvajal-Uruena, I., García-Marcos, L., Busquets-Monge, R., Morales Suárez-Varela, M., García De Andoin, N., Batlles-Garrido, J., Blanco-Quirós, A., López-Silvarrey, A., García-Hernández, G., Guillén-Grimaj, F., González-Díaz, C. & Bellido-Blasco, J. (2005) Geographic variation in the prevalence of asthma symptoms in Spanish children and adolescents. International Study of Asthma and Allergies in Childhood (ISAAC) Phase 3, Spain. *Arch Bronconeumol*, **41**, 659-66.
- Castro-Giner, F., De Cid, R., González, J.R., Jarvis, D., Heinrich, J., Janson, C., Omenaas, E.R., Matheson, M.C., Pin, I., Anto, J.M., Wjst, M., Estivill, X. & Kogevinas, M. (2010) Positionally cloned genes and age-specific effects in asthma and atopy: an international population-based cohort study (ECRHS). *Thorax*, **65**, 124-31.
- Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A. (1994) The history and geography of human genes. Princeton, NJ: Princeton University Press.
- Chanock, S.J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D.J., Thomas, G., Hirschhorn, J.N., Abecasis, G., Altshuler, D., Bailey-Wilson, J.E., Brooks, L.D., Cardon, L.R., Daly, M., Donnelly, P., Fraumeni, J.F., Jr., Freimer, N.B., Gerhard, D.S., Gunter, C., Guttmacher, A.E., Guyer, M.S., Harris, E.L., Hoh, J., Hoover, R., Kong, C.A., Merikangas, K.R., Morton, C.C., Palmer, L.J., Phimister, E.G., Rice, J.P., Roberts, J., Rotimi, C., Tucker, M.A., Vogan, K.J., Wacholder, S., Wijsman, E.M., Winn, D.M. & Collins, F.S. (2007) Replicating genotype-phenotype associations. *Nature*, **447**, 655-60.

## BIBLIOGRAFÍA

---

- Chatila, T.A. (2004) Interleukin-4 receptor signaling pathways in asthma pathogenesis. *Trends Mol Med*, **10**, 493-9.
- Chejne, A. (1974) Muslim Spain. Its History and Culture. Minneapolis, MN: University of Minnesota Press.
- Choudhry, S., Ung, N., Avila, P.C., Ziv, E., Nazario, S., Casal, J., Torres, A., Gorman, J.D., Salari, K., Rodríguez-Santana, J.R., Toscano, M., Sylvia, J.S., Alioto, M., Castro, R.A., Salazar, M., Gomez, I., Fagan, J.K., Salas, J., Clark, S., Lilly, C., Matallana, H., Selman, M., Chapela, R., Sheppard, D., Weiss, S.T., Ford, J.G., Boushey, H.A., Drazen, J.M., Rodriguez-Cintron, W., Silverman, E.K. & Burchard, E.G. (2005) Pharmacogenetic differences in response to albuterol between Puerto Ricans and Mexicans with asthma. *Am J Respir Crit Care Med*, **171**, 563-70.
- Clancy, R.M., Marion, M.C., Kaufman, K.M., Ramos, P.S., Adler, A., Harley, J.B., Langefeld, C.D. & Buyon, J.P. (2010) Identification of candidate loci at 6p21 and 21q22 in a genome-wide association study of cardiac manifestations of neonatal lupus. *Arthritis Rheum*, **62**, 3415-24.
- Clayton, D. & Mckeigue, P.M. (2001) Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet*, **358**, 1356-60.
- Cohen, J. (2002) The immunopathogenesis of sepsis. *Nature*, **420**, 885-91.
- Conde, L., Vaquerizas, J.M., Dopazo, H., Arbiza, L., Reumers, J., Rousseau, F., Schymkowitz, J. & Dopazo, J. (2006) PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes. *Nucleic Acids Res*, **34**, W621-5.
- Consortium, T.G.P. (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061-73.
- Contopoulos-Ioannidis, D.G., Manoli, E.N. & Ioannidis, J.P. (2005) Meta-analysis of the association of beta2-adrenergic receptor polymorphisms with asthma phenotypes. *J Allergy Clin Immunol*, **115**, 963-72.
- Cookson, W. (2004) The immunogenetics of asthma and eczema: a new focus on the epithelium. *Nat Rev Immunol*, **4**, 978-88.
- Cookson, W.O., Sharp, P.A., Faux, J.A. & Hopkin, J.M. (1989) Linkage between immunoglobulin E responses underlying asthma and rhinitis and chromosome 11q. *Lancet*, **1**, 1292-5.
- Cox, D.R. & Mccullagh, P. (1982) Some aspects of analysis of covariance. *Biometrics*, **38**, 541-61.
- Craddock, N., Hurles, M.E., Cardin, N., Pearson, R.D., Plagnol, V., Robson, S., Vukcevic, D., Barnes, C., Conrad, D.F., Giannoulatou, E., Holmes, C., Marchini, J.L., Stirrups, K., Tobin, M.D., Wain, L.V., Yau, C., Aerts, J., Ahmad, T., Andrews, T.D., Arbury, H., Attwood, A., Auton, A., Ball, S.G., Balmforth, A.J., Barrett, J.C., Barroso, I., Barton, A., Bennett, A.J., Bhaskar, S., Blaszczyk, K., Bowes, J., Brand, O.J., Braund, P.S., Bredin, F., Breen, G., Brown, M.J., Bruce, I.N., Bull, J., Burren, O.S., Burton, J., Byrnes, J., Caesar, S., Clee, C.M., Coffey, A.J., Connell, J.M., Cooper, J.D., Dominiczak, A.F., Downes, K., Drummond, H.E., Dudakia, D., Dunham, A., Ebbs, B., Eccles, D., Edkins, S., Edwards, C., Elliot, A., Emery, P., Evans, D.M., Evans, G., Eyre, S., Farmer, A., Ferrier, I.N., Feuk, L., Fitzgerald, T., Flynn, E., Forbes, A., Forty, L., Franklyn, J.A., Freathy, R.M., Gibbs, P., Gilbert, P., Gokumen, O., Gordon-Smith, K., Gray, E., Green, E., Groves, C.J., Grozeva, D., Gwilliam, R., Hall, A., Hammond, N., Hardy, M., Harrison, P., Hassanali, N., Hebaishi, H., Hines, S., Hinks, A., Hitman, G.A., Hocking, L., Howard, E., Howard, P., Howson, J.M., Hughes, D., Hunt, S., Isaacs, J.D., Jain, M., Jewell, D.P., Johnson, T., Jolley, J.D., Jones, I.R., Jones, L.A.,

- et al. (2010) Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*, **464**, 713-20.
- Cui, T., Wang, L., Wu, J. & Xie, J. (2003) The association analysis of FcepsilonR1beta with allergic asthma in a Chinese population. *Chin Med J (Engl)*, **116**, 1875-8.
- Cuturi, M.C., Murphy, M., Costa-Giomi, M.P., Weinmann, R., Perussia, B. & Trinchieri, G. (1987) Independent regulation of tumor necrosis factor and lymphotoxin production by human peripheral blood lymphocytes. *J Exp Med*, **165**, 1581-94.
- Daley, D., Lemire, M., Akhbir, L., Chan-Yeung, M., He, J.Q., McDonald, T., Sandford, A., Stefanowicz, D., Tripp, B., Zamar, D., Bosse, Y., Ferretti, V., Montpetit, A., Tessier, M.C., Becker, A., Kozyrskyj, A.L., Beilby, J., Mccaskie, P.A., Musk, B., Warrington, N., James, A., Laprise, C., Palmer, L.J., Pare, P.D. & Hudson, T.J. (2009) Analyses of associations with asthma in four asthma population samples from Canada and Australia. *Hum Genet*, **125**, 445-59.
- De Pablos-Velasco, P., Martínez-Martín, F.J., Rodríguez Pérez, F., Urioste, L.M. & García Robles, R. (2002) Prevalence, awareness, treatment and control of hypertension in a Canarian population. Relationship with glucose tolerance categories. The Guia Study. *J Hypertens*, **20**, 1965-71.
- Denham, S., Koppelman, G.H., Blakey, J., Wjst, M., Ferreira, M.A., Hall, I.P. & Sayers, I. (2008) Meta-analysis of genome-wide linkage studies of asthma and related traits. *Respir Res*, **9**, 38.
- Devitt, A., Moffatt, O.D., Raykundalia, C., Capra, J.D., Simmons, D.L. & Gregory, C.D. (1998) Human CD14 mediates recognition and phagocytosis of apoptotic cells. *Nature*, **392**, 505-9.
- Dickson, S.P., Wang, K., Krantz, I., Hakonarson, H. & Goldstein, D.B. (2010) Rare variants create synthetic genome-wide associations. *PLoS Biol*, **8**, e1000294.
- Donnadieu, E., Jouvin, M.H., Rana, S., Moffatt, M.F., Mockford, E.H., Cookson, W.O. & Kinet, J.P. (2003) Competing functions encoded in the allergy-associated F(c)epsilonR1beta gene. *Immunity*, **18**, 665-74.
- Drayton, D.L., Chan, K., Lesslauer, W., Lee, J., Ying, X.Y. & Ruddle, N.H. (2002) Lymphocyte traffic in lymphoid organ neogenesis: differential roles of Ltalpha and LTalphabeta. *Adv Exp Med Biol*, **512**, 43-8.
- Duffy, D.L., Martin, N.G., Battistutta, D., Hopper, J.L. & Mathews, J.D. (1990) Genetics of asthma and hay fever in Australian twins. *Am Rev Respir Dis*, **142**, 1351-8.
- Eleftherohorinou, H., Wright, V., Hoggart, C., Hartikainen, A.L., Jarvelin, M.R., Balding, D., Coin, L. & Levin, M. (2009) Pathway analysis of GWAS provides new insights into genetic susceptibility to 3 inflammatory diseases. *PLoS One*, **4**, e8068.
- Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*, **14**, 2611-20.
- Evans, D.M., Visscher, P.M. & Wray, N.R. (2009) Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum Mol Genet*, **18**, 3525-31.
- Ferreira, M.A., Matheson, M.C., Duffy, D.L., Marks, G.B., Hui, J., Le Souef, P., Danoy, P., Baltic, S., Nyholt, D.R., Jenkins, M., Hayden, C., Willemsen, G., Ang, W., Kuokkanen, M., Beilby, J., Cheah, F., De Geus, E.J., Ramasamy, A., Vedantam, S., Salomaa, V., Madden, P.A., Heath, A.C., Hopper, J.L., Visscher, P.M., Musk, B., Leeder, S.R., Jarvelin, M.R., Pennell, C., Boomsma, D.I., Hirschhorn, J.N., Walters, H., Martin, N.G., James, A., Jones, G., Abramson, M.J., Robertson, C.F., Dharmage, S.C.,

