



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Programa de Maestría y Doctorado en Ciencias Médicas, Odontológicas y de la Salud
Facultad de Medicina
Ciencias de la Salud: Epidemiología

Identificación de variantes genéticas de contribución ancestral amerindia asociadas a enfermedad pulmonar obstructiva crónica (EPOC) en fumadores mexicanos mestizos mediante análisis de genoma completo

Que para optar el Grado de
Maestro en Ciencias de la Salud

Presenta:
Biól. Luis Alberto López Flores

Director de tesis:

Dr. en C. Ramcés Falfán Valencia
Instituto Nacional de Enfermedades Respiratorias
Ismael Cosío Villegas

Comité tutor:

Dra. Alejandra Ramírez Venegas
Instituto Nacional de Enfermedades Respiratorias
Ismael Cosío Villegas

M. en C. Juan Carlos Fernández López
Instituto Nacional de Medicina Genómica

Ciudad Universitaria, CD. MX. Agosto 2019



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Contenido

1	Introducción	5
1.1	EPOC	8
1.1.1	Epidemiología de la EPOC.....	9
1.2	La paradoja hispana.....	14
1.3	Variación genética humana	16
1.4	Ancestría genética humana	18
1.5	Mexicano mestizo.....	21
1.6	Genómica de la EPOC.....	23
1.6.1	GWAS en EPOC.....	25
2	Antecedentes	30
3	Planteamiento del problema	34
4	Justificación	35
5	Hipótesis	36
6	Objetivos	37
6.1	Objetivo general	37
6.2	Objetivos específicos	37
7	Materiales y métodos	38
7.1	Consideraciones éticas	38
7.2	Diseño de estudio	38
7.3	Población de estudio.....	38
7.4	Operacionalización y medición de las variables	39
7.4.1	Datos demográficos.....	40
7.4.2	Ancestría autoreportada	40
7.4.3	Antecedentes heredo patológicos respiratorios.....	40
7.4.4	Tabaquismo	40
7.4.5	Otras exposiciones	41
7.4.6	Función pulmonar.....	41
7.5	Selección y procesamiento de las muestras	41
7.5.1	Selección de muestras	41
7.5.2	Extracción de ADN	42
7.5.3	Control de calidad del ADN	43
7.6	Microarreglo Infinium® Multi-Ethnic AMR/AFR BeadChip	43
7.6.1	Características generales.....	43
7.6.2	Genotipificación	44
7.6.3	Control de calidad.....	44
7.7	Análisis de ancestría	45
7.8	Asociación de genoma completo	45
7.9	Anotación genética.....	47
7.10	Análisis estadístico	47
7.11	Visualización de datos.....	48
8	Resultados	49
8.1	Descripción de variables	49
8.2	Ancestría	52
8.2.1	Análisis de correlación.....	54
8.3	Análisis de genoma completo	55
8.3.1	Modelo de regresión.....	55
8.3.2	Estratificación de variables en el modelo.....	64
8.3.3	Regiones cromosómicas	75
8.4	Relación de variantes con el componente amerindio.....	76
9	Discusión	81
10	Conclusiones	96
11	Bibliografía	97

Cuadros

Cuadro 1.1. Clasificación de la severidad por limitación al flujo de aire en EPOC por estadio GOLD.....	9
Cuadro 1.2. Determinantes de la salud en EPOC mediante modelo de Dahlgren-Whitehead.....	10
Cuadro 7.1. Definición de población de estudio.....	38
Cuadro 7.2. Criterios de inclusión.....	39
Cuadro 7.3. Criterios de exclusión.....	39
Cuadro 7.4. Operacionalización de las variables.....	39
Cuadro 7.5. Paquetes de R empleados en análisis de datos.....	48
Cuadro 7.6. Paquetes de R empleados en la visualización de datos.....	48

Tablas

Tabla 1.1. Reportes de los componentes ancestrales en el mexicano mestizo en distintas regiones geográficas.....	22
Tabla 1.2. Loci con mayor número de variantes asociadas a EPOC y función pulmonar en distintas poblaciones.....	28
Tabla 1.3. Estudios con estrategia de GWAS en EPOC y función pulmonar.....	29
Tabla 1.4. Meta-análisis que han utilizado la técnica GWAS con variables respiratorias.....	29
Tabla 2.1. Variantes genéticas asociadas con variables respiratorias en población hispana por medio de GWAS.....	31
Tabla 2.2. Proporción de los componentes ancestrales en cohortes de población hispana reportados en GWAS con EPOC.....	32
Tabla 8.1. Variables demográficas, de tabaquismo, antecedentes y de función pulmonar de los grupos de estudio.....	51
Tabla 8.2. Variables clínicas en los pacientes con EPOC.....	51
Tabla 8.3. Variables de ajuste en el modelo.....	56
Tabla 8.4. Variantes asociadas a EPOC en grupo de estudio de acuerdo al modelo sin ajustar.....	61
Tabla 8.5. Variantes asociadas a EPOC en grupo de estudio de acuerdo al modelo ajustado.....	61
Tabla 8.6. Variables de ajuste en el modelo estratificado en hombres.....	66
Tabla 8.7. Variables de ajuste en el modelo estratificado en mujeres.....	66
Tabla 8.8. Variantes asociadas a EPOC de acuerdo al modelo ajustado estratificado por sexo.....	69
Tabla 8.9. Variables de ajuste en el modelo estratificado en fumadores actuales.....	72
Tabla 8.10. Variables de ajuste en el modelo estratificado en exfumadores.....	72
Tabla 8.11. Variantes asociadas a EPOC de acuerdo al modelo ajustado estratificado por estatus de tabaquismo.....	74
Tabla 8.12. Variantes asociadas en el modelo sin ajustar y su relación respecto al componente AMI en casos y controles.....	78
Tabla 8.13. Variantes asociadas en el modelo ajustado y su relación respecto al componente AMI en casos y controles.....	78

Figuras

Figura 1.1. Árbol filogenético de las poblaciones humanas y su relación con lenguajes.....	18
Figura 7.1. Algoritmo de selección de muestras.....	42
Figura 7.2. Algoritmo de variables de ajuste para el GWAS por modelo de regresión logística.....	46
Figura 7.3. Algoritmo de las variantes genéticas asociadas en el GWAS.....	46
Figura 8.1. Proporción de los componentes ancestrales en cada sujeto de grupo de estudio.....	53
Figura 8.2. Análisis de componentes principales entre PC1 y PC2 en los grupos de estudio.....	53
Figura 8.3. Corr plot de las variables cuantitativas de los grupos de estudio.....	55
Figura 8.4. QQ-plot del modelo sin ajustar (modelo 0).....	59
Figura 8.5. Manhattan plot del modelo sin ajustar (modelo 0).....	59
Figura 8.6. QQ-plot del modelo ajustado.....	60
Figura 8.7. Manhattan plot del modelo ajustado.....	60
Figura 8.8. Gráfico de asociación regional en modelo sin ajustar (modelo 0).....	62
Figura 8.9. Gráfico de asociación regional en modelo ajustado.....	63
Figura 8.10. Manhattan plot del modelo ajustado estratificado en hombres.....	67
Figura 8.11. Manhattan plot del modelo ajustado estratificado en mujeres.....	67
Figura 8.12. Gráfico de asociación regional en modelo ajustado estratificado en hombres.....	68
Figura 8.13. Gráfico de asociación regional en modelo ajustado estratificado en mujeres.....	68
Figura 8.14. Manhattan plot del modelo ajustado estratificado en fumadores actuales.....	73
Figura 8.15. Manhattan plot del modelo ajustado estratificado en exfumadores.....	73
Figura 8.16. Gráfico de asociación regional en modelo ajustado estratificado por estatus de tabaquismo.....	74
Figura 8.17. Circle plot de las regiones cromosómicas de los SNP asociados de acuerdo al modelo de regresión realizado.....	766
Figura 8.18. Distribución de los alelos asociados en el modelo sin ajustar respecto a la ancestría amerindia.....	79
Figura 8.22. Distribución de los alelos asociados en el modelo ajustado respecto a la ancestría amerindia.....	80

Lista de abreviaturas

1kGP	1000 Genomes Project
A1AT	Alfa 1 antitripsina
AFR	Africana
AMI	Amerindia
CPD	Cigarros por día
CVF	Capacidad vital forzada
EPOC	Enfermedad pulmonar obstructiva crónica
EUR	Europea
GOLD	Estrategia Global para el Diagnóstico, Manejo y Prevención de la EPOC
GWAS	Estudio de asociación de genoma completo
GWMAS	Estudio de metaanálisis de asociación de genoma completo
HGDP	Human Genome Diversity Project
IBD	Identidad por descendiente
IT	Índice tabaquico
OR	Razón de momios
PCA	Análisis de componentes principales
PCn	Componente principal n
PLATINO	Proyecto Latinoamericano de Investigación en Obstrucción Pulmonar
RIC	Rango intercuatílico
SNP	Polimorfismo de un sólo nucleótido
VEF ₁	Volumen espiratorio forzado al primer segundo

Resumen

La enfermedad pulmonar obstructiva crónica se encuentra como la cuarta causa de mortalidad global. A pesar de que se han identificado determinantes de salud y factores de riesgo ambientales para su desarrollo, los factores biológicos implicados en la etiología no son del todo claros. Los factores genéticos presentan un papel importante en las enfermedades comunes y complejas como la EPOC, sin embargo, se ha reportado que la distribución de estos factores son particulares en las distintas poblaciones humanas. Estrategias novedosas, como estudios de asociación de genoma completo han tenido la capacidad de identificar factores genéticos de riesgo en regiones en el genoma no contempladas previamente, sin embargo, han sido realizados mayoritariamente en poblaciones de componente ancestral europeo, africano y asiático. El objetivo de este estudio fue identificar variantes genéticas de contribución ancestral amerindia asociadas con EPOC secundaria a tabaquismo en pacientes mexicanos mestizos. Se realizó un estudio de asociación de genoma completo mediante el microarreglo Infinium® Multi-Ethnic AMR/AFR BeadChip, en el cual se incluyeron 575 participantes en total, 221 fumadores con EPOC y 344 fumadores sin EPOC, de los cuales se pudieron evaluar 738,822 SNP. Se determinaron las proporciones de los componentes ancestrales de los individuos, en los cuales se muestra que los controles tienen una mediana de 50.3% (36.9-62.3) de componente amerindio y los casos 46.6% (35-58.3). Posteriormente, mediante un análisis de regresión logística multinomial en un modelo sin ajustar, ajustado por variables de controles y estratificando por sexo y estatus de tabaquismo se obtuvo significancia sugestiva de genoma completo ($p < 10 \times 10^{-5}$) en un total de 23 SNPs asociados, de las cuales 17 sugieren cierta plausibilidad biológica con la EPOC y 2 se asocian al componente amerindio. Empleando la estrategia de análisis de genoma completo, se identificaron variantes genéticas asociadas a EPOC en fumadores mexicanos mestizos, que sugieren plausibilidad biológica en la enfermedad. Los resultados obtenidos están sujetos a validación poblacional, para confirmar o descartar factores genéticos de riesgo en la EPOC.

1 Introducción

La enfermedad pulmonar obstructiva crónica (EPOC) se encuentra en los primeros lugares de mortalidad en el mundo, situación que ha dado importancia para intervenir en la prevención y tratamiento mediante la comprensión de su compleja etiología. A pesar de que se han identificado determinantes de salud y factores de riesgo ambientales para su desarrollo, los factores biológicos implicados en la etiología no son del todo claros. Los factores genéticos presentan un papel importante en las enfermedades comunes y complejas como la EPOC, sin embargo, se ha reportado que la distribución de estos factores son particulares en las distintas poblaciones humanas.

Evidencias en otros países, como el fenómeno de la Paradoja Hispana en EE. UU., nos demuestra que existen diferencias en salud entre distintos grupos raciales. En este sentido, destacan las enfermedades respiratorias crónicas de vías inferiores, como la EPOC, la cual tiene una prevalencia menor en México en comparación con otras poblaciones del mundo, inclusive que otros países de Latinoamérica, lo cual podrían ser, entre otros factores, reflejo de la mayor proporción del componente ancestral amerindio en esta población mestiza.

Estrategias novedosas, como estudios de asociación de genoma completo (GWAS) han tenido la capacidad de identificar factores genéticos de riesgo en regiones en el genoma no contempladas previamente, sin embargo, han sido realizados mayoritariamente en poblaciones de componente ancestral europeo, africano y asiático. A pesar de existir algunos reportes en poblaciones hispanas, las cuales son muy heterogéneas, hasta ahora, no existe un GWAS exclusivo en sujetos mexicanos con EPOC que haga énfasis en el componente amerindio.

1.1 EPOC

La EPOC es definida como una enfermedad no transmisible, común, prevenible y tratable, caracterizada por una limitación persistente al flujo de aire debido a la mezcla de enfermedad de vías aéreas (bronquitis crónica) y/o destrucción del parénquima (enfisema), causado por la exposición crónica a gases y/o partículas nocivas en las vías aéreas y pulmones. ¹ En 1977 se describió la historia natural de la enfermedad (conocida en ese entonces como enfisema) en sujetos fumadores mayores de 40 años, los cuales presentaban alteraciones pulmonares, y síntomas respiratorios. ²

La Estrategia Global para el Diagnóstico, Manejo y Prevención de la EPOC (GOLD, por sus siglas en inglés) es el organismo internacional encargado de coleccionar y presentar evidencia actual confiable respecto a la evaluación, diagnóstico y tratamiento de la EPOC. ¹ Según la GOLD, para una sospecha de EPOC, se deben tomar en cuenta distintos factores en el individuo, entre estos están ser mayor de 40 años; presentar síntomas respiratorios que incluyen disnea, tos recurrente y producción de esputo; así como exposición crónica a gases y/o partículas nocivas. Sin embargo, con la exploración clínica no es suficiente, por lo cual existe la prueba de espirometría para confirmar el diagnóstico de EPOC. ¹

La espirometría es una prueba no invasiva que mide la función pulmonar, principalmente mediante dos variables: el volumen máximo de aire exhalado de los pulmones en una maniobra de esfuerzo máximo (capacidad vital forzada, CVF) y el volumen máximo de aire exhalado en el primer segundo de la maniobra (volumen espiratorio forzado al primer segundo, VEF₁), así como la relación de estas variables en el cociente VEF₁/CVF. ³ Los parámetros de las pruebas de función pulmonar presentan variabilidad interindividual y dependen de las características antropométricas del paciente (sexo, edad, talla, peso y raza). Es por ello que se utilizan valores de referencia para cada individuo, obtenidos de ecuaciones de referencia específicos en cada población, los cuales tienen que representar los valores normales en individuos clínicamente sanos (particularmente a nivel

respiratorio) de acuerdo a las variables de sexo, edad, talla, peso y raza, obteniendo valores predichos acordes para cada individuo. ⁴⁻⁶

La EPOC es una enfermedad progresiva si no es correctamente tratada, la cual presenta una escala de severidad, clasificada según GOLD en los estadios descritos en el **Cuadro 1.1**.

Cuadro 1.1 Clasificación de la severidad por limitación al flujo de aire en EPOC (basado en VEF₁ post-broncodilatador) por estadio GOLD.

Estadio GOLD	Severidad	Parámetros de referencia
I	Leve	VEF ₁ ≥ 80% del valor predicho
II	Moderado	50% ≤ VEF ₁ < 80% del valor predicho
III	Grave	30% ≤ VEF ₁ < 50% del valor predicho
IV	Muy grave	VEF ₁ < 30% del valor predicho

El primer sistema afectado en la EPOC es el respiratorio, comprometiendo principalmente la función pulmonar en VEF₁ con una pérdida superior a 90 mL al año. ^{7,8} Sin embargo, a pesar de que la función pulmonar naturalmente disminuye a partir de los 40 años (aproximadamente 22-62 ml al año en VEF₁) ⁹⁻¹², se ha descrito que varía de acuerdo al componente racial, la cual podría explicarse por factores genéticos, ^{13,14} más adelante abordados.

1.1.1 Epidemiología de la EPOC

La EPOC se considera un gran problema de salud pública, debido a la alta mortalidad que ocasiona, así como los altos costos económicos que origina, todo esto a pesar de ser una enfermedad altamente prevenible. Actualmente, se estima como la cuarta causa de mortalidad global, ¹⁵ sin embargo, se proyecta que se coloque en el tercer puesto en 2020. ¹⁶ La carga global causada por la enfermedad es bastante discutida, ya que existen distintas metodologías utilizadas para su abordaje y no se ha podido llegar a un consenso, sin embargo, recientemente se han calculado estimaciones confiables con base en la información disponible de 1990 a 2015, en los cuales se reporta que ha habido un aumento en la mortalidad y prevalencia del 11.8% y 44.2%, respectivamente. ¹⁷

La EPOC es una enfermedad compleja y multifactorial, causada por una red compleja de factores extrínsecos e intrínsecos que interaccionan en función del tiempo. Estos factores, de acuerdo con Dahlgren y Whitehead ¹⁸, se pueden agrupar en distintos niveles según sea la contribución al individuo, desde el contexto individual hasta el global. Entre estas agrupaciones, denominadas determinantes de la salud, se presentan diferencias o desbalances que concluyen en el desarrollo de EPOC en el individuo. Los determinantes de la salud en la EPOC se resumen en el **Cuadro 1.2**.

Cuadro 1.2 *Determinantes de la salud en EPOC mediante modelo de Dahlgren-Whitehead.*

Nivel	Determinantes de la salud	Determinantes de la salud en EPOC
1	Variación biológica	-Variantes genéticas -Etnicidad ^{5,6,13,14,19-21} -Edad ^{9-12,22} -Sexo ²²⁻²⁵ -Desarrollo pulmonar gestacional ^{26,27} -Bajo peso al nacimiento ²⁶
2	Factores individuales de estilo de vida	-Tabaquismo activo ^{2,22} -Exposiciones a alérgenos en la infancia ²⁶ -Historial familiar con factores de riesgo ambientales ^{23,28} -Hospitalizaciones en infancia ^{29,30}
3	Redes sociales y comunitarias	-Realización de actividades adoptadas por un género, determinadas por la cultura de una población ^{23,29,31,32} -Nivel de educación ²⁹ -Bajo nivel socioeconómico ³³ -Falta de capacitación médica en el diagnóstico de EPOC ³⁴ -Educación farmacológica para el uso y adherencia de tratamientos ³⁵
4	Condiciones laborales y de vivienda	-Vivienda urbana-rural ^{22,36} -Uso de combustión de biomasa ^{31,32} -Exposiciones ocupacionales por agentes orgánicos (molienda de harina o grano, procesamiento de algodón, silvicultura), inorgánicos (asbesto, aluminio, carbón o roca dura, construcción de túneles, fundición) y gases irritantes (soldadura, quema, fabricación de químicos o plásticos, transporte público y químicos de limpieza en seco) ²⁹
5	Condiciones políticas, sociales y económicas	-Pertener a un país de mediano y bajos ingresos ¹⁵ -Políticas anti-tabaco (MPOWER ³⁷ , Convenio Marco para el Control de Tabaco ³⁸) -Estrategias para riesgos ambientales ³⁹ -Guías para la calidad del aire intramuros ⁴⁰

Estos determinantes de la salud pueden ser clasificados en biológicos y no biológicos. Los factores no biológicos, de naturaleza ajena al individuo, tienen una relación estrechamente directa en el estilo de vida del individuo, hasta otros con una relación más indirecta, pero también con participación individual como las relaciones sociales entre individuos y poblaciones, condiciones laborales y de vivienda en el entorno social, y por último, esto inmerso en condiciones sociales, económicas y políticas con influencia en la enfermedad del individuo.

El estilo de vida en el que el individuo se desarrolla es el contacto más inmediato para presentar factores de riesgo a desarrollar EPOC. Desde la infancia, el individuo se encuentra vulnerable a exposiciones en el ambiente, cuya principal vía de entrada al organismo es por las vías respiratorias. Se ha reportado que las exposiciones a alérgenos e infecciones en la infancia están asociadas a hospitalizaciones por enfermedades respiratorias crónicas en la adultez. ^{26,30}

La OMS reporta que casi el 90% de las muertes por EPOC ocurren en países de medianos y bajos ingresos. ¹⁵ Además, algunos fenómenos sociales, como la equidad de género, han logrado emparejar condiciones socio-culturales entre hombres y mujeres, lo que ha aumentado la incidencia de EPOC en mujeres por factores de riesgo como tabaquismo y exposiciones ocupacionales, equiparándose con la de los hombres, usualmente en países de medianos y altos ingresos. ^{15,36} Por el contrario, en países en los cuales la equidad de género es baja (usualmente de ingresos medianos y bajos), prevalece el uso de combustión de biomasa como medio de energía doméstica, la cual ha demostrado ser factor de riesgo para EPOC, con hallazgos predominantemente en mujeres con bajo nivel socioeconómico. ^{31,32}

Entre los factores de riesgo ambiental que influyen en el desarrollo y progresión de la EPOC se encuentran principalmente tres clasificaciones de exposiciones por partículas nocivas: exposiciones ocupacionales por partículas orgánicas y/o minerales, combustión de biomasa y tabaquismo. ¹ El tabaquismo es el factor de riesgo ambiental más estudiado por los múltiples daños a la salud, debido a la acción en el organismo de los más de 5,000 componentes químicos contenidos en el cigarro. ⁴¹⁻⁴³ La manifestación de los primeros síntomas de la EPOC debido al

tabaquismo suceden tras un proceso crónico de exposición, sin embargo, al ignorarse estos síntomas podrían empeorar, incluso promover la manifestación de otras enfermedades más letales, como distintos tipos de cáncer, principalmente de pulmón, laringe, riñón, vejiga, estómago, colon, cavidad oral, esófago y leucemia.⁴⁴ A pesar de esto, se ha reportado que no todos los fumadores desarrollarán EPOC, ya que aproximadamente solo el 13% de los sujetos fumadores llegan a desarrollar la enfermedad.^{2,45,46}

Las políticas de salud relacionadas en la EPOC presentan un panorama de estrategias preventivas en los principales factores de riesgos ambientales,³⁹ aunque algunas se enfocan específicamente en tabaquismo como la iniciativa MPOWER³⁷ (por sus siglas en inglés) que cubre 6 ejes principales: monitorear el uso del tabaco, proteger a la gente del humo del tabaco, ofrecer ayuda para dejar de fumar, advertir de las consecuencias del tabaquismo, cumplir las prohibiciones de su promoción y patrocinio comercial, y aumentar los impuestos del tabaco, así como otras políticas que resaltan la importancia de prevenir la contaminación intramuros en viviendas debido a combustión por productos de biomasa y/o fósiles.⁴⁰

Sin embargo, estos determinantes no biológicos por sí solos no podrían provocar la EPOC debido a su estructura inerte, por lo cual interaccionan entre sí con factores biológicos de naturaleza propia del individuo que incluyen edad, sexo, etnicidad y genética estructural en función del tiempo.⁴⁷ Primeramente, antes de ser un organismo plenamente formado, el individuo está sujeto a la formación intrauterina, condicionada por las exposiciones que la madre tenga en este periodo gestacional. Se ha reportado que exposiciones ambientales como tabaquismo y combustión por biomasa en la madre durante la gestación, tiene efectos nocivos directamente en la salud pulmonar del feto, los cuales se reflejan en una función pulmonar alterada en la adultez.^{48,49}

La EPOC es una enfermedad que usualmente se presenta a partir de los 40 años, con historial crónico de exposiciones mencionadas anteriormente.⁹⁻¹² Sin embargo, pueden existir sujetos que la padezcan a edades más tempranas, debido a la

deficiencia genética de la proteína alfa 1 antitripsina (A1AT), incluso sin tener los factores de riesgo ambientales. El sexo biológico, es un factor importante, ya que se ha evidenciado que la EPOC secundaria a tabaquismo es más frecuente en hombres que mujeres, explicado por el patrón de consumo. ⁴⁷ Sin embargo, se ha reportado que el sexo femenino está asociado a menor función pulmonar, inicio más temprano de síntomas y mayor severidad con menor exposición al tabaquismo, en comparación con el sexo opuesto. ²⁵ Esto puede ser explicado por la anatomía pulmonar propia de las mujeres, reflejado en menor grosor y lumen de las vías respiratorias. ²⁴

Con base en lo anterior, se puede concluir que existe una red de interacción compleja entre determinantes de salud y factores de riesgo tanto ambientales como biológicos en el desarrollo de la EPOC. A pesar de ello se ha descrito que existen otros factores desconocidos involucrados en su desarrollo y progresión, por lo que se sugiere que el factor genético está ampliamente relacionado. ⁵⁰

1.1.1.1 Epidemiología de la EPOC en México

Tomando en consideración estos determinantes y factores de riesgo en el desarrollo de la EPOC, resalta el caso de México, país con particularidades en sus determinantes no biológicos y biológicos propios de su población, que han sido escasamente evaluados.

México pertenece al grupo de países de medianos ingresos, según datos del Banco Mundial, ⁵¹ más del 90% de las muertes por EPOC ocurre en estos países. ¹⁵ Sin embargo, a pesar de que se ha reportado que en estos países la incidencia de EPOC aumente debido a las principales exposiciones ambientales (tabaquismo, combustión de biomasa y exposiciones ocupacionales), en México no ocurre este patrón recíproco.

El proyecto PLATINO (Proyecto Latinoamericano de Investigación en Obstrucción Pulmonar) ⁵², un estudio con alta validez metodológica, realizado en 5 ciudades de Latinoamérica, ha colocado a la Ciudad de México con una prevalencia de 7.8%, ¹⁹ siendo la más baja de las ciudades de estudiadas (Montevideo, São Paulo,

Santiago, Caracas). Sin embargo, debido al limitado alcance que se presenta en los estudios de prevalencia para establecer causalidad, PLATINO reporta algunas variables que podrían sugerir nuevas hipótesis de investigación: Además de la baja prevalencia de EPOC en México, PLATINO reporta que de los individuos analizados, los mexicanos, presentan la menor proporción de fumadores (8%), valores de función pulmonar más elevados, así como la menor proporción de individuos por gravedad en grados GOLD. De acuerdo al grupo étnico autoreportado, los mexicanos contienen la mayor proporción de individuos mestizos (52.5%) y amerindios (indígenas) (17.4%) en el estudio.¹⁹ En cambio, la población de Montevideo, Uruguay, presenta la mayor prevalencia de EPOC (19.7%), así como la mayor proporción de fumadores (22%) y la mayor proporción de individuos por gravedad en grados GOLD. Interesantemente, es la población con mayor proporción de individuos autoreportados como europeos (89.6%).¹⁹ Estos resultados podrían sugerir que estas variables están involucradas en el desarrollo de la EPOC.

Según otros reportes, con hallazgos en la misma dirección, la prevalencia nacional de tabaquismo activo en México está entre 16-18% en la población entre 12 y 65 años, de los cuales el 6.4% son fumadores diarios, con un consumo de 7.5 cigarros por día.^{53,54} De acuerdo a estas cifras con otras poblaciones, los mexicanos son una población con un patrón de bajo consumo de tabaco, comparando con reportes de poblaciones de origen Europeo²⁰ y latinoamericanas^{19,21}, lo cual parecería ser la explicación para la baja prevalencia de EPOC, sin embargo, se ha demostrado que es poco probable que la intensidad de fumar solo explique la “protección” de los hispanos contra la EPOC.⁵⁵

1.2 La paradoja hispana

En EE. UU., un país con poblaciones biológica y socialmente distintas, interaccionan en un mismo lugar geográfico, se presenta un escenario único para evaluar fenómenos epidemiológicos. La distribución de las enfermedades es variable, y en algunos casos es dependiente de factores poblacionales particulares, como la raza o etnia a la que pertenece la población. Se ha reportado que la población

denominada “hispana” (personas hispano-hablantes, predominantemente latinoamericanas que residen en EE. UU.) ⁵⁶, presenta menor mortalidad y mayor esperanza de vida cuando se compara con el resto de las poblaciones (caucásica, afroamericana y asiática), a pesar de tener un nivel socioeconómico más bajo y menor acceso a servicios de salud ⁵⁷, fenómeno denominado “paradoja hispana”. ⁵⁸ A pesar de que revisiones sistemáticas y meta-análisis lo han confirmado ⁵⁹, este fenómeno no se ha logrado explicar en su totalidad, por lo cual se han planteado algunas propuestas que tratan de aproximarse a su posible explicación epidemiológica:

- Efecto del migrante saludable. Migrantes hispanos que logran entrar (ilegalmente) a EE. UU., son más saludables, ya que son más jóvenes y físicamente más aptos. ⁶⁰
- Efecto de migración de retorno (sesgo del salmón). Los migrantes hispanos establecidos en EE. UU., regresan a su país de origen después de obtener el suficiente trabajo, jubilación o enfermar; lo que hace que no sean contabilizadas sus defunciones y se conviertan en “estadísticamente inmortales”. ⁶¹
- Ventaja étnica del enclave (ventaja del barrio). Los hispanos habitan en un enclave (vecindarios con altas proporciones de otros hispanos con antecedentes culturales similares) con una alta estructura de soporte social, la cual beneficia a la salud de sus habitantes. ⁶²
- Errores sistemáticos de información. Sesgos en la recopilación de datos, como la información sobre el origen étnico en los certificados de defunción, dando, aparentemente, una mejor salud en hispanos. ^{63,64}

Sin embargo, en enfermedades pulmonares crónicas de las vías inferiores, este fenómeno llega a un 50% menos comparado con población general, ⁵⁷ lo que plantea que existen fenómenos biológicos particulares de esta población que podrían esclarecer el fenómeno.

Se ha argumentado que la baja prevalencia de EPOC en hispanos es debido al bajo consumo de cigarrillos. Sin embargo, se ha demostrado que después de controlar

estas variables la prevalencia de síntomas de obstrucción pulmonar continúan siendo significativos. ^{65,66}

Por otro lado, se sabe que la población hispana es el grupo étnico mayoritario en EE. UU., comprendiendo más de 57 millones de individuos (~17% de la población total) ⁶⁷. Sin embargo, el término hispano es genérico, e incluye una mezcla de poblaciones latinoamericanas con distinto origen y condiciones sociales, entre las cuales se encuentran los mexicanos con una mayor proporción (63.5%) que el resto (puertorriqueños, 9.5%; cubanos, 3.9%; salvadoreños, 3.8%; dominicanos, 3.3%; y guatemaltecos, 2.5%), ⁶⁷ lo que podría estar sobrerrepresentado a la población mexicana en el resultado de la paradoja hispánica.

Sin embargo, la paradoja hispana es un hallazgo epidemiológico, el cual cumple los requisitos para ser una “caja negra”, ⁶⁸ de la cual se han discutido explicaciones biológicas que en combinación con sociales (descritas anteriormente), podrían formular una vía más integral de efecto del fenómeno. Por un lado, la alimentación ha sido motivo de discusión. Se ha propuesto que el alto consumo de leguminosas (frijol, habas) y legumbres altas en fibra en la dieta de estas poblaciones, contienen nutrientes asociados a una preservación pulmonar a agresiones por el tabaquismo, (como la inflamación crónica), reflejadas en una mejor función pulmonar. ⁶⁹⁻⁷¹ Por otro lado, quizá el argumento más valioso en esta tesis es la ancestría amerindia, compartida entre las poblaciones hispanas, la cual tendrían un efecto directo en la conformación de la arquitectura pulmonar “protectora”. Se ha asociado con un mayor componente amerindio con menor probabilidad de padecer EPOC, menor decline la función pulmonar, así como mayores valores de VEF₁ y VEF₁/CVF ^{55,65,72-74}, sin embargo, las causas biológicas aún son desconocidas.

1.3 Variación genética humana

La variación humana, además de ser cultural, es biológica y se fundamenta por cambios en genoma entre un individuo y otro. El genoma humano tiene una longitud aproximada de más de 6,000 Mb (1 Mb = 10⁶ bases), del cual se ha estimado que contiene 20,418 genes codificantes en 23 pares de cromosomas. ⁷⁵ Las variaciones genéticas más comunes son las variantes de un sólo nucleótido (*single-nucleotide*

variant, SNV), entre los cuales se encuentra los SNPs (que tienen una frecuencia mayor al 1% en las poblaciones humanas), se ha estimado que se encuentran, en promedio, en una tasa de 1/1 Kb ⁷⁶ y se han encontrado más de 38 millones. ⁷⁷ Además de estas variaciones, se pueden encontrar inserciones, deleciones y otras variantes estructurales más grandes, como variantes en el número de copias (*copy number variants*, CNVs) a lo largo del genoma.

Debido a su gran longitud, las diferencias en el genoma humano parecen ser mínimas, por ejemplo al compararlo con la especie más cercana, el chimpancé (*Pan troglodytes*), tiene una diferencia de apenas el 4%. ⁷⁸ De igual manera, entre humanos existen diferencias “mínimas”, que se han generado a lo largo de la evolución humana, las cuales difieren principalmente en los componentes ancestrales, acentuándose en grupos étnicos socialmente más cerrados, existiendo una correlación entre estas poblaciones y el lenguaje que hablan (**Figura 1.1**). ⁷⁹ Esta variación intraespecífica en las diversas poblaciones humanas conforma la denominada estructura poblacional humana. ⁸⁰ Sobre los cambios que existen en la estructura de una población comparada con otra, se sostiene que podrían existir diferencias en rasgos de importancia biomédica. ⁸¹

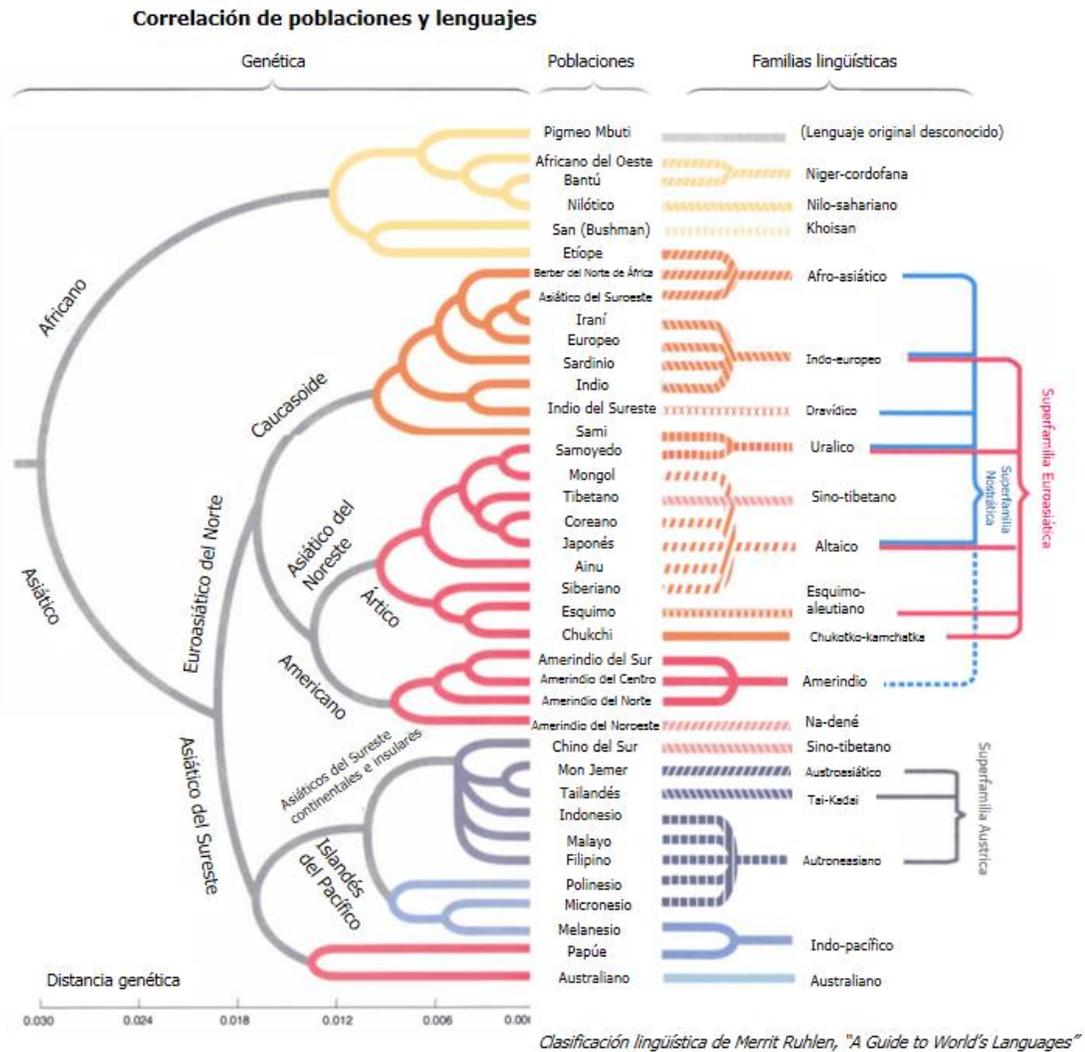


Figura 1.1 Árbol filogenético de las poblaciones humanas y su relación con lenguajes (Modificado de Cavalli-Sforza, 1991⁸²)

1.4 Ancestría genética humana

En estudios epidemiológicos, analizar poblaciones homogéneas conlleva a robustecer la validez de los resultados y conclusiones obtenidos, sin embargo, las poblaciones humanas presentan una variación intrínseca incuestionable, la cuál ha sido categorizada en grupos raciales o étnicos, términos esenciales en la determinación de la ancestría.

El término ancestría suscita a una larga discusión sobre su significado y relación con los términos "raza" y "etnia", utilizados indistintamente, sin embargo, disputar

sobre ello no es la intención de esta tesis, por lo que utilizaré definiciones concretas para estos términos.

La ancestría se define como el origen compartido entre individuos de distintos grados de parentesco ⁸³, utilizando este mismo significado entonces, la ancestría genética se puede definir como la inferencia sobre poblaciones originarias de un individuo con base en su variación genética. Diferenciando de su significado, la raza es un término de carácter social, definido como cada uno de los grupos en que se subdividen algunas especies biológicas y cuyos caracteres diferenciales se perpetúan por herencia. ⁵⁶ De igual manera, el término etnia se define como comunidad humana definida por afinidades raciales, lingüísticas, culturales, entre otras. ⁵⁶ Teniendo los términos bien definidos, se utilizará su significado a lo largo de la tesis.

El grupo racial como variable categórica autoreportada es bastante utilizado en estudios epidemiológicos, sin embargo, no refleja la variación fenotípica que posee el individuo, independientemente del conocimiento de sus antecedentes genealógicos, por lo cual se requieren mediciones menos subjetivas por medio de metodologías válidas de la ancestría. ^{84,85}

La ancestría de una población de individuos puede clasificarse en dos tipos: local y global ⁸⁶. La ancestría local se refiere a la mezcla de distintas poblaciones en un locus dado de un individuo. Por otra parte, la ancestría global representa la contribución de las poblaciones de referencia en todo el genoma de la población mestizada. Por esta razón, la ancestría genética se puede calcular en distintas regiones del genoma humano con distintos enfoques: en el DNA mitocondrial, existen distintas regiones (HV, D-Loop) que han servido para clasificar las poblaciones humanas por vía materna ⁸⁷; en el cromosoma sexual Y se han identificado haplogrupos específicos de grupos poblacionales, heredados por vía paterna. ⁸⁸ Sin embargo, a pesar de que estos dos grupos de marcadores son genéticamente más conservados (tienen menor tasa de mutación y no recombinan), sólo informan de alguna vía parental, por lo que carecen del panorama ancestral

total de un individuo. Debido a estas limitaciones, se ha optado por utilizar a los autosomas para una mejor evaluación de la ancestría genética.