## BIBLIOGRAFÍA

---

- Brown, M.A., Montgomery, G.W. & Thompson, P.J. (2011a) Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. *Lancet*, **378**, 1006-14.
- Ferreira, M.A., Mccrae, A.F., Medland, S.E., Nyholt, D.R., Gordon, S.D., Wright, M.J., Henders, A.K., Madden, P.A., Visscher, P.M., Wray, N.R., Heath, A.C., Montgomery, G.W., Duffy, D.L. & Martin, N.G. (2011b) Association between ORMDL3, IL1RL1 and a deletion on chromosome 17q21 with asthma risk in Australia. *Eur J Hum Genet*, **19**, 458-64.
- Ferreira, M.A., Zhao, Z.Z., Thomsen, S.F., James, M., Evans, D.M., Postmus, P.E., Kyvik, K.O., Backer, V., Boomsma, D.I., Martin, N.G., Montgomery, G.W. & Duffy, D.L. (2009) Association and interaction analyses of eight genes under asthma linkage peaks. *Allergy*, **64**, 1623-8.
- Fields, P.E., Kim, S.T. & Flavell, R.A. (2002) Cutting edge: changes in histone acetylation at the IL-4 and IFN-gamma loci accompany Th1/Th2 differentiation. *J Immunol*, **169**, 647-50.
- Finkelstein, Y., Bournissen, F.G., Hutson, J.R. & Shannon, M. (2009) Polymorphism of the *ADRB2* gene and response to inhaled beta-agonists in children with asthma: a meta-analysis. *J Asthma*, **46**, 900-5.
- Flores, C., Larruga, J.M., González, A.M., Hernández, M., Pinto, F. & M., C.V. (2001) The origin of the Canary Island aborigines and their contribution to the modern population: a molecular genetics perspective. *Current Anthropology*, **42**, 749-755.
- Flores, C., Ma, S.F., Pino-Yanes, M., Wade, M.S., Pérez-Méndez, L., Kittles, R.A., Wang, D., Papaiahgari, S., Ford, J.G., Kumar, R. & García, J.G.N. (2011) African Ancestry is associated with asthma risk in African Americans. *PLoS One*, en prensa.
- Flores, C., Maca-Meyer, N., González, A.M., Oefner, P.J., Shen, P., Pérez, J.A., Rojas, A., Larruga, J.M. & Underhill, P.A. (2004) Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. *Eur J Hum Genet*, **12**, 855-63.
- Flores, C., Maca-Meyer, N., Pérez, J.A., González, A.M., Larruga, J.M. & Cabrera, V.M. (2003) A predominant European ancestry of paternal lineages from Canary Islanders. *Ann Hum Genet*, **67**, 138-52.
- Folster-Holst, R., Moises, H.W., Yang, L., Fritsch, W., Weissenbach, J. & Christophers, E. (1998) Linkage between atopy and the IgE high-affinity receptor gene at 11q13 in atopic dermatitis families. *Hum Genet*, **102**, 236-9.
- Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M.M., Tsui, S.K., Xue, H., Wong, J.T., Galver, L.M., Fan, J.B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.F., Phillips, M.S., Roumy, S., Sallee, C., Verner, A., Hudson, T.J., Kwok, P.Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.C., Mak, W., Song, Y.Q., Tam, P.K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., et al. (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851-61.

- Fregel, R., Gomes, V., Gusmao, L., González, A.M., Cabrera, V.M., Amorim, A. & Larruga, J.M. (2009) Demographic history of Canary Islands male gene-pool: replacement of native lineages by European. *BMC Evol Biol*, **9**, 181.
- Fregel, R., Maca-Meyer, N., Cabrera, V.M., González, A.M. & Larruga, J.M. (2005) Description of a simple multiplex PCR-SSCP method for ABO genotyping and its application to the peopling of the Canary Islands. *Immunogenetics*, **57**, 572-8.
- Galanter, J.M., Torgerson, D., Gignoux, C.R., Sen, S., Roth, L.A., Via, M., Aldrich, M.C., Eng, C., Huntsman, S., Rodríguez-Santana, J., Rodríguez-Cintron, W., Chapelá, R., Ford, J.G. & Burchard, E.G. (2011) Cosmopolitan and ethnic-specific replication of genetic risk factors for asthma in 2 Latino populations. *J Allergy Clin Immunol*, **128**, 37-43 e12.
- García-Closas, M. & Lubin, J.H. (1999) Power and sample size calculations in case-control studies of gene-environment interactions: comments on different approaches. *Am J Epidemiol*, **149**, 689-92.
- García-Martín, E., García-Menaya, J., Sánchez, B., Martínez, C., Rosendo, R. & Agúndez, J.A. (2007) Polymorphisms of histamine-metabolizing enzymes and clinical manifestations of asthma and allergic rhinitis. *Clin Exp Allergy*, **37**, 1175-82.
- Gauchat, J.F., Lebman, D.A., Coffman, R.L., Gascan, H. & De Vries, J.E. (1990) Structure and expression of germline epsilon transcripts in human B cells induced by interleukin 4 to switch to IgE production. *J Exp Med*, **172**, 463-73.
- Gayán, J., Galan, J.J., González-Pérez, A., Sáez, M.E., Martínez-Larrad, M.T., Zabena, C., Rivero, M.C., Salinas, A., Ramírez-Lorca, R., Morón, F.J., Royo, J.L., Moreno-Rey, C., Velasco, J., Carrasco, J.M., Molero, E., Ochoa, C., Ochoa, M.D., Gutiérrez, M., Reina, M., Pascual, R., Romo-Astorga, A., Susillo-González, J.L., Vázquez, E., Real, L.M., Ruiz, A. & Serrano-Ríos, M. (2010) Genetic structure of the Spanish population. *BMC Genomics*, **11**, 326.
- Gelfand, E.W. (2008) Is asthma in childhood different from asthma in adults? Why do we need special approaches to asthma in children? *Allergy Asthma Proc*, **29**, 99-102.
- Genuneit, J., Cantelmo, J.L., Weinmayr, G., Wong, G.W., Cooper, P.J., Riiikjarv, M.A., Gotua, M., Kabesch, M., Von Mutius, E., Forastiere, F., Crane, J., Nystad, W., El-Sharif, N., Batlles-Garrido, J., García-Marcos, L., García-Hernández, G., Morales-Suárez-Varela, M., Nilsson, L., Braback, L., Saraclar, Y., Weiland, S.K., Cookson, W.O., Strachan, D. & Moffatt, M.F. (2009) A multi-centre study of candidate genes for wheeze and allergy: the International Study of Asthma and Allergies in Childhood Phase 2. *Clin Exp Allergy*, **39**, 1875-88.
- Ghoreschi, K., Thomas, P., Breit, S., Dugas, M., Mailhammer, R., Van Eden, W., Van Der Zee, R., Biedermann, T., Prinz, J., Mack, M., Mrowietz, U., Christophers, E., Schlondorff, D., Plewig, G., Sander, C.A. & Rocken, M. (2003) Interleukin-4 therapy of psoriasis induces Th2 responses and improves human autoimmune disease. *Nat Med*, **9**, 40-6.
- González, A.M., Brehm, A., Pérez, J.A., Maca-Meyer, N., Flores, C. & Cabrera, V.M. (2003) Mitochondrial DNA affinities at the Atlantic fringe of Europe. *Am J Phys Anthropol*, **120**, 391-404.
- González, J.R., Armengol, L., Sole, X., Guino, E., Mercader, J.M., Estivill, X. & Moreno, V. (2007) SNPAssoc: an R package to perform whole genome association studies. *Bioinformatics*, **23**, 644-5.
- Granada, M., Wilk, J.B., Tuzova, M., Strachan, D.P., Weidinger, S., Albrecht, E., Gieger, C., Heinrich, J., Himes, B.E., Hunninghake, G.M., Celedon, J.C., Weiss, S.T., Cruikshank, W.W., Farrer, L.A., Center, D.M. & O'connor, G.T. (2011) A genome-wide