En la estimación de la ancestría genética se emplean distintos marcadores, estos van desde los cada vez menos utilizados STRs (*Short tandem Repeats*), hasta los más novedosos, como los polimorfismos de un solo nucleótido (SNPs) que son utilizados como marcadores informativos de ancestría. Mediante estos marcadores se puede estimar la ancestría genética por diversos métodos, entre los cuales destacan el de asociación estructurada y el análisis de componentes principales.

El análisis de asociación estructurada utiliza una aproximación basada en estadística Bayesiana, empleando cadenas de Markov Monte Carlo, en la cual estima simultáneamente un vector multidimensional de todas las frecuencias alélicas para todas las poblaciones en todos los *loci*, vectores de poblaciones ancestrales para cada alelo para cada individuo.⁸⁹

Por otro lado, el análisis de componentes principales (PCA) se basa en análisis de varianza, simplificando la variación de datos de gran dimensión, conservando tendencias y patrones, transformándolos en variación más reducida de la dimensión de los datos, denominados componentes principales.⁹⁰ Si existiera una estructura subyacente en la población, se separaría de acuerdo a los componentes principales, confluyendo en *clusters* espacialmente homogéneos en aquellos grupos genéticamente similares.⁹¹

La ancestría genética suele ser una variable confusora que a menudo no se contempla en estudios epidemiológicos. Incluso suele ser poco medida en estudios de asociación genética debido a que se deben contemplar un conjunto de AIMs o SNPs contenidos en el microarreglo, adicionalmente de las variantes genéticas a estudiar. Según Liu et al.,⁹¹ la confusión por estratificación poblacional puede ocurrir por varias razones:

- Una variante con diferencia en frecuencias alélicas entre poblaciones, que a su vez está asociada a una enfermedad, la cual es más frecuente en una población, podría resultar en asociaciones espurias.

- Las frecuencias alélicas de una variante causal de la enfermedad es sustancialmente distinta entre poblaciones
- Las proporciones poblacionales no están pareadas en casos y controles
- La estructura poblacional no es tomada en cuenta para el modelo estadístico de asociación genética

La estratificación poblacional puede cuantificarse mediante el factor de inflación genómico (λ), el cual, según la hipótesis nula, debe ser cercano a 1, es decir, que no existe estratificación en la población. ⁹²

1.5 Mexicano mestizo

La población mexicana presenta una estructura genética poblacional particular y distinta a otras poblaciones humanas. La principal característica es su grado de mestizaje, resultado de paulatinos encuentros históricos, entre individuos indígenas nativos del continente Americano (amerindios), con individuos españoles peninsulares (europeos) y en menor proporción con africanos y asiáticos, cuyas interacciones constituyen al denominado mexicano mestizo, conservando características biológicas y culturales de sus poblaciones ancestrales. ⁹³

A pesar de considerarse una población homogénea, el mexicano mestizo presenta variabilidad genética de acuerdo al origen geográfico de sus habitantes (**Tabla 1.1**). Se han estimado mediante técnicas moleculares los componentes ancestrales del mexicano mestizo en autosomas, siendo el amerindio el mayor (35-85%), seguido por el europeo (13-62%) y en menor proporción el africano (1-7%), adicionalmente se han reportado vestigios de ancestría compartida con asiáticos del este ($\leq 1\%$). ^{81,94-96} Sin embargo, es evidente que la diversidad genética depende de la región geográfica del País a manera de gradiente. Mientras que en los estados del norte (Durango, Sonora, y Tamaulipas) el componente europeo es mayor (43-62%) y el amerindio (35-52%) es menor; en los de occidente (Guanajuato, Jalisco, Zacatecas) el europeo (37-46%) y amerindio (50-59%) son menos heterogéneos. En cambio, en los estados del centro y sur (Campeche, Ciudad de México, Guerrero, Oaxaca, Veracruz y Yucatán) las proporciones ocurren de manera inversa para el

componente europeo (13-39%) y amerindio (56-85%), respectivamente. ^{81,94-96} Por otro lado, el componente africano es el menor, pero no sigue una frecuencia gradual entre los estados, así como el asiático que fue reportado únicamente en un estudio.

94

De acuerdo a la proporción de estos componentes ancestrales, se ha propuesto que el componente genético amerindio (el de mayor proporción) en mexicanos mestizos, podría ser un factor protector en el desarrollo y progresión de la EPOC ^{55,97,98}, reflejado en la baja prevalencia y mortalidad causada por la EPOC respecto a otras poblaciones en el mundo. ^{19,21,73,99}

Tabla 1.1 Reportes de los componentes ancestrales en el mexicano mestizo en distintas regiones geográficas.

Estudio	n	Muestra	Lugar de residencia	SNP anc	Ancestría (%)		
					AMI	EUR	AFR
Silva-Zolezzi et al., 2008 ⁹⁴	300	Universitarios	6 estados MX	1,814	55	42	2
			Guanajuato, MX		58	40	1
			Guerrero, MX		66	28	4
			Sonora, MX		36	62	1
			Veracruz, MX		61	36	2
			Yucatán, MX		59	39	1
			Zacatecas, MX		51	46	2
Bryc et al., 2010 ⁹⁶	112	Controles hospitalarios	Jalisco, MX	73,901	50	44	6
Johnson et al., 2011 ¹⁰⁰	984	Padres de niños con asma	Ciudad de México, MX	482,906	65	31	3
Ruiz-Linares et al., 2014 ⁹⁵	905	Universitarios	Ciudad de México, MX	5,000	56	39	5
Moreno-Estrada et al., 2014 ^{77,81}	49	Padres de niños sanos	Los Ángeles, EEUA	71,581	48	48	4
Moreno-Estrada et al., 2014 ⁸¹	49		Sonora, MX	71,581	35	61	3
	19		Durango, MX		44	50	5
	17		Tamaulipas, MX		52	43	5
	50		Zacatecas, MX		51	44	5
	50		Jalisco, MX		51	45	4
	48		Guanajuato, MX		59	37	4
	50		Veracruz, MX		62	33	4
	50		Guerrero, MX		68	25	7
	18		Oaxaca, MX		85	13	2
	20		Campeche, MX		80	18	2
	49		Yucatán, MX		60	37	3

SNP anc: SNP utilizados para determinar ancestría; MX: México*; AMI: Amerindia, EUR: Europea, AFR: Africana

1.6 Genómica de la EPOC

El componente genético ha sido ampliamente discutido como componente etiológico en distintas enfermedades, el cual se clasifica en modelos de herencia:

- Enfermedades monogénicas o mendelianas, en donde existe un defecto causado por una mutación de un gen específico, el cual tiene un efecto muy marcado en originar el rasgo o enfermedad de interés (alta penetrancia), con poco aporte del ambiente. Las variantes que causan estas enfermedades tienen una baja prevalencia (<1%) y se heredan en grupos poblacionales muy cerrados o en familias. Estas variantes se transmiten mediante patrones de herencia dominantes, recesivos, codominantes o ligados al sexo. ¹⁰¹
- Enfermedades comunes o complejas, en donde no existe un gen causal, sino más bien una predisposición genética, determinada por un conjunto de variantes genéticas en genes involucrados en las distintas vías de la patología interaccionando entre sí y con el ambiente. Las variantes genéticas o polimorfismos aportan generalmente efectos discretos en promover el rasgo o enfermedad de interés con un aporte importante del ambiente, es por eso que el riesgo de poseer determinada variante no determina a desarrollar la enfermedad. Este modelo se adapta más a enfermedades comunes en la población general. No existe un patrón de herencia específico para la transmisión de estas variantes. ¹⁰²

De acuerdo a estos modelos de clasificación, la EPOC puede ser monogénica o compleja, sin embargo, se tendrá una consideración más adecuada como común y compleja debido a las interacciones ambientales mencionadas anteriormente. El primer factor genético y el más ampliamente estudiado en la EPOC, es la deficiencia de la proteína A1AT, descubierta en 1964 ¹⁰³, considerado un rasgo monogénico, causado por mutaciones en el gen *SERPINA1*, principalmente los alelos alterados *Pi*S* y *Pi*Z*, entre otras menos frecuentes. Estos alelos y otros de baja frecuencia, alteran a la proteína y causan la enfermedad a una menor edad que por exposición ambiental, incluso sin exponerse a los factores de riesgo. Sin embargo, las mutaciones en *SERPINA1* se distribuyen en poblaciones particularmente

caucásicas europeas, siendo casi inexistentes en otras poblaciones. Sin embargo, los casos de AAT explican apenas el 1% de los casos de EPOC, ¹ por lo que se ha propuesto que existen otras regiones en el genoma involucradas en la susceptibilidad al desarrollo de la misma.

Durante el siglo XX, algunas metodologías lograron identificar *loci* de susceptibilidad distintas a *SERPINA1*. Primeramente, con estudios en gemelos lograron identificar factores genéticos heredables en la EPOC. ¹⁰⁴ Posteriormente, estudios de ligamiento en familias y tríos fueron capaces de identificar *loci* específicos implicados en susceptibilidad al desarrollo de la EPOC. ^{105,106} Sin embargo, fue hasta finales del siglo pasado y principios del siglo XXI, cuando mediante tecnologías más eficientes de genotipificación, se lograron identificar variantes genéticas asociadas a EPOC por medio de estudios de gen candidato, fundamentados por una hipótesis *a priori* en productos proteicos involucrados en la enfermedad, los cuales son codificados por genes de interés o “candidatos”, sin embargo, esta estrategia está limitada al conocimiento previo en la literatura sobre la enfermedad y sus vías patológicas.

Debido a ello, se ha optado por la implementación de nuevas herramientas que permitan una cobertura más amplia del genoma, enfocado en variantes comunes que difieren en su frecuencia del 5-95% en las poblaciones. ¹⁰⁷ Entre estas nuevas tecnologías destacan los GWAS, los cuales se fundamentan en genotipificar una gran cantidad de variantes genéticas (>500,000) distribuidos por todo el genoma de individuos con algún rasgo y compararlos con individuos sin el rasgo, logrando descubrir hallazgos genéticos en distintas vías involucrados en el rasgo o enfermedad.

Desde 2009 hasta la fecha actual, se han identificado por significancia de genoma completo a variantes genéticas en alrededor de 360 genes en todos los autosomas en 44 GWAS y meta-análisis (GWMAS) en EPOC directamente, o indirectamente con tabaquismo y función pulmonar. Se presume que la mayoría de estas variantes pudieran tener algún efecto en las vías patológicas conocidas como respuesta al estrés oxidativo, arquitectura pulmonar, desequilibrio proteasa-antiproteasa,

hipersecreción de moco, muerte celular, inflamación crónica, respuesta a la vía aérea, función pulmonar, oxigenación y vías de neurotransmisión cerebral, sin embargo, otras permanecen con efectos que no se han logrado determinar por ensayos biológicos debido a su localización en el gen, pero podrían contribuir en la expresión y/o regulación de otros genes.

Las variantes genéticas identificadas presentan distintas frecuencias alélicas según la población estudiada, por ello una de las características del diseño de los GWAS se limita a realizarse en poblaciones sin estratificación poblacional. Por lo tanto, al enfocarse en los hallazgos de otras poblaciones restringe la posibilidad de encontrar otras regiones en el genoma no antes contempladas y que estén involucradas en la enfermedad en esa población. Ya que se espera que las variantes comunes intrapoblacionales sean las que se asocian al rasgo de interés (EPOC), el cual tiene un comportamiento desigual en la población.

1.6.1 GWAS en EPOC

De los reportes realizados a la fecha, se han evaluado distintas poblaciones, denominadas poblaciones de descubrimiento, en las que se han realizado importantes hallazgos, entre los cuales 26 han sido en poblaciones de ancestría caucásica, 12 en africana, 5 en asiática y 4 en hispana, esto sin contar los hallazgos en diversas poblaciones de replicación. Adicionalmente, se han publicado más de 10 meta-análisis, confirmando la asociación de más de 50 genes.

Entre los hallazgos reportados, los *loci* que han mostrado al menos cinco replicaciones por distintos reportes en diversas poblaciones (**Tabla 1.2**) han sido en *HHIP* en europeos (rs13147758, rs1828591, rs13118928, rs12504628, rs1980057, rs1032295, rs1032296, rs13141641, rs138641402, rs13118083)^{108–122}, afroamericanos (rs13141641)¹¹⁷ e hispanos (rs1980057).¹²³ *FAM13A* en europeos (rs1903003, rs2869967, rs6830970, rs1964516, rs7671167, rs4416442, rs2045517, rs13110699)^{110–112,115,117,118,120,122,124,125}, asiáticos (rs2609255, rs2609260, rs2609261, rs2609262, rs2609264, rs2609265, rs1246642, rs2464528)¹²⁶, afroamericanos (rs4416442).¹¹⁷ e hispanos (rs2869967).¹²³ *CHRNA3/5* en europeos (rs1051730, rs12914385, rs114205691, rs56077333, rs4416442,

rs9788721, rs17486278, rs7180002, rs16969968, rs951266, rs138544659)^{108,110,117-121,127-131} y afroamericanos (rs56077333, rs12914385, rs17486278).^{119,120,124} Los *loci* cercanos *GSTCD-INTS12-NPNT* en europeos (rs11097901, rs11728716, rs10516526, rs17331332, rs17036341, rs34712979, rs2047409, rs11735213).^{109,115,120,122,124,129,132,133} e hispanos (rs11727189, rs10516526).¹²³ *HTR4* en europeos (rs7733088, rs4597955, rs3995090, rs6889822, rs7735184, rs11168048)^{109,120,122,124,127,129,132,134} e hispanos (rs2571445, rs3995090).¹²³ *TGFB2* en europeos (rs993925, rs1690789, rs4846480)^{110,117,119,120,122,124,131,132}, afroamericanos (rs4846480)¹¹⁷ e hispanos (rs993925).¹²³ *AGER/PPT2* en europeos (rs2070600, rs10947233, rs9391855)^{109,112,115,120,128,135}, asiáticos (rs10947233).^{126,135} e hispanos (rs2070600). *RIN3* en europeos (rs754388, rs117068593, rs10498635)^{110,112,118,124,129,136} y afroamericanos (rs754388).¹²⁹ *ADGRG6* en europeos (rs3817928, rs7776375, rs6937121, rs11155242, rs148274477)^{115,120,122,124,136}, africanos (rs3817928)¹²⁰ e hispanos rs262113, rs3817928).¹²³ *ADAM19* en europeos (rs2277027, rs1422795, rs11744671, rs11134779, rs1422795)^{115,120,122,124,127}, afroamericanos (rs11134779).¹²⁰ e hispanos (rs2277027).¹²³ *MMP12* en caucásicos (rs17368659, rs17368582, rs626750, rs72981684, rs615098).^{112,117,119,120,124} *THSD4* en europeos (rs12899618, rs7167951, rs12591467, rs10851839)^{109,111,120,122,137} e hispanos (rs12899618). *ARMC2* en europeos (rs2798641, rs2806356)^{120,122,124,131,132} e hispanos (rs2798641).¹²³ *PTCH1* en europeos (rs16909898, rs16909859)^{115,120,124,132,133} e hispanos (rs16909898).¹²³ *MECOM/EVI1* en europeos (rs56341938, rs1344555).^{111,120,122,124,132} e hispanos (rs1344555).¹²³ *HDAC4* en europeos (rs12477314)^{112,120,122,124,132} e hispanos (rs12477314).¹²³ *RARB* en europeos (rs1529672)^{112,120,122,124,132} e hispanos (rs1529672).¹²³ *MMP15* (rs2304488, rs12447804, rs150232756)^{114,124,132,136} e hispanos (rs12447804).¹²³ *KCNE2* en europeos (rs9978142)^{120,122,124,132} e hispanos (rs28593428).¹²³ *CFDP1* en europeos (rs2865531, rs7186831, rs8051407)^{120,124,132}, africanos (rs2865531)¹²⁰ e hispanos (rs2865531).¹²³ Estos hallazgos se han obtenido por plataformas de las marcas Affymetrix (Genome-Wide Human SNP Array 5.0-6.0, UK BiLEVE array, GeneChip Human Mapping 500K Array Set) e Illumina (Human Omni Express

Exome BeadChip, Human660W Quad v.1_A BeadChip, Human610-Quad, Omni 2.5M array, Human Omni Express chip y HumanHap550 BeadChip v3), los cuales tienen una cobertura de variantes enfocada en poblaciones europeas.

Estas poblaciones analizadas corresponden a estudios con diseños y enfoques clínicos distintos, resumidos cronológicamente en la **Tabla 1.3**. Estos estudios van desde aquellos que desde un inicio tuvieron el objetivo de evaluar parámetros respiratorios en pacientes con enfermedad pulmonar (Norway GenKOLS ^{138,139}; The Lung Health Study, LHS ¹⁴⁰; The Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points Study, ECLIPSE ¹⁴¹; The National Emphysema Treatment Trial, NETT ¹⁴²; Subpopulations and Intermediate Outcomes in COPD Study, SPIROMICS ¹⁴³; Genetic Epidemiology of COPD, COPDGene ¹⁴⁴; Boston Early-Onset COPD Study, BEOCOPD ²³; Transcontinental COPD Genetics Study, TCGS ¹⁴⁵; Costa Rica ⁷³, Netherlands-Leuven Longkanker Screenings Onderzoek, NELSON ¹⁴⁶ y Treatment Of Emphysema With A Selective Retinoid Agonist Study, TESRA ¹⁴⁷), así como de aquellos en los que se incluyó población general con otros objetivos, pero adicionalmente evaluaron variables respiratorias (Tercera generación del Framingham Heart Study, FHS-3 ¹⁴⁸; UK Biobank ¹⁴⁹; The Multi-Ethnic Study of Atherosclerosis/SNP Health Association Resource, MESA/SHARe ¹⁵⁰; Healthy-Twin Study, HTS ¹⁵¹; The Korean Association Resource Project, KARE; The Normative Aging Study, NAS ¹⁵²; Long Life Family Study, LLFS ¹⁵³; The Hispanic Community Health Study/Study of Latinos, HCHS/SOL ¹⁵⁴). Dentro de estos últimos estudios, surgieron otras subpoblaciones que adecuaron variables pulmonares y les dieron seguimiento (UK BiLEVE ¹²⁹, MESA Lung/ SHARe ¹³⁵, KARE3 ¹²⁶).

De los 17 estudios, 9 han sido en cohortes, seguido de 4 que tuvieron un inicio de ensayo clínico aleatorizado, pero posteriormente se siguieron como cohorte; otros 4 estudios han sido de tipo casos y controles.

A estos proyectos se suman consorcios que incorporan distintos estudios a manera de meta-análisis que han logrado concentrar grandes cantidades de sujetos, entre ellos se encuentra el International COPD Genetics Consortium (ICGC) ¹⁵⁵, SpiroMeta Consortium ¹³⁶ y Cohorts for Heart and Aging Research in Genomic

Epidemiology (CHARGE) ¹⁵⁶, descritos en la **Tabla 1.4**, los cuales han validado asociaciones descubiertas previamente en distintas poblaciones individualmente.

Además, como parte del análisis sucesivo de los GWAS, los hallazgos obtenidos en las poblaciones de descubrimiento han sido replicados en otras decenas de estudios, validando que su asociación en el rasgo o enfermedad no haya sido por un sesgo de selección debido a un efecto de la población analizada, determinando variantes causales en las poblaciones.

Tabla 1.2 Loci con mayor número de variantes asociadas a EPOC y función pulmonar en distintas poblaciones.

Locus/loci	Poblaciones			
	EUR	AFA	ASI	HIS
<i>HHIP</i>	15 (10)	1 (1)		1 (1)
<i>CHRNA3/5</i>	12 (11)	3 (3)		
<i>FAM13A</i>	10 (8)	1 (1)	1 (8)	1 (1)
<i>GSTCD-INTS12-NPNT</i>	8 (8)			1 (2)
<i>HTR4</i>	8 (6)			1 (2)
<i>TGFB2</i>	8 (3)	1 (1)		1 (1)
<i>AGER/PPT2</i>	6 (3)		2 (1)	1 (1)
<i>RIN3</i>	6 (3)	1 (1)		
<i>ADGRG6 (GPR126)</i>	5 (6)	2 (1)		1 (2)
<i>ADAM19</i>	5 (5)	1 (1)		1 (1)
<i>MMP12</i>	5 (5)			
<i>THSD4</i>	5 (4)			1 (1)
<i>ARMC2</i>	5 (2)			1 (1)
<i>PTCH1</i>	5 (2)			1 (1)
<i>MECOM/EVI1</i>	5 (2)			
<i>HDAC4</i>	5 (1)			1 (1)
<i>RARB</i>	5 (1)			1 (1)
<i>MMP15</i>	4 (3)			1 (1)
<i>KCNE2</i>	4 (1)			1 (1)
<i>CFDP1</i>	3 (3)	1 (1)		1 (1)

Valores presentados en absolutos, número de reportes del *locus/loci* (variantes distintas en total), EUR: Europea, AFA: Afroamericana, ASI: Asiática, HIS: Hispana

Tabla 1.3 Estudios con estrategia de GWAS en EPOC y función pulmonar.

Estudio ^{Referencia}	Tipo de estudio	n [§]	Población(es) incluida(s) (País)	Edad	GWAS con EPOC-FP*
GenKOLS ^{138,139}	CaCo	1,929	EUR (Noruega)	40-80	11 ^{108,110,116,117,121,124,125,131,157-159}
FHS-3 ¹⁴⁸	Cohorte	4,095	EUR (EE. UU.)	19->60	2 ^{114,137}
ECLIPSE ¹⁴¹	Cohorte	1,719	(12)	40-75	11 ^{110,116-118,121,124,125,157-160}
NETT ¹⁴²	ECA/ Cohorte	1,218	EUR, AFA (EE. UU.)	ND	7 ^{116,117,124,125,131,158,159}
NAS ¹⁵²	Cohorte	2,032	ND (EE. UU.)	40->60	4 ^{116,117,125,131}
LHS ¹⁴⁰	ECA/ Cohorte	5,887	EUR (EE. UU., Canadá)	35-60	3 ^{130,161,162}
COPDGene ¹⁴⁴	Cohorte	~10,000	EUR, AFA (EE. UU.)	40-80	17 ^{110,116-121,124,125,128,131,157,158,160,163}
MESA ¹⁵⁰	Cohorte	8,409	EUR, AFA, ASI, HIS (EE. UU.)	45-84	
MESA Lung/ SHARE ¹³⁵	Cohorte	7,914	EUR, AFA, ASI, HIS (EE. UU.)	45-84	2 ^{135,164}
KARE ¹⁶⁵	Cohorte	10,038	ASI (Corea del Sur)	43-74	1 ¹⁶⁶
KARE3 ¹²⁶	Cohorte	6,223	ASI (Corea del Sur)	43-74	1 ¹²⁶
UK Biobank ¹⁴⁹	Cohorte	502,682	EUR (Reino Unido)	40-69	
UK BiLEVE ¹²⁹	CaCo	50,008	EUR (Reino Unido)	ND	1 ¹²⁹
Costa Rica ⁷³	CaCo	578	HIS (Costa Rica)	21->70	1 ¹⁶⁴
LLFS ¹⁵³	Cohorte familiar	4,559	EUR (EE. UU.-Dinamarca)	31-110	1 ¹⁶⁷
TCGS ¹⁴⁵	CaCo	979	EUR (Polonia), ASI (Corea del Sur)	40-80	2 ^{122,128}
BEOCOPOD ¹⁶⁸	Cohorte familiar	949	EUR (EE. UU.)	30-70	2 ^{122,128}
SPIROMICS ¹⁴³	ECA/ Cohorte	2,981	EUR (EE. UU.)	40-80	1 ¹¹²
HTS ¹⁵¹	Cohorte	2,720	ASI (Corea del Sur)	17-81	1 ¹²⁶
NELSON ¹⁴⁶	ECA/ Cohorte	7,557	EUR (Países Bajos-Bélgica)	>70	1 ¹⁶⁹
HCHS/SOL ¹⁵⁴	Cohorte	15,825	HIS (EE. UU.)	18-74	1 ¹²³
TESRA ¹⁴⁷	ECA/ Cohorte	492	EUR	ND	1 ¹³¹

EUR: Europea, AFR: Afroamericana, ASI: Asiática, HIS: Hispana, §: al inicio del estudio, *Incluyendo meta-análisis, ECA: Ensayo clínico aleatorizado, CaCo: Casos y controles, FP: Función pulmonar

Tabla 1.4 Meta-análisis que han utilizado la técnica GWAS con variables respiratorias.

Estudio ^{Referencia}	n*	Estudios incluidos	Poblaciones incluidas	GWAS con EPOC-FP *
ICGC ^{155,170}	63,192	26	EUR, AFA, ASI	2 ^{122,128}
SpiroMeta ¹³⁶	38,199	17	EUR	7 ^{109,127,132,133,136,171,172}
CHARGE ¹⁷³	90,715	22	EUR, AFA, ASI, HIS	6 ^{115,127,132,133,171,173}

*Analizados por GWAS

2 Antecedentes

Los GWAS que han sido realizados en individuos de origen hispano son escasos, además de incluir poblaciones heterogéneas entre ellos. El primer reporte sobre hallazgos de un GWAS con EPOC en una población hispana fue en el estudio MESA¹³⁵, el cual en su diseño se incluye a individuos de origen mexicano (51%), dominicano (15%), puertorriqueño (14%), sudamericano (8%), centroamericano (7%), cubano (4%), entre otros (1%), en un total de 1,374 sujetos no relacionados.¹⁷⁴ Posteriormente, se realizó un meta-análisis del estudio MESA, junto con el estudio Costa Rica, el cual evalúa a 527 individuos no relacionados oriundos de San José, Costa Rica.¹⁶⁴ Por último el estudio HCHS/SOL, el cual incluye participantes de origen mexicano (41%), puertorriqueño (17%), cubano (15%), centroamericano (11%), dominicano (9%) y sudamericano (7%), en un total de 15,825 sujetos.¹⁷⁵

Estos estudios han logrado identificar hallazgos únicos en poblaciones hispanas, ya que se analizan como un grupo único (**Tabla 2.1**). Estos hallazgos se han asociado a un menor riesgo de padecimientos pulmonares. El rs10411619 en el gen *MAN2B1* se ha asociado a una menor relación del lóbulo pulmonar superior-inferior por tomografía en pacientes con EPOC¹⁷³; el rs858249, que se encuentra entre los genes *NUPL2* y *KLHL7*, así como el rs286499 en *DLG2* se encontraron asociados a un menor riesgo de EPOC¹⁶⁴; los rs145174011, rs4791658 y rs6746679 en los genes *HAL*, *ZSWIM7* y *PABPC1P2*, respectivamente, se han encontrado asociados a variables de función pulmonar (VEF_1/CVF , VEF_1 y CVF).^{123,173} En meta-análisis en donde se han incluido poblaciones hispanas, se han obtenido algunos hallazgos significativos, los cuales, interesantemente, han sido, en su mayoría, a betas negativas, dirección a menor riesgo del rasgo.

Tabla 2.1 Variantes genéticas asociadas con variables respiratorias en población hispana por medio de GWAS.

SNP	Alelo asociado	Gen más cercano	Región cromosómica	Variable asociada	β	Frecuencia del alelo asociado	Estudio(s) Referencia
rs10411619	T	<i>MAN2B1</i>	19p13.13	RLPSI	-0.177	0.96	MESALS ¹³⁵
rs7957346	C	<i>SNRPF</i>	12q23.1	PE	-0.055	0.5	MESALS ¹³⁵
rs10947233	T	<i>AGER/PPT2</i>	6p21.32	PE	-0.150	0.01	MESALS ¹³⁵
rs858249 [§]	A	<i>NUPL2/ KLHL7</i>	7p15.3	EPOC	-1.2/-1.1	0.09	CR + MESALS ¹⁶⁴
rs286499 [§]	G	<i>DLG2</i>	11q14.1	EPOC	-0.7/-1.2	0.14	CR + MESALS ¹⁶⁴
rs145174011	T	<i>HAL</i>	12q23.1	VEF ₁ /CVF	4.85	0.99	HCHS/SOL ¹²³
rs4791658	G	<i>ZSWIM7</i>	17p12	VEF ₁	33.35	0.54	HCHS/SOL ¹²³
rs6746679	A	<i>PABPC1P2</i>	2q22.3	CVF	-37.36	0.56	HCHS/SOL + CHARGE ¹⁷³

[§]: Hallazgo de meta-análisis, MESALS: MESA Lung/ SHARe, CR: Costa Rica, RLPSI: Relación lóbulo pulmonar superior-inferior, PE: Porcentaje de enfisema

Sin embargo, como se revisó anteriormente, las distintas poblaciones que conforman a este grupo son genéticamente heterogéneas, debido a la distinta mezcla racial que adquirieron históricamente.^{174,176} Las proporciones genéticas de estas poblaciones se distribuyen alrededor de tres componentes ancestrales: africano, europeo y amerindio, descritos en la **Tabla 2.2**. En estos estudios, el componente africano se encuentra más representado entre sujetos de República Dominicana (42-43%) y Puerto Rico (16-25%), seguidos de Cuba (7-21%) y Centroamérica (10-18%), finalmente Sudamérica (4-11%), México (4%) y Costa Rica (3%) presentan las menores proporciones. El componente europeo es mayor en sujetos de Cuba (73-84%) y Puerto Rico (62-72%), seguido de individuos de Costa Rica (62%), República Dominicana (50%), y Sudamérica (47-50%), siendo México (47%) y Centroamérica (37-45%) los de menor proporción. Finalmente, el componente amerindio se encuentra en mayor proporción en sujetos de México (48-49%), seguido por Centroamérica (45%), Sudamérica (40-47%) y Costa Rica (35%), los que presentan menor porcentaje de este componente son Puerto Rico (13%), República Dominicana (6-8%) y Cuba (5-6%).^{73,174,176}

Tabla 2.2 Proporción de los componentes ancestrales en cohortes de población hispana reportados en GWAS con EPOC.

Población	MESA/SHARE ¹⁷⁴			HCHS/SOL ^{*176}			Costa Rica ^{§ 73,164}		
	AMI	EUR	AFR	AMI	EUR	AFR	AMI	EUR	AFR
Mexicana	48	47	4	49	47	4			
Dominicana	6	50	43	7.5	50	42.5			
Puertorriqueña	13	62	25	12.5	71.5	16			
Cubana	6	73	21	5	87.5	7.5			
Centroamericana	45	37	18	45	45	10			
Sudamericana	40	50	11	47	49	4			
Costarricense							35	62	3

Valores presentados en porcentajes. *: Valores en mediana de autosomas, §: Valores en medias de autosomas, AMI: Amerindia, CAU: Caucásica, AFR: Africana

De acuerdo con esto, se evidencia que el grupo denominado hispano no es genéticamente homogéneo, y de agruparlo, no reflejaría los componentes ancestrales de cada grupo poblacional, por lo cual se tendrían que analizar individualmente, como el estudio Costa Rica. Sin embargo, en más del 40% en estos dos estudios (MESA/SHARE y HCHS/SOL) se incluyen mexicanos y sus proporciones de componente ancestral amerindio son las más altas (48-49%), por

lo cual el análisis en conjunto de las demás poblaciones podría desestimar los resultados.

Estos estudios han incluido mayoritariamente individuos latinos que residen en EE. UU., los cuales no representan a sus respectivas poblaciones.⁸⁵ Incluso se ha evidenciado que en el caso del mexicano, los que residen en EE. UU. presentan una estructura genética distinta al compararse sujetos de su país de origen.¹⁰⁰

A pesar de no existir GWAS en EPOC exclusivamente en sujetos mexicanos de México, se han obtenido hallazgos en estudios de gen candidato con base en hallazgos en otras poblaciones. Se han encontrado variantes en genes involucrados en vías relacionadas con el comportamiento de fumar (*HTR2A*, *CYP2A6*, *CHRN5*, *NRXN1*, *DRD4*)¹⁷⁷⁻¹⁷⁹, inflamación (*TNF*, *IL6*, *CSF2*)¹⁸⁰⁻¹⁸² y arquitectura pulmonar (*ADAM19*, *SERPINA1*)^{183,184}, cuyas frecuencias alélicas y genotípicas difieren a las reportadas en otras poblaciones. Sin embargo, estos hallazgos solamente reflejan el riesgo a desarrollar la EPOC en regiones conocidas del genoma, por lo que podrían existir otras regiones de interés para la población mexicana mestiza.

3 Planteamiento del problema

El fenómeno epidemiológico de la paradoja hispana no ha logrado explicar porque los individuos de origen latino presentan menores tasas de mortalidad en enfermedades respiratorias crónicas de vías inferiores, como la EPOC. Por lo que se argumenta que existen factores genéticos en las poblaciones latinas en común, las cuales generan ventaja al no desarrollo o desarrollo más tardío de la enfermedad. En la literatura, se utiliza el término hispano o latino generalizando a todas las poblaciones de Latinoamérica, las cuales han demostrado tener componentes ancestrales muy distintos entre cada población, en el caso particular de México, se ha evidenciado que la mayoría de sus habitantes, de origen mestizo, poseen de las más altas proporciones de componente ancestral amerindio. Es por ello que en el mexicano mestizo, podrían existir variantes genéticas de menor riesgo a padecer EPOC secundaria a tabaquismo.

4 Justificación

La estrategia asociación de genoma completo, se basa en analizar cientos de miles de variantes genéticas distribuidas a lo largo de todo el genoma, lo cual permitirá identificar regiones previamente no reportadas en la EPOC, particulares de la población mexicana. Los reportes sobre variantes genéticas de riesgo para el desarrollo de EPOC son predominantemente en poblaciones europeas, genéticamente muy distintas al mexicano mestizo, además de que muy pocos estudios contemplan el papel del componente ancestral amerindio en la enfermedad. Adicionalmente, las asociaciones reportadas en la literatura, se basan en tecnologías de microarreglos que identifican variantes con coberturas en poblaciones europeos, descartando las variantes que puedan encontrarse en poblaciones no europeas. Es por ello que proponemos que los hallazgos obtenidos en este estudio, permitirán generar y comprobar hipótesis sobre el involucramiento de factores genéticos de componentes ancestrales en la EPOC en mexicanos mestizos.

5 Hipótesis

Existen variantes genéticas de contribución ancestral amerindia que se asocian a riesgo de EPOC secundaria a tabaquismo en fumadores mexicanos mestizos.

6 Objetivos

6.1 Objetivo general

Identificar variantes genéticas de contribución ancestral amerindia asociadas con EPOC secundaria a tabaquismo en pacientes mexicanos mestizos.

6.2 Objetivos específicos

-Identificar variantes genéticas asociadas con EPOC secundaria a tabaquismo en pacientes mexicanos mestizos.

-Evaluar la contribución ancestral de las variantes genéticas asociadas en EPOC secundaria a tabaquismo.

7 Materiales y métodos

7.1 Consideraciones éticas

El presente estudio fue realizado mediante el protocolo de investigación titulado *"Participación de la contribución ancestral amerindia en la adaptación a hipoxemia por grandes altitudes y sus implicaciones en las enfermedades respiratorias obstructivas en mexicanos mestizos e indígenas"*, sometido y aprobado por los comités de Investigación, Ética en Investigación y de Bioseguridad del INER, por medio del documento B14-17 (**Anexo I-III**). Los participantes firmaron un consentimiento informado antes de realizarse el cuestionario y la toma de muestra, en el cual se describen los objetivos del estudio y los posibles riesgos de la toma de muestra.

7.2 Diseño de estudio

Se plantea un estudio de casos y controles, los cuales se definen en el **Cuadro 7.1**.

Cuadro 7.1 Definición de población de estudio.

Condición	Definición
Caso	Pacientes con diagnóstico clínico y espirométrico de EPOC secundaria a tabaquismo.
Control	Individuos fumadores con diagnóstico espirométrico y clínico descartado de EPOC.

7.3 Población de estudio

Los participantes fueron reclutados del Departamento de Investigación en Tabaquismo y EPOC del INER por medio de consentimiento informado (**Anexo IV-V**), así como de las campañas contra tabaquismo realizadas por el mismo departamento. Los datos de función pulmonar (espirometría) de los pacientes fueron recolectados a partir del expediente que correspondía al cuestionario previamente realizado entre 2009 al 2018. Los criterios de inclusión y exclusión de la muestra empleados se describen en el **Cuadro 7.2** y **7.3**, respectivamente.

El INER es un centro de referencia perteneciente al tercer nivel de atención ¹⁸⁵, en el cual se pueden atender pacientes de todo el País. Sin embargo, se establece en

la Ciudad de México, en el cual se concentra en la mayoría de sujetos atendidos, principalmente del centro de País.

Cuadro 7.2 Criterios de inclusión.

Variable	Criterio
Edad	40-80 años
Sexo	Indistinto
Etnicidad	Mexicanos mestizos (auto-referencia)
Tabaquismo	-Historial de consumo diario de cigarro -Edad de inicio al tabaquismo en la adolescencia
Diagnóstico	-Valoración médica -Espirometría basal

Cuadro 7.3 Criterios de exclusión.

Variable	Criterio
Otros productos de tabaco	Consumo de puro, pipa o hookkah
Drogas	Consumo de drogas
Otras exposiciones ambientales	-Humo por combustión de biomasa ≥ 100 horas-año de exposición -Químicos industriales -Polvos
Otras enfermedades respiratorias	-Asma por respuesta positiva a broncodilatador espirometría según las guías GINA ¹⁸⁶ -Fibrosis pulmonar -Tuberculosis

7.4 Operacionalización y medición de las variables

Las variables consideradas y su operacionalización se resumen en el **Cuadro 7.4**.

Cuadro 7.4. Operacionalización de las variables.

Variable	Definición conceptual	Definición operacional	Indicador	Tipo de variable
EPOC	Persistencia de síntomas respiratorios y limitación al flujo de aire en las vías aéreas, debido a anomalías alveolares ¹	-Valor menor a 0.7 en la relación VEF ₁ y CVF por espirometría post broncodilatador ¹ -Valoración clínica por el médico	-VEF ₁ (L) -CVF (L)	Dicotómica (EPOC, no EPOC)
Ancestría genética	Variación genética compartida particulares entre individuos de una población	Componentes poblacionales que conforman una nueva población.	Marcadores	-Amerindia -Europea -Africana (Categorica)
Variante genética	Alteración en la secuencia de nucleótidos de ADN con una frecuencia ≥ 1 % en la población	Variante genética en una secuencia de ADN	SNP	Dicotómica (Alelo asociado, alelo no asociado)
Tabaquismo	Consumo habitual de la planta de tabaco y sus productos	Índice tabáquico	-Cigarros consumidos por día -Años fumando	Continua

7.4.1 Datos demográficos

Los datos demográficos fueron colectados mediante el cuestionario en **Anexo V-VI**. Se preguntó a los individuos nombre completo, lugar de nacimiento, edad, sexo y ocupación.

7.4.2 Ancestría autoreportada

La ancestría fue autoreportada según el estado en México en el que nacieron los sujetos de estudio, sus padres y sus abuelos (**Anexo VI**). Debido a los relativamente recientes fenómenos de migración de españoles en México (colonización y migración masiva)¹⁸⁷, se optó por incluir sujetos que reportaran al menos una pareja de abuelos oriundos de México.

7.4.3 Antecedentes heredo patológicos respiratorios

Los antecedentes heredo patológicos fueron interrogados con el mismo cuestionario (**Anexo VII**). Se le preguntó al individuo la causa de muerte de sus padres y abuelos, según haya sido el caso; posteriormente se le mencionaron una serie de patologías respiratorias y no respiratorias y si algún familiar de primer o segundo grado padecía o padeció la enfermedad. Se tomó como antecedente al responder afirmativo a “EPOC” o “enfisema” en algún familiar de primer o segundo grado (abuelos, padres, hermanos).

7.4.4 Tabaquismo

Las variables de tabaquismo se colectaron mediante el mismo cuestionario (**Anexo VII**). Se incluyeron las variables de estatus de tabaquismo, cigarros fumados por día, edad de inicio al tabaquismo, años fumando de manera continua, cesación al tabaquismo y tabaquismo pasivo (omitida en fumadores).

Se emplearon las fórmulas para índice tabáquico (IT) de acuerdo a Wain et al., 2015¹²⁹ para los siguientes casos:

Para fumadores actuales:

$$\left(\frac{N^{\circ} \text{ de cigarros por día}}{20}\right) \times (\text{edad de reclutamiento} - \text{edad de inicio})$$

Para exfumadores:

$$\left(\frac{N^{\circ} \text{ de cigarros por día}}{20}\right) \times (\text{edad de cesación} - \text{edad de inicio})$$

7.4.5 Otras exposiciones

Se cuestionó al individuo sobre su exposición a otras variables ambientales en el mismo cuestionario (**Anexo VII**). Se puso énfasis en la combustión por biomasa crónica (≥ 100 horas-año) y otros compuestos orgánicos, inorgánicos y gases irritantes.

7.4.6 Función pulmonar

Se siguieron los criterios propuestos por la ATS/ERS.¹⁸⁸ La función pulmonar fue medida mediante la prueba de espirometría según los valores de referencia de Pérez-Padilla et al., 2007¹⁸⁹ para población mexicana. Antes de realizar la prueba, a los sujetos se les midió peso y estatura, así como realizó una toma de presión arterial mediante baumanómetro aneroide por personal capacitado de salud, ya que la espirometría es una prueba de esfuerzo máximo que eleva la presión arterial. Posteriormente se indicaron las instrucciones al sujeto por un técnico especialista en pruebas de función respiratoria. Se utilizó el equipo EasyOne® (nnd Medizintechnik AG, Zurich). Se realizó la prueba hasta obtener 3 resultados de calidad A y/o B (Muy buena y buena calidad) en un máximo de 8 intentos. Posteriormente, se administraron 200 μg de Salbutamol en el paciente y se esperaron de 15 a 20 min para volver a realizar la prueba con las mismas indicaciones anteriores. Se capturaron los mejores resultados de VEF₁, CVF y relación VEF₁/CVF que el equipo haya indicado.

7.5 Selección y procesamiento de las muestras

7.5.1 Selección de muestras

Las utilizadas son parte del Biobanco del laboratorio HLA del INER y fueron seleccionadas según el algoritmo de la **Figura 7.1**.

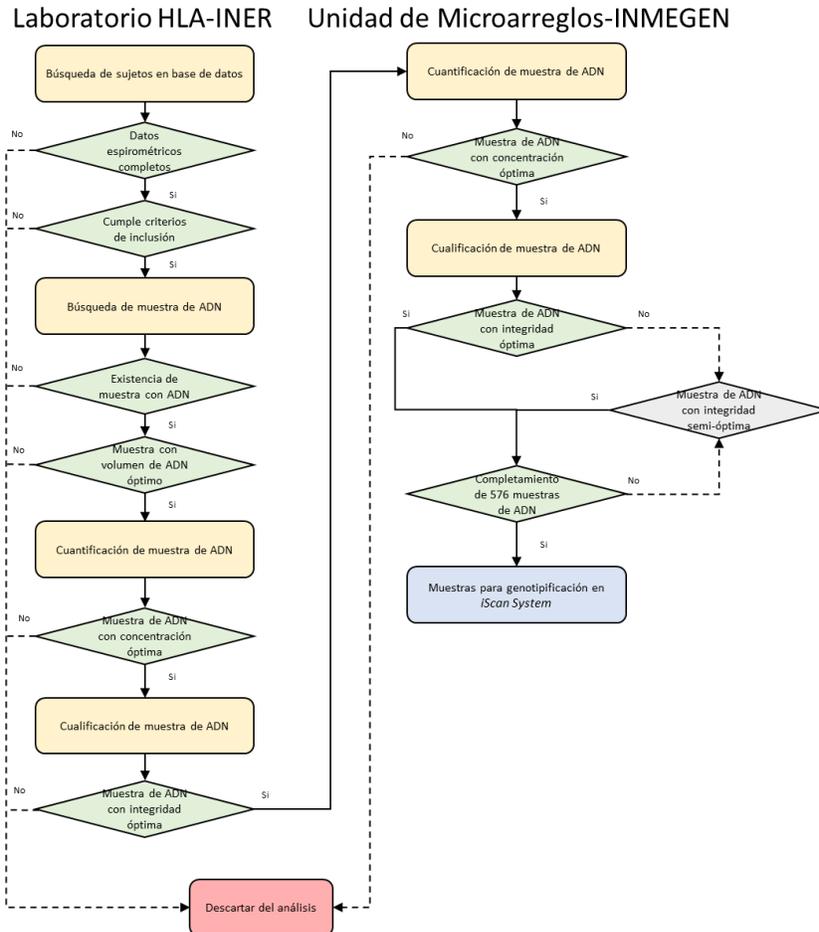


Figura 7.1 Algoritmo de selección de muestras

7.5.2 Extracción de ADN

Se extrajeron 15 mL de sangre por venopunción en antebrazo en tubos con EDTA como anticoagulante (S-Monovette®. Sarsted, Nümbrecht, Germany) Los tubos fueron centrifugados durante 5 min a 3,000 rpm. Posteriormente el plasma fue retirado de la fase celular, y fue agregada solución de lisis de eritrocitos hasta completar un volumen de 30 mL. Se centrifugaron a 3,000 rpm por 10 min, se decantó la mitad del sobrenadante y se adicionó solución RBC hasta 30 mL, se mezcló vigorosamente y se centrifugó a 3,000 rpm por 5 min (se repitieron dos veces este paso), se decantó el sobrenadante. A partir de este paso se empleó el kit comercial de extracción de ácidos nucleicos *BD tract DNA isolation kit* (Maxim Biotech Inc., CA, USA): Se agregaron 4 mL de solución BD1 mezclándose vigorosamente, se centrifuga a 3,000 rpm por 5 min. Se decantó todo el sobrenadante y se agregaron 4 mL de solución BD2 mezclándose vigorosamente,

se centrifugó a 3,000 rpm por 5 min. Se decantó el sobrenadante y se agregó 1 ml de solución BD3 mezclándose vigorosamente, se incubó en baño maría a 65°C durante 24 horas. Pasado el tiempo, se agregaron 500 µL de la mezcla en dos tubos microtubos (Sarsted, Nümbrecht, Germany) de 2 mL, a los cuales se les adicionaron 200 µL de solución BD4 mezclándose vigorosamente y se centrifugaron a 14,000 rpm por 10 min. Se decantó el sobrenadante en otros dos tubos microtubos de 2 mL y se agregaron 400 µL de isopropanol grado biología molecular (Sigma-Aldrich Co., MO, USA); se centrifugaron a 14,000 rpm durante 5 min. Se decantó el sobrenadante y se dejaron secar los tubos durante 3-4 h a temperatura ambiente. Se rehidrataron las muestras con amortiguador TE (10 mM Tris, 1mM EDTA, pH=8) (Invitrogen, CA, USA) y se incubaron en baño maría a 65°C durante 12 h. Posteriormente se conservaron a 4°C hasta el momento en que se usaron.

7.5.3 Control de calidad del ADN

Se cuantificaron las muestras por medio de espectrofotometría por luz UV en el equipo Nanodrop 2000 (Thermoscientific, MA, USA). Se consideraron muestras útiles aquellas que tienen concentración ≥ 50 ng/µl y pureza de relación 260/280 ≥ 1.7 .

Para evaluar la integridad del ADN genómico, realizó electroforesis en geles de agarosa al 1.3% con TAE 1X (Tris Acetate EDTA) (Cleaver Scientific Ltd, UK) a 300V, 300mA y 300W durante 45 min. Se utilizó marcador de peso molecular de 250-25 K pb (CSL-MDNA-HR DNA Ladder RTU, Cleaver Scientific Ltd, UK) para referenciar los pesos molecular del DNA y calidad de las muestras de participantes. Al observarse la banda de 25 Kb, 10 Kb u 8 Kb se tomaba como ADN genómico de buena calidad.

7.6 Microarreglo Infinium® Multi-Ethnic AMR/AFR BeadChip

7.6.1 Características generales

La genotipificación de variantes en el ADN fue realizada con el microarreglo *Infinium® Multi-Ethnic AMR/AFR BeadChip* contiene 1'430,141 variantes genéticas, según las especificaciones del proveedor (Illumina Inc. CA, USA). Entre estas

variantes se incluyen SNPs, mutaciones de línea germinal, inserciones, deleciones y CNVs. Este microarreglo contiene distintos marcadores en el genoma nuclear y extranuclear, con variantes relevantes en estudios clínicos y de ancestría, lo que permite la identificación de asociaciones genéticas con rasgos comunes y raros, lo que proporciona información multiétnica incluyendo variantes propias de poblaciones americanas y africanas mestizadas. Las variantes contenidas en el microarreglo son de interés biomédico reportadas en distintos proyectos, incluidos la fase 3 de 1000 Genomes Project (1kGP) ¹⁹⁰, Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) ¹⁹¹, Population Architecture using Genomics and Epidemiology (PAGE) ¹⁹², T2D-Genes Consortium ¹⁹³, Online Mendelian Inheritance in Man (OMIM) ¹⁹⁴, ClinVar ¹⁹⁵, American College of Medical Genetics and Genomics (ACMG) ¹⁹⁶, paneles de detecciones del operador y otros recursos, para más detalles revisar los **Anexos VIII-IX**.

7.6.2 Genotipificación

El microarreglo se analizó por medio del equipo *iScan System* (Illumina Inc. CA, USA) realizado por personal capacitado del Instituto Nacional de Medicina Genómica (INMEGEN), utilizando los requerimientos por Illumina®.

7.6.3 Control de calidad

Fue realizado mediante el software PLINK (versión 1.07) ¹⁹⁷. Se depuraron SNPs cuyos genotipos faltantes fueron superiores al 5%, así como aquellos que no se encontraran en algún cromosoma según la anotación del producto (3,963). Se excluyeron SNPs que estuvieran en cromosoma Y (2,360), región pseudoautosómica humana (3,753) y mitocondrial (645). Debido a que se buscan variantes comunes en la población, se eliminaron aquellas con una frecuencia menor al 1% ($\approx 500,000$).

De los individuos remanentes, se contempló un umbral del 95% en *call rate*, por lo cual se excluyeron a 14 sujetos que no cumplieron con ello. Para evitar la presunta relación familiar directa de los participantes, se excluyeron del análisis a sujetos cuyo coeficiente de coancestría (θ) por identidad por descendiente (IBD) fuera

mayor a 0.125, siendo un total de 23 individuos eliminados por tener algún grado directo de parentesco.

7.7 Análisis de ancestría

Para las poblaciones de referencia, se consideraron como europeos (EUR) a 60 residentes de Utah, EE.UU. con ancestría del Norte y Oeste de Europa (CEU) del Proyecto HapMap ¹⁹⁰; como africanos (AFR) a 60 sujetos etnia Yoruba en Ibadan, Nigeria (YRI) del Proyecto HapMap ¹⁹⁰; se consideraron amerindios (AMI) a 70 sujetos, 21 Zapotecas de Oaxaca, México, 26 Mayas de Campeche, México y 23 Tepehuanos de Durango, México reclutados en Moreno-Estrada et al., 2014. ⁸¹

Mediante el software ADMIXTURE (versión 1.3) ⁸⁶ fueron calculadas las proporciones de las poblaciones ancestrales con $K = 3$, asumiendo que son 3 poblaciones ancestrales que dan origen a la población mestizada de estudios (casos y controles).

El PCA fue realizado mediante el software EIGENSOFT (versión 6.1.4) ^{198,199} para estimar por los componentes principales en el grupo de estudio.

7.8 Asociación de genoma completo

Se evaluó como variable dependiente tener diagnóstico de EPOC secundaria a tabaquismo. Siguiendo el algoritmo de la **Figura 7.2**, se evaluaron variables reportadas en la literatura que hubieran sido ajustadas más de una vez en algún estudio publicado de la **Tabla 1.3** en estudios que evaluaban EPOC secundaria a tabaquismo como variable dependiente. Se realizó una regresión logística multinomial mediante el software PLINK (versión 1.07) ¹⁹⁷. La ecuación del modelo fue la siguiente:

$$\log \frac{\theta(x)}{1 - \theta(x)} = \beta_0 + \beta_1 x_1 + \beta_n x_n$$

Donde:

$\theta = \text{EPOC} = 1$

$x_1 = \text{SNP}$

$x_n = \text{variables de ajuste}$

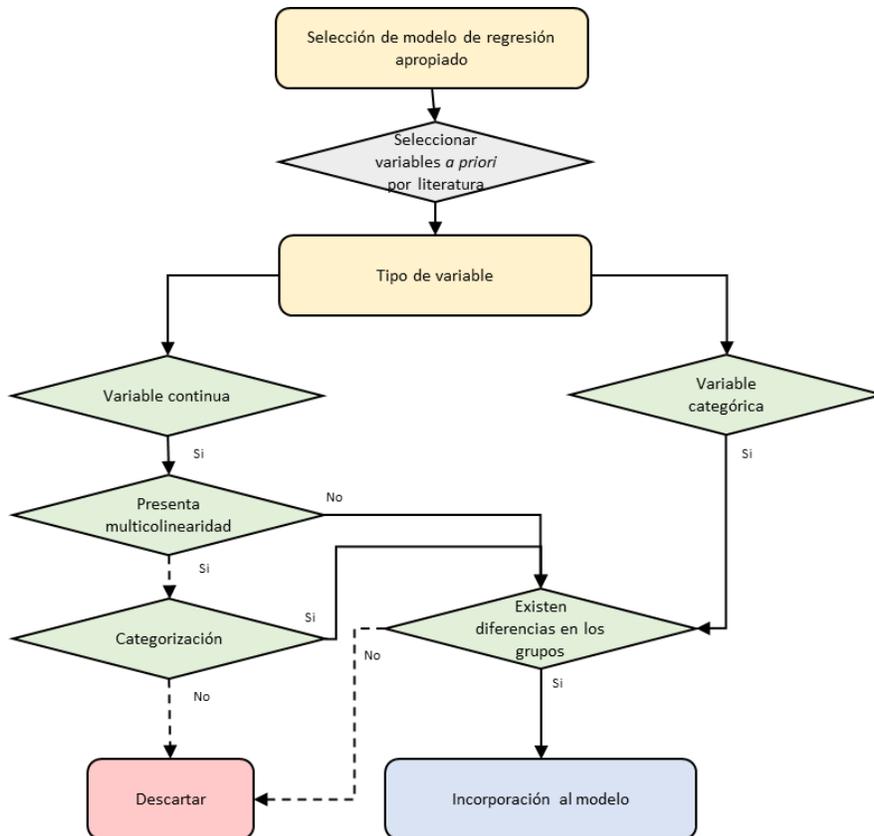


Figura 7.2 Algoritmo de variables de ajuste para el GWAS por modelo de regresión logística.

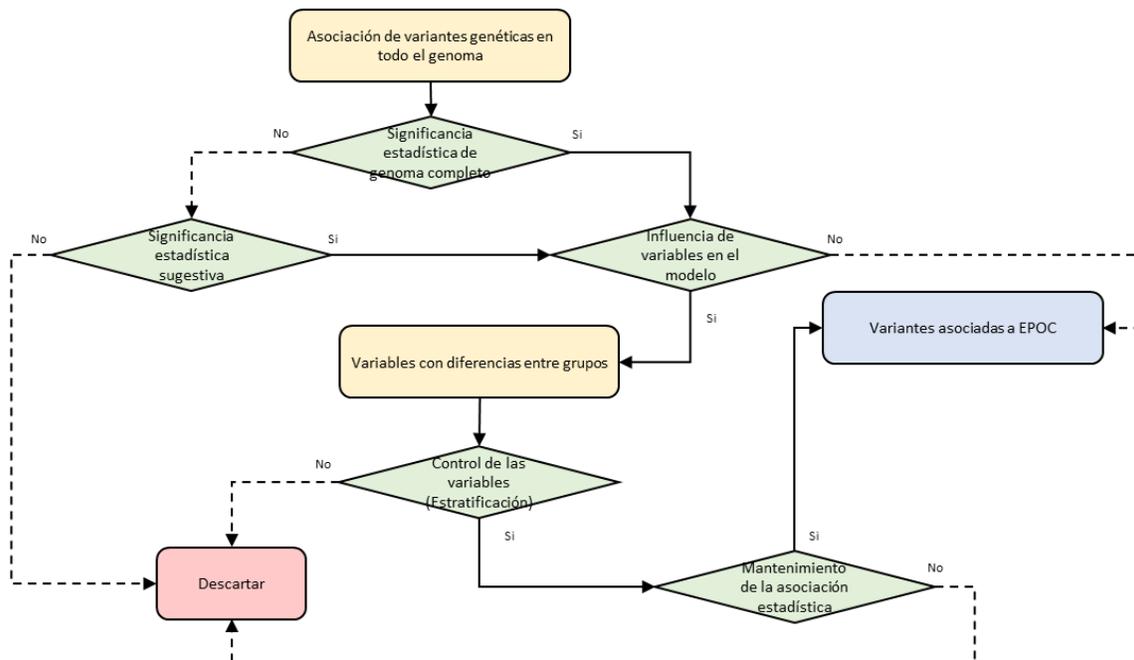


Figura 7.3 Algoritmo de las variantes genéticas asociadas en el GWAS.

Para análisis de regresión logística se siguió el algoritmo de la **Figura 7.3**. Posteriormente, se graficaron los cuantiles de valores de p de cada variante respecto a los teóricos en un *QQ-plot*, para determinar la distribución y poder estadístico de estos valores, respecto a los valores teóricos esperados.

7.9 Anotación genética

Las variantes se identificaron en UCSC Genome Browser ²⁰⁰ (Human Dec. 2013 (GRCh38/hg38) Assembly) para su localización física (región cromosómica), así como si se encontraban dentro o cercano a algún gen. Las frecuencias alélicas de las variantes asociados fueron referenciadas de 1kGP ¹⁹⁰, que incluye la frecuencia global, en europeos y americanos.

En caso de que la variante tuviera un identificador validado (i.e. *rs*), se buscó con esa identificación; en cambio, si no contaba con identificador validado (i.e. *JHU*), se buscó por la localización en el cromosoma en pares de bases, referenciada en el microarreglo.

7.10 Análisis estadístico

Se utilizó el software estadístico R (versión 3.2.3). ²⁰¹ Para el análisis descriptivo, se verificó la normalidad de las variables cuantitativas mediante la prueba Shapiro-Wilk ($\alpha = 0.95$) y se graficaron en histogramas para verificar medidas de simetría. En caso de corresponder a variables con una distribución normal, se reportaron en medias y desviación estándar; en caso de no cumplir con distribución normal, se reportaron en mediana y rangos intercuartílicos.

Para todas las variables cuantitativas se realizó un diagrama de correlación por medio del coeficiente correlación de Pearson (ρ) con $\alpha = 0.95$, para determinar entre que variables se encuentra correlación y tomar decisiones para el modelo de asociación.

Para la elección de las variables de ajuste en el modelo de regresión se eligieron aquellas que tenían diferencias significativas entre ellas y estuvieran reportada en la literatura, las cuales fueron comparadas mediante la prueba de t de Student para medidas independientes y Wilcoxon-Mann-Whitney en caso de ser paramétricas y

no paramétricas, respectivamente. Los datos fueron analizados mediante los paquetes en R descritos en el **Cuadro 7.5** para distintas etapas del análisis.

Cuadro 7.5. Paquetes de R empleados en análisis de datos.

Paquete de R	Análisis
STATS ²⁰¹	Estadística descriptiva (pruebas de normalidad) y comparación de grupos (t de Student, Wilcoxon-Mann-Whitney)
corrplot ²⁰²	Correlación (ρ de Pearson)
lsr ²⁰³	Magnitud media del efecto (D de Cohen)

7.11 Visualización de datos

Se utilizó el software estadístico R (versión 3.2.3) ²⁰¹ para graficar los datos obtenido en los distintos análisis, utilizando los paquetes del **Cuadro 7.6**.

Para el análisis de asociación de las variantes en regiones cromosómicas se empleó el software LocusZoom ²⁰⁴, el cual se utilizó la anotación de Americanos del 1kGP, (hg19/1000 Genomes Nov 2014 AMR).

Cuadro 7.6 Paquetes de R empleados en la visualización de datos.

Paquete de R	Tipo de gráfico
ggplot2 ²⁰⁵	<i>Boxplot</i> , gráfico de barras, <i>scatterplot</i> , histogramas
corrplot ²⁰²	Gráfico de correlación
qqman ²⁰⁶	<i>Manhattan plot</i> , <i>Q-Q plot</i> para variantes genéticas
circlize ²⁰⁷	<i>Circleplot</i>
ggpubr ²⁰⁸	Histograma de densidad

8 Resultados

De acuerdo a los criterios de inclusión, se incluyeron 575 participantes en total, 221 que cumplían la definición de caso y 344 que cumplían la de control. Sin embargo, después del control de calidad de las muestras para el microarreglo, se obtuvieron 216 casos y 320 controles con razón 1:1.5, sumando un total de 536 individuos con 738,822 variantes genotipificadas exitosamente. Los datos se resumen en el Las variables cuantitativas de los participantes se graficaron en histogramas (**Anexos X-XII**) para determinar su tipo de distribución, sin embargo, ya que la mayoría no cumple distribución normal se describen en medianas (RIC) (**Tabla 8.1** y en los **Anexos XIII-XVI**).

8.1 Descripción de variables

La edad de los casos tiene una mediana de 66 años, en comparación de unos controles más jóvenes con una mediana de edad de 57 años. Respecto al sexo los grupos se encuentran dispares, mientras que el 77.3% (n = 167) de los casos prevalece el sexo masculino, en los controles el 53.6% (n = 171) son del sexo femenino. En el IMC existen diferencias entre los grupos, en los casos se presenta una mediana (24.8) y rangos en normopeso (21.7-28.3); los controles tienen una mediana mayor (27.6), pero con rangos en sobrepeso (24.8-30.4). Respecto al tabaquismo, existen algunas diferencias entre los grupos: la proporción de fumadores actuales en el momento del cuestionario fue de 34.3% en casos, comparado con 79.1% de los controles; en los cigarros fumados por día, los casos presentan una mediana de 20, con RIC de 13-20, distinto a los controles que tienen una mediana de 14 y un RIC más amplio (8-20); en el tiempo fumando en años los casos presentan una mediana (42) y RIC (35.7-48.2) mayores a los controles [37 (32-43)]; al igual que CPD y años fumando, las variables que componen el IT, de manera acumulada los casos han fumado más tiempo que los controles, con una mediana de 40.25 (25-55) y 23 (13.2-37), respectivamente; respecto a la edad que los sujetos reportan iniciaron su consumo de tabaco, presentan valores idénticos [17 (15-20)].

En el cuestionario se incluyeron preguntas clave sobre los antecedentes heredopatológicos y familiares, que de no considerarse previamente, podrían sesgar los resultados. A manera de filtro previo a la ancestría de los sujetos del estudio, se les cuestionó si alguno de sus abuelos era de origen español, sin embargo, en casos (5.5%) y controles (5.6%) el número de sujetos que lo reportaban fue similar; también se recopiló la respuesta de si algún familiar directo (abuelos, padres, hermanos e hijos) del sujeto de estudio padece o padeció EPOC, siendo mayor en casos (17.1%) que en controles (11.2%). Se midió la función pulmonar en los sujetos de estudio, parámetro que discrimina el padecimiento de la EPOC según GOLD, en lo cual se observa claramente que los casos poseen peores niveles en mediana y RIC en VEF₁ en litros [1.4 (0.9-2)] y porcentaje [58 (39.5-74)], CVF en litros [2.9 (2.2-3.5)] y porcentaje [83 (69.5-98)], y por consiguiente se encuentran obstruidos de acuerdo a la relación VEF₁/CVF en porcentaje [53.6 (40-63.1)], en comparación de los niveles controles en VEF₁ en litros (2.6 (2.1-3.1)) y porcentaje [98.5 (88-109)], CVF en litros [3.3 (2.7-3.9)] y porcentaje [94 (85.5-105)], así como en VEF₁/CVF en porcentaje [80.5 (77-84)]. Analizando solamente a los casos (**Tabla 8.2, Anexo I**) presentan una mediana de edad de diagnóstico de 63 años (57-68), y respecto a la severidad de la enfermedad, los sujetos con el estadio GOLD II comprende el 45.8% del total de casos, seguido de GOLD III (27.3%), I (15.3%) y con menor proporción el estadio más severo, GOLD IV (11.1%).

Tabla 8.1 Variables demográficas, de tabaquismo, antecedentes y de función pulmonar de los grupos de estudio.

Variable	Casos	Controles
n	216	320
Edad (años)	66 (60-72)	57 (51-63)
Sexo masculino, n (%)	167 (77.3)	149 (46.4)
IMC (kg/m ²)	24.8 (21.7-28.3)	27.6 (24.8-30.4)
<i>Tabaquismo</i>		
Fumador actual, n (%)	74 (34.3)	254 (79.1)
Cigarros fumados por día (n)	20 (13.7-20)	14 (8-20)
Tiempo fumando (años)	42 (35.7-48.2)	37 (32-43)
IT (paquetes-año)	40.25 (25-55)	23 (13.2-37)
Edad de inicio (años)	17 (15-20)	17 (15-20)
<i>Antecedentes</i>		
Antecedentes con EPOC, n (%)	37 (17.1)	36 (11.2)
Abuelos españoles, n (%)	12 (5.5)	18 (5.6)
<i>Función pulmonar*</i>		
VEF ₁ (L)	1.4 (0.9-2)	2.6 (2.1-3.1)
VEF ₁ (%)	58 (39.5-74)	98.5 (88-109)
CVF (L)	2.9 (2.2-3.5)	3.3 (2.7-3.9)
CVF (%)	83 (69.5-98)	94 (85.5-105)
VEF ₁ /CVF, (%)	53.6 (40-63.1)	80.5 (77-84)
<i>Ancestría</i>		
AMI (%)	46.6 (35-58.3)	50.3 (36.9-62.3)
EUR (%)	49.3 (37.9-59.7)	45.5 (34.5-59.1)
AFR (%)	3.1 (2.3-4.5)	3.22 (2-4.5)

Valores presentados en Mediana (RIC); n (%); *: valores post-broncodilatador

Tabla 8.2 Variables clínicas en los pacientes con EPOC.

Variables clínicas	Casos
Edad de diagnóstico (años)	63 (57-68)
Estadio GOLD*, n (%)	
I	33 (15.3)
II	99 (45.8)
III	59 (27.3)
IV	24 (11.1)

Valores presentados en Mediana (RIC); n (%); *: Determinado por VEF₁ post-broncodilatador

8.2 Ancestría

Entre las tres poblaciones de referencia y los participantes del estudio (casos y controles), se identificaron 147,197 SNPs en común, los cuales permitieron determinar el análisis de ancestría y PCA. Las proporciones de los componentes ancestrales por individuo se muestra en la **Figura 8.1** y en el **Anexo XVI** el promedio, en la cual se observa el amplio espectro de variación en la estructura poblacional del mexicano mestizo, sin embargo, estas proporciones difieren entre los grupos: se observa que los controles tienen una mediana de 50.3% (36.9-62.3) de componente amerindio y los casos 46.6% (35-58.3), existiendo diferencia estadísticamente significativa ($p = 0.047$). En cambio, respecto al componente europeo se observa que los casos tienen una mediana de 49.3% (37.9-59.7) a diferencia del 45.5% (34.5-59.1) de los controles, con una diferencia estadística limítrofe ($p = 0.05$); en cuanto al componente africano, ambos grupos son similares ($p = 0.43$), con una mediana de 3.1% (2.3-4.5) y 3.2% (2-4.5) en casos y controles, respectivamente.

En la evaluación de ancestría por PCA, en la se observa agregación de las poblaciones de referencia. Al comparar los primeros PC, que explican la mayoría de variabilidad en la población, se obtuvieron los valores de 43.63 en PC1 y 21.39 en PC2 (**Figura 2**), en los que se puede observar que existe una separación respecto a los ejes de variación entre la población africana y las no africanas (amerindia, europea y mexicano mestizo). De igual manera, se puede observar que el grupo de casos y controles se localiza entre la población caucásica y amerindia, llegando incluso a tener individuos extremos que se sobrelapan en las poblaciones de referencia. En la población caucásica se localizan controles ($n = 2$); en la amerindia casos ($n = 6$) y controles ($n = 18$), al igual que algunos acercándose a la africana tanto casos ($n = 3$) como controles ($n = 4$). Debido a que el PC3 (**Anexo II**) presenta un valor de 1.83, los PC que siguen presentarán valores más bajos, siendo menos informativos.

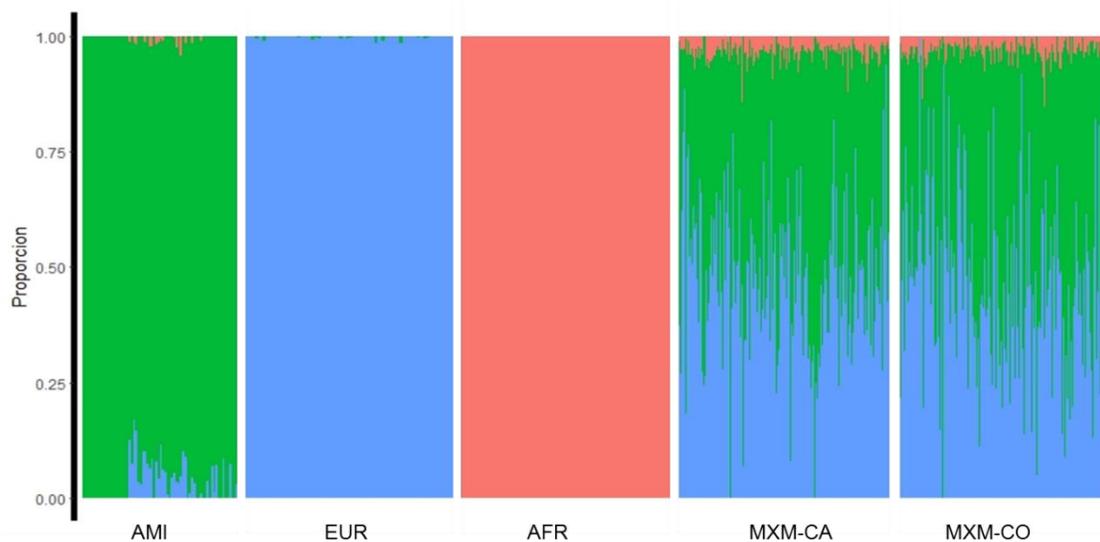


Figura 8.1 Proporción de los componentes ancestrales en cada sujeto de grupo de estudio. MXM-CA: Casos mexicano mestizos, MXM-CO: Controles mexicanos mestizos

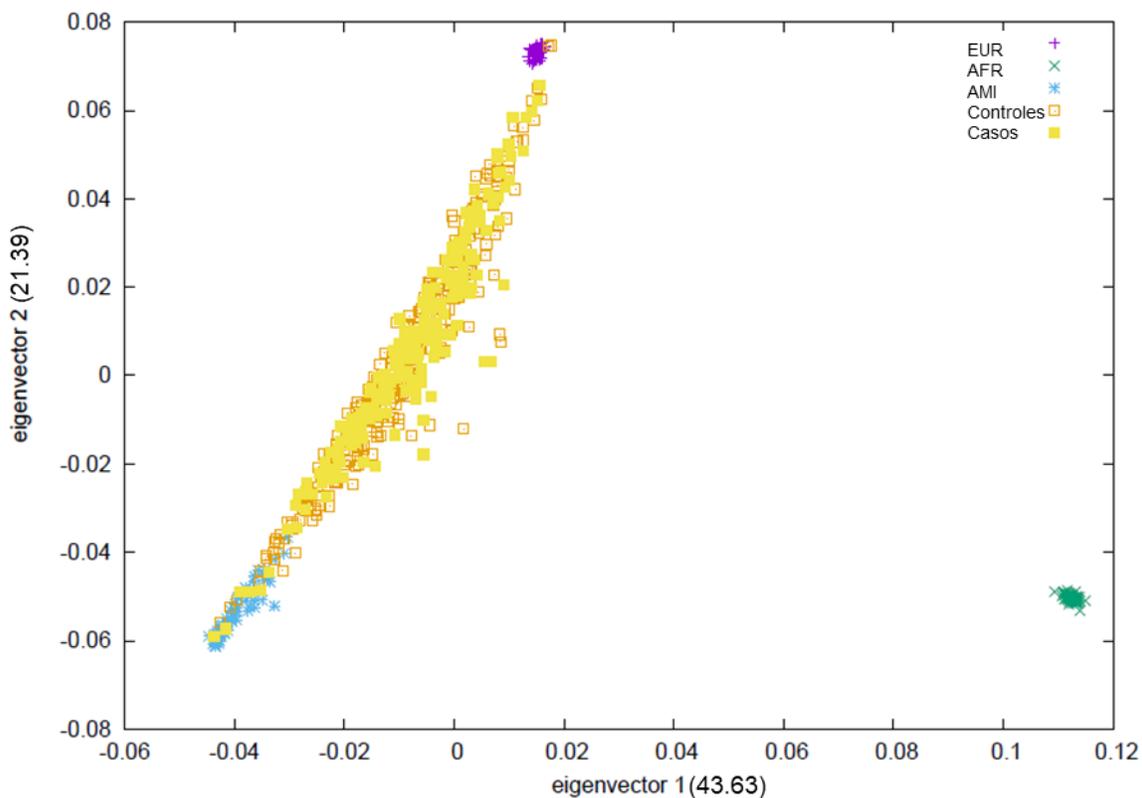


Figura 8.2 Análisis de componentes principales entre PC1 y PC2 en los grupos de estudio

8.2.1 Análisis de correlación

Las correlaciones entre las variables se pueden observar en la **Figura 8.3**, en dónde se presentan las correlaciones que fueron estadísticamente significativas ($p < 0.05$). La edad es la variable que tiene más correlación con otras variables. La edad tiene una correlación positiva moderada con tiempo fumando en años ($\rho = 0.49$) e IT ($\rho = 0.29$); también presenta correlación negativa en variables espirométricas, en CVF por porcentaje ($\rho = -0.14$) y litros ($\rho = -0.33$) es menor y aumenta inversamente en VEF₁ en porcentaje ($\rho = -0.36$) y litros ($\rho = -0.51$), así como en la relación VEF₁/CVF ($\rho = -0.46$), esta última también presenta una correlación positiva con el IMC ($\rho = 0.37$). Respecto a las variables relacionadas con el tabaquismo, el IT es una variable compuesta calculada por medio de CPD ($\rho = 0.92$) y años fumando ($\rho = 0.39$), presentado una correlación positiva; debido a que los años fumando se calculan con base en la edad de inicio del participante, existe correlación negativa ($\rho = -0.46$). Las variables de función pulmonar presentan correlación entre las variables que la componen, sin embargo, presentan correlación negativa respecto al tabaquismo en CPD (CVF %, $\rho = -0.12$; VEF₁ L, $\rho = -0.11$, VEF₁ %, $\rho = -0.23$; VEF₁/CVF %, $\rho = -0.23$), años fumando (VEF₁ L, $\rho = -0.18$, VEF₁ %, $\rho = -0.18$; VEF₁/CVF %, $\rho = -0.21$) e IT (CVF %, $\rho = -0.16$; VEF₁ L, $\rho = -0.19$, VEF₁ %, $\rho = -0.28$; VEF₁/CVF %, $\rho = -0.29$). De la misma manera, existieron correlaciones respecto a los componentes ancestrales y los componentes principales (PC) en ambos sentidos, la ancestría AMI presentó una correlación negativa con ancestría EUR ($\rho = -0.99$), AFR ($\rho = -0.23$), PC1 ($\rho = -0.98$) y PC2 ($\rho = -0.99$); la ancestría EUR mostró valores positivos respecto a PC1 ($\rho = 0.96$) y PC2 ($\rho = 1.00$); la ancestría AFR mostró correlación con PC1 ($\rho = 0.42$). No se identificaron correlaciones entre los componentes ancestrales y PC respecto a las variables clínicas.