## BIBLIOGRAFÍA

---

- association study of plasma total IgE concentrations in the Framingham Heart Study. *J Allergy Clin Immunol*, en prensa.
- Gray, P.W., Aggarwal, B.B., Benton, C.V., Bringman, T.S., Henzel, W.J., Jarrett, J.A., Leung, D.W., Moffat, B., Ng, P., Svedersky, L.P. & Et Al. (1984) Cloning and expression of cDNA for human lymphotoxin, a lymphokine with tumour necrosis activity. *Nature*, **312**, 721-4.
- Green, S.L., Gaillard, M.C., Song, E., Dewar, J.B. & Halkas, A. (1998) Polymorphisms of the beta chain of the high-affinity immunoglobulin E receptor (Fcepsilon RI-beta) in South African black and white asthmatic and nonasthmatic individuals. *Am J Respir Crit Care Med*, **158**, 1487-92.
- Gu, W., Dong, H., Jiang, D.P., Zhou, J., Du, D.Y., Gao, J.M., Yao, Y.Z., Zhang, L.Y., Wen, A.Q., Liu, Q., Wang, Z.G. & Jiang, J.X. (2008) Functional significance of CD14 promoter polymorphisms and their clinical relevance in a Chinese Han population. *Crit Care Med*, **36**, 2274-80.
- Guan, W., Liang, L., Boehnke, M. & Abecasis, G.R. (2009) Genotype-based matching to correct for population stratification in large-scale case-control genetic association studies. *Genetic Epidemiol*, **33**, 508-17.
- Gupta, D., Kirkland, T.N., Viriyakosol, S. & Dziarski, R. (1996) CD14 is a cell-activating receptor for bacterial peptidoglycan. *J Biol Chem*, **271**, 23310-6.
- Halapi, E., Gudbjartsson, D.F., Jonsdottir, G.M., Bjornsdottir, U.S., Thorleifsson, G., Helgadóttir, H., Williams, C., Koppelman, G.H., Heinzmann, A., Boezen, H.M., Jonasdóttir, A., Blondal, T., Gudjonsson, S.A., Jonasdóttir, A., Thorlacius, T., Henry, A.P., Altmueller, J., Krueger, M., Shin, H.D., Uh, S.T., Cheong, H.S., Jonsdottir, B., Ludviksson, B.R., Ludviksdottir, D., Gislason, D., Park, C.S., Deichmann, K., Thompson, P.J., Wjst, M., Hall, I.P., Postma, D.S., Gislason, T., Kong, A., Jonsdottir, I., Thorsteinsdottir, U. & Stefansson, K. (2010) A sequence variant on 17q21 is associated with age at onset and severity of asthma. *Eur J Hum Genet*, **18**, 902-8.
- Hamburg, M.A. & Collins, F.S. (2010) The path to personalized medicine. *N Engl J Med*, **363**, 301-4.
- Hammad, H. & Lambrecht, B.N. (2011) Dendritic cells and airway epithelial cells at the interface between innate and adaptive immune responses. *Allergy*, **66**, 579-87.
- Hancock, D.B., Romieu, I., Shi, M., Sienna-Monge, J.J., Wu, H., Chiu, G.Y., Li, H., Del Rio-Navarro, B.E., Willis-Owen, S.A., Weiss, S.T., Raby, B.A., Gao, H., Eng, C., Chapela, R., Burchard, E.G., Tang, H., Sullivan, P.F. & London, S.J. (2009) Genome-wide association study implicates chromosome 9q21.31 as a susceptibility locus for asthma in Mexican children. *PLoS Genet*, **5**, e1000623.
- Heath, S.C., Gut, I.G., Brennan, P., McKay, J.D., Bencko, V., Fabianova, E., Foretova, L., Georges, M., Janout, V., Kabesch, M., Krokan, H.E., Elvestad, M.B., Lissowska, J., Mates, D., Rudnai, P., Skorpén, F., Schreiber, S., Soria, J.M., Syvanen, A.C., Meneton, P., Herberg, S., Galan, P., Szeszenia-Dabrowska, N., Zaridze, D., Genin, E., Cardon, L.R. & Lathrop, M. (2008) Investigation of the fine structure of European populations with applications to disease association studies. *Eur J Hum Genet*, **16**, 1413-29.
- Henckaerts, L., Nielsen, K.R., Steffensen, R., Van Steen, K., Mathieu, C., Giulietti, A., Wouters, P.J., Milants, I., Vanhorebeek, I., Langouche, L., Vermeire, S., Rutgeerts, P., Thiel, S., Wilmer, A., Hansen, T.K. & Van Den Berghe, G. (2009) Polymorphisms in innate immunity genes predispose to bacteremia and death in the medical intensive care unit. *Crit Care Med*, **37**, 192-201, e1-3.

- Hershey, G.K., Friedrich, M.F., Esswein, L.A., Thomas, M.L. & Chatila, T.A. (1997) The association of atopy with a gain-of-function mutation in the alpha subunit of the interleukin-4 receptor. *N Engl J Med*, **337**, 1720-5.
- Himes, B.E., Hunninghake, G.M., Baurley, J.W., Rafaels, N.M., Sleiman, P., Strachan, D.P., Wilk, J.B., Willis-Owen, S.A., Klanderman, B., Lasky-Su, J., Lazarus, R., Murphy, A.J., Soto-Quirós, M.E., Avila, L., Beaty, T., Mathias, R.A., Ruczinski, I., Barnes, K.C., Celedon, J.C., Cookson, W.O., Gauderman, W.J., Gilliland, F.D., Hakonarson, H., Lange, C., Moffatt, M.F., O'connor, G.T., Raby, B.A., Silverman, E.K. & Weiss, S.T. (2009) Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene. *Am J Hum Genet*, **84**, 581-93.
- Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S. & Manolio, T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*, **106**, 9362-7.
- Hirota, T., Takahashi, A., Kubo, M., Tsunoda, T., Tomita, K., Doi, S., Fujita, K., Miyatake, A., Enomoto, T., Miyagawa, T., Adachi, M., Tanaka, H., Niimi, A., Matsumoto, H., Ito, I., Masuko, H., Sakamoto, T., Hizawa, N., Taniguchi, M., Lima, J.J., Irvin, C.G., Peters, S.P., Himes, B.E., Litonjua, A.A., Tantisira, K.G., Weiss, S.T., Kamatani, N., Nakamura, Y. & Tamari, M. (2011) Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat Genet*, **43**, 893-6.
- Holberg, C.J., Halonen, M., Wright, A.L. & Martínez, F.D. (1999) Familial aggregation and segregation analysis of eosinophil levels. *Am J Respir Crit Care Med*, **160**, 1604-10.
- Holgate, S.T. (2011) The sentinel role of the airway epithelium in asthma pathogenesis. *Immunol Rev*, **242**, 205-19.
- Holtzman, N.A. & Marteau, T.M. (2000) Will genetics revolutionize medicine? *N Engl J Med*, **343**, 141-4.
- Huang, L., Wang, C. & Rosenberg, N.A. (2009) The relationship between imputation error and statistical power in genetic association studies in diverse populations. *Am J Hum Genet*, **85**, 692-8.
- Hunninghake, G.M., Weiss, S.T. & Celedon, J.C. (2006) Asthma in Hispanics. *Am J Respir Crit Care Med*, **173**, 143-63.
- Ioannidis, J.P., Gwinn, M., Little, J., Higgins, J.P., Bernstein, J.L., Boffetta, P., Bondy, M., Bray, M.S., Brenchley, P.E., Buffler, P.A., Casas, J.P., Chokkalingam, A., Danesh, J., Smith, G.D., Dolan, S., Duncan, R., Gruis, N.A., Hartge, P., Hashibe, M., Hunter, D.J., Jarvelin, M.R., Malmer, B., Maraganore, D.M., Newton-Bishop, J.A., O'brien, T.R., Petersen, G., Riboli, E., Salanti, G., Seminara, D., Smeeth, L., Taioli, E., Timpson, N., Uitterlinden, A.G., Vineis, P., Wareham, N., Winn, D.M., Zimmern, R. & Khoury, M.J. (2006) A road map for efficient and reliable human genome epidemiology. *Nat Genet*, **38**, 3-5.
- Isidoro-García, M., Dávila, I., Laffond, E., Moreno, E., Lorente, F. & González-Sarmiento, R. (2005a) Interleukin-4 (IL4) and Interleukin-4 receptor (IL4RA) polymorphisms in asthma: a case control study. *Clin Mol Allergy*, **3**, 15.
- Isidoro-García, M., Dávila, I., Moreno, E., Lorente, F. & González-Sarmiento, R. (2005b) Analysis of the leukotriene C4 synthase A-444C promoter polymorphism in a Spanish population. *J Allergy Clin Immunol*, **115**, 206-7.
- Isidoro-García, M., Roa-Gomez, S., Davila, I., Lorente, F. & González-Sarmiento, R. (2003) Lack of association between the 7888 C/T polymorphism in the AID gene and atopy in a Spanish population. *J Allergy Clin Immunol*, **112**, 460; author reply 460-1.



## BIBLIOGRAFÍA

---

- Janssens, A.C., Aulchenko, Y.S., Elefante, S., Borsboom, G.J., Steyerberg, E.W. & Van Duijn, C.M. (2006) Predictive testing for complex diseases using multiple genes: fact or fiction? *Genet Med*, **8**, 395-400.
- Juliá-Serdá, G., Cabrera-Navarro, P., Acosta-Fernandez, O., Martín-Pérez, P., Losada-Cabrera, P., García-Bello, M.A., Carrillo-Díaz, T. & Anto-Boque, J. (2011) High prevalence of asthma and atopy in the Canary Islands, Spain. *Int J Tuberc Lung Dis*, **15**, 536-41.
- Juliá Serdá, G., Cabrera Navarro, P., Acosta Fernandez, O., Martín Pérez, P., Batista Martín, J., Alamo Santana, F., Rodríguez De Castro, F. & Anto Boque, J.M. (2005) High prevalence of asthma symptoms in the Canary Islands: climatic influence? *J Asthma*, **42**, 507-11.
- Kawashima, T., Noguchi, E., Arinami, T., Yamakawa-Kobayashi, K., Nakagawa, H., Otsuka, F. & Hamaguchi, H. (1998) Linkage and association of an interleukin 4 gene polymorphism with atopic dermatitis in Japanese families. *J Med Genet*, **35**, 502-4.
- Kedda, M.A., Lose, F., Duffy, D., Bell, E., Thompson, P.J. & Upham, J. (2005) The CD14 C-159T polymorphism is not associated with asthma or asthma severity in an Australian adult population. *Thorax*, **60**, 211-4.
- Kelly-Welch, A.E., Hanson, E.M., Boothby, M.R. & Keegan, A.D. (2003) Interleukin-4 and interleukin-13 signaling connections maps. *Science*, **300**, 1527-8.
- Kim, Y.K., Oh, S.Y., Oh, H.B., Chun, S.Y., Cho, S.H., Koh, Y.Y., Min, K.U. & Kim, Y.Y. (2002) Coding single nucleotide polymorphism in the high-affinity immunoglobulin E receptor b chain (FcepsilonRI-beta) gene is associated with immunoglobulin E receptor-mediated histamine release from basophils. *Clin Exp Allergy*, **32**, 751-5.
- Kim, Y.K., Park, H.W., Yang, J.S., Oh, S.Y., Chang, Y.S., Shin, E.S., Lee, J.E., Kim, S., Gho, Y.S., Cho, S.H., Min, K.U. & Kim, Y.Y. (2007) Association and functional relevance of E237G, a polymorphism of the high-affinity immunoglobulin E-receptor beta chain gene, to airway hyper-responsiveness. *Clin Exp Allergy*, **37**, 592-8.
- Kinet, J.P. (1999) The high-affinity IgE receptor (Fc epsilon RI): from physiology to pathology. *Annu Rev Immunol*, **17**, 931-72.
- Klassert, T.E., Sanchez, J.J., Almeida, T.A., Candenias, L., Pinto, F., Acosta, O. & Hernández, M. (2010) Common variants of the neuropeptide expressing tachykinin genes and susceptibility to asthma: a case-control study. *J Neuroimmunol*, **227**, 202-7.
- Knight, J.C., Keating, B.J. & Kwiatkowski, D.P. (2004) Allele-specific repression of lymphotoxin-alpha by activated B cell factor-1. *Nat Genet*, **36**, 394-9.
- Knight, J.C., Keating, B.J., Rockett, K.A. & Kwiatkowski, D.P. (2003) In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. *Nat Genet*, **33**, 469-75.
- Kobayashi, K., Hernández, L.D., Galan, J.E., Janeway, C.A., Jr., Medzhitov, R. & Flavell, R.A. (2002) IRAK-M is a negative regulator of Toll-like receptor signaling. *Cell*, **110**, 191-202.
- Kobilka, B.K., Dixon, R.A., Frielle, T., Dohlman, H.G., Bolanowski, M.A., Sigal, I.S., Yang-Feng, T.L., Francke, U., Caron, M.G. & Lefkowitz, R.J. (1987) cDNA for the human beta 2-adrenergic receptor: a protein with multiple membrane-spanning domains and encoded by a gene whose chromosomal location is shared with that of the receptor for platelet-derived growth factor. *Proc Natl Acad Sci U S A*, **84**, 46-50.

- Kotanides, H. & Reich, N.C. (1996) Interleukin-4-induced STAT6 recognizes and activates a target site in the promoter of the interleukin-4 receptor gene. *J Biol Chem*, **271**, 25555-61.
- Kraft, M. (2011) Asthma Phenotypes and Interleukin-13--Moving Closer to Personalized Medicine. *N Engl J Med*, **365**:1141-4.
- Kraft, P., Wacholder, S., Cornelis, M.C., Hu, F.B., Hayes, R.B., Thomas, G., Hoover, R., Hunter, D.J. & Chanock, S. (2009) Beyond odds ratios--communicating disease risk based on genetic profiles. *Nat Rev Genet*, **10**, 264-9.
- Kruse, S., Forster, J., Kuehr, J. & Deichmann, K.A. (1999a) Characterization of the membrane-bound and a soluble form of human IL-4 receptor alpha produced by alternative splicing. *Int Immunol*, **11**, 1965-70.
- Kruse, S., Japha, T., Tedner, M., Sparholt, S.H., Forster, J., Kuehr, J. & Deichmann, K.A. (1999b) The polymorphisms S503P and Q576R in the interleukin-4 receptor alpha gene are associated with atopy and influence the signal transduction. *Immunology*, **96**, 365-71.
- Kumar, R., Seibold, M.A., Aldrich, M.C., Williams, L.K., Reiner, A.P., Colangelo, L., Galanter, J., Gignoux, C., Hu, D., Sen, S., Choudhry, S., Peterson, E.L., Rodríguez-Santana, J., Rodríguez-Cintron, W., Nalls, M.A., Leak, T.S., O'meara, E., Meibohm, B., Kritchevsky, S.B., Li, R., Harris, T.B., Nickerson, D.A., Fornage, M., Enright, P., Ziv, E., Smith, L.J., Liu, K. & Burchard, E.G. (2010) Genetic ancestry in lung-function predictions. *N Engl J Med*, **363**, 321-30.
- Laayouni, H., Calafell, F. & Bertranpetit, J. (2010) A genome-wide survey does not show the genetic distinctiveness of Basques. *Hum Genet*, **127**, 455-8.
- Laitinen, T., Ollikainen, V., Lázaro, C., Kauppi, P., De Cid, R., Anto, J.M., Estivill, X., Lokki, H., Mannila, H., Laitinen, L.A. & Kere, J. (2000) Association study of the chromosomal region containing the FCER2 gene suggests it has a regulatory role in atopic disorders. *Am J Respir Crit Care Med*, **161**, 700-6.
- Laitinen, T., Polvi, A., Rydman, P., Vendelin, J., Pulkkinen, V., Salmikangas, P., Makela, S., Rehn, M., Pirskanen, A., Rautanen, A., Zucchelli, M., Gullsten, H., Leino, M., Alenius, H., Petays, T., Haahtela, T., Laitinen, A., Laprise, C., Hudson, T.J., Laitinen, L.A. & Kere, J. (2004) Characterization of a common susceptibility locus for asthma-related traits. *Science*, **304**, 300-4.
- Laporte, S.L., Juo, Z.S., Vaclavikova, J., Colf, L.A., Qi, X., Heller, N.M., Keegan, A.D. & García, K.C. (2008) Molecular and structural basis of cytokine receptor pleiotropy in the interleukin-4/13 system. *Cell*, **132**, 259-72.
- Lasky-Su, J., Lyon, H.N., Emilsson, V., Heid, I.M., Molony, C., Raby, B.A., Lazarus, R., Klanderma, B., Soto-Quirós, M.E., Avila, L., Silverman, E.K., Thorleifsson, G., Thorsteinsdottir, U., Kronenberg, F., Vollmert, C., Illig, T., Fox, C.S., Levy, D., Laird, N., Ding, X., Mcqueen, M.B., Butler, J., Ardlie, K., Papoutsakis, C., Dedoussis, G., O'donnell, C.J., Wichmann, H.E., Celedon, J.C., Schadt, E., Hirschhorn, J., Weiss, S.T., Stefansson, K. & Lange, C. (2008) On the replication of genetic associations: timing can be everything! *Am J Hum Genet*, **82**, 849-58.
- Lázaro, C., De Cid, R., Sunyer, J., Soriano, J., Giménez, J., Alvarez, M., Casals, T., Anto, J.M. & Estivill, X. (1999) Missense mutations in the cystic fibrosis gene in adult patients with asthma. *Hum Mutat*, **14**, 510-9.
- Le Beau, M.M., Lemons, R.S., Espinosa, R., 3rd, Larson, R.A., Arai, N. & Rowley, J.D. (1989) Interleukin-4 and interleukin-5 map to human chromosome 5 in a region encoding growth factors and receptors and are deleted in myeloid leukemias with a del(5q). *Blood*, **73**, 647-50.

## BIBLIOGRAFÍA

---

- Le Souef, P.N., Candelaria, P. & Goldblatt, J. (2006) Evolution and respiratory genetics. *Eur Respir J*, **28**, 1258-63.
- Lee, D.K., Currie, G.P., Hall, I.P., Lima, J.J. & Lipworth, B.J. (2004) The arginine-16 beta2-adrenoceptor polymorphism predisposes to bronchoprotective subsensitivity in patients treated with formoterol and salmeterol. *Br J Clin Pharmacol*, **57**, 68-75.
- Lemanske, R.F., Jr. & Busse, W.W. (2010) Asthma: clinical expression and molecular mechanisms. *J Allergy Clin Immunol*, **125**, S95-102.
- Leslie, S., Donnelly, P. & Mcvean, G. (2008) A statistical method for predicting classical HLA alleles from SNP data. *Am J Hum Genet*, **82**, 48-56.
- Lester, L.A., Rich, S.S., Blumenthal, M.N., Togias, A., Murphy, S., Malveaux, F., Miller, M.E., Dunston, G.M., Solway, J., Wolf, R.L., Samet, J.M., Marsh, D.G., Meyers, D.A., Ober, C. & Bleecker, E.R. (2001) Ethnic differences in asthma and associated phenotypes: collaborative study on the genetics of asthma. *J Allergy Clin Immunol*, **108**, 357-62.
- Levan, T.D., Bloom, J.W., Bailey, T.J., Karp, C.L., Halonen, M., Martínez, F.D. & Vercelli, D. (2001) A common single nucleotide polymorphism in the CD14 promoter decreases the affinity of Sp protein binding and enhances transcriptional activity. *J Immunol*, **167**, 5838-44.
- Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L. & Myers, R.M. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, **319**, 1100-4.
- Li, Q. & Yu, K. (2008) Improved correction for population stratification in genome-wide association studies by identifying hidden population structures. *Genetic Epidemiol*, **32**, 215-26.
- Li, X., Howard, T.D., Zheng, S.L., Haselkorn, T., Peters, S.P., Meyers, D.A. & Bleecker, E.R. (2010a) Genome-wide association study of asthma identifies RAD50-IL13 and HLA-DR/DQ regions. *J Allergy Clin Immunol*, **125**, 328-335 e11.
- Li, X., Zhang, Y., Zhang, J., Xiao, Y., Huang, J., Tian, C., He, C., Deng, Y., Yang, Y. & Fan, H. (2010b) Asthma susceptible genes in Chinese population: a meta-analysis. *Respir Res*, **11**, 129.
- Lin, D.Y., Hu, Y. & Huang, B.E. (2008) Simple and efficient analysis of disease association with missing genotype data. *Am J Hum Genet*, **82**, 444-52.
- Lin, S., Cicala, C., Scharenberg, A.M. & Kinet, J.P. (1996) The Fc(epsilon)RIbeta subunit functions as an amplifier of Fc(epsilon)RIgamma-mediated cell activation signals. *Cell*, **85**, 985-95.
- Little, J., Higgins, J.P., Ioannidis, J.P., Moher, D., Gagnon, F., Von Elm, E., Khoury, M.J., Cohen, B., Davey-Smith, G., Grimshaw, J., Scheet, P., Gwinn, M., Williamson, R.E., Zou, G.Y., Hutchings, K., Johnson, C.Y., Tait, V., Wiens, M., Golding, J., Van Duijn, C., Mclaughlin, J., Paterson, A., Wells, G., Fortier, I., Freedman, M., Zecevic, M., King, R., Infante-Rivard, C., Stewart, A. & Birkett, N. (2009) Strengthening the reporting of genetic association studies (STREGA): an extension of the STROBE Statement. *Hum Genet*, **125**, 131-51.
- Loza, M.J. & Chang, B.L. (2007) Association between Q551R IL4R genetic variants and atopic asthma risk demonstrated by meta-analysis. *J Allergy Clin Immunol*, **120**, 578-85.
- Maca-Meyer, N., Sánchez-Velasco, P., Flores, C., Larruga, J.M., González, A.M., Oterino, A. & Leyva-Cobián, F. (2003) Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain). *Ann Hum Genet*, **67**, 329-39.