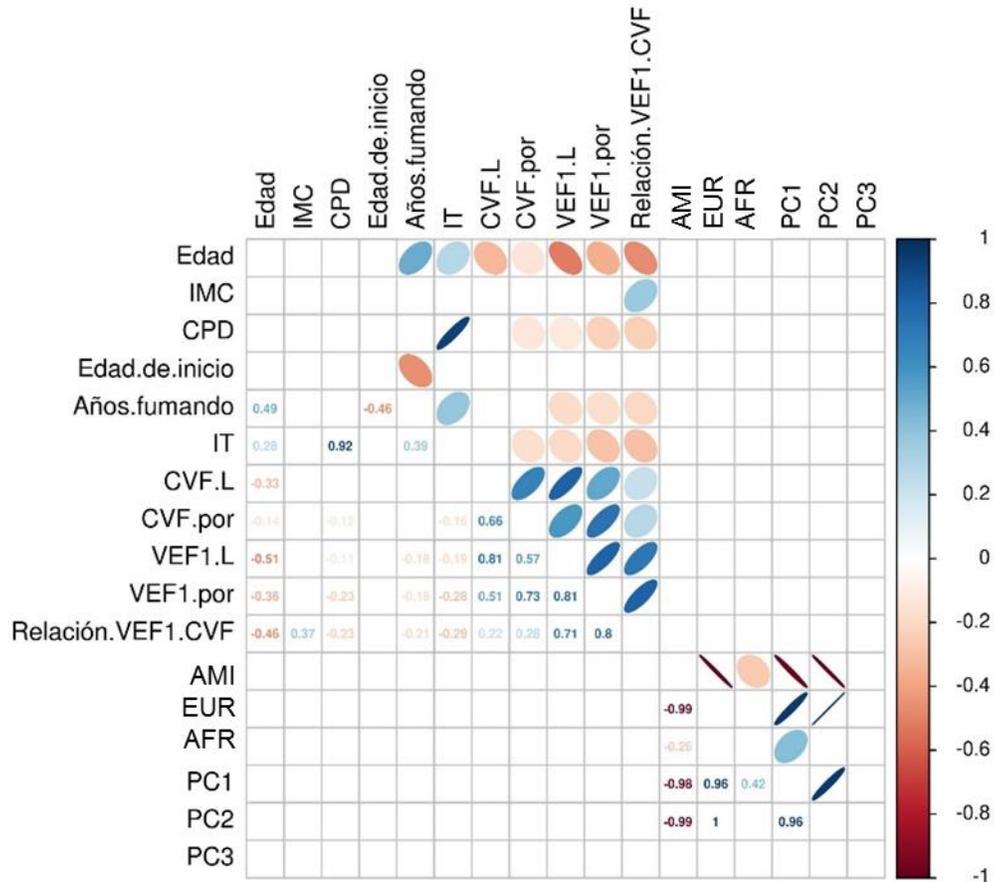


Figura 8.3 Corr plot de las variables cuantitativas de los grupos de estudio. Se presentan las correlaciones que fueron estadísticamente significativas. Se muestran graficadas en elipses en la mitad diagonal superior y su valor de coeficiente de correlación de Pearson en la mitad diagonal inferior.

8.3 Análisis de genoma completo

8.3.1 Modelo de regresión

Con el propósito de relacionar susceptibilidad de variantes genéticas con la enfermedad, se optó por un modelo de regresión logística con padecer EPOC como variable dependiente. Con base en la literatura y las correlaciones en las variables de importancia, se decidió ajustar por edad, sexo estatus de tabaquismo, IT y PC1 (**Tabla 8.3**). Debido a la correlación en edad e IT, y la importancia que muestran en la EPOC, se decidió tomar la edad en estratos de 10 años e IT de manera continua como exposición acumulada de tabaquismo.

Tabla 8.3 Variables de ajuste en el modelo.

Variable	Casos	Controles	p
n	216	320	
Edad (años)	66 (60-72)	57 (51-63)	<0.001
40-50 años, n (%)	5 (2.3)	64 (20)	<0.001
51-60 años, n (%)	50 (23.1)	146 (45.6)	
61-70 años, n (%)	95 (44)	86 (26.9)	
≥ 71 años, n (%)	66 (30.5)	24 (7.5)	
Sexo masculino, n (%)	167 (77.3)	149 (46.4)	<0.001
<i>Tabaquismo</i>			
IT (paquetes-año)	20 (13.7-20)	14 (8-20)	<0.001
<i>Ancestría</i>			
AMI (%)	46.6 (35-58.3)	50.3 (36.9-62.3)	0.047
EUR (%)	49.3 (37.9-59.7)	45.5 (34.5-59.1)	0.05
AFR (%)	3.1 (2.3-4.5)	3.22 (2-4.5)	0.43

Valores presentados en Mediana (RIC); n (%);

Se estableció un valor de significancia de genoma completo de $p = 6.77 \times 10^{-8}$ basado en 738,822 variantes totales que pasaron control de calidad previa y un valor sugestivo de significancia de genoma completo ($p < 1 \times 10^{-5}$). Se realizó un modelo sin ajustar, tomando en cuenta todas las variantes independientemente con tener o no EPOC, en el cual las variantes significativas menores a 1×10^{-5} se resumen en la **Tabla 8.4**. En el *QQ plot* se observa que la estratificación poblacional es mínima ($\lambda = 1.06$) (**Figura 8.4**). Se muestran las variantes graficadas en un *Manhattan plot* en la **Figura 8.5**. La variante con un valor de significancia mayor, fue el alelo C de rs6110505 [OR = 1.87 (1.44-2.41), $p = 1.88 \times 10^{-6}$] en *MACROD2*. El alelo A rs12518621 cerca del gen *LINC01170* presentó riesgo a EPOC [OR = 0.55 (0.43-0.71), $p = 3.06 \times 10^{-6}$]. El alelo C, de la variante 16:67993643-T-C en *DPEP2* obtuvo un menor riesgo a EPOC [OR = 0.46 (0.33-0.64), $p = 4.89 \times 10^{-6}$]. Las variantes localizadas en *DPEP3*, exm-rs255049 [OR = 0.47 (0.33-0.65), $p = 7.08 \times 10^{-6}$] y 16:67977382-T-C [OR = 0.47 (0.33-0.65), $p = 8.78 \times 10^{-6}$] presentaron menor riesgo a EPOC. Cerca de *AL445430* se asoció el alelo A de rs4711289 con un riesgo incrementado de EPOC [OR = 1.9 (1.43-2.5), $p = 7.37 \times 10^{-6}$]. El alelo C de las variantes 16:67833171-T-C y exm2264618, localizados en *CENTP* y *RANBP10*, respectivamente, presentaron menor riesgo a EPOC [(OR = 0.45 (0.32-0.64), $p = 9.44 \times 10^{-6}$)].

En el análisis de asociación regional, se visualizaron las variantes asociadas (*lead* SNP) en su región cromosómica y si presentaban desequilibrio de ligamiento con otras variantes de su alrededor que estuvieran en un mismo bloque. De acuerdo con el valor de significancia de la variante respecto a su región cromosómica, en el modelo sin ajustar (**Figura 8.8**), se encontró que en 20p12.1 (**Figura 8.8-A**), el rs6110505, a pesar de encontrarse en un bloque dentro del gen, no presentaba desequilibrio de ligamiento con las demás variantes. En la región cromosómica 5q23.2 (**Figura 8.8-B**), el rs12518621 no presenta desequilibrio de ligamiento y se encuentra a 200 Kb de *ZNF608*. La región 16q22.1 (**Figura 8.8-C**) se visualiza con una gran cantidad de genes en una corta distancia, sin embargo, apenas existen 11 variantes en 600 Kb, en las cuales destaca el rs6499143 en *CENPT* con un valor moderado de r^2 (0.6-0.7) con una variante en *DUS2*. En la región 6p24.2-.1 (**Figura 8.8-D**) rs4711289 presentó una r^2 (0.6-0.7) con dos variantes, aparentemente en un pequeño bloque, cercano a *NEDD9*.

Posteriormente, se realizó el modelo ajustado con las variables propuestas en la literatura, así como que tuvieran diferencia estadística en el análisis multivariado. El modelo de regresión fue ajustado por edad (categórica), sexo, IT y PC1. Para evitar multicolinealidad, se determinó que el factor de inflación de la varianza fuera menor a 10 en cada variante. En el QQ *plot* se puede observar que no hay evidencia de estratificación poblacional ($\lambda = 1.008$) (**Figura 8.6**). Se identificaron 8 variantes asociadas en el modelo, resumidas en la **Tabla 8.5**. Se encontraron dos variantes cerca de *AGR3*, el alelo A de rs4559136 [OR = 0.30 (0.12-0.41), $p = 8.64 \times 10^{-7}$] y el alelo C de rs4320447 [OR = 0.27 (0.15-0.47), $p = 6.95 \times 10^{-6}$] asociadas a menor riesgo de EPOC. También, se encontró al alelo G de JHU_15.94073449 [OR = 10.97 (3.94-30.53), $p = 4.50 \times 10^{-6}$] cercano a *MCTP2* asociado a mayor riesgo de EPOC. La variante rs2901070, cercano a *SERPINI1* por medio del alelo T se asoció a riesgo de EPOC [OR = 4.00 (2.21-7.25), $p = 4.67 \times 10^{-6}$]. Cerca de *ADGRB3*, se encontraron dos variantes asociadas a riesgo de EPOC, el alelo T de rs352523 [OR = 5.49 $\times 10^{-6}$ (1.62-3.37), $p = 5.49 \times 10^{-6}$] y el alelo T de rs9363933 [OR = 8.97 $\times 10^{-6}$ (1.62-3.49), $p = 8.97 \times 10^{-6}$]. Las variantes rs17257107 y 1:26450101-A-G en *PDIK1L* y *DHDDS*,

respectivamente, por medio del alelo G se asociaron a riesgo [OR = 3.04, (1.86-4.96), $p = 9.00 \times 10^{-6}$].

Las regiones cromosómicas de los *lead* SNPs asociados a EPOC en el modelo ajustado se muestran en la **Figura 8.9**. En la región 7q21.1 (**Figura 8.9-A**), la variante rs4559136 muestra un alto desequilibrio de ligamiento ($r^2 \geq 0.8$) con rs4320447, ambas cercanas a *AGR3* y en mismo bloque de aproximadamente 100 Kb. El rs2901070 se localiza en la región 3q26.1 (**Figura 8.9-B**), entre *SERPINI1* y *LINCO1330*. En la región 6q12 (**Figura 8.9-C**) se localiza el rs352523 físicamente muy cercano a rs9363933, con alto desequilibrio de ligamiento ($r^2 \geq 0.8$) con otra variante, los cuales aparentemente se encuentran cercanos a *BAI13*. La región 1p36.11 (**Figura 8.9-D**) localiza a rs17257107 en *PDIK1L* con alto desequilibrio de ligamiento ($r^2 \geq 0.8$) con otra variante en *TRIM63*, las cuales se encuentran en un bloque de aproximadamente 100 Kb.

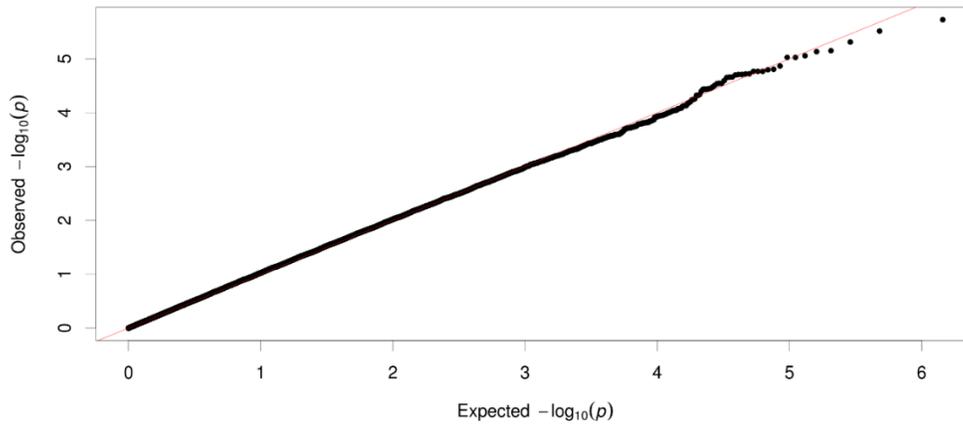


Figura 8.4 QQ-plot del modelo sin ajustar (modelo 0).

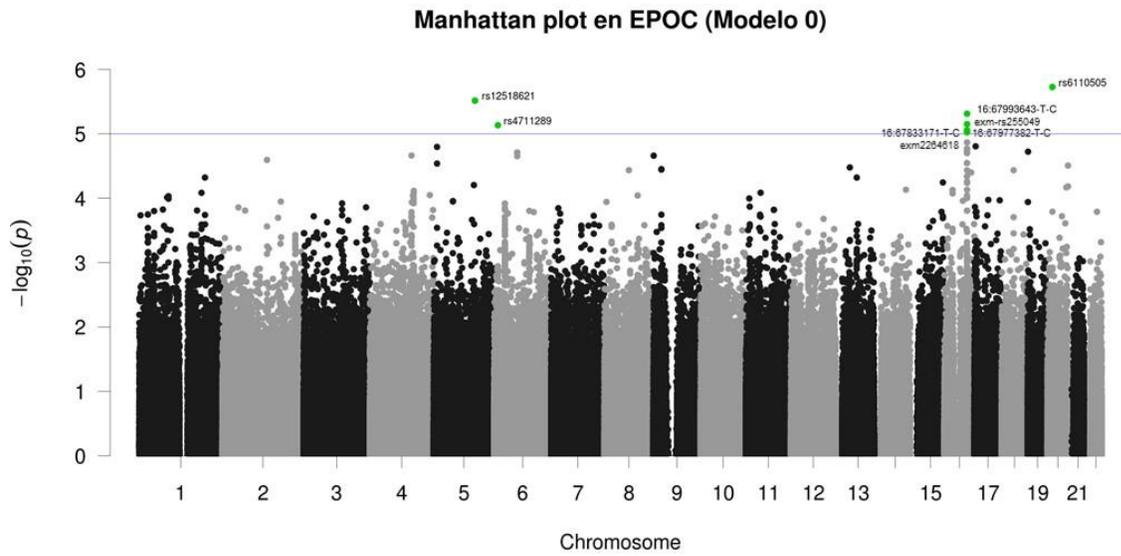


Figura 8.5 Manhattan plot del modelo sin ajustar (modelo 0).

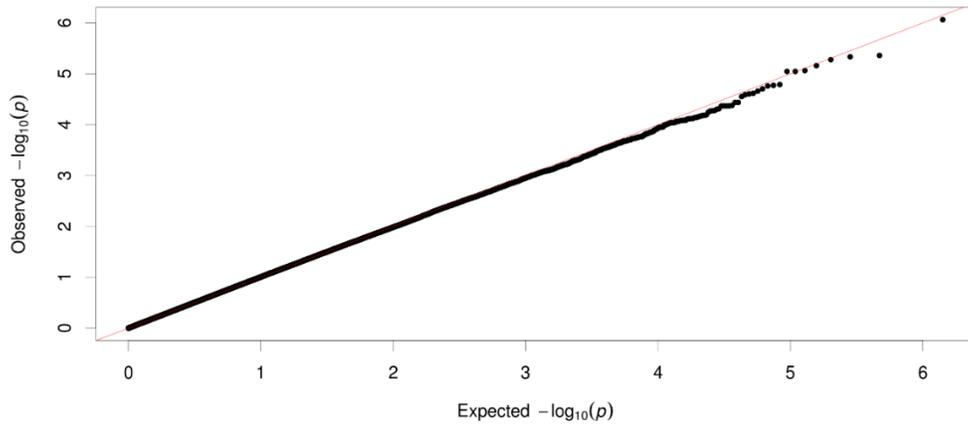


Figura 8.6 Q-Q-plot del modelo ajustado.

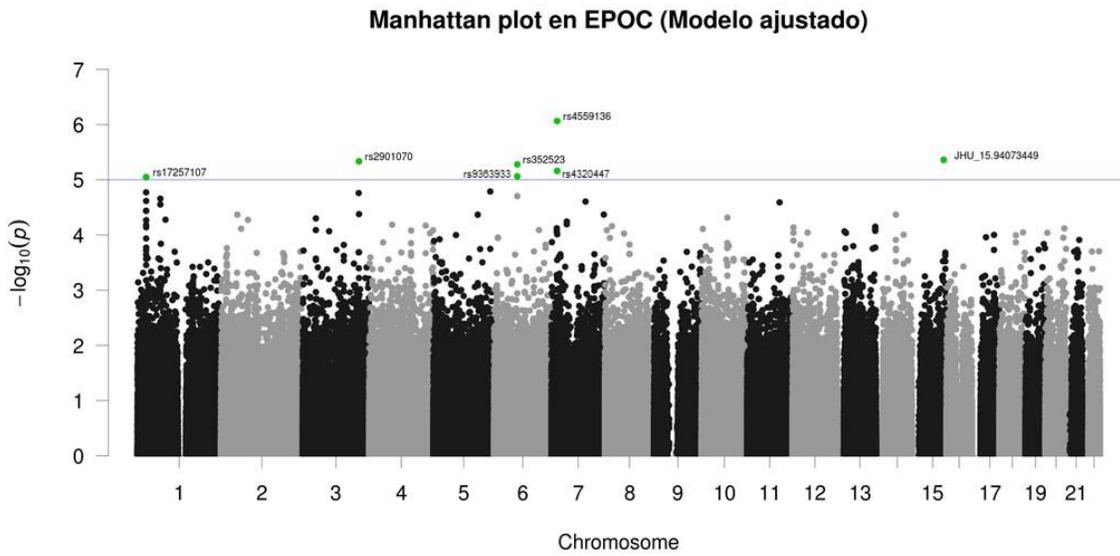


Figura 8.7 Manhattan plot del modelo ajustado.

Tabla 8.4 Variantes asociadas a EPOC en grupo de estudio de acuerdo al modelo sin ajustar.

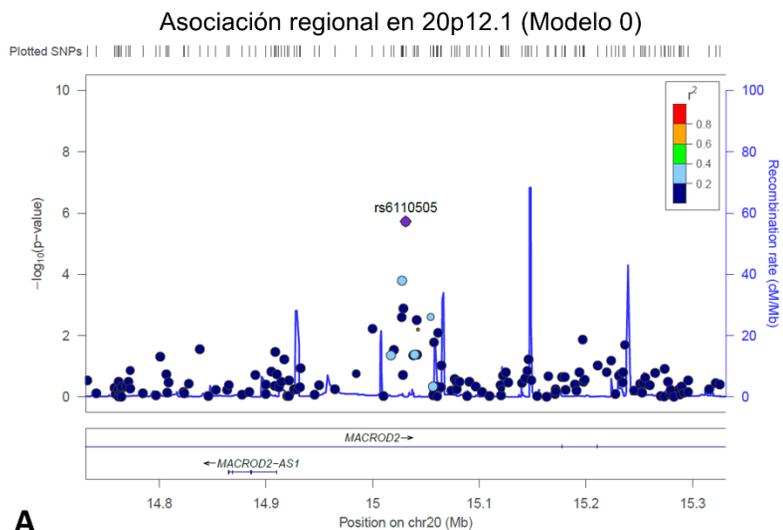
Región	SNP	Alelo asociado	Frecuencia del alelo	Gen	p	OR	IC (95%)
20p12.1	rs6110505	C	0.57/0.42	<i>MACROD2</i>	1.88x10 ⁻⁶	1.87	1.44-2.41
5q23.2	rs12518621	A	0.35/0.5	<i>LINC01170</i>	3.06x10 ⁻⁶	0.55	0.43-0.71
16q22.1	16:67993643-T-C	C	0.14/0.26	<i>DPEP2</i>	4.89x10 ⁻⁶	0.46	0.33-0.64
16q22.1	exm-rs255049	C	0.14/0.26	<i>DPEP3</i>	7.08x10 ⁻⁶	0.47	0.33-0.65
6p24.2-.1	rs4711289	A	0.33/0.21	<i>AL445430*</i>	7.37x10 ⁻⁶	1.9	1.43-2.5
16q22.1	16:67977382-T-C	C	0.13/0.25	<i>DPEP3</i>	8.78x10 ⁻⁶	0.47	0.33-0.65
16q22.1	16:67833171-T-C	C	0.12/0.23	<i>CENPT</i>	9.44x10 ⁻⁶	0.45	0.32-0.64
16q22.1	exm2264618	C	0.12/0.23	<i>RANBP10</i>	9.44x10 ⁻⁶	0.45	0.32-0.64

Frecuencia de alelo: Casos/Controles, *: Cerca del gen

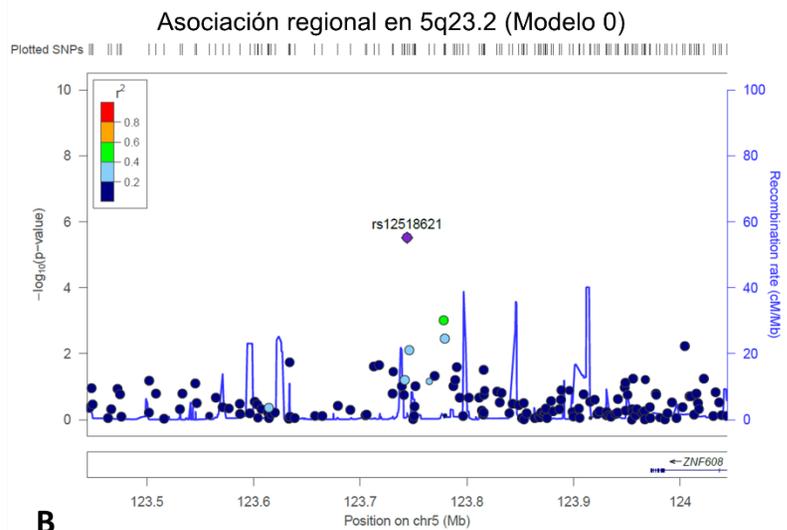
Tabla 8.5 Variantes asociadas a EPOC en grupo de estudio de acuerdo con el modelo ajustado.

Región	SNP	Alelo asociado	Frecuencia del alelo	Gen	p	OR	IC (95%)
7q21.11	rs4559136	A	0.07/0.15	<i>AGR3*</i>	8.64x10 ⁻⁷	0.30	0.12-0.41
15q26.2	JHU_15.94073449	G	0.06/0.01	<i>MCTP2*</i>	4.50x10 ⁻⁶	10.97	3.94-30.53
3q26.1	rs2901070	T	0.12/0.06	<i>SERPINI1*</i>	4.67x10 ⁻⁶	4.00	2.21-7.25
6q12	rs352523	T	0.30/0.19	<i>ADGRB3*</i>	5.49x10 ⁻⁶	2.33	1.62-3.37
7p21.1	rs4320447	C	0.08/0.15	<i>AGR3*</i>	6.95x10 ⁻⁶	0.27	0.15-0.47
6q12	rs9363933	T	0.27/0.16	<i>ADGRB3*</i>	8.97x10 ⁻⁶	2.38	1.62-3.49
1p36.11	rs17257107	G	0.16/0.08	<i>PDIK1L</i>	9.00x10 ⁻⁶	3.04	1.86-4.96
1p36.11	1:26450101-A-G	G	0.16/0.08	<i>DHDDS</i>	9.00x10 ⁻⁶	3.04	1.86-4.96

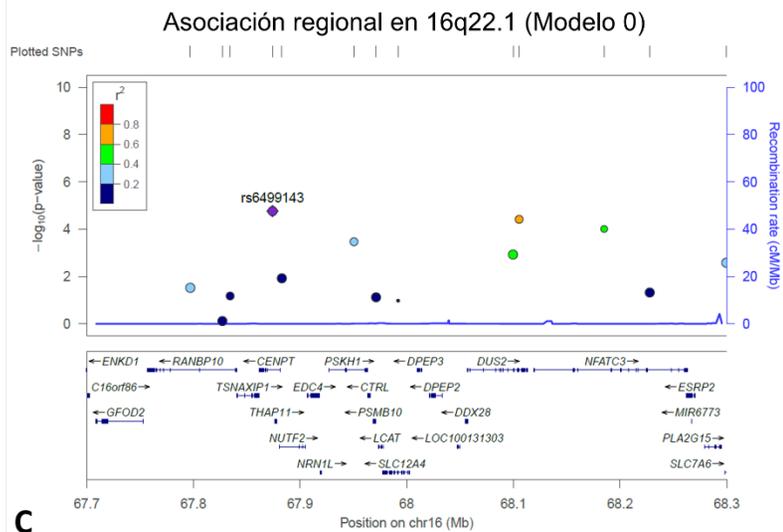
Frecuencia de alelo: Casos/Controles, *: Cerca del gen



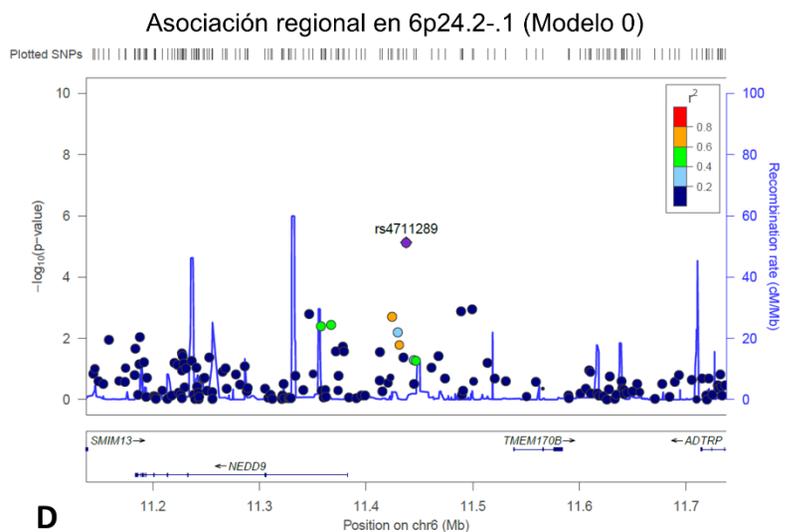
A



B

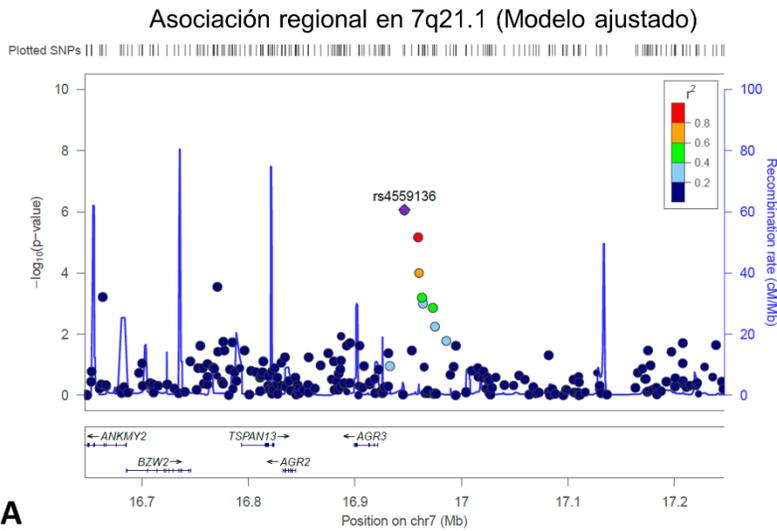


C

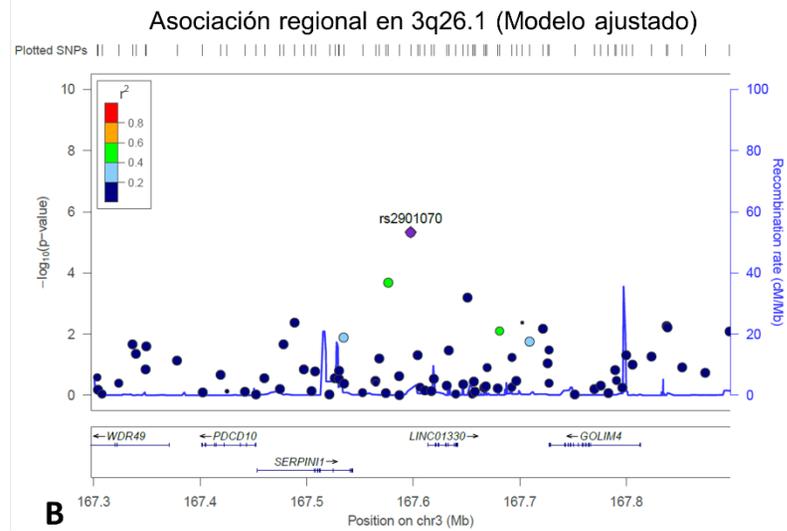


D

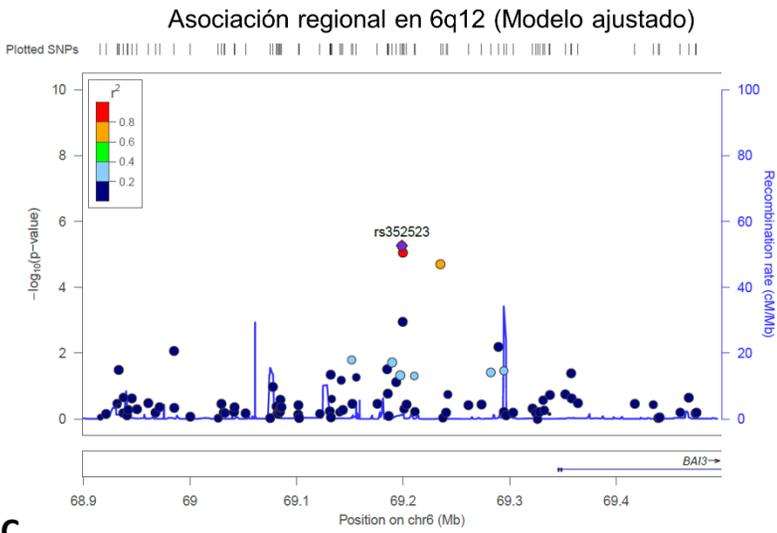
Figura 8.8 Gráfico de asociación regional en modelo sin ajustar (modelo 0). **A.** Región 20p12.1, **B.** Región 5q23.2, **C.** Región 16q22.1, **D.** Región 6p24.2-.1.



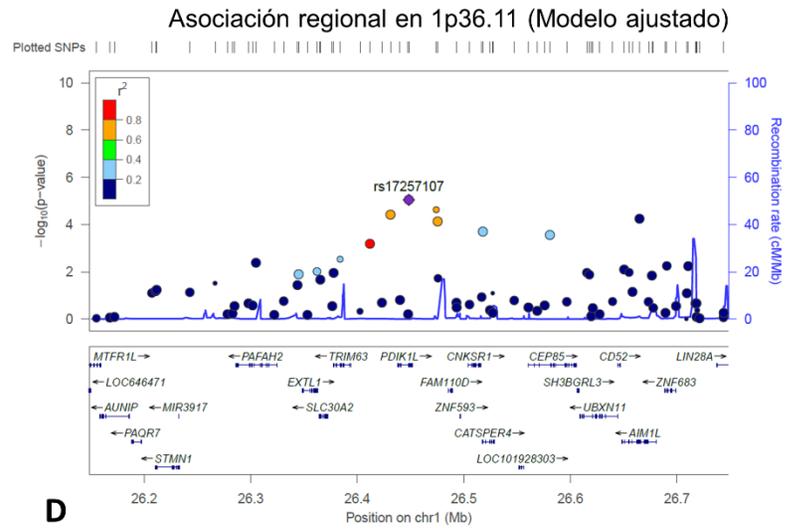
A



B



C



D

Figura 8.9. Gráfico de asociación regional en modelo ajustado. **A.** Región 7q21.11, **B.** Región 3q26.1, **C.** Región 6q12, **D.** Región 1p36.11.

8.3.2 Estratificación de variables en el modelo

8.3.2.1 Estratificación por sexo

Debido a las diferencias estadísticas respecto al sexo, se estratificaron a los grupos por sexo y se ajustó el modelo por las mismas covariables (edad, IT, estatus de tabaquismo y PC1). Las variables de ajuste en hombres se resumen la **Tabla 8.6**, en la cual la razón casos:controles fue 1:0.9. Los casos fueron significativamente de mayor edad ($p < 0.001$), por lo cual los estratos de 61-70 (41.3%) y ≥ 71 años (34.1%) contenían la mayor cantidad de sujetos, comparado con los estratos 51-60 (49%) y 61-70 años (27.5%) que contienen a más de $\frac{3}{4}$ del grupo. El estrato de 40-50 años fue menor en casos (3%) que en controles (15.4%). Respecto al tabaquismo, la proporción de fumadores actuales fue menor en casos (29.9%) que en controles (77.2%), con valores estadísticamente significativos ($p < 0.001$); así como valores mayores en IT en casos con una mediana de 44 (27.5-60), que en controles 28.9 (17.8-41.6), también con diferencias estadísticamente significativas ($p < 0.001$). Los componentes ancestrales, no fueron diferentes entre grupos, sin embargo, el componente AMI fue el mayor en proporción, con valores de mediana de 50.3% (38.2-63.1) y 48.4% (39.4-61.3) respectivamente en controles y casos, seguido del EUR y AFR, con valores en mediana 47.6% (35.9-57.2) y 3.1% (2.2-4.5) en casos y 45% (33.4-57.6) y 3.4% (2-4.7) en controles.

En mujeres, las comparaciones con las variables de ajuste se resumen en la **Tabla 8.7**, donde la relación casos:controles fue 1:4.5. La edad en el grupo de casos fue mayor, con mayor proporción el estrato de 61-70 años (53.1%), seguido de 51-60 años (28.6%) y por último ≥ 71 años (18.4%), difiriendo de los controles, en los que el estrato de 51-60 años (42.7%) fue mayor, seguido de 61-70 años (26.3%), 40-50 años (24.1%) y ≥ 71 años (8%). El grupo de controles presenta una mayor proporción de fumadores actuales (49%) comparado con los casos (80.7%), con diferencias estadísticamente significativas ($p < 0.001$). En la exposición acumulada, el grupo de casos presenta mayores valores de IT en mediana 31 (21-42) comparado con 18 (10.1-30.4) en controles, existiendo diferencias en los grupos ($p < 0.001$). Respecto a los componentes ancestrales, resultaron con significancia estadística ($p < 0.001$) EUR y AMI, mientras que para EUR la mayor proporción fue

en casos (59.1% (47.4-67.2)) comparado con controles (45.4% (34.9-60.4)), para AMI fue de manera inversa con una mayor proporción en controles (50.6% (36.2-60.4)) que en casos (37.6% (29.8-48.1)), en cambio en AFR el grupo de casos (3% (2.3-4.6)) no obtuvo diferencias con controles (3.1% (2-4.4)).

En el análisis GWAS, graficado en *Mahattan plot* (**Figura 8.10**), se obtuvieron tres variantes con significancia sugestiva de genoma completo, dos en hombres y una en mujeres (**Tabla 8.8**). En el caso de los hombres, se obtuvo un valor significativo en el alelo A de rs2762924, ubicado cerca de *CYP24A1* se asoció a menor riesgo de EPOC [OR = 0.33 (0.21-0.53), $p = 4.73 \times 10^{-6}$]. El alelo A del rs10420331 [OR = 2.62 (1.73-3.98), $p = 5.58 \times 10^{-6}$] y el alelo C del rs11084307 [OR = 2.83 (1.80-4.45), $p = 6.46 \times 10^{-6}$], ubicados en *CACNG8*, se asociaron a mayor riesgo de EPOC. En mujeres, se asociaron los alelos C de JHU_12.3109691 en *TSPAN9* y rs3825365 en *TEAD4* a menor riesgo de EPOC [OR = 0.16 (0.07-0.36), $p = 7.79 \times 10^{-6}$].

En el gráfico de asociación por regiones, respecto a los *lead* SNPs en hombres (**Figura 8.12**), se localizó el rs10420331 en la región 19q13.42 (**Figura 8.12-A**) en *CACNG8*, sin presentar desequilibrio de ligamiento. Por otro lado, en la región 20q13.2 (**Figura 8.12-B**) se encontró el rs2762924, cercano a *CYP24A1* en un bloque con un *cluster* de variantes, sin presentar desequilibrio de ligamiento. En mujeres, la variante rs3825365 se ubica en *TEAD4*, localizado en la región 12p13.33 (**Figura 8.13**), muy cercana físicamente a otra variante, pero con desequilibrio de ligamiento moderado ($r^2 = 0.6-0.8$).

Tabla 8.6 Variables de ajuste en el modelo estratificado en hombres.

Variable	Hombres		p
	Casos	Controles	
n	167	149	
Edad (años)	67 (61-73.5)	58 (52-63)	<0.001
40-50 años, n (%)	5 (3)	23 (15.4)	<0.001
51-60 años, n (%)	36 (21.6)	73 (49)	
61-70 años, n (%)	69 (41.3)	41 (27.5)	
≥ 71 años, n (%)	57 (34.1)	12 (8)	
<i>Tabaquismo</i>			
Fumador actual, n (%)	50 (29.9)	115 (77.2)	<0.001
IT (paquetes-año)	44 (27.5-60)	28.9 (17.8-41.6)	<0.001
<i>Ancestría</i>			
AMI (%)	48.4 (39.4-61.3)	50.3 (38.2-63.1)	0.423
EUR (%)	47.6 (35.9-57.2)	45 (33.4-57.6)	0.483
AFR (%)	3.1 (2.2-4.5)	3.4 (2-4.7)	0.706

Valores presentados en Mediana (RIC); n (%);

Tabla 8.7 Variables de ajuste en el modelo estratificado en mujeres.

Variable	Mujeres		p
	Casos	Controles	
n	49	171	
Edad (años)	64 (60-69)	56 (51-63)	<0.001
40-50 años, n (%)	0 (0)	41 (24.1)	<0.001
51-60 años, n (%)	14 (28.6)	73 (42.7)	
61-70 años, n (%)	26 (53.1)	45 (26.3)	
≥ 71 años, n (%)	9 (18.4)	12 (8)	
<i>Tabaquismo</i>			
Fumador actual, n (%)	24 (49)	138 (80.7)	<0.001
IT (paquetes-año)	31 (21-42)	18 (10.1-30.4)	<0.001
<i>Ancestría</i>			
AMI (%)	37.6 (29.8-48.1)	50.6 (36.2-60.4)	<0.001
EUR (%)	59.1 (47.4-67.2)	45.4 (34.9-60.4)	<0.001
AFR (%)	3 (2.3-4.6)	3.1 (2-4.4)	0.545

Valores presentados en Mediana (RIC); n (%);

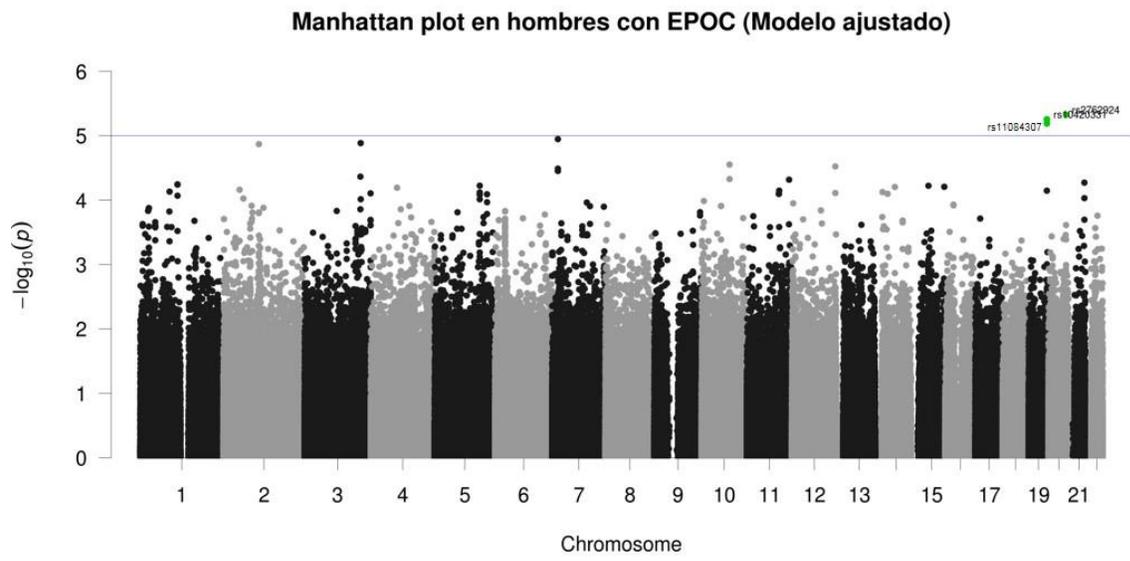


Figura 8.10 Manhattan plot del modelo ajustado estratificado en hombres.