- Maca-Meyer, N., Villar, J., Pérez-Méndez, L., Cabrera De León, A. & Flores, C. (2004) A tale of aborigines, conquerors and slaves: Alu insertion polymorphisms and the peopling of Canary Islands. *Ann Hum Genet*, **68**, 600-5.
- Macgregor, S., Craddock, N. & Holmans, P.A. (2006) Use of phenotypic covariates in association analysis by sequential addition of cases. *Eur J Hum Genet*, **14**, 529-34.
- Maller, J., George, S., Purcell, S., Fagerness, J., Altshuler, D., Daly, M.J. & Seddon, J.M. (2006) Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nat Genet*, **38**, 1055-9.
- Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorf, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F., McCarroll, S.A. & Visscher, P.M. (2009) Finding the missing heritability of complex diseases. *Nature*, **461**, 747-53.
- Marchini, J., Cardon, L.R., Phillips, M.S. & Donnelly, P. (2004) The effects of human population structure on large genetic association studies. *Nat Genet*, **36**, 512-7.
- Marchini, J. & Howie, B. (2010) Genotype imputation for genome-wide association studies. *Nat Rev Genet*, **11**, 499-511.
- Marsh, D.G., Neely, J.D., Breazeale, D.R., Ghosh, B., Freidhoff, L.R., Ehrlich-Kautzky, E., Schou, C., Krishnaswamy, G. & Beaty, T.H. (1994) Linkage analysis of IL4 and other chromosome 5q31.1 markers and total serum immunoglobulin E concentrations. *Science*, **264**, 1152-6.
- Mathias, R.A., Grant, A.V., Rafaels, N., Hand, T., Gao, L., Vergara, C., Tsai, Y.J., Yang, M., Campbell, M., Foster, C., Gao, P., Togiias, A., Hansel, N.N., Diette, G., Adkinson, N.F., Liu, M.C., Faruque, M., Dunston, G.M., Watson, H.R., Bracken, M.B., Hoh, J., Maul, P., Maul, T., Jedlicka, A.E., Murray, T., Hetmanski, J.B., Ashworth, R., Ongaco, C.M., Hetrick, K.N., Doheny, K.F., Pugh, E.W., Rotimi, C.N., Ford, J., Eng, C., Burchard, E.G., Sleiman, P.M., Hakonarson, H., Forno, E., Raby, B.A., Weiss, S.T., Scott, A.F., Kabesch, M., Liang, L., Abecasis, G., Moffatt, M.F., Cookson, W.O., Ruczinski, I., Beaty, T.H. & Barnes, K.C. (2010) A genome-wide association study on African-ancestry populations for asthma. *J Allergy Clin Immunol*, **125**, 336-346 e4.
- Matsumoto, M., Lo, S.F., Carruthers, C.J., Min, J., Mariathasan, S., Huang, G., Plas, D.R., Martin, S.M., Geha, R.S., Nahm, M.H. & Chaplin, D.D. (1996) Affinity maturation without germinal centres in lymphotoxin-alpha-deficient mice. *Nature*, **382**, 462-6.
- Messi, M., Giacchetto, I., Nagata, K., Lanzavecchia, A., Natoli, G. & Sallusto, F. (2003) Memory and flexibility of cytokine gene expression as separable properties of human T(H)1 and T(H)2 lymphocytes. *Nat Immunol*, **4**, 78-86.
- Meyers, D.A. (2010) Genetics of asthma and allergy: what have we learned? *J Allergy Clin Immunol*, **126**, 439-46; quiz 447-8.
- Michel, S., Liang, L., Depner, M., Klopp, N., Ruether, A., Kumar, A., Schedel, M., Vogelberg, C., Von Mutius, E., Von Berg, A., Bufe, A., Rietschel, E., Heinzmann, A., Laub, O., Simma, B., Frischer, T., Genuneit, J., Gut, I.G., Schreiber, S., Lathrop, M., Illig, T. & Kabesch, M. (2010) Unifying candidate gene and GWAS approaches in asthma. *PLoS One*, **5**, e13894.
- Migita, O., Noguchi, E., Koga, M., Jian, Z., Shibasaki, M., Migita, T., Ito, S., Ichikawa, K., Matsui, A. & Arinami, T. (2005) Haplotype analysis of a 100 kb region spanning TNF-LTA identifies a polymorphism in the LTA promoter region that is associated with atopic asthma susceptibility in Japan. *Clin Exp Allergy*, **35**, 790-6.

## BIBLIOGRAFÍA

---

- Minty, A., Chalon, P., Derocq, J.M., Dumont, X., Guillemot, J.C., Kaghad, M., Labit, C., Leplatois, P., Liauzun, P., Miloux, B. & Et Al. (1993) Interleukin-13 is a new human lymphokine regulating inflammatory and immune responses. *Nature*, **362**, 248-50.
- Moffatt, M.F., Gut, I.G., Demenais, F., Strachan, D.P., Bouzigon, E., Heath, S., Von Mutius, E., Farrall, M., Lathrop, M. & Cookson, W.O. (2010) A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med*, **363**, 1211-21.
- Moffatt, M.F., Kabesch, M., Liang, L., Dixon, A.L., Strachan, D., Heath, S., Depner, M., Von Berg, A., Bufer, A., Rietschel, E., Heinzmann, A., Simma, B., Frischer, T., Willis-Owen, S.A., Wong, K.C., Illig, T., Vogelberg, C., Weiland, S.K., Von Mutius, E., Abecasis, G.R., Farrall, M., Gut, I.G., Lathrop, G.M. & Cookson, W.O. (2007) Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature*, **448**, 470-3.
- Munthe-Kaas, M.C., Torjussen, T.M., Gervin, K., Lodrup Carlsen, K.C., Carlsen, K.H., Granum, B., Hjorthaug, H.S., Undlien, D. & Lyle, R. (2010) CD14 polymorphisms and serum CD14 levels through childhood: a role for gene methylation? *J Allergy Clin Immunol*, **125**, 1361-8.
- Muro, M., Marín, L., Torio, A., Pagan, J.A. & Álvarez-López, M.R. (2008) CCL5/RANTES chemokine gene promoter polymorphisms are not associated with atopic and nonatopic asthma in a Spanish population. *Int J Immunogenet*, **35**, 19-23.
- Nagelkerke, N.J.D. (1991) A note on a general definition of the coefficient of determination. *Biometrika*, **78**, 691-692.
- Nakashima, K., Hirota, T., Obara, K., Shimizu, M., Jodo, A., Kameda, M., Doi, S., Fujita, K., Shirakawa, T., Enomoto, T., Kishi, F., Yoshihara, S., Matsumoto, K., Saito, H., Suzuki, Y., Nakamura, Y. & Tamari, M. (2006) An association study of asthma and related phenotypes with polymorphisms in negative regulator molecules of the TLR signaling pathway. *J Hum Genet*, **51**, 284-91.
- Neale, B.M. & Sham, P.C. (2004) The future of association studies: gene-based analysis and replication. *Am J Hum Genet*, **75**, 353-62.
- Nedwin, G.E., Naylor, S.L., Sakaguchi, A.Y., Smith, D., Jarrett-Nedwin, J., Pennica, D., Goeddel, D.V. & Gray, P.W. (1985) Human lymphotoxin and tumor necrosis factor genes: structure, homology and chromosomal localization. *Nucleic Acids Res*, **13**, 6361-73.
- Ng, S.C., Tsoi, K.K., Kamm, M.A., Xia, B., Wu, J., Chan, F.K. & Sung, J.J. (2011) Genetics of inflammatory bowel disease in Asia: Systematic review and meta-analysis. *Inflamm Bowel Dis*, en prensa.
- Nicolae, D., Cox, N.J., Lester, L.A., Schneider, D., Tan, Z., Billstrand, C., Kuldane, S., Donfack, J., Kogut, P., Patel, N.M., Goodenbour, J., Howard, T., Wolf, R., Koppelman, G.H., White, S.R., Parry, R., Postma, D.S., Meyers, D., Bleecker, E.R., Hunt, J.S., Solway, J. & Ober, C. (2005) Fine mapping and positional candidate studies identify HLA-G as an asthma susceptibility gene on chromosome 6p21. *Am J Hum Genet*, **76**, 349-57.
- Nicolae, D.L. (2006) Testing untyped alleles (TUNA)-applications to genome-wide association studies. *Genetic Epidemiol*, **30**, 718-27.
- Nicolae, D.L. & Ober, C. (2009) (Too) great expectations: the challenges in replicating asthma disease genes. *Am J Respir Crit Care Med*, **179**, 1078-9.
- Nishimura, F., Shibasaki, M., Ichikawa, K., Arinami, T. & Noguchi, E. (2006) Failure to find an association between CD14-159C/T polymorphism and asthma: a family-based association test and meta-analysis. *Allergol Int*, **55**, 55-8.

- Noguchi, E., Yokouchi, Y., Zhang, J., Shibuya, K., Shibuya, A., Bannai, M., Tokunaga, K., Doi, H., Tamari, M., Shimizu, M., Shirakawa, T., Shibasaki, M., Ichikawa, K. & Arinami, T. (2005) Positional identification of an asthma susceptibility gene on human chromosome 5q33. *Am J Respir Crit Care Med*, **172**, 183-8.
- O'byrne, P.M. & Inman, M.D. (2003) Airway hyperresponsiveness. *Chest*, **123**, 411S-6S.
- Ober, C. (2005) Perspectives on the past decade of asthma genetics. *J Allergy Clin Immunol*, **116**, 274-8.
- Ober, C. & Hoffjan, S. (2006) Asthma genetics 2006: the long and winding road to gene discovery. *Genes Immun*, **7**, 95-100.
- Ober, C., Leavitt, S.A., Tsalenko, A., Howard, T.D., Hoki, D.M., Daniel, R., Newman, D.L., Wu, X., Parry, R., Lester, L.A., Solway, J., Blumenthal, M., King, R.A., Xu, J., Meyers, D.A., Bleecker, E.R. & Cox, N.J. (2000a) Variation in the interleukin 4-receptor alpha gene confers susceptibility to asthma and atopy in ethnically diverse populations. *Am J Hum Genet*, **66**, 517-26.
- Ober, C., Tsalenko, A., Parry, R. & Cox, N.J. (2000b) A second-generation genomewide screen for asthma-susceptibility alleles in a founder population. *Am J Hum Genet*, **67**, 1154-62.
- Ober, C., Tsalenko, A., Willadsen, S., Newman, D., Daniel, R., Wu, X., Andal, J., Hoki, D., Schneider, D., True, K., Schou, C., Parry, R. & Cox, N. (1999) Genome-wide screen for atopy susceptibility alleles in the Hutterites. *Clin Exp Allergy*, **29** Suppl 4, 11-5.
- Ober, C. & Yao, T.C. (2011) The genetics of asthma and allergic disease: a 21st century perspective. *Immunol Rev*, **242**, 10-30.
- Ozaki, K. & Tanaka, T. (2005) Genome-wide association study to identify SNPs conferring risk of myocardial infarction and their functional analyses. *Cell Mol Life Sci*, **62**, 1804-13.
- Pascual, M., Sanz, C., Isidoro-García, M., Dávila, I., Moreno, E., Laffond, E. & Lorente, F. (2008) (CCTTT)<sub>n</sub> polymorphism of NOS2A in nasal polyposis and asthma: a case-control study. *J Investig Allergol Clin Immunol*, **18**, 239-44.
- Pepe, M.S. (2003) *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York, NY: Oxford University Press.
- Pereira, L., Cunha, C., Alves, C. & Amorim, A. (2005) African female heritage in Iberia: a reassessment of mtDNA lineage distribution in present times. *Hum Biol*, **77**, 213-29.
- Pernis, A.B. & Rothman, P.B. (2002) JAK-STAT signaling in asthma. *J Clin Invest*, **109**, 1279-83.
- Pino-Yanes, M., Corrales, A., Basaldúa, S., Hernández, A., Guerra, L., Villar, J. & Flores, C. (2011) North African influences and potential bias in case-control association studies in the Spanish population. *PLoS One*, **6**, e18389.
- Powell, R.M., Wicks, J., Holloway, J.W., Holgate, S.T. & Davies, D.E. (2004) The splicing and fate of ADAM33 transcripts in primary human airways fibroblasts. *Am J Respir Cell Mol Biol*, **31**, 13-21.
- Price, A.L., Butler, J., Patterson, N., Capelli, C., Pascali, V.L., Scarnicci, F., Ruiz-Linares, A., Groop, L., Saetta, A.A., Korkolopoulou, P., Seligsohn, U., Waliszewska, A., Schirmer, C., Ardlie, K., Ramos, A., Nemesh, J., Arbeitman, L., Goldstein, D.B., Reich, D. & Hirschhorn, J.N. (2008) Discerning the ancestry of European Americans in genetic association studies. *PLoS Genet*, **4**, e236.

## BIBLIOGRAFÍA

---

- Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. & Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*, **38**, 904-9.
- Pritchard, J.K. & Rosenberg, N.A. (1999) Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet*, **65**, 220-8.
- Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-59.
- Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R. & Willer, C.J. (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*, **26**, 2336-7.
- Pulit, S.L., Voight, B.F. & De Bakker, P.I. (2010) Multiethnic genetic association studies improve power for locus discovery. *PLoS One*, **5**, e12600.
- Punnonen, J., Aversa, G., Cocks, B.G., McKenzie, A.N., Menon, S., Zurawski, G., De Waal Malefyt, R. & De Vries, J.E. (1993) Interleukin 13 induces interleukin 4-independent IgG4 and IgE synthesis and CD23 expression by human B cells. *Proc Natl Acad Sci U S A*, **90**, 3730-4.
- Purcell, S., Daly, M.J. & Sham, P.C. (2007) WHAP: haplotype-based association analysis. *Bioinformatics*, **23**, 255-6.
- R Development Core Team (2008) R: A Language and Environment for Statistical Computing. Viena: R Foundation for Statistical Computing.
- Radstake, T.R., Gorlova, O., Rueda, B., Martin, J.E., Alizadeh, B.Z., Palomino-Morales, R., Coenen, M.J., Vonk, M.C., Voskuyl, A.E., Schuerwegh, A.J., Broen, J.C., Van Riel, P.L., Van 'T Slot, R., Italiaander, A., Ophoff, R.A., Riemekasten, G., Hunzelmann, N., Simeon, C.P., Ortego-Centeno, N., González-Gay, M.A., González-Escribano, M.F., Airo, P., Van Laar, J., Herrick, A., Worthington, J., Hesselstrand, R., Smith, V., De Keyser, F., Houssiau, F., Chee, M.M., Madhok, R., Shiels, P., Westhovens, R., Kreuter, A., Kiener, H., De Baere, E., Witte, T., Padykov, L., Klareskog, L., Beretta, L., Scorza, R., Lie, B.A., Hoffmann-Vold, A.M., Carreira, P., Varga, J., Hinchcliff, M., Gregersen, P.K., Lee, A.T., Ying, J., Han, Y., Weng, S.F., Amos, C.I., Wigley, F.M., Hummers, L., Nelson, J.L., Agarwal, S.K., Assassi, S., Gourh, P., Tan, F.K., Koeleman, B.P., Arnett, F.C., Martin, J. & Mayes, M.D. (2010) Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. *Nat Genet*, **42**, 426-9.
- Rando, J.C., Cabrera, V.M., Larruga, J.M., Hernández, M., González, A.M., Pinto, F. & Bandelt, H.J. (1999) Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. *Ann Hum Genet*, **63**, 413-28.
- Rando, J.C., Pinto, F., González, A.M., Hernández, M., Larruga, J.M., Cabrera, V.M. & Bandelt, H.J. (1998) Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Ann Hum Genet*, **62**, 531-50.
- Randolph, A.G., Lange, C., Silverman, E.K., Lazarus, R. & Weiss, S.T. (2005) Extended haplotype in the tumor necrosis factor gene cluster is associated with asthma and asthma-related phenotypes. *Am J Respir Crit Care Med*, **172**, 687-92.
- Reihnsaus, E., Innis, M., Macintyre, N. & Liggett, S.B. (1993) Mutations in the gene encoding for the beta 2-adrenergic receptor in normal and asthmatic subjects. *Am J Respir Cell Mol Biol*, **8**, 334-9.
- Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona,

- G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Rienzo, A., Novelletto, A., Oppenheim, A., Norby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Scozari, R., Torroni, A. & Bandelt, H.J. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet*, **67**, 1251-76.
- Roberts, D. & Hiorns, R. (1965) Methods of analysis of the genetic composition of a hybrid population. *Hum Biol*, **37**, 38-43.
- Rodríguez-Ezpeleta, N., Alvarez-Busto, J., Imaz, L., Regueiro, M., Azcarate, M.N., Bilbao, R., Iriando, M., Gil, A., Estonba, A. & Aransay, A.M. (2010) High-density SNP genotyping detects homogeneity of Spanish and French Basques, and confirms their genomic distinctiveness from other European populations. *Hum Genet*, **128**, 113-117.
- Rogers, A.J., Raby, B.A., Lasky-Su, J.A., Murphy, A., Lazarus, R., Klanderman, B.J., Sylvia, J.S., Ziniti, J.P., Lange, C., Celedon, J.C., Silverman, E.K. & Weiss, S.T. (2009) Assessing the reproducibility of asthma candidate gene associations, using genome-wide data. *Am J Respir Crit Care Med*, **179**, 1084-90.
- Rosenberg, N.A., Li, L.M., Ward, R. & Pritchard, J.K. (2003) Informativeness of genetic markers for inference of ancestry. *Am J Hum Genet*, **73**, 1402-22.
- Salari, K., Choudhry, S., Tang, H., Naqvi, M., Lind, D., Avila, P.C., Coyle, N.E., Ung, N., Nazario, S., Casal, J., Torres-Palacios, A., Clark, S., Phong, A., Gomez, I., Matallana, H., Pérez-Stable, E.J., Shriver, M.D., Kwok, P.Y., Sheppard, D., Rodríguez-Cintron, W., Risch, N.J., Burchard, E.G. & Ziv, E. (2005) Genetic admixture and asthma-related phenotypes in Mexican American and Puerto Rican asthmatics. *Genetic Epidemiol*, **29**, 76-86.
- Sánchez-Lerma, B., Morales-Chirivella, F.J., Penuelas, I., Blanco Guerra, C., Mesa Lugo, F., Aguinaga-Ontoso, I. & Guillén-Grima, F. (2009) High prevalence of asthma and allergic diseases in children aged 6 and 7 years from the Canary Islands: the International Study of Asthma and Allergies in Childhood. *J Investig Allergol Clin Immunol*, **19**, 383-90.
- Sandford, A.J., Shirakawa, T., Moffatt, M.F., Daniels, S.E., Ra, C., Faux, J.A., Young, R.P., Nakamura, Y., Lathrop, G.M., Cookson, W.O. & Et Al. (1993) Localisation of atopy and beta subunit of high-affinity IgE receptor (Fc epsilon RI) on chromosome 11q. *Lancet*, **341**, 332-4.
- Sanz, C., Isidoro-García, M., Dávila, I., De Pedro, M.P., Méndez Sde, A., Padron, J., Ruiz-Sanfrancisco, A. & Lorente, F. (2009) A new PTGDR promoter polymorphism in a population of children with asthma. *Pediatr Allergy Immunol*, **20**, 151-6.
- Sanz, C., Isidro-García, M., Dávila, I., Moreno, E., Laffond, E. & Lorente, F. (2006) Analysis of 927T > C CYSLTR1 and -444A > C LTC4S polymorphisms in patients with asthma. *J Investig Allergol Clin Immunol*, **16**, 331-7.
- Sasieni, P.D. (1997) From genotypes to genes: doubling the sample size. *Biometrics*, **53**, 1253-61.
- Scharenberg, A.M. & Kinet, J.P. (1994) Initial events in Fc epsilon RI signal transduction. *J Allergy Clin Immunol*, **94**, 1142-6.
- Schmid-Grendelmeier, P., Altnauer, F., Fischer, B., Bizer, C., Straumann, A., Menz, G., Blaser, K., Wuthrich, B. & Simon, H.U. (2002) Eosinophils express functional IL-13 in eosinophilic inflammatory diseases. *J Immunol*, **169**, 1021-7.
- Selmaj, K., Raine, C.S., Farooq, M., Norton, W.T. & Brosnan, C.F. (1991) Cytokine cytotoxicity against oligodendrocytes. Apoptosis induced by lymphotoxin. *J Immunol*, **147**, 1522-9.