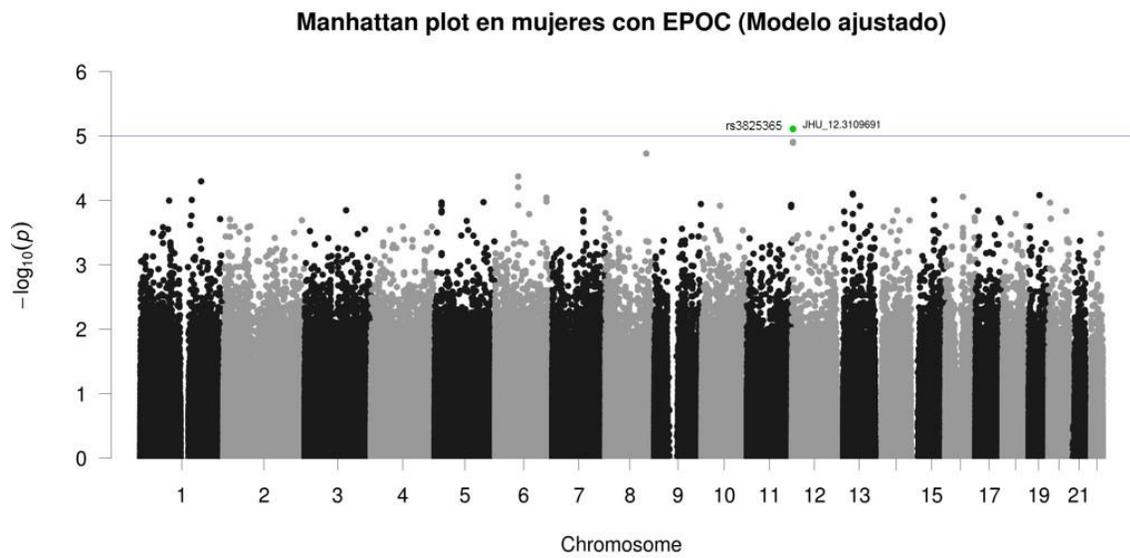
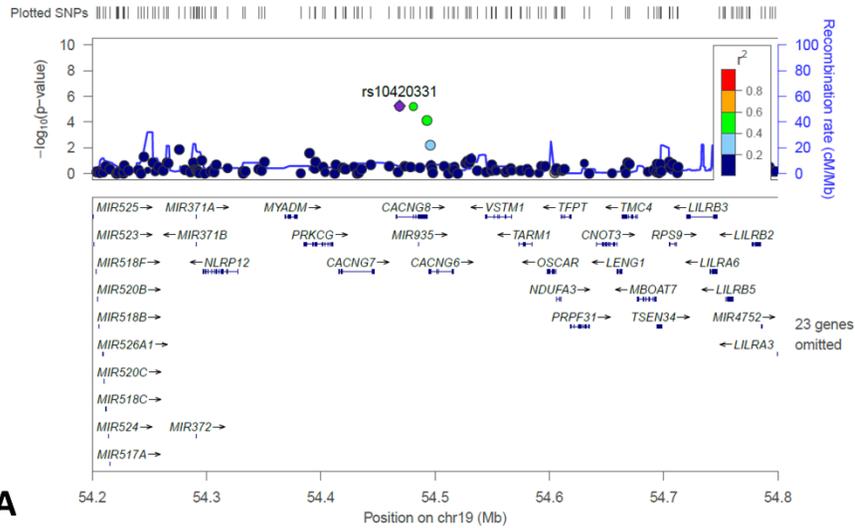


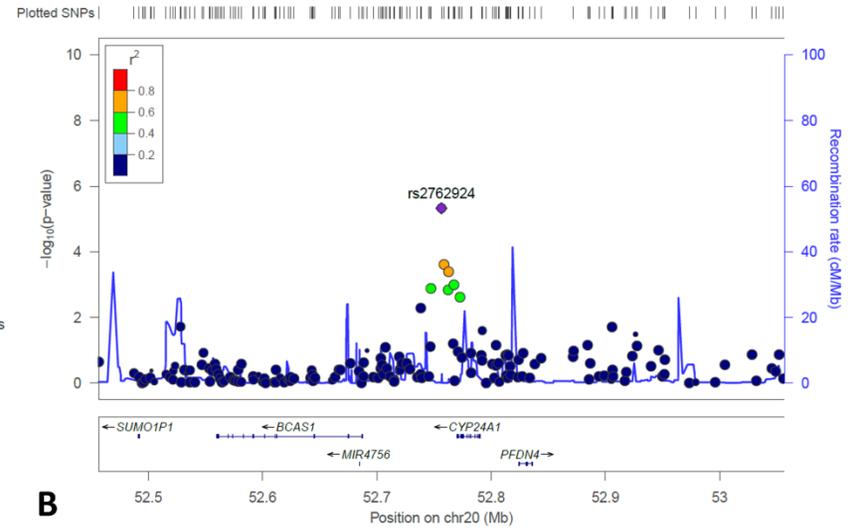
Figura 8.11 Manhattan plot del modelo ajustado estratificado en mujeres.

Asociación regional en 19q13.42 (Hombres)



A

Asociación regional en 20q13.2 (Hombres)



B

Figura 8.12 Gráfico de asociación regional en modelo ajustado estratificado en hombres. **A.** Región 13q13.42, **B.** Región 20q13.2.

Asociación regional en 12p13.33 (Mujeres)

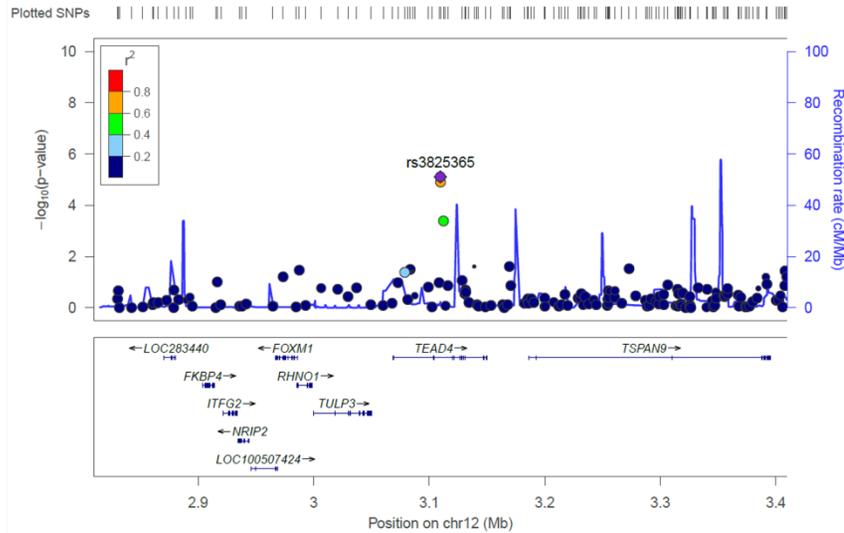


Figura 8.13 Gráfico de asociación regional en modelo ajustado estratificado en mujeres.

Tabla 8.8 Variantes asociadas a EPOC de acuerdo con el modelo ajustado estratificado por sexo.

Región	SNP	Alelo asociado	Frecuencia del alelo	Gen	p	OR	IC (95%)
Hombres							
20q13.2	rs2762924	A	0.26/0.32	<i>CYP24A1*</i>	4.73x10 ⁻⁶	0.33	0.21-0.53
19q13.42	rs10420331	A	0.43/0.33	<i>CACNG8</i>	5.58x10 ⁻⁶	2.62	1.73-3.98
19q13.42	rs11084307	C	0.32/0.23	<i>CACNG8</i>	6.46x10 ⁻⁶	2.83	1.80-4.45
Mujeres							
12p13.33	JHU_12.3109691	C	0.38/0.46	<i>TSPAN9</i>	7.79x10 ⁻⁶	0.16	0.07-0.36
12p13.33	rs3825365	C	0.38/0.47	<i>TEAD4</i>	7.79x10 ⁻⁶	0.16	0.07-0.36

Frecuencia de alelo: Casos/Controles, *: Cerca del gen

8.3.2.2 Estratificación por estatus de tabaquismo

Se estratificaron los grupos de acuerdo con el estatus de tabaquismo en fumadores actuales y exfumadores. En los fumadores actuales, se resumen las variables de ajuste en la **Tabla 8.9**. En los grupos, se presentó una razón casos:controles 1:34. Respecto a la edad, existieron diferencias significativas ($p < 0.001$), en donde el grupo de casos fue mayor, siendo los estratos de 61-70 (41.9%) y 51-60 años (35.1%) los que comprenden la mayoría de los sujetos y menormente los de ≥ 71 (17.6%) y 40-50 años (5.4%). Por su parte, los controles que presentaron mayoría en el estrato de 51-60 años (47.4%) y reduciéndose en 61-70 (26.1%), 40-50 (21.3%) y ≥ 71 años (5.1%). Respecto al sexo, el grupo de casos presentó más hombres (67.6%) y el de controles más mujeres (54.6%), existiendo diferencias significativas ($p < 0.001$). En cuanto al tabaquismo evaluado por IT, los casos presentaron una mediana de 40.7 paquetes-año (29.2-52.5) mayor a los controles con una mediana de 24.6 paquetes-año (13.5-38), que fueron significativas ($p < 0.001$). Existieron diferencias respecto a los componentes AMI ($p = 0.022$) y EUR ($p = 0.021$), en los controles el componente AMI fue mayor (49.7 (36.2-60.4)) que en casos [44.1 (31.3-53.2)], e inversamente, el EUR fue mayor en casos [51.8 (42.3-64.4)] que en controles [46.7 (35.7-60.5)], sin embargo, en el AFR se presentó muy similar en casos [3.1 (2.3-4.7)] y controles [3.2 (2-4.5)].

Por otro lado, en el análisis en exfumadores (**Tabla 8.10**), resultó una razón casos:controles de 1:0.5. El grupo de casos tuvo mayor edad con diferencias significativas ($p < 0.001$), en las que en el estrato de 61-70 (45.1%) y ≥ 71 (37.3%) concentra la mayoría de sujetos, disminuyendo en el de 51-60 (16.9%) y 40-50 años (0.7%); en cambio, en controles el estrato que albergan la mayor proporción son el de 51-60 (38.8%), seguido de 61-70 (29.8%), ≥ 71 (16.4%) y 40-50 (14.9%). Existieron diferencias de acuerdo con el sexo ($p < 0.001$), mientras que en los casos fue una proporción de 82.4% hombres, para controles fue 50.7%. Acorde con el historial acumulado de tabaquismo en IT existieron diferencias significativas entre los grupos ($p < 0.001$), siendo mayor en los casos (40 (23.6-57)), comparado con controles (18.5 (9-30.5)). Respecto al componente ancestral, se presentan diferencias significativas: en la comparación del componente AMI ($p < 0.017$), se

presenta aumentado en controles (56.4 (41.9-68.5)) que en casos (47.6 (37.3-60.6)); en cambio en el EUR ocurre lo contrario, se encuentra disminuido en controles (40.9 (28.1-53.8)) y aumentado en casos (47.8 (35.9-58.5)); el componente AFR fue muy similar en ambos grupos, 3.1 (2.2-4.3) en casos y 3.1 (1.6-4.2) en controles.

En el GWAS, se obtuvieron dos variables con significancia estadística (**Tabla 8.11**), una en la comparación de fumadores actuales y otra en exfumadores, graficadas en *Mahattan plot* en **Figura 8.14** y **Figura 8.15**. En el grupo de fumadores actuales, se encontró el alelo C de JHU_19.57835849, localizado en *ZNF587B*, asociado a mayor riesgo de EPOC [OR = 3.15 (1.93-5.15), $p = 4.29 \times 10^{-6}$]. Por otro lado, en exfumadores el alelo C de rs7644995 cercano a *RTP3* se asoció a menor riesgo de EPOC [OR = 0.17 (0.08-0.37), $p = 8.82 \times 10^{-6}$].

De acuerdo al análisis de asociación regional en fumadores actuales, no se pudo analizar la variante asociada JHU_19.57835849 en la región cromosómica 19q13.42 (**Figura 8.16**), sin embargo, al graficar la región donde se encontraría, existe un bloque de aproximadamente 400 Kb en el cual el *lead* SNP fue rs129720 en *ZNF671*, que presenta alto desequilibrio de ligamiento ($r^2 \geq 0.8$) con dos variantes, una en *ZNF552* y *ZNF587*, gen en el que se ubica JHU_19.57835849 (**Figura 8.16-A**). En la asociación regional en exfumadores, se encontró el *lead* SNP rs7644995 en la región 3p21.31, entre los genes *LTF* y *RTP3*, sin embargo, no presentó desequilibrio de ligamiento con alguna otra variante (**Figura 8.16-B**).

Tabla 8.9 Variables de ajuste en el modelo estratificado en fumadores actuales.

Variable	Fumadores actuales		p
	Casos	Controles	
n	74	253	
Edad (años)	63 (57.2-67)	56 (51-62)	<0.001
40-50 años, n (%)	4 (5.4)	54 (21.3)	<0.001
51-60 años, n (%)	26 (35.1)	120 (47.4)	
61-70 años, n (%)	31 (41.9)	66 (26.1)	
≥ 71 años, n (%)	13 (17.6)	13 (5.1)	
Sexo masculino, n (%)	50 (67.6)	115 (45.4)	<0.001
<i>Tabaquismo</i>			
IT (paquetes-año)	40.7 (29.2-52.5)	24.6 (13.5-38)	<0.001
<i>Ancestría</i>			
AMI (%)	44.1 (31.3-53.2)	49.7 (36.2-60.4)	0.022
EUR (%)	51.8 (42.3-64.4)	46.7 (35.7-60.5)	0.021
AFR (%)	3.1 (2.3-4.7)	3.2 (2-4.5)	0.555

Valores presentados en Mediana (RIC); n (%);

Tabla 8.10 Variables de ajuste en el modelo estratificado en exfumadores.

Variable	Exfumadores		p
	Casos	Controles	
n	142	67	
Edad (años)	68 (63-74)	59 (54-67)	<0.001
40-50 años, n (%)	1 (0.7)	10 (14.9)	<0.001
51-60 años, n (%)	24 (16.9)	26 (38.8)	
61-70 años, n (%)	64 (45.1)	20 (29.8)	
≥ 71 años, n (%)	53 (37.3)	11 (16.4)	
Sexo masculino, n (%)	117 (82.4)	34 (50.7)	<0.001
<i>Tabaquismo</i>			
IT (paquetes-año)	40 (23.6-57)	18.5 (9-30.5)	<0.001
<i>Ancestría</i>			
AMI (%)	47.6 (37.3-60.6)	56.4 (41.9-68.5)	0.017
EUR (%)	47.8 (35.9-58.5)	40.9 (28.1-53.8)	0.023
AFR (%)	3.1 (2.2-4.3)	3.1 (1.6-4.2)	0.276

Valores presentados en Mediana (RIC); n (%);

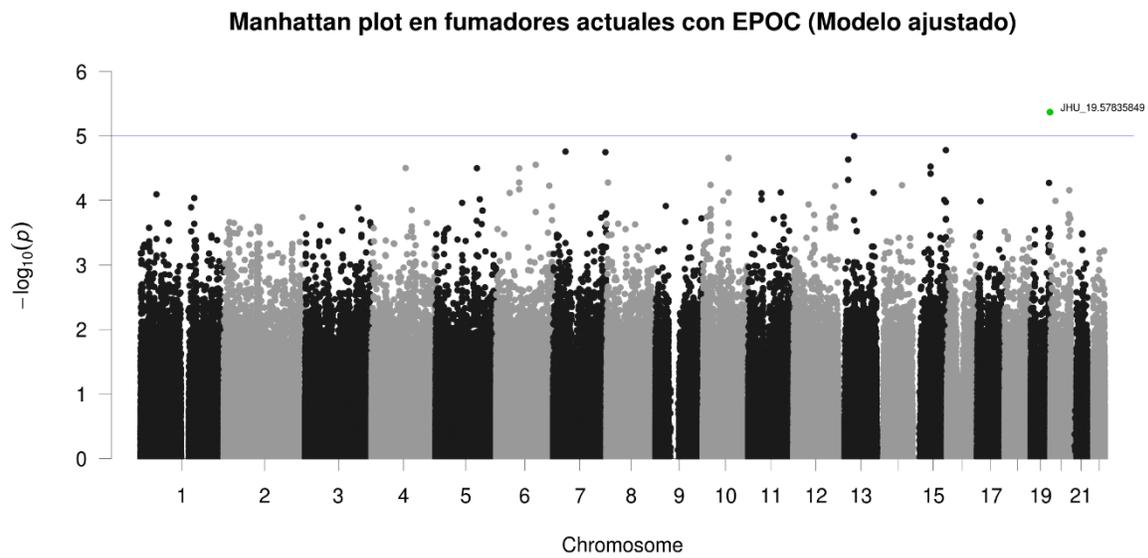


Figura 8.14 Manhattan plot del modelo ajustado estratificado en fumadores actuales.

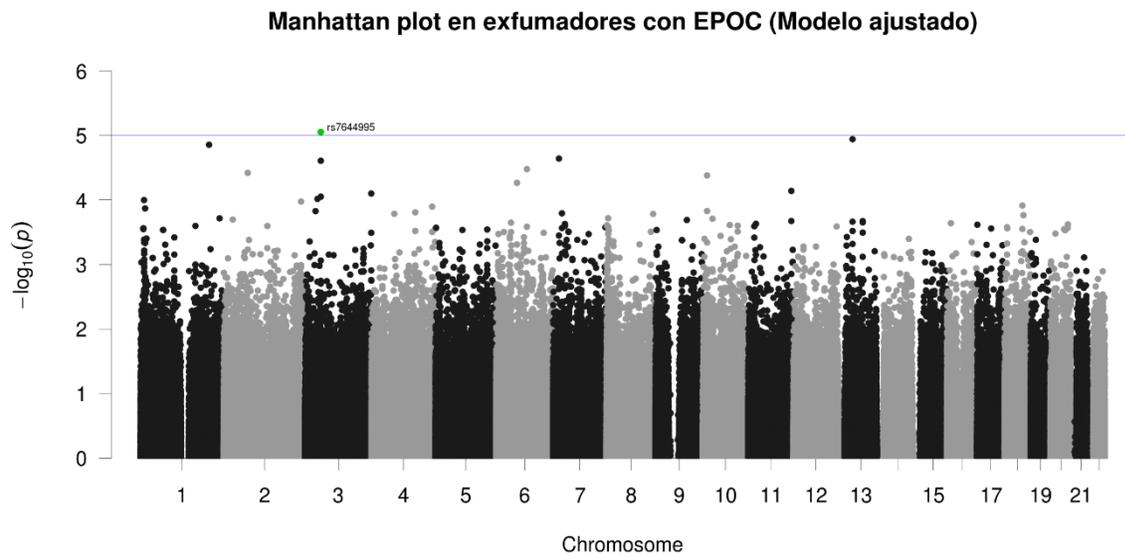


Figura 8.15 Manhattan plot del modelo ajustado estratificado en exfumadores.

Tabla 8.11 Variantes asociadas a EPOC de acuerdo al modelo ajustado estratificado por estatus de tabaquismo.

Región	SNP	Alelo asociado	Frecuencia del alelo	Gen	p	OR	IC (95%)
Fumadores actuales							
19q13.43	JHU_19.57835849	C	0.28/0.21	ZNF587B	4.29×10^{-6}	3.15	1.93-5.15
Exfumadores							
3p21.31	rs7644995	C	0.11/0.14	RTP3*	8.82×10^{-6}	0.17	0.08-0.37

Frecuencia de alelo: Casos/Controles, *: Cerca del gen

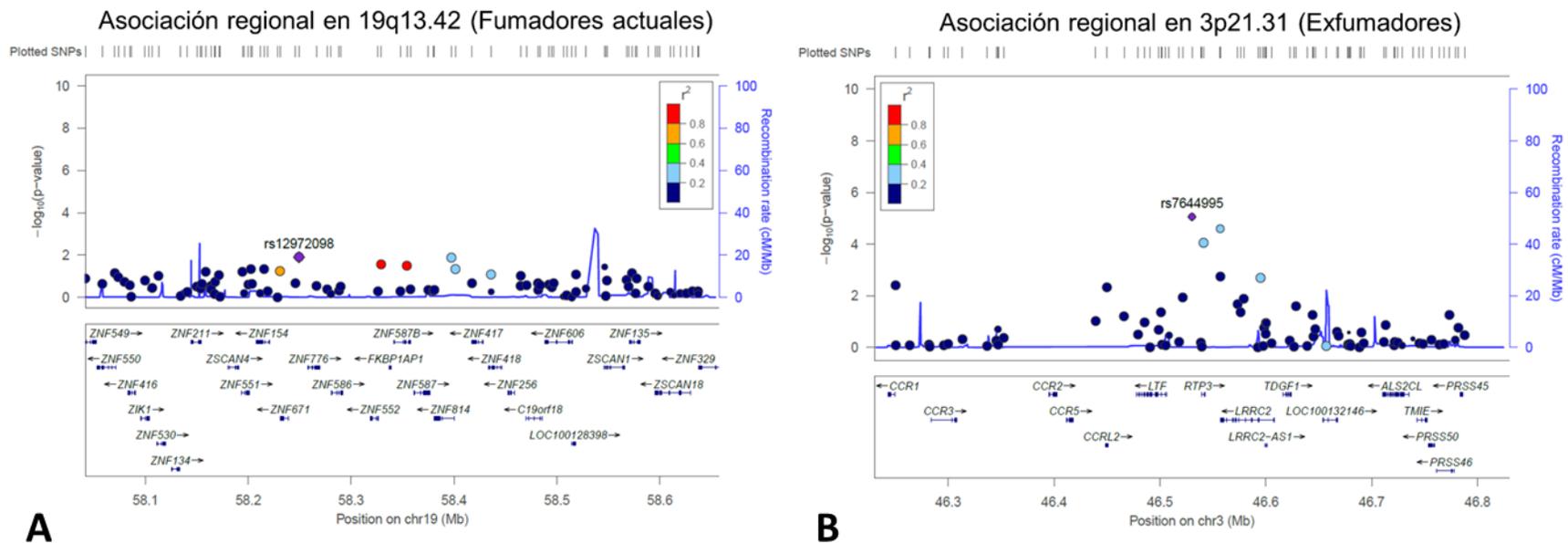


Figura 8.16 Gráfico de asociación regional en modelo ajustado estratificado por estatus de tabaquismo. **A.** Fumadores actuales en región 19p13.42, **B.** Exfumadores en región 3p21.31.

8.3.3 Regiones cromosómicas

Las variantes que se encontraron asociadas a EPOC se encuentran en los cromosomas 1, 3, 5, 6, 7, 12, 15, 16, 19 y 20 (**Figura 8.17**). En el *circleplot* se puede observar que la región con más variantes asociadas corresponde a 16q22.1 con 5 por el modelo sin ajustar. Posteriormente, la región comprendida entre 19q13.42-.43 presenta 2 variantes asociadas entre el modelo estratificado en hombres y en fumadores actuales. Las regiones 1p36, 6q12 y 12q13.33 resultaron con 2 hallazgos cada una de acuerdo al modelo ajustado y estratificado en mujeres. Las demás regiones presentaron únicamente una variante asociada, siendo en el modelo ajustado 3q26.1, 7p21.1, 7p21.11 y 12p13.33; en el modelo sin ajustar 5q23.2, 6p24.2-.1 y 20p12.1; y en los modelos estratificados en exfumadores y hombres, en 3p21.31 y 20q13.2, respectivamente.

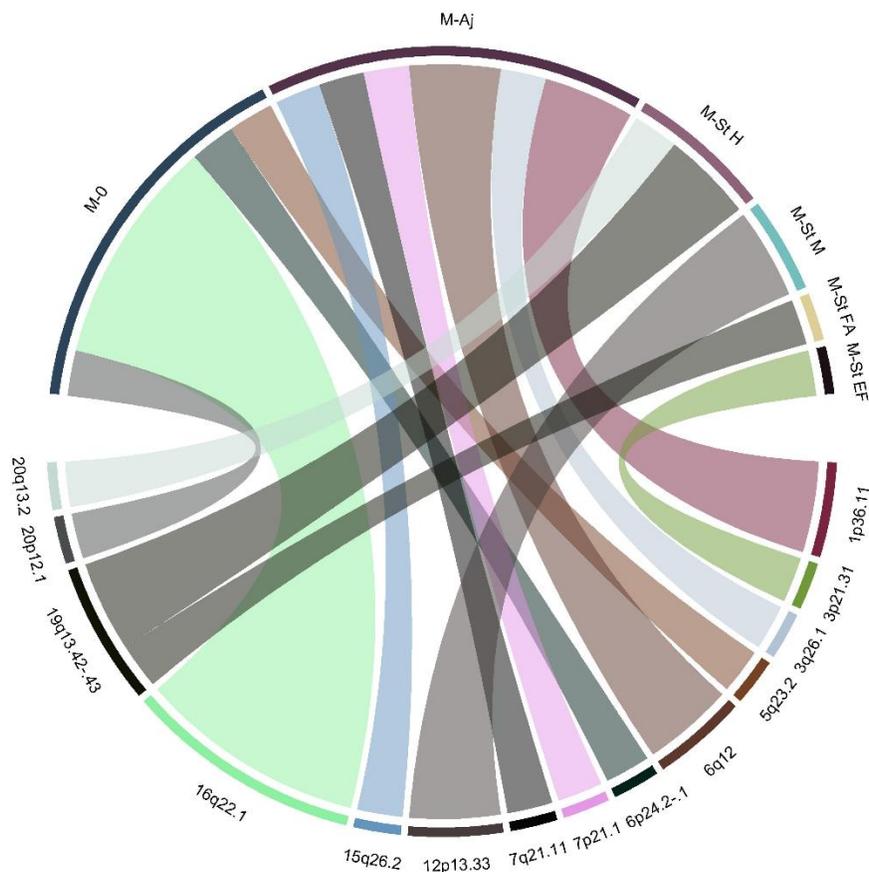


Figura 8.17 Circle plot de las regiones cromosómicas de los SNP asociados de acuerdo al modelo de regresión realizado. M-: Modelo de regresión, 0: Sin ajuste, Aj: Ajustado, St: Estratificado, H: Hombres, M: Mujeres, FA: Fumador actual, EF: Exfumador.

8.4 Relación de variantes con el componente amerindio

Se compararon las medias de ancestría AMI en individuos que tuvieran la variante asociada entre casos y controles, de acuerdo con el modelo estadístico en que se haya encontrado, para después estimar la magnitud del efecto de esa variante respecto al componente AMI, es decir, si la variante se direcciona a mayor/menor componente AMI en casos o en controles.

En el modelo sin ajustar, se describen las variantes ordenadas de acuerdo con la región cromosómica en la **Tabla 8.12**, todas las variantes resultaron cumplir una distribución aproximadamente normal respecto al componente AMI y la media en controles fue mayor. En el alelo C de rs12518621 (**Figura 8.18-A**), los casos tuvieron una media de componente AMI (0.439) significativamente ($p = 0.046$)

menor a la de los controles (0.469), con una magnitud media del efecto muy pequeña ($d = 0.167$); en las demás variantes (**Figura 8.18-B-H**) no se obtuvieron diferencias estadísticamente significativas, a pesar de que en controles la media del componente AMI era mayor.

Respecto a las variantes asociadas por medio del modelo ajustado (**Tabla 8.13**), todas presentan una distribución aproximadamente normal. Respecto al alelo T de rs9363933 (**Figura 8.19-F**), existieron diferencias estadísticamente significativas ($p = 0.016$) en la media del componente AMI, en el cual fue mayor en controles (0.496) que en casos (0.464), presentando una magnitud media del efecto de 0.169; en las demás variantes (**Figura 8.19-A,C-E,G-H**) no existieron diferencias estadísticamente significativas, a pesar de que la media del componente AMI fue mayor en controles que en casos, exceptuando al alelo G de JHU_15.94073449, en el cual fue menor el componente AMI en controles (0.392) que en casos (0.428) (**Figura 8.19-B**).

Las variantes asociadas en el modelo ajustado estratificado tanto por sexo y por estatus de tabaquismo de acuerdo al componente AMI (**Anexo XXI**) se compararon entre casos y controles, los cuales cumplieron con una distribución aproximadamente normal y los alelos asociados de controles tuvieron una frecuencia mayor en componente AMI que en casos tanto en sexo (**Anexo XXII**) como en estatus de tabaquismo (**Anexo XXII**), sin llegar a ser significativas.

Tabla 8.12 Variantes asociadas en el modelo sin ajustar y su relación respecto al componente AMI en casos y controles.

SNP	Alelo asociado	p de Normalidad*	Media AMI en alelo asociado (casos/controles)	p de diferencia de medias [§]	Magnitud media del efecto
rs12518621	C	0.157	0.439/0.469	0.046	0.167
rs4711289	A	0.371	0.473/0.496	0.303	0.129
exm2264618	C	0.644	0.482/0.495	0.698	0.063
16:67833171-T-C	C	0.644	0.482/0.495	0.698	0.063
16:67977382-T-C	C	0.645	0.470/0.492	0.466	0.101
16:67993643-T-C	C	0.520	0.469/0.489	0.478	0.105
exm-rs255049	C	0.457	0.462/0.491	0.304	0.154
rs6110505	C	0.208	0.462/0.464	0.923	0.008

*: Prueba de Shapiro-Wilk, §: T-Student para muestras independientes

Tabla 8.13 Variantes asociadas en el modelo ajustado y su relación respecto al componente AMI en casos y controles.

SNP	Alelo asociado	p de Normalidad*	Media AMI en alelo asociado (casos/controles)	p de diferencia entre medias [§]	Magnitud media del efecto
rs4559136	A	0.666	0.463/0.521	0.135	0.178
JHU_15.94073449	G	0.913	0.428/0.392	0.606	0.178
rs2901070	T	0.141	0.527/0.576	0.187	0.023
rs352523	T	0.769	0.484/0.517	0.145	0.284
rs4320447	C	0.555	0.450/0.520	0.068	0.188
rs9363933	T	0.166	0.464/0.496	0.016	0.169
rs17257107	G	0.126	0.425/0.456	0.365	0.317
1:26450101-A-G	G	0.126	0.425/0.456	0.365	0.382

*: Prueba de Shapiro-Wilk, §: T-Student para muestras independientes

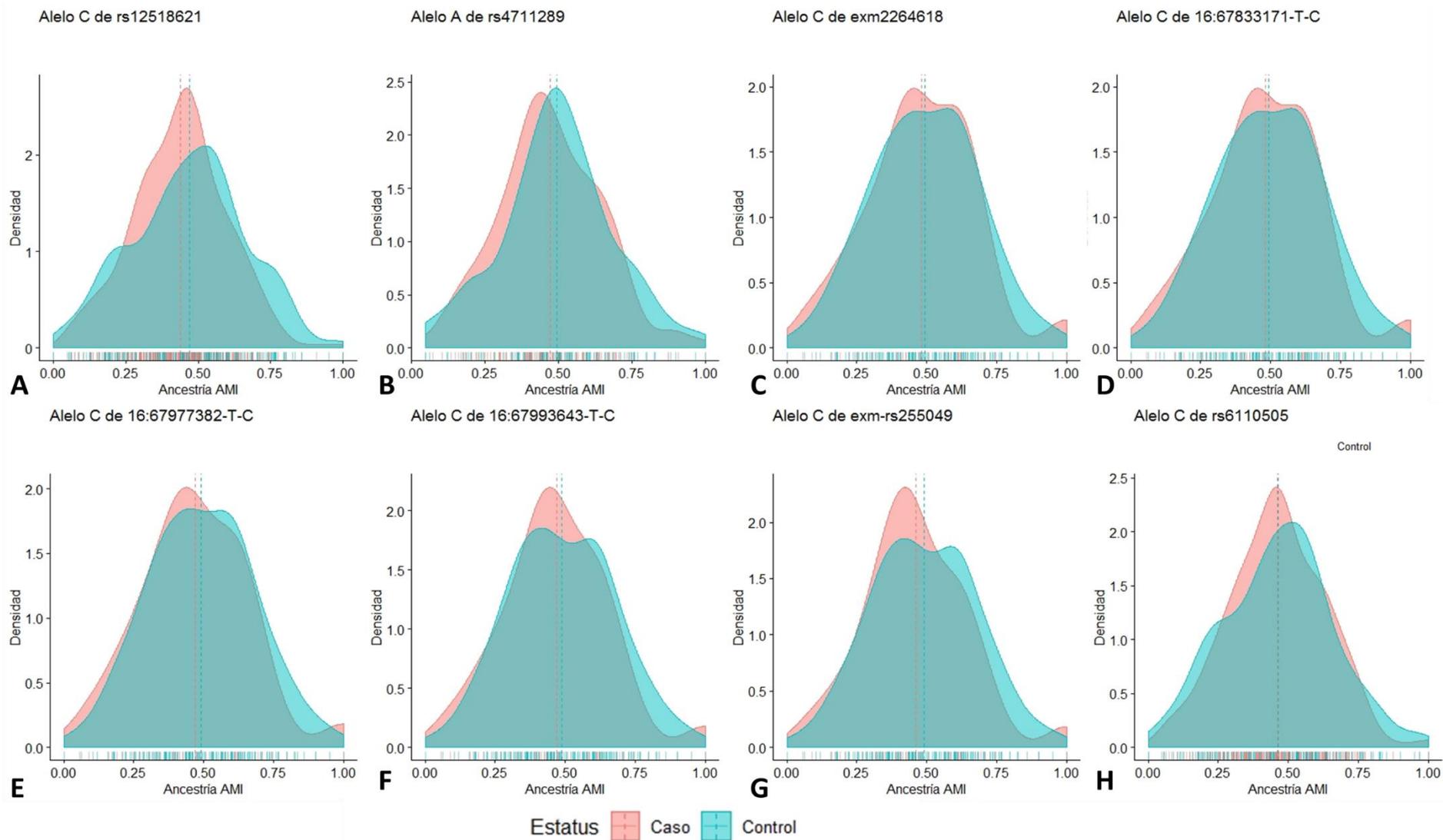


Figura 8.18 Distribución de los alelos asociados en el modelo sin ajustar respecto a la ancestría amerindia. **A.** rs12518621, **B.** rs4711289, **C.** exm2264618, **D.** 16:67833171-T-C, **E.** 16:67977382-T-C, **F.** 16:67993643-T-C, **G.** exm-rs255049, **H.** rs6110505.

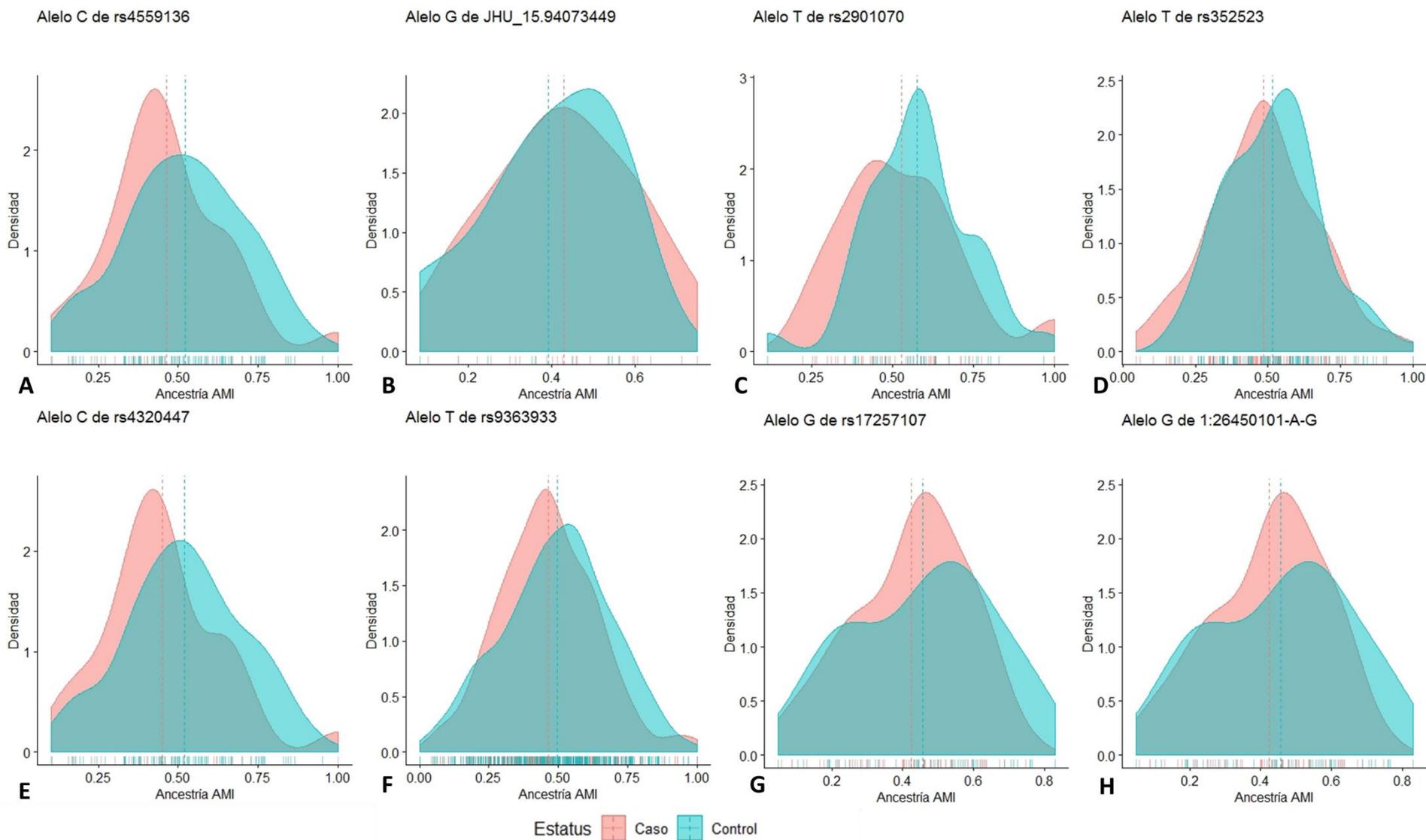


Figura 8.19 Distribución de los alelos asociados en el modelo ajustado respecto a la ancestría amerindia. **A.** rs4559136, **B.** JHU_15.94073449, **C.** rs2901070, **D.** rs352523, **E.** rs4320447, **F.** rs9363933, **G.** rs17257107, **H.** 1:26450101-A-G.

9 Discusión

El presente trabajo, según nuestro conocimiento, es el primero que evalúa susceptibilidad genética a EPOC en población en México mediante la estrategia de GWAS empleando un microarreglo enriquecido con variantes americanas.

Se pudieron evaluar 738,822 variantes genéticas de tipo SNP que, por definición, estuvieran con una frecuencia $\geq 1\%$ en la población, de las cuales 23 se asociaron a EPOC por significancia sugestiva de genoma completo ($p < 10 \times 10^{-5}$). De estas, 17 muestran asociación que sugiere cierta plausibilidad biológica con la EPOC y solamente 2 se asocian al componente AMI. Estas variantes, denominadas comunes, presumen tener un efecto discreto en las enfermedades comunes y complejas como la EPOC, que adicionalmente con la interacción de variables ambientales, aumentan el riesgo para desarrollar en algún momento la enfermedad.

209,210

A pesar de que existe la discusión en cuanto a si los GWAS deberían tener o no una hipótesis *a priori*,²¹¹ nuestro trabajo se enfocó en identificar factores riesgo a EPOC, con el fin de dar una respuesta biológica al fenómeno epidemiológico de la paradoja hispana, a pesar de que la mayoría de estudios se enfocan en encontrar variantes de mayor riesgo para la enfermedad. Sin embargo, no descartamos que existan variantes de mayor riesgo en nuestra población, las cuales posiblemente sean mayormente de contribución europea, como la deficiencia de A1AT.

En este estudio, se incluyeron fumadores y exfumadores con consumo diario de cigarros, sin embargo, estos difieren con los de la encuesta nacional de tabaquismo en México (ENCODAT).⁵⁴ Mientras que en la ENCODAT los consumidores diarios de cigarros reportan un promedio de 7.4, en nuestro estudio reportamos una mediana de 20 en sujetos con EPOC y 14 en fumadores sin la enfermedad; en el mismo sentido, reportamos una mediana de 17 años en edad de inicio en ambos grupos, comparado con el reporte de 19.3 en promedio. Por otro lado, se ha reportado que solamente el 5% de los fumadores diarios en México, consume más de 20 cigarros al día.²¹² Debido a estas diferencias, se podría sospechar que los

sujetos incluidos poseen un nivel de dependencia mayor al reportado en la población general, ya que se ha reportado que los fumadores diarios reconocen tener un problema de autocontrol en su consumo, en comparación de fumadores ocasionales.²¹³

Asociaciones por modelo sin ajustar

Respecto al modelo sin ajustar por variables, 8 señales asociadas a EPOC con significancia sugestiva, las cuales en caso de no haber tenido diferencias significativas entre los grupos de comparación, podrían ser más fiables, por lo que se argumentará que su asociación no haya sido espuria mediante la revisión de la literatura.

La variante con mayor valor de significancia fue rs6110505 ($p = 1.88 \times 10^{-6}$) la cual no ha sido asociadas previamente a EPOC, sin embargo, en *MACROD2* (*Mono-ADP Ribosylhydrolase 2*), gen en el que se localiza, se han reportado algunas asociaciones con la EPOC y la función pulmonar. En un GWAS del estudio SpiroMeta, la variante rs204652 en *MACROD2* se asoció a disminución de VEF1 analizado en todos los participantes ($p = 6.81 \times 10^{-5}$, $\beta = -0.187$) y empeorando en fumadores ($p = 1.67 \times 10^{-5}$, $\beta = -0.251$).¹⁷² Por medio de NGS (*Next Generation Sequencing*) se ha asociado rs41275442 a mayor riesgo (OR = 2.1 (1.0-4.5)) a padecer EPOC.²¹⁴ *MACROD2* codifica para una desacetilasa, que remueve las ADP-ribosas de residuos de Aspartato y Glutamato en proteínas que contienen un solo residuo ADP-ribosa.²¹⁵ Se ha encontrado en la mayoría de tejidos, sin embargo, el transcrito muestra una expresión aumentada en cerebro, órganos sexuales masculinos y pulmón²¹⁶, además también se ha asociado a autismo.^{217,218}

La siguiente variante con mayor valor de significancia fue el rs12518621, localizado en *LINC01170* (*Long intergenic non-protein coding RNA 1170*), el cual no se ha asociado a ninguna patología respiratoria; sin embargo, en pacientes con carcinoma endometrial, se encuentra sobrepresado.²¹⁹ En la región cromosómica 16q22.1 resultaron 5 variantes asociadas en 4 distintos genes. Las variantes exm-rs255049 y 16:67977382-T-C en *DPEP3* y 16:67993643-T-C en *DPEP2* no se han reportado en patologías respiratorias, no obstante, los genes *DPEP* (*Dipeptidase*) presentan

otras variantes asociadas a distintos fenotipos, como *DPEP3* asociado a sobreexpresión en células de carcinoma escamoso ²²⁰, así como en el aumento de niveles de HDL-C en conjunto con *DPEP2*, ²²¹ gen que se ha reportado subexpresado en células sanguíneas en la producción de IL-6 ²²², así como sobreexpresado en la presencia de infecciones virales como miocarditis inducida por virus Coxsackie B3 ²²³; los productos de estos dos genes (*DPEP3* y *DPEP2*) pertenecen a la familia de dipeptidasas unidas a membrana, enzimas con la función de hidrolizar diversos dipéptidos como leucotrienos, el anillo betalactámico de algunos antibióticos, entre otros. ²²⁴ Dentro de esta región también se asoció la variante 16:67833171-T-C en *CENPT* (*Centromere protein T*), factor necesario en el ensamblado del centrómero en la mitosis, el cual participa en eventos de *splicing* alternativo en células de cáncer de pulmón metastásico, ²²⁵ se encontró expresión alterada en el desarrollo pulmonar intrauterino, afectado por exposición a alérgenos, como ovoalbúmina en modelo murino. ²²⁶ Además de estos, la variante exm2264618 en *RANBP10* (*RAN Binding Protein 10*) también se encontró asociada en esta región cromosómica, este gen se ha evidenciado como supresor de tumores ²²⁷, el cual se ha reportado como blanco de miR-361, afectando la supervivencia en cáncer de pulmón. ²²⁸ Por último, la variante rs4711289 cercana a *AL445430*, no ha tenido previas asociaciones y se presume que el gen no codifica ningún transcrito. Sin embargo, permanece la interrogante del papel de las variantes en los genes encontrados en la región 16q22.1 y su papel en el desarrollo de la EPOC de acuerdo a su función, en la inflamación, exacerbaciones, desarrollo pulmonar o posiblemente en la respuesta a fármacos.

Asociaciones por modelo ajustado

Posteriormente, el modelo se ajustó por las variables de edad, sexo, estatus de tabaquismo, IT y PC1, con base en que fueron las más empleadas en otros GWAS en donde la variable dependiente era la EPOC. Estas variables fueron elegidas con el objetivo de incrementar la validez interna del estudio. ²²⁹ Además, existieron diferencias importantes en estas variables ajustadas en el modelo entre los grupos. Primeramente, la edad juega un papel determinante en la EPOC, ya que la función pulmonar disminuye naturalmente con la edad, debido a cambios estructurales y

fisiológicos en el pulmón. ⁶ Con base en esto, no incluimos sujetos mayores a 80 años, ya que por lo general tienden a ser los casos más graves. ¹⁶⁸ Sin embargo, a pesar de ello, el grupo de casos tuvo una edad más avanzada que los controles. Con el objetivo de controlar mejor la edad, y debido a la correlación positiva que se obtuvo con el tabaquismo medido en IT, se categorizó en estratos de 10 años. El IT mide la exposición acumulativa de tabaquismo, el cual muestra una correlación negativa respecto a la obstrucción pulmonar ²³⁰, hallazgo encontrado en los participantes del estudio. Por otro lado, la variable de años fumando de manera continua resultó no ser un buen predictor de la disminución de la función pulmonar, a diferencia de lo propuesto por Bhatt et al., 2018. ²³¹

El rs4559136 fue el SNP asociado a EPOC con mayor valor de significancia, el cual también mostró alto desequilibrio de ligamiento ($r^2 = 0.8$) con el rs4320447, ubicados en la región 7p21, cerca de *AGR3*. El gen *AGR3* (*Anterior gradient protein 3*) codifica para la proteína de igual nombre perteneciente a la familia *Anterior gradient*, que incluye a *AGR2*. Las *AGR* pertenecen a la superfamilia de proteínas disulfuro isomerasas, las cuales contienen el dominio del núcleo de tiorredoxina, catalizando actividades como redox de grupos tiol, intercambio de disulfuro y como chaperonas en el retículo endoplasmático. ²³² *AGR3* se encuentra expresada en tejidos pulmonares (pulmón y bronquios) ²³³ Debido a que el fenotipo evaluado en este estudio fue la EPOC, el interés de la plausibilidad biológica para el hallazgo es con base en la fisiología respiratoria. Las proteínas *AGR* se han relacionado estrechamente con las mucinas, debido a su función biológica de disulfuro isomerasa en el retículo endoplasmático de células productoras de mucina. ²³⁴ Las mucinas (especialmente *MUC5AC* y *MUC5B*) son de los principales componentes del moco en las vías aéreas debido a sus propiedades reológicas. ²³⁵ Se ha reportado que una de las funciones de *AGR3* y *AGR2* es contribuir en la regulación de las células epiteliales multiciliadas en la vía aérea. ²³⁶ En la EPOC, la hipersecreción de moco es un factor de mal pronóstico, asociada a mayor número de hospitalizaciones y exacerbaciones. ²³⁷ Respecto a esto, *AGR3* se ha encontrado alterada en muestras de tejido pulmonar en pacientes con EPOC, más abundantemente en pacientes con exacerbaciones infrecuentes, que con

frecuentes, lo que podría sugerir que se encuentra relacionada con un proceso infeccioso temprano, quizá con menor severidad a la enfermedad.²³⁸ Las funciones fisiológicas de estas proteínas también se han evidenciado en la regulación y diferenciación celular, principalmente asociado al desarrollo embrionario y carcinogénesis.^{239,240} Respecto al desarrollo embrionario, se ha reportado que la expresión de *AGR3* se encuentra regulada por la vía de señalización Wnt, siendo diferenciada en la etapa pseudoglandular y canalicular del desarrollo embrionario del pulmón.²⁴¹ En relación a la función que desempeña en la carcinogénesis, se han reportado 231 variantes de *AGR3* en el proyecto GTEx²⁴², identificadas en cáncer en tejidos de muestras de pulmón, piel (expuesta y no expuesta al sol), tiroides, colon, próstata y páncreas.²¹⁶ En relación a las variantes genéticas en *AGR3*, un GWMAS ha validado la asociación de rs1636744 ($\beta = 0.091$, $p = 7.1 \times 10^{-9}$) a melanoma cutáneo maligno.²⁴³ Específicamente, la variante asociada rs4559136, ha sido previamente asociada al fenotipo ludópata en población alemana, sin embargo, no alcanzó significancia de genoma completo ($p = 7.26 \times 10^{-5}$, OR = 0.388, IC95% 0.243-0.619).²⁴⁴ Con base en la literatura previa, y con el efecto a protección que muestran las dos variables encontradas, podríamos proponer que los cambios en *AGR3*, los cuales no son estructurales sino más bien de expresión a la baja, alteran la expresión de *MUC5AC* y *MUC5B*, que se manifiesta por una baja producción de moco en fumadores.

La variante JHU_15.94073449 no ha sido asignada, sin embargo, se encuentra localizada cerca de *MCTP2* (*Multiple C2 Domains-Containing Transmembrane Protein 2*), gen que se ha relacionado principalmente con patologías cardíacas y algunas respiratorias. La función del producto de *MCTP2* se relaciona con la señalización del Ca^{+2} en la membrana celular.²⁴⁵ En un principio se evidenció que desempeña un papel importante en el sistema cardiovascular, ya que se asoció en atresia mitral, por lo cual su función estaba relacionada con el desarrollo del flujo de salida cardíaco.²⁴⁶ Sin embargo, el gen ha mostrado pleiotropía por fenotipos respiratorios. Se reportó la variante rs12910412 asociada ($p = 5.22 \times 10^{-7}$) a mayor riesgo (OR = 1.30, IC95%= 1.17-1.44) de EPOC en fumadores con el fenotipo clínico de bronquitis crónica en un GWMAS.¹¹⁸ Posteriormente, *MCTP2* se encontró

diferencialmente expresado en células mesenquimales de pulmón y epidermis ²⁴⁷, y en eosinófilos de esputo y lavado bronquioalveolar ²⁴⁸, así como hipometilado (*fold change* = -1.64) en el epitelio de la vía aérea, en fumadores usuarios de pipa de agua. ²⁴⁹ Con base en esta información y de acuerdo con los resultados obtenidos, ser podría proponer que la variante JHU_15.94073449, posee un papel regulador de la expresión de *MCTP2*, que conlleva a un mayor riesgo (OR = 10.97, IC95%= 3.94-30.53) de disminuir su expresión debido al daño causado por la inflamación crónica en la vía área de fumadores.

Otra variante que resultó significativa (4.67×10^{-6}) de fue la variante rs2901070 se localiza cerca de *SERPINI1* (*Serine Protease Inhibitor, Clade I, Member 1*), perteneciente a la familia de inhibidores de serin proteasa como A1AT, el cual codifica para la enzima neuroserpina, un factor neurotrófico del activador de plasminógeno del tipo tisular, activador de plasminógeno uroquinasa y plasmina. ²⁵⁰ *SERPINI1* reporta una alta expresión del transcrito y proteína en el cerebro, específicamente en la corteza cerebral, hipocampo y núcleo caudado. ²¹⁶ Se ha reportado con expresión diferencial aumentada en cáncer de pulmón (media $\log_2 \sim 3.8$) ^{251,252} y tabaquismo (1.55 veces mayor) ²⁵³, así como disminuida en hipertensión pulmonar (2.55 veces menor). ²⁴⁷ Probablemente su asociación en este trabajo, pudo haber sido indirecta, y más bien se asocie a la exposición crónica de tabaquismo, explicada por el mecanismo en el cual los niveles de activador de plasminógeno en el cerebro son aumentados por el tabaquismo (reflejado en niveles aumentados en sangre debido a la nicotina). ²⁵⁴ Por lo que aumentaría la expresión de *SERPINI1* y con ello, la producción de neuroserpina, tratando de inhibir el activador de plasminógeno. De esta manera, la variante rs2901070, debido a su cercanía física (~ 40 Kb) con *SERPINI1*, y que se asocia al riesgo, podría encontrarse en una secuencia reguladora que disminuya la expresión del transcrito.

Otro hallazgo de importancia, fueron las variantes rs352523 y rs9363933, que se encuentran cercanas a *ADGRB3* (*Adhesion G Protein-Coupled Receptor B3*), también denominado *BAI3* (*Brain-Specific Angiogenesis Inhibitor 3*). Este gen participa en la regulación de diversos aspectos del sistema nervioso central, como

guía de axón, mielinización y sinapsis, ²⁵⁵ por lo que se ha reportado con expresión aumentada en regiones del sistema nervioso central, como en la corteza cerebral, glándula pituitaria, hipotálamo, hipocampo, cerebelo y núcleo caudado. ²¹⁶ Debido a las funciones que desempeña y las regiones en las que se ha encontrado, se ha reportado asociado a esquizofrenia ²⁵⁶, trastorno bipolar ²⁵⁷ y adicción a las drogas ^{258,259}, por lo que se sospecha que también su asociación con EPOC haya sido indirecta, y más bien se relacione con la adicción a la nicotina. La superfamilia de los receptores acoplados a proteína G, contiene otras proteínas codificadas por genes que se han involucrado directamente en el pulmón como *ADGRA2*, *ADGRC1*, *ADGRE5*, *ADGRF5* y *ADGRG6*, ²⁵⁵ en los cuales no identificamos ninguna asociación. Sin embargo, distintas variantes en *ADGRG6* (rs3817928, rs7776375, rs6937121, rs11155242, rs148274477, rs262113) han sido asociadas a menor función pulmonar en europeos, africanos e hispanos, ^{115,120,122–124,136} en cambio en nuestros sujetos solamente se encontró el rs3817928 con una frecuencia alélica menor que los grupos reportados y sin ser significativo.

En la región cromosómica 1p36.11 se obtuvieron dos variantes asociadas, sin embargo, esta región es característica por alojar una gran cantidad de genes, por lo cual se podría confundir con aparentes asociaciones indirectas, que más bien sean producto de la corta distancia y podrían no tener una explicación biológica en la EPOC. Primeramente, la variante rs17257107 localizada en *PDIK1L* (*PDLIM1-Interacting Kinase 1-Like*) no ha tenido reportes en ningún rasgo particular. Sin embargo, su producto tiene función de proteína con actividad serina/treonina quinasa, por lo que en GWAS se ha asociado a fenotipos dispersos, como el tamaño del esqueleto en humanos ²⁶⁰ y mejor calidad de composición mineral en leche de vaca ^{261,262}, además de reportar una expresión alterada en sujetos tras una cirugía cardiotorácica. ²⁶³ Los reportes de este gen no muestran una clara función, sin embargo, esta variante se encontró con alto desequilibrio de ligamiento con una variante en *TRIM63* (*Tripartite Motif-Containing Protein 63*), por lo que se sospecha de una asociación indirecta. Este gen, también denominado *MURF1* (*Muscle-Specific Ring Finger Protein 1*) ha sido ampliamente evidenciado con atrofia del músculo esquelético ²⁶⁴, una de las afectaciones clínicas sistémicas de la EPOC. ²⁶⁵

En sujetos sanos, *MURF1* se expresa en tejido músculo esquelético y cardíaco, ²¹⁶ sin embargo, se ha reportado que en pacientes con EPOC y atrofia del músculo esquelético, la expresión de *MURF1* se eleva hasta 517% ²⁶⁶ en pacientes estables, y aún más cuando presentan exacerbaciones agudas, ²⁶⁷ además de comprobarse en ensayos *in vivo* ²⁶⁸, sin embargo, no aumentan los niveles de la proteína. ²⁶⁶ También, se ha reportado que la hipoxia tiene un efecto similar, aumentando la expresión de *MURF1*. ²⁶⁹ Esto sugiere que la hipertrofia del músculo, además de un ambiente bajo en oxígeno, es regulada por un conjunto de factores de transcripción y señalización, en la denominada vía FoxO-1/Atrogin-1/MuRF1, de la cual permanece mucha información sin explicar. De esta manera se argumentaría que la variante cercana a *TRIM63*, por su localización cerca del gen tendría un papel relevante en la expresión y/o regulación de los factores de transcripción, aunque no tuvimos la significancia estadística para identificarla, quizá por no incorporar al IMC en el modelo, debido a que es más relacionada con la función pulmonar que con la EPOC en sí.

Por último, en el modelo ajustado, la variante con mayor nivel de significancia ($p = 9.00 \times 10^{-6}$), fue 1:26450101-A-G, la cual tiene valores idénticos a rs17257107, lo que podría sugerir una localización física muy cercana, sin embargo, al no ser una variante validada no se pudo graficar en *Locus Zoom*. Esta variante se encuentra en *DHDDS* (*Dehydrololichyl Diphosphate Synthase*), cuya proteína codificada cataliza el alargamiento de la cadena cis-prenil para producir el esqueleto polipropilénico del dolicol, un lípido portador glucosílico requerido para la biosíntesis de varias clases de glucoproteínas. ²⁷⁰ En relación con fenotipos relacionados con fisiología respiratoria, recientemente se asociaron en un GWAS de 400,102 individuos con ancestría europea, a dos variantes en *DHDDS*, rs9438626 con CVF (8.06×10^{-10}) y rs12096239 con VEF₁ (2.08×10^{-12}). ²⁷¹ Sin embargo, parece tratarse de un gen pleiotrópico, el cual también se ha asociado a otros fenotipos, principalmente con mutaciones puntuales en exón, las cuales son de baja frecuencia y alta penetrancia, produciendo fenotipos como desorden congénito de glicosilación tipo I ²⁷² y renitis pigmentosa. ^{273,274}

Los alelos asociados de los rs12518621 y rs9363933 presentan diferencia en la frecuencia respecto al componente AMI en controles en comparación con los casos, sin embargo, como se menciona en la literatura ²⁷⁵, se cuantificó esta diferencia mediante la *d* de Cohen, para evidenciar una diferencia del efecto, la cual resultó ser pequeña quizá debido al bajo número de sujetos con el alelo, además del hecho que todos los sujetos son mexicanos mestizos seleccionados precisamente con esta característica. En la literatura, se han reportado otras variantes propias de poblaciones AMI no encontradas en otras poblaciones, como algunos alelos de *HLA-DRB1* ²⁷⁶, así como variantes más prevalentes en poblaciones mestizadas con alto componente AMI, como el rs148533712 en *RORA* y rs11216230 en *SIK3* relacionados con mayor dislipidemia; ²⁷⁷ rs964184 cercana a los genes *BUD13*, *ZNF259* y *APOA5* y asociada a altos niveles de triglicéridos; ²⁷⁸ así como recientemente se asoció a rs4133185 en *AGMO* con disminución del VEF₁. ¹²³

Asociaciones por modelo estratificado

Debido al sesgo que podría ocurrir por la diferencia existente en las variables dicotómicas de ajuste en el modelo, se decidió estratificar, con el fin de nulificar la diferencia de variables que se encuentren desiguales y obtener resultados de un estrato en específico. Sin embargo, los grupos se redujeron en número, por lo que adicionalmente analizar múltiples variantes pueden haber generado asociaciones espurias. Entre los hallazgos en hombres, se asoció la variante rs2762924, cercana a *CYP24A1* (*Cytochrome P450, Family 24, Subfamily A, Polypeptide 1*), perteneciente a la gran familia de los catalizadores biológicos CYP450, del cual *CYP24A1* en particular, es la enzima encargada del catabolismo de la vitamina D a su forma activa y posterior degradación. ²⁷⁹ La deficiencia de vitamina D en suero, se ha correlacionado con disminución de la función pulmonar. ²⁸⁰ Además se ha reportado que en EPOC, la expresión de *CYP24A1* está incrementada por TGF- β 1, estos niveles aumentados disminuyen la expresión de mediadores constitutivos del hospedero, aumentando el daño tisular en la vía aérea. ²⁸¹

En el modelo estratificado en hombres se obtuvieron dos variantes asociadas (rs10420331 y rs11084307) en *CACNG8* (*Calcium Channel, Voltage-Dependent*,

Gamma-8 Subunit), gen que recientemente se ha encontrado con expresión alterada en leucocitos en asma atópico en adolescentes.²⁸² Además, en el estratificado en mujeres se asociaron dos variantes físicamente muy próximas, sin embargo, ya que una variante no está validada, no se pudo relacionar sí se encontraba en desequilibrio de ligamiento. La JHU_12.3109691, localizada en *TSPAN9 (Tetraspanin 9)*, el cual se ha visto expresado en la activación plaquetas²⁸³, así como modulador de endosomas en infecciones virales²⁸⁴, sin embargo, hasta ahora no existe una clara relación con EPOC. La otra variante asociada fue rs3825365 en *TEAD4 (Tea Domain Family Member 4)*, un gen cuyo transcrito parece tener una participación en la hipoxia, debido a su rol como estimulador transcripcional de *VEGF (Vascular endothelial growth factor)* bajo condiciones de bajo oxígeno²⁸⁵, relacionado con la vía de HIF-1, los cuales se han encontrado en niveles aumentados en tejido pulmonar de pacientes con EPOC.²⁸⁶ Estratificando en fumadores actuales, se reportó la variante JHU_19.57835849 en *ZNFP587B*, la cual existe poca información. Sin embargo, presenta un alto desequilibrio de ligamiento con otra variante en *ZNF154 (Zinc Finger Protein 154)*, la cual se ha asociado a la regulación en procesos de cáncer.^{287,288} Por último, en la estratificación por fumadores, se asoció rs7644995 cercano a *RTP3 (Receptor-Transporting Protein 3)*, el cual ha tenido reportes en un rasgo en específico.

De las 97 variantes en 20 *loci* con más reportes en la literatura encontramos 34 en 17 *loci* en nuestra muestra (**Anexo XX**), sin embargo, a pesar de que no alcanzaron valores de significancia de genoma completo, algunas variantes muestran significancia menor a 0.05 y frecuencias alélicas similares a las reportadas en la literatura, por lo que la falta de poder estadístico debido al reducido número de participantes incluidos, entre otros factores, podría ser la causa de estos resultados. Las variantes del gen *FAM13A* han sido replicadas mayormente en europeos, posiblemente por ello las variantes rs1903003 [OR = 0.70 (0.51-0.97), p = 0.033], rs2869967 [OR = 1.44 (1.04-1.99), p = 0.025], rs7671167 [OR = 0.64 (0.47-0.89), p = 0.007] mostraron frecuencias alélicas similares a las reportadas en estas poblaciones, así como valores de riesgo en la misma dirección^{108,112,115,118,125}, confirmando el hallazgo en nuestra población; de la misma manera, las variantes

rs2609255 [OR = 1.47 (1.02-2.12), $p = 0.041$] y rs2609264 [OR = 1.51 (1.05-2.17), $p = 0.028$], reportadas previamente en población asiática ¹²⁶, muestran una dirección similar del riesgo, aunque las frecuencias alélicas difieren respecto a los asiáticos, pero no comparado con europeos. La variante rs2277027 en *ADAM19* [OR = 0.66 (0.48-0.90), $p = 0.012$] presenta una frecuencia alélica menor a la reportada en otras poblaciones del 1kGP ²⁸⁹, así como un efecto inverso al reportado en europeos e hispanos. ^{115,123} La variante rs615098 en *MMP12* mostró asociación [OR = 0.56 (0.32-0.98), $p = 0.043$] a EPOC, sin embargo la dirección del efecto es contrario al reportado en europeos ¹¹² y las frecuencias alélicas reportadas son el doble a las obtenidas en nuestra muestra. ²⁸⁹ Respecto a las demás variantes reportadas en la literatura en poblaciones europeas e incluso algunas replicadas en hispanos, no se obtuvieron diferencias estadísticas menores a $p = 0.05$, sin embargo, las frecuencias alélicas oscilan de manera general entre un 10-20% mayor o menor, respecto a las poblaciones europeas; así como de manera general de un 3-5% respecto a las poblaciones americanas de referencia. A pesar de que muchas de estas previas asociaciones han sido replicadas en población hispana ¹²³, no se tiene conocimiento de que se hayan reportado solamente en mexicanos. Además, la inclusión de otras poblaciones no mexicanas con mayor componente caucásico (cubanos) o africano (puertorriqueños) dentro de los hispanos, el gran número de sujetos de estudio, entre otros factores, podrían ser las causas de no haber replicado en nuestro análisis.

La estratificación poblacional es un fenómeno que, de no contemplarse, puede inducir asociaciones tanto de falsos positivos, como falsos negativos. ²⁹⁰ Por lo que se han implementado métodos para poder evaluarla, entre los cuales utilizamos tres de los mayormente utilizados por su validez y confiabilidad. Primeramente, utilizamos el método de asociación estructurada por medio de ADMIXTURE ⁸⁶ para poder determinar la proporción de componentes ancestrales. Según la literatura, existe un gradiente respecto a los componentes ancestrales en las regiones del País: mientras en los estados de la región Centro-Sureste aumenta el AMI, disminuye el EUR; inversamente a la región Norte, donde aumenta el EUR y disminuye el AMI. ^{81,94} Sin embargo, de acuerdo a esto, se sospecharía que los pacientes con EPOC del grupo de casos pertenecieran a la región Norte y los

fumadores del grupo control más al Centro-Sureste, lo cual de ser cierto, se reflejaría en valores elevados de λ .

Las diferencias en la función pulmonar por el grupo racial o étnico han sido un tema ampliamente discutido en la neumología.^{291,292} Debido a esto se han creado ecuaciones de referencia que incorporan características inherentes de la población estudiada.⁵ En el caso de las poblaciones mestizadas, ha sido complejo asignar un grupo racial específico a la variabilidad inherente de estos sujetos, lo que podría provocar sesgos de información y una incorrecta asignación de valores de referencia.⁵ Por lo que en estas poblaciones, específicamente hispanas, se han creado valores de referencia^{293,294}, por lo cual en este estudio consideramos valores para individuos adultos de México¹⁸⁹, a diferencia de otros estudios que utilizan referencias estándar, sin considerar el grupo étnico o racial.¹⁶⁴

Entre las poblaciones hispanas, se ha evidenciado que los valores de función pulmonar aumentan en poblaciones con mayor proporción de individuos de ancestría AMI y disminuyen respecto a la ancestría EUR.¹⁹ Por lo que el componente AMI en estas poblaciones podría ser el causante del aumento de valores de función pulmonar, evidenciados en estudios con enfermedades respiratorias obstructivas como EPOC^{55,72,73} y asma.^{81,295,296} En el presente estudio, nuestro objetivo se centró demostrar un aumento en el componente AMI en individuos fumadores sin EPOC en sujetos de una misma población, sin embargo, nuestra variable de desenlace fue la obstrucción pulmonar dicotomizada, por lo que posteriormente sería deseable evaluar a la función pulmonar como variable desenlace respecto al componente AMI en los mismo sujetos.

Al comparar las proporciones de los componentes ancestrales de nuestros sujetos de estudio con los participantes de origen mexicano de MESA¹⁷⁴ y HCHS/SOL¹⁷⁶ (entre los cuales no existe una diferencia tan marcada), sin embargo, presentan mayor componente AMI (48-49%) que nuestros casos (46.6%) y mayor EUR (47%) que nuestros controles (45.5%), así como un aumento en el AFR (4%), lo cual tendría diversas posibles explicaciones. Primero, la cantidad de marcadores en común con las poblaciones de referencia es variable, en nuestro estudio se

obtuvieron 147,197 variantes en común con las poblaciones de referencia, mientras que en MESA ¹⁷⁴ y en HCHS/SOL ¹⁷⁶ ocuparon respectivamente 144,564 y 431,143 SNP en común altamente informativos con sus poblaciones de referencia. Segundo, las poblaciones de referencia utilizadas en este estudio no fueron las mismas a ellos, ya que en este estudio se utilizaron como referencia, en población europea solamente a CEU y africana solamente a YRI del HapMap ¹⁹⁰, así como la combinación de Zapotecas de Oaxaca, Mayas de Campeche y Tepehuanos de Durango del *Native Mexican Diversity Panel* ⁸¹ como referencia AMI; en MESA ¹⁷⁴ se utilizaron poblaciones del HGDP (*Human Genome Diversity Project*) ²⁹⁷, en el cual la contribución AMI fue referida de 5 poblaciones (Karitiana, Surui, colombianos, Mayas y Pimas), la EUR con 8, (orcadianos, adigués, rusos, vascos, franceses, italianos, sardos y toscanos) y la AFR con 6 (Bantués, Mandenka, Yorubas, San, pigmeos Mbuti, pigmeos Biaka). Tercero, los estudios MESA ¹⁷⁴ y HCHS/SOL ¹⁷⁶ incluyen sujetos de origen mexicano en EE. UU. de las regiones con mayor concentración de migrantes latinos (MESA ¹⁵⁰: Los Ángeles, CA; Forsythn, NC; St. Paul, MN; Chicago, IL; New York, NY; Baltimore, MD. HCHS/SOL ¹⁵⁴: San Diego, CA; Miami, FL; Chicago, IL; Bronx, NY), los cuales, en el caso de los mexicanos, no representan a la generalidad de la población del País, ya que según cifras de la Secretaría de Gobernación de México, las entidades de Michoacán, Guanajuato, Ciudad de México, Jalisco y San Luis Potosí suman más del 40% del origen de emigrantes hacia las ciudades mencionadas anteriormente en EE. UU. ²⁹⁸ Los estudios MESA Lung/ SHARe ¹³⁵, Costa Rica ⁷³ y HCHS/SOL ¹⁷⁶ han sido los únicos estudios que han evaluado variantes genéticas por GWAS en individuos de origen hispano con EPOC, sin embargo, con excepción del Costa Rica, los agrupan como un único grupo, lo cual no sería homogéneo en sus componentes ancestrales.

Adicionalmente, se utilizó el método de PCA, el cual es el más utilizado en los GWAS, ya que minimiza la variabilidad total de los componentes ancestrales de cada individuo. El PC1 es elegido para minimizar la varianza mayor en los sujetos de estudio, seguido del PC2 y así respectivamente, hasta que el PC n explicaría la menor variación entre la población. ⁹⁰ En nuestros resultados, el PC1 pudo determinar el 43.53% de la varianza en la ancestría en la población de estudio, por

lo cual se tomó una variable de ajuste en el modelo de regresión logística. Al graficarse en PC1 y PC2, los fumadores y pacientes con EPOC mostraban un ordenamiento similar a lo reportado por el Proyecto de Diversidad del Mexicano ⁹⁴, lo cual nos indica que los individuos del estudio tienen características de la población general mexicana.

También se determinó por control genómico el factor de inflación λ , el cual debe ser <1.05 para considerar que no existe estratificación poblacional ²⁹⁹; en nuestro estudio, obtuvimos un valor de 1.06 en el modelo sin ajuste, sin embargo después de ajustar por las demás variables, incluido PC1, este valor se redujo a 1.008.

Por medio del coeficiente de coancestría (θ) por IBD fuimos capaces de identificar a 23 individuos que tenían alguna relación directa hasta segundo grado con los participantes, lo cual, de no haberse identificado pudo haber causado un sesgo de estratificación poblacional debido a la relación críptica. ²⁹⁹

El presente estudio no está exento de limitantes, una de las principales fue tecnología del microarreglo en sí misma, ya que incluye variantes que tienen una cobertura mayor en poblaciones africanas y americanas en conjunto, por lo cual alrededor de 500,000 variantes tuvieron una frecuencia menor al 1% o nula en mexicana, en parte, debido al enriquecimiento del microarreglo con variantes de poblaciones africanas. Además, 6 de las variantes asociadas no tuvieron una identificación asignada en bases de consulta internacional, ya que las poblaciones mestizas han sido de las menos evaluadas en GWAS.

El tamaño de muestra empleado fue reducido, comparado con los estudios previamente mencionados, ya que diversos métodos empleados, como PCA, control genómico, además de las múltiples variantes evaluadas, son dependientes del tamaño de muestra, y entre más amplio sea, valores más válidos pueden resultar. ¹⁰⁷

La manera en la que se definió a la EPOC fue mediante lo indicado por GOLD 2017 ¹, sin embargo, existen otros criterios que pueden definir a la EPOC ³⁰⁰, como la indicada por ATS/ERS 2004. ³⁰¹ Recientemente, se ha reportado que las variantes

asociadas difieren respecto a la definición de EPOC utilizadas, incluso en la misma población.¹¹³ Además, dicotomizar una variable como la obstrucción pulmonar, tiende a disminuir su variabilidad, limitando estadísticamente su análisis.³⁰² Además, existen algunos participantes con limitación al flujo de aire, que se encuentran en valores limítrofes respecto a la obstrucción (relación VEF_1/CVF post broncodilatador < 0.7), los cuales de continuar fumando se ha reportado que pudieran desarrollar la enfermedad o disminuir su función pulmonar [$\beta = -4.45 (-8.6,-0.36)$].³⁰³ Por otro lado, también existen pacientes con EPOC GOLD I, estadio menos severo de la clasificación de severidad GOLD¹, los cuales se ha reportado que tras la cesación al tabaquismo, el VEF_1 puede aumentar significativamente 2% al año, pudiendo llegar a niveles normales de acuerdo a la edad de cesación.³⁰⁴

A pesar de ser una técnica relativamente novedosa para el estudio de enfermedades complejas, los GWAS presentan algunas desventajas. Puede dar paso a potenciales falsos positivos debido a las múltiples pruebas realizadas por cada variante, lo que conlleva a utilizar pruebas estadísticas más astringentes como Bonferroni.³⁰⁵ Sin embargo, estos niveles rigurosos de significancia estadística dejarían atrás variantes que pudieran ser relevantes, pero con efectos más discretos.

Algunos de los genes en que se localizaban las variantes asociadas se han relacionado con cáncer. Además de la EPOC, el tabaquismo crónico produce diferentes tipos de cáncer, principalmente cáncer de pulmón, laringe, riñón, vejiga, estómago, colon, cavidad oral, esófago y leucemia,⁴⁴ De hecho, la EPOC secundaria a tabaquismo es un factor de riesgo a cáncer de pulmón [HR = 5.0 (1.5–17.1)]³⁰⁶, sin embargo, este desenlace no fue evaluado.

Además de esta estrategia, estratificamos por sexo y estatus de tabaquismo, sin embargo, se redujo el número de individuos por grupo, posiblemente al ajustar por estas variables se pudo haber *sobreajustado* los resultados, creando posiblemente asociaciones espurias o bien dejando fuera asociaciones verdaderas.³⁰⁷

Permanecen todavía las variantes asociadas sujetas a una replicación en una población de estudio independiente, ya que los resultados de los GWAS no finalizan con las variantes asociadas a nivel de significancia propuesto, se necesita una robusta replicación de las variantes en otra población de estudio para validar el hallazgo, además de la existencia de plausibilidad biológica del efecto de la variante en el rasgo. ^{107,308}

10 Conclusiones

Empleando la estrategia de análisis de genoma completo empleando un microarreglo enriquecido con variantes americanas, se identificaron variantes genéticas asociadas a EPOC en fumadores mexicanos mestizos, que sugieren plausibilidad biológica en la enfermedad. Existen diferencias en el componente amerindio entre fumadores con EPOC, respecto a los que no la presentan. El componente amerindio parece tener un efecto ambivalente en la enfermedad, debido a que tomándolo en consideración tiene un efecto de protección, sin embargo, algunas variantes asociadas al mayor riesgo se encuentran relacionadas en individuos con mayor componente amerindio. Existen variantes previamente reportadas en la literatura, las cuales se asociaron a EPOC en nuestra muestra, presentando tendencia al riesgo en el mismo sentido del reportado, sin presentar significancia debido al reducido tamaño de muestra. Los resultados obtenidos están sujetos a validación poblacional, para confirmar o descartar factores genéticos de riesgo en la EPOC.