## BIBLIOGRAFÍA

---

- Semino, O., Passarino, G., Oefner, P.J., Lin, A.A., Arbuzova, S., Beckman, L.E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A.S., Cavalli-Sforza, L.L. & Underhill, P.A. (2000) The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science*, **290**, 1155-9.
- Sharma, S., Sharma, A., Kumar, S., Sharma, S.K. & Ghosh, B. (2006) Association of TNF haplotypes with asthma, serum IgE levels, and correlation with serum TNF-alpha levels. *Am J Respir Cell Mol Biol*, **35**, 488-95.
- Shirai, T., Yamaguchi, H., Ito, H., Todd, C.W. & Wallace, R.B. (1985) Cloning and expression in *Escherichia coli* of the gene for human tumour necrosis factor. *Nature*, **313**, 803-6.
- Shirakawa, T., Mao, X.Q., Sasaki, S., Enomoto, T., Kawai, M., Morimoto, K. & Hopkin, J. (1996) Association between atopic asthma and a coding variant of Fc epsilon RI beta in a Japanese population. *Hum Mol Genet*, **5**, 1129-30.
- Simpson, A., John, S.L., Jury, F., Niven, R., Woodcock, A., Ollier, W.E. & Custovic, A. (2006) Endotoxin exposure, CD14, and allergic disease: an interaction between genes and the environment. *Am J Respir Crit Care Med*, **174**, 386-92.
- Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D., Boutin, P., Vincent, D., Belisle, A., Hadjadj, S., Balkau, B., Heude, B., Charpentier, G., Hudson, T.J., Montpetit, A., Pshezhetsky, A.V., Prentki, M., Posner, B.I., Balding, D.J., Meyre, D., Polychronakos, C. & Froguel, P. (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*, **445**, 881-5.
- Sleiman, P.M., Flory, J., Imielinski, M., Bradfield, J.P., Annaiah, K., Willis-Owen, S.A., Wang, K., Rafaels, N.M., Michel, S., Bonnelykke, K., Zhang, H., Kim, C.E., Frackelton, E.C., Glessner, J.T., Hou, C., Otieno, F.G., Santa, E., Thomas, K., Smith, R.M., Glaberson, W.R., Garris, M., Chiavacci, R.M., Beaty, T.H., Ruczinski, I., Orange, J.M., Allen, J., Spergel, J.M., Grundmeier, R., Mathias, R.A., Christie, J.D., Von Mutius, E., Cookson, W.O., Kabesch, M., Moffatt, M.F., Grunstein, M.M., Barnes, K.C., Devoto, M., Magnusson, M., Li, H., Grant, S.F., Bisgaard, H. & Hakonarson, H. (2010) Variants of DENND1B associated with asthma in children. *N Engl J Med*, **362**, 36-44.
- Smyth, D.J., Howson, J.M., Payne, F., Maier, L.M., Bailey, R., Holland, K., Lowe, C.E., Cooper, J.D., Hulme, J.S., Vella, A., Dahlman, I., Lam, A.C., Nutland, S., Walker, N.M., Twells, R.C. & Todd, J.A. (2006) Analysis of polymorphisms in 16 genes in type 1 diabetes that have been associated with other immune-mediated diseases. *BMC Med Genet*, **7**, 20.
- Soderhall, C., Marenholz, I., Kerscher, T., Ruschendorf, F., Esparza-Gordillo, J., Worm, M., Gruber, C., Mayr, G., Albrecht, M., Rohde, K., Schulz, H., Wahn, U., Hubner, N. & Lee, Y.A. (2007) Variants in a novel epidermal collagen gene (COL29A1) are associated with atopic dermatitis. *PLoS Biol*, **5**, e242.
- Soriano, J.B., De Cid, R., Estivill, X., Anto, J.M., Sunyer, J., Otero, D., Roca, J., Rodríguez-Roisin, R., Morell, F., Rodrigo, M.J., Ercilla, G., Beaty, T.H. & Lázaro, C. (2000) Association study of proposed candidate genes/regions in a population of Spanish asthmatics. *Eur J Epidemiol*, **16**, 745-50.
- Soriano, J.B., Ercilla, G., Sunyer, J., Real, F.X., Lázaro, C., Rodrigo, M.J., Estivill, X., Roca, J., Rodríguez-Roisin, R., Morell, F. & Anto, J.M. (1997) HLA class II genes in soybean epidemic asthma patients. *Am J Respir Crit Care Med*, **156**, 1394-8.
- Stephens, M., Smith, N.J. & Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*, **68**, 978-89.
- Storey, J.D. & Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A*, **100**, 9440-5.

- Strachan, D.P. (1989) Hay fever, hygiene, and household size. *BMJ*, **299**, 1259-60.
- Sun, X., Ma, S.F., Wade, M.S., Flores, C., Pino-Yanes, M., Moitra, J., Ober, C., Kittles, R., Hussain, A., Ford, J.G. & García, J.G.N. (2010) Functional variants of sphingosine-1-phosphate receptor 1 gene associate with asthma susceptibility. *J Allergy Clin Immunol*, **126**, 241-9.
- Szefler, S.J., Ando, R., Cicutto, L.C., Surs, W., Hill, M.R. & Martin, R.J. (1991) Plasma histamine, epinephrine, cortisol, and leukocyte beta-adrenergic receptors in nocturnal asthma. *Clin Pharmacol Ther*, **49**, 59-68.
- Tellería, J.J., Blanco-Quirós, A., Muntión, S., Antonio Garrote, J., Arranz, E., Armentia, A., Díez, I. & Castro, J. (2006) Tachyphylaxis to beta2-agonists in Spanish asthmatic patients could be modulated by beta2-adrenoceptor gene polymorphisms. *Respir Med*, **100**, 1072-8.
- Tellería, J.J., Blanco-Quirós, A., Varillas, D., Armentia, A., Fernandez-Carvajal, I., Jesus Alonso, M. & Díez, I. (2008) ALOX5 promoter genotype and response to montelukast in moderate persistent asthma. *Respir Med*, **102**, 857-61.
- Teuffel, O., Ethier, M.C., Beyene, J. & Sung, L. (2010) Association between tumor necrosis factor-alpha promoter -308 A/G polymorphism and susceptibility to sepsis and sepsis mortality: a systematic review and meta-analysis. *Crit Care Med*, **38**, 276-82.
- The Spanish Group of the European Asthma Study (1995) The European Asthma Study. The prevalence of asthma-related symptoms in 5 Spanish areas. *Med Clin (Barc)*, **104**, 487-92.
- Thomsen, S.F., Van Der Sluis, S., Kyvik, K.O., Skytthe, A. & Backer, V. (2010) Estimates of asthma heritability in a large twin sample. *Clin Exp Allergy*, **40**, 1054-61.
- Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J.B., Awomoyi, A.A., Bodo, J.M., Doumbo, O., Ibrahim, M., Juma, A.T., Kotze, M.J., Lema, G., Moore, J.H., Mortensen, H., Nyambo, T.B., Omar, S.A., Powell, K., Pretorius, G.S., Smith, M.W., Thera, M.A., Wambebe, C., Weber, J.L. & Williams, S.M. (2009) The genetic structure and history of Africans and African Americans. *Science*, **324**, 1035-44.
- Torgerson, D.G., Ampleford, E.J., Chiu, G.Y., Gauderman, W.J., Gignoux, C.R., Graves, P.E., Himes, B.E., Levin, A.M., Mathias, R.A., Hancock, D.B., Baurley, J.W., Eng, C., Stern, D.A., Celedon, J.C., Rafaels, N., Capurso, D., Conti, D.V., Roth, L.A., Soto-Quirós, M., Togias, A., Li, X., Myers, R.A., Romieu, I., Berg, D.J., Hu, D., Hansel, N.N., Hernández, R.D., Israel, E., Salam, M.T., Galanter, J., Avila, P.C., Avila, L., Rodriguez-Santana, J.R., Chapela, R., Rodríguez-Cintron, W., Diette, G.B., Adkinson, N.F., Abel, R.A., Ross, K.D., Shi, M., Faruque, M.U., Dunston, G.M., Watson, H.R., Mantese, V.J., Ezurum, S.C., Liang, L., Ruczinski, I., Ford, J.G., Huntsman, S., Chung, K.F., Vora, H., Li, X., Calhoun, W.J., Castro, M., Sienra-Monge, J.J., Del Rio-Navarro, B., Deichmann, K.A., Heinzmann, A., Wenzel, S.E., Busse, W.W., Gern, J.E., Lemanske, R.F., Jr., Beaty, T.H., Bleecker, E.R., Raby, B.A., Meyers, D.A., London, S.J., Gilliland, F.D., Burchard, E.G., Martínez, F.D., Weiss, S.T., Williams, L.K., Barnes, K.C., Ober, C. & Nicolae, D.L. (2011) Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet*, **43**, 887-92.
- Torío, A., Sánchez-Guerrero, I., Muro, M., Herrero, N., Pagán, J., Minguela, A., Marín, L., Moya-Quiles, M.R., Sanchís, M.J. & Alvarez-López, M.R. (2000) Analysis of the phenotypic distribution of HLA class I and class II in atopic and non-atopic asthma patients. *Eur J Immunogenet*, **27**, 81-5.
- Torío, A., Sánchez-Guerrero, I., Muro, M., Villar, L.M., Minguela, A., Marín, L., Moya-Quiles, M.R., Montes-Ares, O., Pagan, J. & Alvarez-López, M.R. (2003) HLA class II

## BIBLIOGRAFÍA

---

- genotypic frequencies in atopic asthma: association of DRB1\*01-DQB1\*0501 genotype with *Artemisia vulgaris* allergic asthma. *Hum Immunol*, **64**, 811-5.
- Torres-Galván, S.M., Cumplido, J.A., Buset, N., Caballero-Hidalgo, A., Blanco, C., Hernández, E. & Carrillo, T. (2009) 5-Lipoxygenase pathway gene polymorphisms: lack of association with asthma in a Spanish population. *J Investig Allergol Clin Immunol*, **19**, 453-8.
- Tregouet, D.A., Escolano, S., Tiret, L., Mallet, A. & Golmard, J.L. (2004) A new algorithm for haplotype-based association analysis: the Stochastic-EM algorithm. *Ann Hum Genet*, **68**, 165-77.
- Tsai, H.J., Choudhry, S., Naqvi, M., Rodríguez-Cintron, W., Burchard, E.G. & Ziv, E. (2005) Comparison of three methods to estimate genetic ancestry and control for stratification in genetic association studies among admixed populations. *Hum Genet*, **118**, 424-33.
- Tsai, H.J., Kho, J.Y., Shaikh, N., Choudhry, S., Naqvi, M., Navarro, D., Matallana, H., Castro, R., Lilly, C.M., Watson, H.G., Meade, K., Lenoir, M., Thyne, S., Ziv, E. & Burchard, E.G. (2006) Admixture-matched case-control study: a practical approach for genetic association studies in admixed populations. *Hum Genet*, **118**, 626-39.
- Van Eerdewegh, P., Little, R.D., Dupuis, J., Del Mastro, R.G., Falls, K., Simon, J., Torrey, D., Pandit, S., Mckenny, J., Braunschweiger, K., Walsh, A., Liu, Z., Hayward, B., Folz, C., Manning, S.P., Bawa, A., Saracino, L., Thackston, M., Benchekroun, Y., Capparell, N., Wang, M., Adair, R., Feng, Y., Dubois, J., Fitzgerald, M.G., Huang, H., Gibson, R., Allen, K.M., Pedan, A., Danzig, M.R., Umland, S.P., Egan, R.W., Cuss, F.M., Rorke, S., Clough, J.B., Holloway, J.W., Holgate, S.T. & Keith, T.P. (2002) Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. *Nature*, **418**, 426-30.
- Vercelli, D. (2008) Discovering susceptibility genes for asthma and allergy. *Nat Rev Immunol*, **8**, 169-82.
- Vergara, C., Caraballo, L., Mercado, D., Jimenez, S., Rojas, W., Rafaels, N., Hand, T., Campbell, M., Tsai, Y.J., Gao, L., Duque, C., López, S., Bedoya, G., Ruiz-Linares, A. & Barnes, K.C. (2009) African ancestry is associated with risk of asthma and high total serum IgE in a population from the Caribbean Coast of Colombia. *Hum Genet*, **125**, 565-79.
- Villar, J., Cabrera, N., Casula, M., Flores, C., Valladares, F., Muros, M., Blanch, L., Slutsky, A.S. & Kacmarek, R.M. (2010) Mechanical ventilation modulates Toll-like receptor signaling pathway in a sepsis-induced lung injury model. *Intensive Care Med*, **36**, 1049-57.
- Villar, J., Maca-Meyer, N., Pérez-Méndez, L. & Flores, C. (2004) Bench-to bedside review: understanding genetic predisposition to sepsis. *Crit Care*, **8**, 180-9.
- Vineis, P., Schulte, P. & McMichael, A.J. (2001) Misconceptions about the use of genetic tests in populations. *Lancet*, **357**, 709-12.
- Wang, J. (2003) Maximum-likelihood estimation of admixture proportions from genetic data. *Genetics*, **164**, 747-65.
- Wang, J.Y., Liou, Y.H., Wu, Y.J., Hsiao, Y.H. & Wu, L.S. (2009) An association study of 13 SNPs from seven candidate genes with pediatric asthma and a preliminary study for genetic testing by multiple variants in Taiwanese population. *J Clin Immunol*, **29**, 205-9.
- Wei, Z., Wang, K., Qu, H.Q., Zhang, H., Bradfield, J., Kim, C., Frackleton, E., Hou, C., Glessner, J.T., Chiavacci, R., Stanley, C., Monos, D., Grant, S.F., Polychronakos, C. & Hakonarson, H. (2009) From disease association to risk assessment: an optimistic

- view from genome-wide association studies on type 1 diabetes. *PLoS Genet*, **5**, e1000678.
- Weidinger, S., Klopp, N., Rummeler, L., Wagenpfeil, S., Novak, N., Baurecht, H.J., Groer, W., Darsow, U., Heinrich, J., Gauger, A., Schafer, T., Jakob, T., Behrendt, H., Wichmann, H.E., Ring, J. & Illig, T. (2005) Association of NOD1 polymorphisms with atopic eczema and related phenotypes. *J Allergy Clin Immunol*, **116**, 177-84.
- Weir, B. (1996) Genetic data Analysis II. Sunderland: Sinauer Associates Publishers.
- Wen, X. & Nicolae, D.L. (2008) Association studies for untyped markers with TUNA. *Bioinformatics*, **24**, 435-7.
- Wesche, H., Gao, X., Li, X., Kirschning, C.J., Stark, G.R. & Cao, Z. (1999) IRAK-M is a novel member of the Pelle/interleukin-1 receptor-associated kinase (IRAK) family. *J Biol Chem*, **274**, 19403-10.
- White, J.H., Chiano, M., Wigglesworth, M., Geske, R., Riley, J., White, N., Hall, S., Zhu, G., Maurio, F., Savage, T., Anderson, W., Cordy, J., Ducceschi, M., Vestbo, J. & Pillai, S.G. (2008) Identification of a novel asthma susceptibility gene on chromosome 1qter and its functional evaluation. *Hum Mol Genet*, **17**, 1890-903.
- White, J.M. (2003) ADAMs: modulators of cell-cell and cell-matrix interactions. *Curr Opin Cell Biol*, **15**, 598-606.
- Wigginton, J.E., Cutler, D.J. & Abecasis, G.R. (2005) A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet*, **76**, 887-93.
- Wills-Karp, M. & Ewart, S.L. (2004) Time to draw breath: asthma-susceptibility genes are identified. *Nat Rev Genet*, **5**, 376-87.
- Wright, S.D., Ramos, R.A., Tobias, P.S., Ulevitch, R.J. & Mathison, J.C. (1990) CD14, a receptor for complexes of lipopolysaccharide (LPS) and LPS binding protein. *Science*, **249**, 1431-3.
- Wu, H., Romieu, I., Shi, M., Hancock, D.B., Li, H., Sienra-Monge, J.J., Chiu, G.Y., Xu, H., Del Rio-Navarro, B.E. & London, S.J. (2010) Evaluation of candidate genes in a genome-wide association study of childhood asthma in Mexicans. *J Allergy Clin Immunol*, **125**, 321-327 e13.
- Xu, M., Tantisira, K.G., Wu, A., Litonjua, A.A., Chu, J.H., Himes, B.E., Damask, A. & Weiss, S.T. (2011) Genome Wide Association Study to predict severe asthma exacerbations in children using random forests classifiers. *BMC Med Genet*, **12**, 90.
- Yang, Y., Sun, R., Yang, H., Zheng, F. & Gong, F. (2011) -308 G > A of TNF-alpha gene promoter decreases the risk of multiple sclerosis: a meta-analysis. *Mult Scler*, **17**, 658-65.
- Yoshinaka, T., Nishii, K., Yamada, K., Sawada, H., Nishiwaki, E., Smith, K., Yoshino, K., Ishiguro, H. & Higashiyama, S. (2002) Identification and characterization of novel mouse and human ADAM33s with potential metalloprotease activity. *Gene*, **282**, 227-36.
- Zakharia, F., Basu, A., Absher, D., Assimes, T.L., Go, A.S., Hlatky, M.A., Iribarren, C., Knowles, J.W., Li, J., Narasimhan, B., Sidney, S., Southwick, A., Myers, R.M., Quertermous, T., Risch, N. & Tang, H. (2009) Characterizing the admixed African ancestry of African Americans. *Genome Biol*, **10**, R141.
- Zeggini, E. & Morris, A.P. (2011) Analysis of complex disease association studies: a practical guide. Amsterdam: Elsevier/Academic Press.
- Zhang, Y., Leaves, N.I., Anderson, G.G., Ponting, C.P., Broxholme, J., Holt, R., Edser, P., Bhattacharyya, S., Dunham, A., Adcock, I.M., Pulleyn, L., Barnes, P.J., Harper, J.I.,

## BIBLIOGRAFÍA

---

- Abecasis, G., Cardon, L., White, M., Burton, J., Matthews, L., Mott, R., Ross, M., Cox, R., Moffatt, M.F. & Cookson, W.O. (2003) Positional cloning of a quantitative trait locus on chromosome 13q14 that influences immunoglobulin E levels and asthma. *Nat Genet*, **34**, 181-6.
- Zhang, Y., Tian, C., Zhang, J., Li, X., Wan, H., He, C., Guo, L., Meilang, Q., Peng, C., Duo, L., Huang, J. & Fan, H. (2011) The -159C/T polymorphism in the CD14 gene and the risk of asthma: a meta-analysis. *Immunogenetics*, **63**, 23-32.
- Zhao, L. & Bracken, M.B. (2011) Association of CD14 -260 (-159) C>T and asthma: a systematic review and meta-analysis. *BMC Med Genet*, **12**, 93.
- Zhu, Z., Homer, R.J., Wang, Z., Chen, Q., Geba, G.P., Wang, J., Zhang, Y. & Elias, J.A. (1999) Pulmonary expression of interleukin-13 causes inflammation, mucus hypersecretion, subepithelial fibrosis, physiologic abnormalities, and eotaxin production. *J Clin Invest*, **103**, 779-88.
- Zondervan, K.T. & Cardon, L.R. (2007) Designing candidate gene and genome-wide case-control association studies. *Nat Protoc*, **2**, 2492-501.
- Zou, J., Zhu, F., Liu, J., Wang, W., Zhang, R., Garlisi, C.G., Liu, Y.H., Wang, S., Shah, H., Wan, Y. & Umland, S.P. (2004) Catalytic activity of human ADAM33. *J Biol Chem*, **279**, 9818-30.
- Zurawski, S.M., Vega, F., Jr., Huyghe, B. & Zurawski, G. (1993) Receptors for interleukin-13 and interleukin-4 are complex and share a novel component that functions in signal transduction. *EMBO J*, **12**, 2663-70.