11 Bibliografía

1. GOLD. *Global Strategy for the Diagnosis, Management and Prevention of COPD.*; 2017. <http://goldcopd.org/gold-2017-global-strategy-diagnosis-management-prevention-copd/>. Accessed July 28, 2017.
2. Fletcher C, Peto R, Tinker C, Speizer F. *The Natural History of Chronic Bronchitis and Emphysema.* Oxford: Oxford University Press; 1976.
3. García-Río F, Calle M, Burgos F, et al. Espirometría. *Arch Bronconeumol.* 2013;49(9):388-401. doi:10.1016/j.arbres.2013.04.001
4. Pellegrino R, Viegi G, Brusasco V, et al. Interpretative strategies for lung function tests. *Eur Respir J.* 2005;26(5):948-968. doi:10.1183/09031936.05.00035205
5. Quanjer PH, Stanojevic S, Cole TJ, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: The global lung function 2012 equations. *Eur Respir J.* 2012;40(6):1324-1343. doi:10.1183/09031936.00080312
6. Duong ML, Islam S, Rangarajan S, et al. Global differences in lung function by region (PURE): An international, community-based prospective study. *Lancet Respir Med.* 2013;1(8):599-609. doi:10.1016/S2213-2600(13)70164-4
7. Post W, Heederik D, Houba R. Decline in lung function related to exposure and selection processes among workers in the grain processing and animal feed industry. *Occup Environ Med.* 1998;55(5):349-355. doi:10.1136/oem.55.5.349
8. Mannino DM, Davis KJ. Lung function decline and outcomes in an elderly population. *Thorax.* 2006;61(6):472-477. doi:10.1136/thx.2005.052449
9. Burchfiel CM, Marcus EB, Sharp DS, et al. Characteristics associated with rapid decline in forced expiratory volume. *Ann Epidemiol.* 1996;6(3):217-227. doi:1047279796000075 [pii]
10. Ryan G, Knuiman MW, Divitini ML, James A, Musk AW, Bartholomew HC. Decline in lung function and mortality: The Busselton health study. *J Epidemiol Community Health.* 1999;53(4):230-234. doi:10.1136/jech.53.4.230
11. McKeever TM, Scrivener S, Broadfield E, Jones Z, Britton J, Lewis SA. Prospective Study of Diet and Decline in Lung Function in a General Population. *Am J Respir Crit Care Med.* 2002;165(9):1299-1303. doi:10.1164/rccm.2109030
12. Mannino DM, Reichert MM, Davis KJ. Lung function decline and outcomes in an adult population. *Am J Respir Crit Care Med.* 2006;173(9):985-990. doi:10.1164/rccm.200508-1344OC
13. Kumar R, Seibold MA, Aldrich MC, et al. Genetic Ancestry in Lung-Function Predictions. *N Engl J Med.* 2010;363(4):321-330. doi:10.1056/NEJMoa0907897
14. Ortega VE, Kumar R. The Effect of Ancestry and Genetic Variation on Lung Function Predictions: What Is "Normal" Lung Function in Diverse Human Populations? *Curr Allergy Asthma Rep.* 2015;15(4):16. doi:10.1007/s11882-015-0516-2
15. WHO. *Chronic Obstructive Pulmonary Disease (COPD).* World Health Organization; 2016. <http://www.who.int/mediacentre/factsheets/fs315/en/>. Accessed May 18, 2017.

16. Vestbo J, Hurd SS, Agustí AG, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med.* 2013;187(4):347-365. doi:10.1164/rccm.201204-0596PP
17. GBD 2015 Chronic Respiratory Disease Collaborators JB, Abajobir AA, Abate KH, et al. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Respir Med.* 2017;5(9):691-706. doi:10.1016/S2213-2600(17)30293-X
18. Dahlgren G, Whitehead M. *Policies and Strategies to Promote Social Equity in Health. Background Document to WHO - Strategy Paper for Europe.*; 1991. https://ideas.repec.org/p/hhs/ifswps/2007_014.html. Accessed December 7, 2018.
19. Menezes AMB, Perez-Padilla R, Jardim JRB, et al. Chronic obstructive pulmonary disease in five Latin American cities (the PLATINO study): a prevalence study. *Lancet.* 2005;366(9500):1875-1881. doi:10.1016/S0140-6736(05)67632-5
20. Adams SG, Anzueto A, Pugh JA, Lee S, Hazuda HP. Mexican American elders have similar severities of COPD despite less tobacco exposure than European American elders. *Respir Med.* 2006;100(11):1966-1972. doi:10.1016/j.rmed.2006.02.023
21. Barr RG, Avilés-Santa L, Davis SM, et al. Pulmonary disease and age at immigration among hispanics: Results from the hispanic community health study/study of latinos. *Am J Respir Crit Care Med.* 2016;193(4):386-395. doi:10.1164/rccm.201506-1211OC
22. Halbert RJ, Gano A, Mannino DM, Buist AS, Badamgarav E, Natoli JL. Global burden of COPD: systematic review and meta-analysis. *Eur Respir J.* 2006;28(3):523-532. doi:10.1183/09031936.06.00124605
23. Silverman EK, Weiss ST, Drazen JM, et al. Gender-related differences in severe, early-onset chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* 2000;162(6):2152-2158. doi:10.1164/ajrccm.162.6.2003112
24. Martinez FJ, Curtis JL, Sciruba F, et al. Sex differences in severe pulmonary emphysema. *Am J Respir Crit Care Med.* 2007;176(3):243-252. doi:10.1164/rccm.200606-828OC
25. Sørheim I-C, Johannessen A, Gulsvik A, Bakke PS, Silverman EK, DeMeo DL. Gender differences in COPD: are women more susceptible to smoking effects than men? *Thorax.* 2010;65(6):480-485. doi:10.1136/thx.2009.122002
26. Barker DJ, Godfrey KM, Fall C, Osmond C, Winter PD, Shaheen SO. Relation of birth weight and childhood respiratory infection to adult lung function and death from chronic obstructive airways disease. *BMJ.* 1991;303(6804):671-675. doi:10.1136/bmj.303.6804.671
27. Todisco T, de Benedictis FM, Iannacci L, et al. Mild prematurity and respiratory functions. *Eur J Pediatr.* 1993;152(1):55-58. doi:10.1007/BF02072517
28. McCloskey SC, Patel BD, Hinchliffe SJ, Reid ED, Wareham NJ, Lomas DA. Siblings of patients with severe chronic obstructive pulmonary disease have a significant risk of airflow obstruction. *Am J Respir Crit Care Med.* 2001;164(8 Pt 1):1419-1424. doi:10.1164/ajrccm.164.8.2105002
29. Lamprecht B, McBurnie MA, Vollmer WM, et al. COPD in never smokers: Results from the population-based burden of obstructive lung disease study. *Chest.* 2011;139(4):752-763. doi:10.1378/chest.10-1253

30. Hayden LP, Hobbs BD, Cohen RT, et al. Childhood pneumonia increases risk for chronic obstructive pulmonary disease: The COPDGene study. *Respir Res*. 2015;16(1):115. doi:10.1186/s12931-015-0273-8
31. Dennis RJ, Maldonado D, Norman S, Baena E, Martinez G. Woodsmoke exposure and risk for obstructive airways disease among women. *Chest*. 1996;109(1):115-119. doi:10.1378/chest.109.1.115
32. Pérez-Padilla R, Regalado J, Vedal S, et al. Exposure to biomass smoke and chronic airway disease in Mexican women. A case-control study. *Am J Respir Crit Care Med*. 1996;154(3 Pt 1):701-706. doi:10.1164/ajrccm.154.3.8810608
33. Beran D, Zar HJ, Perrin C, Menezes AM, Burney P. Burden of asthma and chronic obstructive pulmonary disease and access to essential medicines in low-income and middle-income countries. *Lancet Respir*. 2015;3:159-170. doi:10.1016/S2213-2600(15)00004-1
34. Miravittles M, De La Roza C, Naberan K, et al. Problemas con el diagnóstico de la EPOC en atención primaria. *Arch Bronconeumol*. 2006;42(1):3-8. doi:10.1157/13083272
35. Hesso I, Gebara SN, Kayyali R. Impact of community pharmacists in COPD management: Inhalation technique and medication adherence. *Respir Med*. 2016;118:22-30. doi:10.1016/j.rmed.2016.07.010
36. Bang KM, Syamlal G, Mazurek JM. Prevalence of chronic obstructive pulmonary disease in the U.S. working population: An analysis of data from the 19972004 national health interview survey. *COPD J Chronic Obstr Pulm Dis*. 2009;6(5):380-387. doi:10.1080/15412550903140899
37. WHO. *WHO Report on the Global Tobacco Epidemic: The MPOWER Package.*; 2008. https://www.who.int/tobacco/mpower/mpower_report_full_2008.pdf. Accessed December 10, 2018.
38. WHO. *Updated Status of the WHO Framework Convention on Tobacco Control.*; 2005. <http://www.who.int/tobacco/framework/countrylist/en/index.html>.
39. CDC. *Chronic Obstructive Pulmonary Disease Public Health Strategic Framework for Prevention.*; 2011. www.cdc.gov/copd. Accessed December 10, 2018.
40. WHO. *WHO Guidelines for Indoor Air Quality: Household Fuel Combustion.*; 2018. <https://www.who.int/airpollution/guidelines/household-fuel-combustion/en/>. Accessed December 10, 2018.
41. Borgerding M, Klus H. Analysis of complex mixtures--cigarette smoke. *Exp Toxicol Pathol*. 2005;57 Suppl 1:43-73. <http://www.ncbi.nlm.nih.gov/pubmed/16092717>. Accessed August 5, 2015.
42. Talhout R, Schulz T, Florek E, van Benthem J, Wester P, Opperhuizen A. Hazardous compounds in tobacco smoke. *Int J Environ Res Public Health*. 2011;8(2):613-628. doi:10.3390/ijerph8020613
43. Thielen A, Klus H, Müller L. Tobacco smoke: unraveling a controversial subject. *Exp Toxicol Pathol*. 2008;60(2-3):141-156. doi:10.1016/j.etp.2008.01.014
44. U.S. Department of Health and Human Services. *How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking & Attributable Disease: A Report of the Surgeon General*. Atlanta, GA: USA; 2010.
45. Rennard SJ, Vestbo J. COPD: the dangerous underestimate of 15%. *Lancet*. 2006;367(9518):1216-1219. doi:10.1016/S0140-6736(06)68516-4

46. Lundbäck B, Lindberg A, Lindström M, et al. Not 15 But 50% of smokers develop COPD? - Report from the Obstructive Lung Disease in Northern Sweden studies. *Respir Med*. 2003;97(2):115-122. doi:10.1053/rmed.2003.1446
47. Feinleib M, Rosenberg HM, Collins JG, Delozier JE, Pokras R, Chevarley FM. Trends in COPD Morbidity and Mortality in the United States. *Am Rev Respir Dis*. 1989;140(3_pt_2):S9-S18. doi:10.1164/ajrccm/140.3_Pt_2.S9
48. Hayatbakhsh MR, Sadasivam S, Mamun AA, Najman JM, Williams GM, O'Callaghan MJ. Maternal smoking during and after pregnancy and lung function in early adulthood: A prospective study. *Thorax*. 2009;64(9):810-814. doi:10.1136/thx.2009.116301
49. Mishra V, Dai X, Smith KR, Mika L. Maternal exposure to biomass smoke and reduced birth weight in Zimbabwe. *Ann Epidemiol*. 2004;14(10):740-747. doi:10.1016/j.annepidem.2004.01.009
50. Molino NA. Genetics of COPD. *Chest*. 2004;125(5):1929-1940. doi:10.1378/chest.125.5.1929
51. The World Bank Group. *World Bank Country and Lending Groups*; 2017. <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>.
52. Menezes AM, Victora CG, Perez-Padilla R. The Platino project: methodology of a multicenter prevalence survey of chronic obstructive pulmonary disease in major Latin American cities. *BMC Med Res Methodol*. 2004;4(1):15. doi:10.1186/1471-2288-4-15
53. *Encuesta Global de Tabaquismo En Adultos. México, 2015*. Cuernavaca, México; 2015. www.gob.mx/salud. Accessed November 23, 2017.
54. Reynales- Shigematsu L, Zavala-Arciniega L, Paz-Ballesteros W, et al. *Encuesta Nacional de Consumo de Drogas, Alcohol y Tabaco 2016-2017: Reporte de Tabaco*. Ciudad de México, México; 2017. www.conadic.gob.mx.
55. Bruse S, Sood A, Petersen H, et al. New Mexican Hispanic Smokers Have Lower Odds of Chronic Obstructive Pulmonary Disease and Less Decline in Lung Function Than Non-Hispanic Whites. *Am J Respir Crit Care Med*. 2011;184(11):1254-1260. doi:10.1164/rccm.201103-0568OC
56. Real Academia Española. Diccionario de la lengua española. 2019. <https://dle.rae.es/?id=KW1s7dJ>.
57. Dominguez K, Penman-Aguilar A, Chang M-H, Moonesinghe R, Castellanos T, Rodriguez-Lainz A. Vital Signs: Leading Causes of Death, Prevalence of Diseases and Risk Factors, and Use of Health Services Among Hispanics in the United States — 2009–2013. *MMWR*. 2015;64(17):469-478. <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6417a5.htm>. Accessed May 18, 2017.
58. Markides KS, Coreil J. The health of Hispanics in the southwestern United States: an epidemiologic paradox. *Public Health Rep*. 1986;101(3):253-265. doi:10.1016/j.annepidem.2007.09.002
59. Ruiz JM, Steffen P, Smith TB. Hispanic mortality paradox: A systematic review and meta-analysis of the longitudinal literature. *Am J Public Health*. 2013;103(3):e52-60. doi:10.2105/AJPH.2012.301103
60. Stephen EH, Foote K, Hendershot GE, Schoenborn CA. Health of the foreign-born population: United States, 1989-90. *Adv Data*. 1994;14(241):1-12. <http://www.ncbi.nlm.nih.gov/pubmed/10132138>. Accessed May 4, 2018.
61. Pablos-Méndez A. Mortality Among Hispanics. *JAMA J Am Med Assoc*. 1994;271(16):1237-1238. doi:10.1001/jama.1994.03510400023017

62. Eschbach K, Ostir G V, Patel K V, Markides KS, Goodwin JS. Neighborhood context and mortality among older Mexican Americans: is there a barrio advantage? 2004;94(10). doi:10.2105/ajph.94.10.1807
63. Arias E, Eschbach K, Schauman WS, Backlund EL, Sorlie PD. The Hispanic mortality advantage and ethnic misclassification on US death certificates. *Am J Public Health*. 2010;100 Suppl 1(S1):S171-7. doi:10.2105/AJPH.2008.135863
64. Howrey B, Goodwin JS, Eschbach K, Freeman J. Lower Stroke Mortality Among Hispanics: An Exploration of Potential Methodological Confounders. *Med Care*. 2010;48(6):534-539. doi:10.1097/mlr.0b013e3181d686cf
65. Sood A, Stidley CA, Picchi MA, et al. Difference in airflow obstruction between hispanic and non-hispanic white female smokers. *COPD J Chronic Obstr Pulm Dis*. 2008;5(5):274-281. doi:10.1080/15412550802363345
66. Rodriguez J, Jiang R, Johnson WC, MacKenzie BA, Smith LJ, Barr RG. The Association of Pipe and Cigar Use With Cotinine Levels, Lung Function, and Airflow Obstruction. *Ann Intern Med*. 2010;152(4):201. doi:10.7326/0003-4819-152-4-201002160-00004
67. US Census Bureau. *Hispanic Heritage Month 2017*.; 2017. <https://www.census.gov/newsroom/facts-for-features/2017/hispanic-heritage.html>. Accessed May 3, 2018.
68. Savitz DA. In Defense of Black Box Epidemiology. *Epidemiology*. 1994;5(5):550-552. https://journals.lww.com/epidem/Citation/1994/09000/In_Defense_of_Black_Box_Epidemiology.15.aspx.
69. Kan H, Stevens J, Heiss G, Rose KM, London SJ. Dietary fiber, lung function, and chronic obstructive pulmonary disease in the atherosclerosis risk in communities study. *Am J Epidemiol*. 2008;167(5):570-578. doi:10.1093/aje/kwm343
70. Varraso R, Willett WC, Camargo CA. Prospective study of dietary fiber and risk of chronic obstructive pulmonary disease among US women and men. *Am J Epidemiol*. 2010;171(7):776-784. doi:10.1093/aje/kwp455
71. Leng S, Picchi MA, Tesfaigzi Y, et al. Dietary nutrients associated with preservation of lung function in hispanic and Non-Hispanic white smokers from New Mexico. *Int J COPD*. 2017;12:3171-3181. doi:10.2147/COPD.S142237
72. Diaz AA, Come CE, Mannino DM, et al. Obstructive Lung Disease in Mexican Americans and Non-Hispanic Whites: An Analysis of Diagnosis and Survival in the National Health and Nutritional Examination Survey III Follow-up Study. *Chest*. 2014;145(2):282-289. doi:10.1378/CHEST.13-1414
73. Chen W, Brehm JM, Boutaoui N, et al. Native American ancestry, lung function, and COPD in Costa Ricans. *Chest*. 2014;145(4):704-710. doi:10.1378/chest.13-1308
74. Berman SM, Arnall DA, Cornwall MW. Pulmonary function test outcomes in healthy Navajo Native American adolescents. *Am J Respir Crit Care Med*. 1994;150(4):1150-1153. doi:10.1164/ajrccm.150.4.7921451
75. Ensembl. Human assembly and gene annotation. http://www.ensembl.org/Homo_sapiens/Info/Annotation. Published 2019. Accessed February 4, 2019.
76. U.S. National Library of Medicine. What are single nucleotide polymorphisms (SNPs)? - Genetics Home Reference - NIH. <https://ghr.nlm.nih.gov/primer/genomicresearch/snp>. Published 2019. Accessed

May 17, 2019.

77. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56-65. doi:10.1038/nature11632
78. Mikkelsen TS, Hillier LW, Eichler EE, et al. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*. 2005;437(7055):69-87. doi:10.1038/nature04072
79. Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain J. Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proc Natl Acad Sci*. 2006;85(16):6002-6006. doi:10.1073/pnas.85.16.6002
80. Rosenberg NA, Pritchard JK, Weber JL, et al. Genetic structure of human populations. *Science (80-)*. 2002;298(5602):2381-2385. doi:10.1126/science.1078311
81. Moreno-Estrada A, Gignoux CR, Fernandez-Lopez JC, et al. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science (80-)*. 2014;344(6189):1280-1285. doi:10.1126/science.1251688
82. Cavalli-Sforza LL. Genes, peoples and languages. *Sci Am*. 1991;265(5):104-110. doi:10.1038/scientificamerican1191-104
83. Chakravarti A. Being Human: Kinship: Race relations. *Nature*. 2009;457(7228):380-381. doi:10.1038/457380a
84. Klimentidis YC, Miller GF, Shriver MD. Genetic admixture, self-reported ethnicity, self-estimated admixture, and skin pigmentation among hispanics and native americans. *Am J Phys Anthropol*. 2009;138(4):375-383. doi:10.1002/ajpa.20945
85. Healy M, Edgar H, Mosley C, Hunley K. Associations between ethnic identity, regional history, and genomic ancestry in New Mexicans of Spanish-speaking descent. *Biodemography Soc Biol*. 2018;64(2):152-170. doi:10.1080/19485565.2018.1545563
86. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655-1664. doi:10.1101/gr.094052.109
87. Ingman M, Kaessmann H, Gyllensten U, Paabo S. Mitochondrial genome variation and the origin of modern humans. *Nature*. 2000;408:708-713. doi:10.1038/35047064
88. Deka R, Ferrell RE, Saha N, et al. Dispersion of human Y chromosome haplotypes based on five microsatellites in global populations. *Genome Res*. 2007;6(12):1177-1184. doi:10.1101/gr.6.12.1177
89. Barnholtz-Sloan JS, McEvoy B, Shriver MD, Rebbeck TR. Ancestry estimation and correction for population stratification in molecular epidemiologic association studies. *Cancer Epidemiol Biomarkers Prev*. 2008;17(3):471-477. doi:10.1158/1055-9965.EPI-07-0491
90. Lever J, Krzywinski M, Altman N. Points of Significance: Principal component analysis. *Nat Methods*. 2017;14(7):641-642. doi:10.1038/nmeth.4346
91. Liu Y, Nyunoya T, Leng S, Belinsky SA, Tesfaigzi Y, Bruse S. Softwares and methods for estimating genetic ancestry in human populations. *Hum Genomics*. 2013;7(1):1. doi:10.1186/1479-7364-7-1
92. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999;55(4):997-1004. doi:10.1111/j.0006-341X.1999.00997.x
93. López-Beltrán C, García Deister V. Aproximaciones científicas al mestizo mexicano. *História, Ciências,*

Saúde – Manguinhos. 2013;20(2):391-410.

94. Silva-Zolezzi I, Hidalgo-Miranda A, Estrada-Gil J, et al. Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. *Proc Natl Acad Sci*. 2009;106(21):8611-8616. doi:10.1073/pnas.0903045106
95. Ruiz-Linares A, Adhikari K, Acuña-Alonzo V, et al. Admixture in Latin America: Geographic Structure, Phenotypic Diversity and Self-Perception of Ancestry Based on 7,342 Individuals. *PLoS Genet*. 2014;10(9):e1004572. doi:10.1371/journal.pgen.1004572
96. Bryc K, Velez C, Karafet T, et al. Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc Natl Acad Sci*. 2010;107(Supplement_2):8954-8961. doi:10.1073/pnas.0914618107
97. Drummond MB. The Hispanic paradox unraveled? *Am J Respir Crit Care Med*. 2011;184(11):1222-1223. doi:184/11/1222 [pii]\r10.1164/rccm.201109-1684ED
98. Harhay MO. The hispanic paradox and chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2012;185(11):1246. doi:10.1164/ajrccm.185.11.1246
99. Díaz AA, Celli B, Celedón JC. Chronic obstructive pulmonary disease in Hispanics A 9-Year update. *Am J Respir Crit Care Med*. 2018;197(1):15-21. doi:10.1164/rccm.201708-1615PP
100. Johnson NA, Coram MA, Shriver MD, et al. Ancestral components of admixed genomes in a Mexican cohort. Copenhaver GP, ed. *PLoS Genet*. 2011;7(12):e1002410. doi:10.1371/journal.pgen.1002410
101. Chial H. Mendelian genetics: Patterns of inheritance and single-gene disorders. *Nat Educ*. 2008;1(1):63. <https://www.nature.com/scitable/topicpage/mendelian-genetics-patterns-of-inheritance-and-single-966>.
102. Craig J. Complex Diseases: Research and Applications. *Nat Educ*. 2008;1(1):184. <https://www.nature.com/scitable/topicpage/complex-diseases-research-and-applications-748>.
103. Eriksson S. Pulmonary emphysema and alpha1-antitrypsin deficiency. *Acta Med Scand*. 1964;175(2):197-205. <http://www.ncbi.nlm.nih.gov/pubmed/14124635>. Accessed August 27, 2018.
104. Redline S, Tishler P V, Lewitter FI, Tager IB, Munoz A, Speizer FE. Assessment of Genetic and Nongenetic Influences on Pulmonary Function. *Am Rev Respir Dis*. 1987;135(1):217-222. doi:10.1164/arrd.1987.135.1.217
105. Larson RK, Barman ML, Kueppers F, Fudenberg HH. Genetic and environmental determinants of chronic obstructive pulmonary disease. *Ann Intern Med*. 1970;72(5):627-632. doi:10.7326/0003-4819-72-5-627
106. Tager IB, Speizer FE. Risk estimates for chronic bronchitis in smokers: a study of male-female differences. *Am Rev Respir Dis*. 1976;113(5):619-625. doi:10.1164/arrd.1976.113.5.619
107. Stram DO. *Design, Analysis, and Interpretation of Genome-Wide Association Scans*. New York, NY: Springer New York; 2014. doi:10.1007/978-1-4614-9443-0
108. Pillai SG, Ge D, Zhu G, et al. A Genome-Wide Association Study in Chronic Obstructive Pulmonary Disease (COPD): Identification of Two Major Susceptibility Loci. McCarthy MI, ed. *PLoS Genet*. 2009;5(3):e1000421. doi:10.1371/journal.pgen.1000421
109. Repapi E, Sayers I, Wain L V, et al. Genome-wide association study identifies five loci associated with lung function. *Nat Genet*. 2010;42(1):36-44. doi:10.1038/ng.501

110. Hardin M, Cho MH, Sharma S, et al. Sex-based genetic association study identifies CELSR1 as a possible chronic obstructive pulmonary disease risk locus among women. *Am J Respir Cell Mol Biol*. 2017;56(3):332-341. doi:10.1165/rcmb.2016-0172OC
111. Wain L V, Shrine N, Artigas MS, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet*. 2017;49(3):416-425. doi:10.1038/ng.3787
112. Li X, Ortega VE, Ampleford EJ, et al. Genome-wide association study of lung function and clinical implication in heavy smokers. *BMC Med Genet*. 2018;19(1):134. doi:10.1186/s12881-018-0656-z
113. Van Der Plaats DA, Vonk JM, Lahousse L, et al. Limited overlap in significant hits between genome-wide association studies on two airflow obstruction definitions in the same population. *BMC Pulm Med*. 2019;19(1):58. doi:10.1186/s12890-019-0811-0
114. Wilk JB, Chen T, Gottlieb DJ, et al. A Genome-Wide Association Study of Pulmonary Function Measures in the Framingham Heart Study. McCarthy MI, ed. *PLoS Genet*. 2009;5(3):e1000429. doi:10.1371/journal.pgen.1000429
115. Hancock DB, Eijgelsheim M, Wilk JB, et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet*. 2010;42(1):45-52. doi:10.1038/ng.500
116. Cho MH, Castaldi PJ, Wan ES, et al. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum Mol Genet*. 2012;21(4):947-957. doi:10.1093/hmg/ddr524
117. Cho MH, McDonald MLN, Zhou X, et al. Risk loci for chronic obstructive pulmonary disease: A genome-wide association study and meta-analysis. *Lancet Respir Med*. 2014;2(3):214-225. doi:10.1016/S2213-2600(14)70002-5
118. Lee JH, Cho MH, Hersh CP, et al. Genetic susceptibility for chronic bronchitis in chronic obstructive pulmonary disease. *Respir Res*. 2014;15(1):113. doi:10.1186/s12931-014-0113-2
119. Castaldi PJ, Cho MH, Estépar RSJ, et al. Genome-wide association identifies regulatory loci associated with distinct local histogram emphysema patterns. *Am J Respir Crit Care Med*. 2014;190(4):399-409. doi:10.1164/rccm.201403-0569OC
120. Lutz SM, Cho MH, Young K, et al. A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. *BMC Genet*. 2015;16(1):138. doi:10.1186/s12863-015-0299-4
121. Boueiz A, Lutz SM, Cho MH, et al. Genome-wide association study of the genetic determinants of emphysema distribution. *Am J Respir Crit Care Med*. 2017;195(6):757-771. doi:10.1164/rccm.201605-0997OC
122. Busch R, Hobbs BD, Zhou J, et al. Genetic association and risk scores in a chronic obstructive pulmonary disease meta-analysis of 16,707 subjects. *Am J Respir Cell Mol Biol*. 2017;57(1):35-46. doi:10.1165/rcmb.2016-0331OC
123. Burkart KM, Sofer T, London SJ, et al. A genome-wide association study in hispanics/latinos identifies novel signals for lung function the hispanic community health study/study of latinos. *Am J Respir Crit Care Med*. 2018;198(2):208-219. doi:10.1164/rccm.201707-1493OC
124. Cho MH, Castaldi PJ, Hersh CP, et al. A genome-wide association study of emphysema and airway quantitative imaging phenotypes. *Am J Respir Crit Care Med*. 2015;192(5):559-569. doi:10.1164/rccm.201501-0148OC

125. Cho MH, Boutaoui N, Klanderman BJ, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet.* 2010;42(3):200-202. doi:10.1038/ng.535
126. Kim WJ, Lee MK, Shin C, et al. Genome-wide association studies identify locus on 6p21 influencing lung function in the Korean population. *Respirology.* 2014;19(3):360-368. doi:10.1111/resp.12230
127. Wilk JB, Shrine NRG, Loehr LR, et al. Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. *Am J Respir Crit Care Med.* 2012;186(7):622-632. doi:10.1164/rccm.201202-0366OC
128. Hobbs BD, Parker MM, Chen H, et al. Exome array analysis identifies a common Variant in IL27 associated with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* 2016;194(1):48-57. doi:10.1164/rccm.201510-2053OC
129. Wain L V, Shrine N, Miller S, et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): A genetic association study in UK Biobank. *Lancet Respir Med.* 2015;3(10):769-781. doi:10.1016/S2213-2600(15)00283-0
130. Obeidat M, Zhou G, Li X, et al. The genetics of smoking in individuals with chronic obstructive pulmonary disease. *Respir Res.* 2018;19(1):59. doi:10.1186/s12931-018-0762-7
131. Sakornsakolpat P, McCormack M, Bakke P, et al. Genome-wide association analysis of single-breath DL CO. *Am J Respir Cell Mol Biol.* 2019;60(5):523-531. doi:10.1165/rcmb.2018-0384OC
132. Artigas MS, Loth DW, Wain L V, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet.* 2011;43(11):1082-1090. doi:10.1038/ng.941
133. Loth DW, Artigas MS, Gharib SA, et al. Genome-wide association analysis identifies six new loci associated with forced vital capacity. *Nat Genet.* 2014;46(7):669-677. doi:10.1038/ng.3011
134. Hancock DB, Romieu I, Shi M, et al. Genome-wide association study implicates chromosome 9q21.31 as a susceptibility locus for asthma in Mexican children. Gibson G, ed. *PLoS Genet.* 2009;5(8):e1000623. doi:10.1371/journal.pgen.1000623
135. Manichaikul A, Hoffman EA, Smolonska J, et al. Genome-Wide Study of Percent Emphysema on Computed Tomography in the General Population. The Multi-Ethnic Study of Atherosclerosis Lung/SNP Health Association Resource Study. *Am J Respir Crit Care Med.* 2014;189(4):408-418. doi:10.1164/rccm.201306-1061OC
136. Artigas MS, Wain L V., Miller S, et al. Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. *Nat Commun.* 2015;6(1):8658. doi:10.1038/ncomms9658
137. Liao SY, Lin X, Christiani DC. Genome-wide Association and Network Analysis of Lung Function in the Framingham Heart Study. *Genet Epidemiol.* 2014;38(6):572-578. doi:10.1002/gepi.21841
138. Sørheim I-C, Johannessen A, Grydeland TB, Omenaas ER, Gulsvik A, Bakke PS. Case-control studies on risk factors for chronic obstructive pulmonary disease: how does the sampling of the cases and controls affect the results? *Clin Respir J.* 2010;4(2):89-96. doi:10.1111/j.1752-699X.2009.00154.x
139. Zhu G, Warren L, Aponte J, et al. The SERPINE2 gene is associated with chronic obstructive pulmonary disease in two large populations. *Am J Respir Crit Care Med.* 2007;176(2):167-173. doi:10.1164/rccm.200611-1723OC
140. Anthonisen NR, Connett JE, Kiley JP, et al. Effects of Smoking Intervention and the Use of an Inhaled

- Anticholinergic Bronchodilator on the Rate of Decline of FEV1: The Lung Health Study. *JAMA J Am Med Assoc.* 1994;272(19):1497-1505. doi:10.1001/jama.1994.03520190043033
141. Vestbo J, Anderson W, Coxson HO, et al. Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *Eur Respir J.* 2008;31(4):869-873. doi:10.1183/09031936.00111707
 142. Group The National Emphysema Treatment Trial Research. Rationale and Design of the National Emphysema Treatment Trial: A Prospective Randomized Trial of Lung Volume Reduction Surgery. *Chest.* 1999;116(6):1750-1761. doi:10.1378/CHEST.116.6.1750
 143. Couper D, LaVange LM, Han ML, et al. Design of the subpopulations and intermediate outcomes in COPD study (SPIROMICS). *Thorax.* 2014;69(5):491-494. doi:10.1136/thoraxjnl-2013-203897
 144. Regan EA, Hokanson JE, Murphy JR, et al. Genetic epidemiology of COPD (COPDGene) study design. *COPD J Chronic Obstr Pulm Dis.* 2010;7(1):32-43. doi:10.3109/15412550903499522
 145. Zhou X, Baron RM, Hardin M, et al. Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. *Hum Mol Genet.* 2012;21(6):1325-1335. doi:10.1093/hmg/ddr569
 146. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of Lung Nodules Detected by Volume CT Scanning. *N Engl J Med.* 2009;361(23):2221-2229. doi:10.1056/NEJMoa0906085
 147. Jones PW, Rames AD. TESRA (Treatment Of Emphysema With A Selective Retinoid Agonist) Study Results. In: *C91. LATE BREAKING CLINICAL TRIALS SYMPOSIUM.* American Thoracic Society; 2011:A6418-A6418. doi:10.1164/ajrccm-conference.2011.183.1_MeetingAbstracts.A6418
 148. Splansky GL, Corey D, Yang Q, et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: Design, recruitment, and initial examination. *Am J Epidemiol.* 2007;165(11):1328-1335. doi:10.1093/aje/kwm021
 149. UK Biobank. <https://www.ukbiobank.ac.uk/>. Accessed January 2, 2019.
 150. Bild DE, Bluemke DA, Burke GL, et al. Multi-Ethnic Study of Atherosclerosis: Objectives and Design. *Am J Epidemiol.* 2002;156(9):871-881. doi:10.1093/aje/kwf113
 151. Sung J, Cho S II, Lee K, et al. Healthy Twin: A Twin-Family study of Korea - Protocols and current status. *Twin Res Hum Genet.* 2006;9(6):844-848. doi:10.1375/183242706779462822
 152. Bell B, Rose CL, Damon A. The Normative Aging Study: An Interdisciplinary and Longitudinal Study of Health and Aging. *Aging Hum Dev.* 1972;3(1):5-17. doi:10.2190/GGVP-XLB5-PC3N-EF0G
 153. Ko A, Cantor RM, Weissglas-Volkov D, et al. Amerindian-specific regions under positive selection harbour new lipid variants in Latinos. *Nat Commun.* 2014;5(1):63-76. doi:10.1038/ncomms4983
 154. LaVange LM, Kalsbeek WD, Sorlie PD, et al. Sample Design and Cohort Selection in the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol.* 2010;20(8):642-649. doi:10.1016/j.annepidem.2010.05.006
 155. Silverman EK, Vestbo J, Agusti A, et al. Opportunities and Challenges in the Genetics of COPD 2010: An International COPD Genetics Conference Report. *COPD J Chronic Obstr Pulm Dis.* 2011;8(2):121-135. doi:10.3109/15412555.2011.558864
 156. Psaty BM, O'Donnell CJ, Gudnason V, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium design of prospective meta-analyses of genome-wide association studies from 5 Cohorts. *Circ Cardiovasc Genet.* 2009;2(1):73-80. doi:10.1161/CIRCGENETICS.108.829747

157. Lee JH, McDonald M-LL, Cho MH, et al. DNAH5 is associated with total lung capacity in chronic obstructive pulmonary disease. *Respir Res.* 2014;15(1):97. doi:10.1186/s12931-014-0097-y
158. Siedlinski M, Cho MH, Bakke P, et al. Genome-wide association study of smoking behaviours in patients with COPD. *Thorax.* 2011;66(10):894-902. doi:10.1136/thoraxjnl-2011-200154
159. Kong X, Cho MH, Anderson W, et al. Genome-wide association study identifies BICD1 as a susceptibility gene for emphysema. *Am J Respir Crit Care Med.* 2011;183(1):43-49. doi:10.1164/rccm.201004-0541OC
160. McDonald M-LN, Cho MH, Sørheim I-C, et al. Common Genetic Variants Associated with Resting Oxygenation in Chronic Obstructive Pulmonary Disease. *Am J Respir Cell Mol Biol.* 2014;51(5):678-687. doi:10.1165/rcmb.2014-0135OC
161. Hansel NN, Ruczinski I, Rafaels N, et al. Genome-wide study identifies two loci associated with lung function decline in mild to moderate COPD. *Hum Genet.* 2013;132(1):79-90. doi:10.1007/s00439-012-1219-6
162. Hansel NN, Pare PD, Rafaels N, et al. Genome-wide association study identification of novel loci associated with airway responsiveness in chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol.* 2015;53(2):226-234. doi:10.1165/rcmb.2014-0198OC
163. Hayden LP, Cho MH, Raby BA, Beaty TH, Silverman EK, Hersh CP. Childhood asthma is associated with COPD and known asthma variants in COPDGene: A genome-wide association study. *Respir Res.* 2018;19(1):209. doi:10.1186/s12931-018-0890-0
164. Chen W, Brehm JM, Manichaikul A, et al. A genome-wide association study of chronic obstructive pulmonary disease in Hispanics. *Ann Am Thorac Soc.* 2015;12(3):340-348. doi:10.1513/AnnalsATS.201408-380OC
165. Cho YS, Go MJ, Kim YJ, et al. A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat Genet.* 2009;41(5):527-534. doi:10.1038/ng.357
166. Suh Y, Lee C. Genome-wide association study for genetic variants related with maximal voluntary ventilation reveals two novel genomic signals associated with lung function. *Med (United States).* 2017;96(44):e8530. doi:10.1097/MD.00000000000008530
167. Thyagarajan B, Wojczynski M, Minster RL, et al. Genetic variants associated with lung function: The long life family study. *Respir Res.* 2014;15(1):134. doi:10.1186/s12931-014-0134-x
168. Silverman EK, Chapman HA, Drazen JM, et al. Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease: Risk to relatives for airflow obstruction and chronic bronchitis. *Am J Respir Crit Care Med.* 1998;157(6 PART I):1770-1778. doi:10.1164/ajrccm.157.6.9706014
169. Dijkstra AE, Boezen HM, Van Den Berge M, et al. Dissecting the genetics of chronic mucus hypersecretion in smokers with and without COPD. *Eur Respir J.* 2015;45(1):60-75. doi:10.1183/09031936.00093314
170. Hobbs BD, De Jong K, Lamontagne M, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet.* 2017;49(3):426-432. doi:10.1038/ng.3752
171. Tang W, Kowgier M, Loth DW, et al. Large-scale genome-wide association studies and meta-analyses of longitudinal change in adult lung function. Chen L, ed. *PLoS One.* 2014;9(7):e100776.

doi:10.1371/journal.pone.0100776

172. Obeidat M, Wain L V, Shrine N, et al. A comprehensive evaluation of potential lung function associated genes in the SpiroMeta general population sample. *PLoS One*. 2011;6(5):19382. doi:10.1371/journal.pone.0019382
173. Wyss AB, Sofer T, Lee MK, et al. Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. *Nat Commun*. 2018;9(1):2976. doi:10.1038/s41467-018-05369-0
174. Manichaikul A, Palmas W, Rodriguez CJ, et al. Population structure of hispanics in the United States: The multi-Ethnic study of Atherosclerosis. Williams SM, ed. *PLoS Genet*. 2012;8(4):e1002640. doi:10.1371/journal.pgen.1002640
175. U.S. Department of Health and Human Services, National Heart, Lung and BI. *Hispanic Community Health Study/Study of Latinos Data Book: A Report to the Communities*. Bethesda, MD; 2013. <https://www.nhlbi.nih.gov/files/docs/resources/NHLBI-HCHSSOL-English-508.pdf>. Accessed January 2, 2019.
176. Conomos MP, Laurie CACC, Stilp AM, et al. Genetic Diversity and Association Studies in US Hispanic/Latino Populations: Applications in the Hispanic Community Health Study/Study of Latinos. *Am J Hum Genet*. 2016;98(1):165-184. doi:10.1016/j.ajhg.2015.12.001
177. Pérez-Rubio G, Pérez-Rodríguez ME, Fernández-López JC, et al. SNPs in NRXN1 and CHRNA5 are associated to smoking and regulation of GABAergic and glutamatergic pathways. *Pharmacogenomics*. 2016;17(10):1145-1158. doi:10.2217/pgs-2016-0020
178. Pérez-Rubio G, López-Flores LA, Ramírez-Venegas A, et al. Genetic polymorphisms in CYP2A6 are associated with a risk of cigarette smoking and predispose to smoking at younger ages. 2017. doi:10.1016/j.gene.2017.07.051
179. Pérez-Rubio G, Ramírez-Venegas A, Díaz VN, et al. Polymorphisms in HTR2A and DRD4 predispose to smoking and smoking quantity. *PLoS One*. 2017;12(1). doi:10.1371/journal.pone.0170019
180. Reséndiz-Hernández JM, Sansores R, Hernández-Zenteno R de J, et al. Identification of genetic variants in the TNF promoter associated with COPD secondary to tobacco smoking and its severity. *Int J Chron Obstruct Pulmon Dis*. 2015;Volume 10(1):1241. doi:10.2147/COPD.S83298
181. Hernández-Montoya J, Pérez-Ramos J, Montaña M, et al. Genetic polymorphisms of matrix metalloproteinases and protein levels in chronic obstructive pulmonary disease in a Mexican population. *Biomark Med*. 2015;9(10):979-988. doi:10.2217/bmm.15.75
182. Ambrocio-Ortiz E, Pérez-Rubio G, Abarca-Rojano E, et al. Influence of proinflammatory cytokine gene polymorphisms on the risk of COPD and the levels of plasma protein. *Cytokine*. 2018;111:364-370. doi:10.1016/J.CYTO.2018.09.017
183. Pérez-Rubio G, Silva-Zolezzi I, Fernández-López JC, et al. Genetic Variants in *IL6R* and *ADAM19* are Associated with COPD Severity in a Mexican Mestizo Population. *COPD J Chronic Obstr Pulm Dis*. 2016;13(5):610-615. doi:10.3109/15412555.2016.1161017
184. Pérez-Rubio G, Jiménez-Valverde LO, Ramírez-Venegas A, et al. Prevalence of Alpha-1 Antitrypsin High-risk Variants in Mexican Mestizo Population and Their Association With Lung Function Values. *Arch Bronconeumol (English Ed)*. 2015;51(2):80-85. doi:10.1016/j.arbr.2014.11.002
185. Beltrán Brown F. Función de los niveles de atención médica. *Salud Publica Mex*. 1982;24(5):485-487.

<http://saludpublica.mx/index.php/spm/article/view/687/674>. Accessed February 6, 2019.

186. Global Initiative for Asthma. *Global Strategy for Asthma Management and Prevention.*; 2018. www.ginasthma.org.
187. Matute A. A Century of Spanish Emigration to México. *Voices of México*. 2017:61-64. <http://www.revistascisan.unam.mx/Voices/pdfs/4013.pdf>. Accessed February 7, 2019.
188. Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *Eur Respir J*. 2005;26(2):319-338. doi:10.1183/09031936.05.00034805
189. Pérez-Padilla R, Bouscoulet LT, Vázquez-García JC, et al. Valores de referencia para la espirometría después de la inhalación de 200 μ g de salbutamol. *Arch Bronconeumol*. 2007;43(10):530-534. doi:10.1016/S1579-2129(07)60123-2
190. 1000 Genomes Project Consortium A, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393
191. CAAPA. <https://www.caapa-project.org/>. Published 2011. Accessed July 10, 2018.
192. PAGE. <https://www.pagestudy.org/index.php/home>. Accessed July 10, 2018.
193. T2D-GENES. <http://www.type2diabetesgenetics.org/projects/t2dGenes>. Accessed July 10, 2018.
194. McKusick V. OMIM. <https://www.ncbi.nlm.nih.gov/omim?itool=toolbar>. Accessed July 10, 2018.
195. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42(Database issue):D980-5. doi:10.1093/nar/gkt1113
196. American College of Medical Genetics and Genomics (ACMG). <https://www.acmg.net/>. Published 2018. Accessed August 27, 2018.
197. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi:10.1086/519795
198. 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(8):904-909. doi:10.1038/ng1847
199. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006;2(12):2074-2093. doi:10.1371/journal.pgen.0020190
200. James Kent W, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res*. 2002;12(6):996-1006. doi:10.1101/gr.229102. Article published online before print in May 2002
201. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria; 2015. <https://www.r-project.org/>.
202. Wei T, Simko V. R package "corrplot": Visualization of a Correlation. 2017. <https://github.com/taiyun/corrplot>.
203. Navarro D. *Learning Statistics with R: A Tutorial for Psychology Students and Other Beginners.*; 2015. <http://ua.edu.au/ccs/teaching/lr>.
204. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: Regional visualization of genome-wide association scan results. In: *Bioinformatics*. Vol 27. ; 2011:2336-2337. doi:10.1093/bioinformatics/btq419

205. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Springer; 2009. <https://ggplot2.tidyverse.org>. Accessed May 9, 2019.
206. D. Turner S. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J Open Source Softw*. 2018;3(25):731. doi:10.21105/joss.00731
207. Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circlize implements and enhances circular visualization in R. *Bioinformatics*. 2014;30(19):2811-2812. doi:10.1093/bioinformatics/btu393
208. Kassambara A. ggpubr: "ggplot2" Based Publication Ready Plots. 2018.
209. Hemminki K, Försti A, Bermejo JL. The "common disease-common variant" hypothesis and familial risks. Janssens ACJW, ed. *PLoS One*. 2008;3(6):e2504. doi:10.1371/journal.pone.0002504
210. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev*. 2009;19(3):212-219. doi:10.1016/j.gde.2009.04.010
211. Kitsios GD, Zintzaras E. Genome-wide association studies: hypothesis-"free" or "engaged"? *Transl Res*. 2009;154(4):161-164. doi:10.1016/j.trsl.2009.07.001
212. *Encuesta Global de Tabaquismo En Adultos. México, 2009*. Cuernavaca, México; 2010. www.who.int/tobacco/surveillance/gats_rep_mexico.pdf.
213. Berg CJ, Berg GLS, Buchanan TS, et al. Perceptions of addiction, attempts to quit, and successful quitting in nondaily and daily smokers. *Psychol Addict Behav*. 2013;27(4):1059-1067. doi:10.1037/a0033790
214. Matsson H, Söderhäll C, Einarsdottir E, et al. Targeted high-throughput sequencing of candidate genes for chronic obstructive pulmonary disease. *BMC Pulm Med*. 2016;16(1):146. doi:10.1186/s12890-016-0309-y
215. Chen D, Vollmar M, Rossi MN, et al. Identification of macrodomain proteins as novel O-acetyl-ADP-ribose deacetylases. *J Biol Chem*. 2011;286(15):13261-13271. doi:10.1074/jbc.M110.206771
216. Gamazon ER, Segrè A V., van de Bunt M, et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat Genet*. 2018;50(7):956-967. doi:10.1038/s41588-018-0154-4
217. Anney R, Klei L, Pinto D, et al. A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet*. 2010;19(20):4072-4082. doi:10.1093/hmg/ddq307
218. Jones R, Cadby G, Blangero J, Abraham L, Whitehouse A, Moses E. MACROD2 gene associated with autistic-like traits in a general population sample. *Psychiatr Genet*. 2014;24(6):241-248. doi:10.1097/YPG.0000000000000052
219. Zhang Z, Liu X, Xu H, et al. LINC01170 promotes the progression of endometrial carcinoma by activating the AKT pathway. *J BUON*. 2018;23(6):1745-1752. <http://www.ncbi.nlm.nih.gov/pubmed/30610803>. Accessed June 15, 2019.
220. Djureinovic D, Hallström BM, Horie M, et al. Profiling cancer testis antigens in non-small-cell lung cancer. *JCI Insight*. 2016;1(10):e86837. doi:10.1172/jci.insight.86837
221. Yao C, Chen BH, Joehanes R, et al. Integromic analysis of genetic variation and gene expression identifies networks for cardiovascular disease phenotypes. *Circulation*. 2015;131(6):536-549. doi:10.1161/CIRCULATIONAHA.114.010696

222. Lin H, Joehanes R, Pilling LC, et al. Whole blood gene expression and interleukin-6 levels. *Genomics*. 2014;104(6):490-495. doi:10.1016/j.ygeno.2014.10.003
223. Yang X, Yue Y, Xiong S. Dpep2 Emerging as a Modulator of Macrophage Inflammation Confers Protection Against CVB3-Induced Viral Myocarditis. *Front Cell Infect Microbiol*. 2019;9:57. doi:10.3389/fcimb.2019.00057
224. Habib GM, Shi ZZ, Cuevas AA, Lieberman MW. Identification of two additional members of the membrane-bound dipeptidase family. *FASEB J*. 2003;17(10):1313-1315. doi:10.1096/fj.02-0899fje
225. Calverley DC, Phang TL, Choudhury QG, et al. Significant downregulation of platelet gene expression in metastatic lung cancer. *Clin Transl Sci*. 2010;3(5):227-232. doi:10.1111/j.1752-8062.2010.00226.x
226. Carpe N, Mandeville I, Kho AT, et al. Maternal allergen exposure reprograms the developmental lung transcriptome in atopic and normoresponsive rat pups. *Am J Physiol Cell Mol Physiol*. 2012;303(10):L899-L911. doi:10.1152/ajplung.00179.2012
227. Harada N, Yokoyama T, Yamaji R, Nakano Y, Inui H. RanBP10 acts as a novel coactivator for the androgen receptor. *Biochem Biophys Res Commun*. 2008;368(1):121-125. doi:10.1016/j.bbrc.2008.01.072
228. Du L, Borkowski R, Zhao Z, et al. A high-throughput screen identifies miRNA inhibitors regulating lung cancer cell survival and response to paclitaxel. *RNA Biol*. 2013;10(11):1700-1713. doi:10.4161/rna.26541
229. Mehta PD. Control Variables in Research. *Int Encycl Soc Behav Sci*. January 2015:840-843. doi:10.1016/B978-0-08-097086-8.44013-4
230. Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. *Am J Respir Crit Care Med*. 2017;195(5):557-582. doi:10.1164/rccm.201701-0218PP
231. Bhatt SP, Kim Y II, Harrington KF, et al. Smoking duration alone provides stronger risk estimates of chronic obstructive pulmonary disease than pack-years. *Thorax*. 2018;73(5):414-421. doi:10.1136/thoraxjnl-2017-210722
232. Määttänen P, Gehring K, Bergeron JJM, Thomas DY. Protein quality control in the ER: The recognition of misfolded proteins. *Semin Cell Dev Biol*. 2010;21(5):500-511. doi:10.1016/J.SEMCDB.2010.03.006
233. Uhlén M, Fagerberg L, Hallström BM, et al. Tissue-based map of the human proteome. *Science (80-)*. 2015;347(6220):1260419-1260419. doi:10.1126/science.1260419
234. Zhao F, Edwards R, Dizon D, et al. Disruption of Paneth and goblet cell homeostasis and increased endoplasmic reticulum stress in *Agr2*^{-/-} mice. *Dev Biol*. 2010;338(2):270-279. doi:10.1016/j.ydbio.2009.12.008
235. Thornton DJ, Rousseau K, McGuckin MA. Structure and Function of the Polymeric Mucins in Airways Mucus. *Annu Rev Physiol*. 2008;70(1):459-486. doi:10.1146/annurev.physiol.70.113006.100702
236. Bonser LR, Schroeder BW, Ostrin LA, et al. The endoplasmic reticulum resident protein AGR3: Required for regulation of ciliary beat frequency in the airway. *Am J Respir Cell Mol Biol*. 2015;53(4):536-543. doi:10.1165/rcmb.2014-0318OC
237. Burgel PR, Nesme-Meyer P, Chanez P, et al. Cough and sputum production are associated with frequent exacerbations and hospitalizations in COPD subjects. *Chest*. 2009;135(4):975-982.

doi:10.1378/chest.08-2062

238. Sun P, Ye R, Wang C, Bai S, Zhao L. Identification of proteomic signatures associated with COPD frequent exacerbators. *Life Sci.* 2019;230:1-9. doi:10.1016/j.lfs.2019.05.047
239. Aberger F, Weidinger G, Grunz H, Richter K. Anterior specification of embryonic ectoderm: the role of the *Xenopus* cement gland-specific gene XAG-2. *Mech Dev.* 1998;72(1-2):115-130. doi:10.1016/S0925-4773(98)00021-5
240. Liu D, Rudland PS, Sibson DR, Platt-Higgins A, Barraclough R. Human homologue of cement gland protein, a novel metastasis inducer associated with breast carcinomas. *Cancer Res.* 2005;65(9):3796-3805. doi:10.1158/0008-5472.CAN-04-3823
241. Sharma S, Tantisira K, Carey V, et al. A role for wnt signaling genes in the pathogenesis of impaired lung function in asthma. *Am J Respir Crit Care Med.* 2010;181(4):328-336. doi:10.1164/rccm.200907-1009OC
242. Carithers LJ, Moore HM. The Genotype-Tissue Expression (GTEx) Project. *Biopreserv Biobank.* 2015;13(5):307-308. doi:10.1089/bio.2015.29031.hmm
243. Law MH, Bishop DT, Lee JE, et al. Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. *Nat Genet.* 2015;47(9):987-995. doi:10.1038/ng.3373
244. Lang M, Leménager T, Streit F, et al. Genome-wide association study of pathological gambling. *Eur Psychiatry.* 2016;36:38-46. doi:10.1016/j.eurpsy.2016.04.001
245. Shin OH, Hau W, Wang Y, Südhof TC. Evolutionarily conserved multiple C2 domain proteins with two transmembrane regions (MCTPs) and unusual Ca²⁺ binding properties. *J Biol Chem.* 2005;280(2):1641-1651. doi:10.1074/jbc.M407305200
246. Lalani SR, Ware SM, Wang X, et al. MCTP2 is a dosage-sensitive gene required for cardiac outflow tract development. *Hum Mol Genet.* 2013;22(21):4339-4348. doi:10.1093/hmg/ddt283
247. Gaskill C, Marriott S, Pratap S, et al. Shared Gene Expression Patterns in Mesenchymal Progenitors Derived from Lung and Epidermis in Pulmonary Arterial Hypertension: Identifying Key Pathways in Pulmonary Vascular Disease. *Pulm Circ.* 2016;6(4):483-497. doi:10.1086/688314
248. Esnault S, Kelly EA, Schwantes EA, et al. Identification of Genes Expressed by Human Airway Eosinophils after an In Vivo Allergen Challenge. *PLoS One.* 2013;8(7). doi:10.1371/journal.pone.0067560
249. Walters MS, Salit J, Ju JH, et al. Waterpipe smoking induces epigenetic changes in the small airway epithelium. Di YP, ed. *PLoS One.* 2017;12(3):e0171112. doi:10.1371/journal.pone.0171112
250. Osterwalder T, Cinelli P, Baici A, et al. The axonally secreted serine proteinase inhibitor, neuroserpin, inhibits plasminogen activators and plasmin but not thrombin. *J Biol Chem.* 1998;273(4):2312-2321. doi:10.1074/jbc.273.4.2312
251. Zhao P, Wu J, Lu F, et al. The imbalance in the complement system and its possible physiological mechanisms in patients with lung cancer. *BMC Cancer.* 2019;19(1). doi:10.1186/s12885-019-5422-x
252. Ramnefjell M, Aamelfot C, Helgeland L, Akhlen LA. Low expression of SerpinB2 is associated with reduced survival in lung adenocarcinomas. *Oncotarget.* 2017;8(53):90706-90718. doi:10.18632/oncotarget.21456
253. Weng DY, Chen J, Taslim C, et al. Persistent alterations of gene expression profiling of human

- peripheral blood mononuclear cells from smokers. *Mol Carcinog.* 2016;55(10):1424-1437. doi:10.1002/mc.22385
254. Zidovetzki R, Chen P, Fisher M, Hofman FM. Nicotine Increases Plasminogen Activator Inhibitor-1 Production by Human Brain Endothelial Cells via Protein Kinase C–Associated Pathway. *Stroke.* 1999;30(3):651-655. doi:10.1161/01.STR.30.3.651
255. Hamann J, Aust G, Arac D, et al. International Union of Basic and Clinical Pharmacology. XCIV. Adhesion G Protein-Coupled Receptors. *Pharmacol Rev.* 2015;67(2):338-367. doi:10.1124/pr.114.009647
256. Liao HM, Chao YL, Huang AL, et al. Identification and characterization of three inherited genomic copy number variations associated with familial schizophrenia. *Schizophr Res.* 2012;139(1-3):229-236. doi:10.1016/j.schres.2012.05.015
257. McCarthy MJ, Nievergelt CM, Kelsoe JR, Welsh DK. A survey of genomic studies supports association of circadian clock genes with bipolar disorder spectrum illnesses and lithium response. Kato T, ed. *PLoS One.* 2012;7(2):e32091. doi:10.1371/journal.pone.0032091
258. Liu QR, Drgon T, Johnson C, Walther D, Hess J, Uhl GR. Addiction molecular genetics: 639,401 SNP whole genome association identifies many “cell adhesion” genes. *Am J Med Genet Part B Neuropsychiatr Genet.* 2006;141(8):918-925. doi:10.1002/ajmg.b.30436
259. Highfill CA, Baker BM, Stevens SD, Anholt RRH, Mackay TFC. Genetics of cocaine and methamphetamine consumption and preference in *Drosophila melanogaster*. Heckel DG, ed. *PLoS Genet.* 2019;15(5):e1007834. doi:10.1371/journal.pgen.1007834
260. Soranzo N, Rivadeneira F, Chinappan-Horsley U, et al. Meta-analysis of genome-wide scans for human adult stature identifies novel loci and associations with measures of skeletal frame size. *PLoS Genet.* 2009;5(4):e1000445. doi:10.1371/journal.pgen.1000445
261. Buitenhuis B, Poulsen NA, Larsen LB, Sehested J. Estimation of genetic parameters and detection of quantitative trait loci for minerals in Danish Holstein and Danish Jersey milk. *BMC Genet.* 2015;16(1):52. doi:10.1186/s12863-015-0209-9
262. Do DN, Bissonnette N, Lacasse P, et al. Genome-wide association analysis and pathways enrichment for lactation persistency in Canadian Holstein cattle. *J Dairy Sci.* 2017;100(3):1955-1970. doi:10.3168/jds.2016-11910
263. Huang TT, Deoghare H V., Smith BK, et al. Gene expression changes in the human diaphragm after cardiothoracic surgery. *J Thorac Cardiovasc Surg.* 2011;142(5):1214-1222.e20. doi:10.1016/j.jtcvs.2011.02.025
264. Bodine SC, Latres E, Baumhueter S, et al. Identification of ubiquitin ligases required for skeletal muscle atrophy. *Science (80-).* 2001;294(5547):1704-1708. doi:10.1126/science.1065874
265. Agustí AGN, Noguera A, Sauleda J, Sala E, Pons J, Busquets X. Systemic effects of chronic obstructive pulmonary disease. *Eur Respir J.* 2003;21(2):347-360. doi:10.1183/09031936.03.00405703
266. Doucet M, Russell AP, Léger B, et al. Muscle atrophy and hypertrophy signaling in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* 2007;176(3):261-269. doi:10.1164/rccm.200605-704OC
267. Crul T, Testelmans D, Spruit MA, et al. Gene expression profiling in vastus lateralis muscle during an acute exacerbation of COPD. *Cell Physiol Biochem.* 2010;25(4-5):491-500. doi:10.1159/000303054

268. Kamiide Y, Furuya M, Inomata N, Yada T. Chronic exposure to cigarette smoke causes extrapulmonary abnormalities in rats. *Environ Toxicol Pharmacol*. 2015;39(2):864-870. doi:10.1016/j.etap.2015.02.016
269. Martin NRW, Aguilar-Agon K, Robinson GP, et al. Hypoxia Impairs Muscle Function and Reduces Myotube Size in Tissue Engineered Skeletal Muscle. *J Cell Biochem*. 2017;118(9):2599-2605. doi:10.1002/jcb.25982
270. Endo S, Zhang YW, Takahashi S, Koyama T. Identification of human dehydrodolichyl diphosphate synthase gene. *Biochim Biophys Acta - Gene Struct Expr*. 2003;1625(3):291-295. doi:10.1016/S0167-4781(02)00628-0
271. Shrine N, Guyatt AL, Erzurumluoglu AM, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet*. 2019;51(3):481-493. doi:10.1038/s41588-018-0321-7
272. Sabry S, Vuillaumier-Barrot S, Mintet E, et al. A case of fatal Type I congenital disorders of glycosylation (CDG I) associated with low dehydrodolichol diphosphate synthase (DHDDS) activity. *Orphanet J Rare Dis*. 2016;11(1):84. doi:10.1186/s13023-016-0468-1
273. Züchner S, Dallman J, Wen R, et al. Whole-exome sequencing links a variant in DHDDS to retinitis pigmentosa. *Am J Hum Genet*. 2011;88(2):201-206. doi:10.1016/j.ajhg.2011.01.001
274. Venturini G, Koskiniemi-Kuendig H, Harper S, Berson EL, Rivolta C. Two specific mutations are prevalent causes of recessive retinitis pigmentosa in North American patients of Jewish ancestry. *Genet Med*. 2015;17(4):285-290. doi:10.1038/gim.2014.132
275. Cano-Corres R, Sánchez-Álvarez J, Fuentes-Arderiu X. The Effect Size : Beyond Statistical Significance. *J Int Fed Clin Chem Lab Med*. 2012;23(1):1-5. <http://www.ncbi.nlm.nih.gov/pubmed/27683404>. Accessed June 18, 2019.
276. Arnaiz-Villena A, Parga-Lozano C, Moreno E, Areces C, Rey D, Gomez-Prieto P. The Origin of Amerindians and the Peopling of the Americas According to HLA Genes: Admixture with Asian and Pacific People. *Curr Genomics*. 2010;11(2):103-114. doi:10.2174/138920210790886862
277. Ko A, Cantor RM, Weissglas-Volkov D, et al. Amerindian-specific regions under positive selection harbour new lipid variants in Latinos. *Nat Commun*. 2014;5(1):3983. doi:10.1038/ncomms4983
278. Parra EJ, Mazurek A, Gignoux CR, et al. Admixture mapping in two Mexican samples identifies significant associations of locus ancestry with triglyceride levels in the BUD13/ ZNF259/APOA5 region and fine mapping points to rs964184 as the main driver of the association signal. Mersha TB, ed. *PLoS One*. 2017;12(2):e0172880. doi:10.1371/journal.pone.0172880
279. Jones G, Prosser DE, Kaufmann M. 25-Hydroxyvitamin D-24-hydroxylase (CYP24A1): Its important role in the degradation of vitamin D. *Arch Biochem Biophys*. 2012;523(1):9-18. doi:10.1016/j.abb.2011.11.003
280. Jolliffe DA, James WY, Hooper RL, et al. Prevalence, determinants and clinical correlates of vitamin D deficiency in patients with Chronic Obstructive Pulmonary Disease in London, UK. *J Steroid Biochem Mol Biol*. 2018;175:138-145. doi:10.1016/j.jsbmb.2017.01.019
281. Schrupf JA, Ninaber DK, Van Der Does AM, Hiemstra PS. TGF- β 1 Impairs Vitamin D-Induced and Constitutive Airway Epithelial Host Defense Mechanisms. *J Innate Immun*. April 2019:1-16. doi:10.1159/000497415
282. Jiang Y, Gruzieva O, Wang T, et al. Transcriptomics of atopy and atopic asthma in white blood cells

- from children and adolescents. *Eur Respir J*. 2019;53(5):1900102. doi:10.1183/13993003.00102-2019
283. Haining EJ, Matthews AL, Noy PJ, et al. Tetraspanin Tspan9 regulates platelet collagen receptor GPVI lateral diffusion and activation. *Platelets*. 2017;28(7):629-642. doi:10.1080/09537104.2016.1254175
284. Stiles KM, Kielian M. Role of TSPAN9 in Alphavirus Entry and Early Endosomes. *J Virol*. 2016;90(9):4289-4297. doi:10.1128/JVI.00018-16
285. Shie J-L, Wu G, Wu J, et al. RTEF-1, a novel transcriptional stimulator of vascular endothelial growth factor in hypoxic endothelial cells. *J Biol Chem*. 2004;279(24):25010-25016. doi:10.1074/jbc.M403103200
286. Fu X, Zhang F. Role of the HIF-1 signaling pathway in chronic obstructive pulmonary disease. *Exp Ther Med*. 2018;16(6):4553-4561. doi:10.3892/etm.2018.6785
287. Hu Y, Qi M-F, Xu Q-L, et al. Candidate tumor suppressor ZNF154 suppresses invasion and metastasis in NPC by inhibiting the EMT via Wnt/ β -catenin signalling. *Oncotarget*. 2017;8(49):85749-85758. doi:10.18632/oncotarget.20479
288. Um SW, Kim HK, Kim Y, et al. Bronchial biopsy specimen as a surrogate for DNA methylation analysis in inoperable lung cancer. *Clin Epigenetics*. 2017;9(1):131. doi:10.1186/s13148-017-0432-5
289. Gibbs RA, Boerwinkle E, Doddapaneni H, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393
290. Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. *Nat Genet*. 2004;36(5):512-517. doi:10.1038/ng1337
291. Quanjer PH. Lung function, race and ethnicity: a conundrum. *Eur Respir J*. 2013;41(6):1249-1251. doi:10.1183/09031936.00053913
292. Braun L, Wolfgang M, Dickersin K. Defining race/ethnicity and explaining difference in research studies on lung function. *Eur Respir J*. 2013;41(6):1362-1370. doi:10.1183/09031936.00091612
293. LaVange L, Davis SM, Hankinson J, et al. Spirometry reference equations from the HCHS/SOL (Hispanic community health study/study of Latinos). *Am J Respir Crit Care Med*. 2017;196(8):993-1003. doi:10.1164/rccm.201610-1987OC
294. Pérez-Padilla R, Valdivia G, Muiño A, et al. Valores de referencia espirométrica en 5 grandes ciudades de Latinoamérica para sujetos de 40 o más años de edad. *Arch Bronconeumol*. 2006;42(7):317-325. doi:10.1157/13090581
295. Torgerson DG, Gignoux CR, Galanter JM, et al. Case-control admixture mapping in Latino populations enriches for known asthma-associated genes. *J Allergy Clin Immunol*. 2012;130:76-82.e12. doi:10.1016/j.jaci.2012.02.040
296. Pino-Yanes M, Thakur N, Gignoux CR, et al. Genetic ancestry influences asthma susceptibility and lung function among Latinos. *J Allergy Clin Immunol*. 2015;135(1):228-235. doi:10.1016/J.JACI.2014.07.053
297. Cavalli-Sforza LL. The human genome diversity project: Past, present and future. *Nat Rev Genet*. 2005;6(4):333-340. doi:10.1038/nrg1596
298. Unidad de Política Migratoria, Consejo Nacional de Población. *Prontuario Sobre Movilidad y Migración Internacional. Dimensiones Del Fenómeno En México*. México; 2016. http://politicamigratoria.gob.mx/es_mx/SEGOB/Prontuario.

299. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. *Nat Rev Genet.* 2010;11(7):459-463. doi:10.1038/nrg2813
300. Reyes-Garcia A, Torre-Bouscoulet L, Pérez-Padilla R. Controversies and limitations in the diagnosis of chronic obstructive pulmonary disease. *Rev Investig Clin.* 2019;71(1):28-35. doi:10.24875/RIC.18002626
301. Celli BR, MacNee W, ATS/ERS Task Force. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur Respir J.* 2004;23(6):932-946. <http://www.ncbi.nlm.nih.gov/pubmed/15219010>. Accessed June 25, 2019.
302. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ.* 2006;332(7549):1080. doi:10.1136/bmj.332.7549.1080
303. Perez-Padilla R, Wehrmeister FC, De Oca MM, et al. Outcomes for symptomatic non-obstructed individuals and individuals with mild (GOLD stage 1) COPD in a population based cohort. *Int J COPD.* 2018;13:3549-3561. doi:10.2147/COPD.S175527
304. Scanlon PD, Connett JE, Waller LA, et al. Smoking cessation and lung function in mild-to-moderate chronic obstructive pulmonary disease: The lung health study. *Am J Respir Crit Care Med.* 2000;161(2 I):381-390. doi:10.1164/ajrccm.161.2.9901044
305. Hunter DJ, Kraft P. Drinking from the Fire Hose — Statistical Issues in Genomewide Association Studies. *N Engl J Med.* 2007;357(5):436-439. doi:10.1056/NEJMp078120
306. Husebø GR, Nielsen R, Hardie J, et al. Risk factors for lung cancer in COPD – results from the Bergen COPD cohort study. *Respir Med.* 2019;152:81-88. doi:10.1016/J.RMED.2019.04.019
307. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology.* 2009;20(4):488-495. doi:10.1097/EDE.0b013e3181a819a1
308. NCI-NHGRI Working Group on Replication in Association Studies. *Replicating Genotype–Phenotype Associations.* Vol 447.; 2007. <https://www.nature.com/articles/447655a.pdf>. Accessed May 21, 2019.

Anexo I Aprobación del Comité de investigación.



"Año del Centenario de la Promulgación de la Constitución Política de los Estados Unidos Mexicanos"

Comité de Investigación

Ciudad de México a 23 de Mayo de 2017
INER/CI/144/17

Dr. en C. Ramcés Falfán Valencia
Investigador Principal

Asunto: DICTAMEN DEL COMITÉ DE INVESTIGACIÓN.
APROBACIÓN

Título del Proyecto: PARTICIPACIÓN DE LA CONTRIBUCIÓN ANCESTRAL AMERINDIA EN LA ADAPTACIÓN A HIPOXEMIA POR GRANDES ALTITUDES Y SUS IMPLICACIONES EN LAS ENFERMEDADES RESPIRATORIAS OBSTRUCTIVAS EN MEXICANOS MESTIZOS E INDÍGENAS.

Código asignado por el Comité: **B14-17**

Le informamos que su proyecto de referencia ha sido evaluado por el Comité y las opiniones acerca de los documentos presentados se encuentran a continuación:

	Nº y/o Fecha Versión	Decisión
Protocolo	Versión 1, Mayo 2017	APROBADO
Consentimiento Informado	Versión 1, Mayo 2017	APROBADO

Este protocolo tiene vigencia de mayo 2017 a mayo 2020.

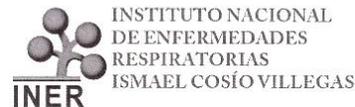
En caso de requerir una ampliación, le rogamos tenga en cuenta que deberá enviar al Comité un reporte de progreso al menos 30 días antes de la fecha de término de su vigencia. El Comité dispone en su página electrónica de un formato estándar que podrá usarse al efecto. Lo anterior forma parte de las obligaciones del Investigador las cuales vienen descritas al reverso de esta hoja.

Atentamente


Dr. Joaquín Zúñiga Ramos
Presidente del Comité

Calzada de Tlalpan No. 4502, Col. Sección XVI, Deleg. Tlalpan, México, D.F., C.P. 14080
Tel. (55) 54 87 17 00 www.iner.salud.gob.mx

Anexo II. Aprobación del Comité de ética en investigación.



"Año del Centenario de la Promulgación de la Constitución Política de los Estados Unidos Mexicanos"

Comité de Ética en Investigación

Ciudad de México a 23 de Mayo de 2017
INER/CEI/178/17

Dr. en C. Ramcés Falfán Valencia
Investigador Principal

Asunto: DICTAMEN DEL COMITÉ DE ÉTICA EN INVESTIGACIÓN.
APROBACIÓN

Título del Proyecto: PARTICIPACIÓN DE LA CONTRIBUCIÓN ANCESTRAL AMERINDIA EN LA ADAPTACIÓN A HIPOXEMIA POR GRANDES ALTITUDES Y SUS IMPLICACIONES EN LAS ENFERMEDADES RESPIRATORIAS OBSTRUCTIVAS EN MEXICANOS MESTIZOS E INDÍGENAS.

Código asignado por el Comité: **B14-17**

Le informamos que su proyecto de referencia ha sido evaluado por el Comité y las opiniones acerca de los documentos presentados se encuentran a continuación:

	Nº y/o Fecha Versión	Decisión
Protocolo	Versión 1, Mayo 2017	APROBADO
Consentimiento Informado	Versión 1, Mayo 2017	APROBADO

Este protocolo tiene vigencia de mayo 2017 a mayo 2020.

En caso de requerir una ampliación, le rogamos tenga en cuenta que deberá enviar al Comité un reporte de progreso al menos 30 días antes de la fecha de término de su vigencia. El Comité dispone en su página electrónica de un formato estándar que podrá usarse al efecto. Lo anterior forma parte de las obligaciones del Investigador las cuales vienen descritas al reverso de esta hoja.

Atentamente

Dra. Rocío Chapela Mendoza
Presidente del Comité

Anexo III Aprobación del Comité de Bioseguridad



COMITÉ DE BIOSEGURIDAD

Ciudad de México, a 30 marzo de 2017

DI/CB/PI/12/17

Estimado: Dr en C. Ramcés Falfán Valencia

Asunto: Dictamen del Comité de Bioseguridad

APROBADO

Título del protocolo: Participación de la contribución ancestral amerindia en la adaptación a hipoxemia por grandes altitudes y sus implicaciones en las enfermedades respiratorias obstructivas en Mexicanos mestizos e indígenas."

FOLIO: 1217

Le informamos que su proyecto ha sido evaluado por este Comité y se ha encontrado que cumple con los requerimientos básicos para llevarse a cabo de manera segura en sus instalaciones.

Es fundamental que durante el desarrollo de este proyecto, se de puntual cumplimiento a lo estipulado en el Reglamento de la Ley General de Salud en Materia de Investigación para la Salud Título Cuarto Capítulo I. La disposición final de los **RPBI y CRETIS** procedentes de las técnicas utilizadas en el proyecto deberá realizarse de acuerdo a la Norma Oficial Mexicana NOM-087-ECOL- SSA-1-2002, Protección Ambiental, Salud Ambiental, Residuos Peligrosos Biológico Infecciosos-Clasificación y especificación de manejo, que se generan en establecimientos que presentan atención médica, coordinado para su efecto por la SEMARNAT.

Atentamente

Edgar Enrique Sevilla Reyes Ph.D.

Presidente del Comité de Bioseguridad

Instituto Nacional de Enfermedades Respiratorias Ismael Cosío Villegas





México D. F. a _____ de _____ 201__

Consentimiento Informado

Protocolo: *Participación de la contribución ancestral amerindia en la adaptación a hipoxemia por grandes altitudes y sus implicaciones en las enfermedades respiratorias obstructivas en Mexicanos mestizos e indígenas*

Descripción del protocolo

El objetivo de este estudio es identificar variantes genéticas que participen en la adaptación a hipoxemia por vivir a grandes altitudes y su participación en la disminución de oxígeno provocada por enfermedades respiratorias obstructivas en mexicanos mestizos e indígenas. Adicionalmente, se busca que la información generada por el presente protocolo, en un futuro, nos ayude a comprender la fisiopatología de estas enfermedades, así como identificar grupos de riesgo, que podrían ayudar en la prevención y tratamiento de enfermedades relacionadas a la hipoxemia.

Procedimiento de participación en el protocolo

Consiste en tres partes:

Primera: Toma de medidas clínicas no invasivas: IMC, presión arterial, oximetría de pulso, óxido nítrico exhalado y espirometría.

Segunda: Toma de una muestra de 21 ml. de sangre obtenida por punción en la vena del antebrazo. La cual será utilizada para la extracción del material genético (ADN y ARN) presente en las células nucleadas de la sangre, el cual, a su vez, se convierte en el material de trabajo de los procesos necesarios en la identificación de variantes genéticas que constituyen riesgo o protección a la aparición, desarrollo a hipoxemia. Las muestras de material genético procedentes de su sangre se manejan con claves de identificación, con la finalidad de resguardar la información relacionada a su persona.

Tercera: Recolección de datos en un formato escrito, en donde le solicitamos responder a un cuestionario con información acerca de sus antecedentes médicos, estado de salud, residencia actual y previa de hasta 2 generaciones anteriores a usted.

Riesgos probables

En la toma de muestra sanguínea (venopunción) puede sentir molestia a la punción con la aguja, un ligero sangrado, y equimosis (aparición de un moretón). La muestra de sangre es tomada por personal capacitado para tal fin, lo que disminuye estos riesgos probables. Adicionalmente existe disponibilidad de tratamiento médico para el procedimiento de toma de muestra e indemnización a que legalmente tendría derecho, por parte de la institución, en el caso de daños que la ameriten, directamente causados por la venopunción y que, si existen gastos adicionales, éstos serán absorbidos por el presupuesto de la investigación.

Aceptación de participación en el protocolo

Al aceptar participar en el presente protocolo, después de haber leído y pensado su decisión, usted está confirmando las siguientes aseveraciones:

Acepto de manera voluntaria donar esta muestra de sangre para todos los estudios genéticos que corresponden a este protocolo y recibiré una copia de este formato de mi consentimiento informado para mi registro personal.

Entiendo que de la muestra de sangre extraerán el material genético para realizar estudios sobre diferentes genes que pueden ocasionar hipoxemia; además se me ha explicado que las muestras obtenidas serán almacenadas de manera segura y en condiciones óptimas para su preservación en un área especial dentro del laboratorio de HLA del INER.

Se me ha explicado a cerca de las complicaciones que puedo llegar a presentar como consecuencia de la punción.

Estoy satisfecho con la información que se me ha proporcionado y mis preguntas han sido contestadas satisfactoriamente.

He leído y me han explicado claramente en que consiste esta forma de consentimiento informado con respecto al uso de mi muestra de sangre en el presente protocolo de investigación.

Anexo V Consentimiento informado, parte 2/2.



"Año del Centenario de la Promulgación de la Constitución Política de los Estados Unidos Mexicanos"

Entiendo mis derechos y requerimientos y sé que por lo tanto si decido no participar lo puedo hacer sin necesidad de justificar mi decisión y sin que por este motivo se vea alterado mi trato como paciente.

Entiendo que estos estudios son realizados con fines de investigación y no ofrecen mejoría en el tratamiento.

Estoy de acuerdo que mi identidad permanecerá anónima en caso de que mi historial médico sea revisado por personas ajenas a estos estudios de investigación.

Entiendo que al firmar esta forma de consentimiento no estoy cediendo ninguno de mis derechos legales, también entiendo que nada en este consentimiento intenta cambiar ninguna ley aplicable en relación a los consentimientos informados.

De igual manera, al donar mi muestra de sangre esta puede ser usada para otros estudios relacionados a enfermedades respiratorias y/o como control de referencia (creación de un banco de muestras). Dichas muestras pueden ser almacenadas hasta por 15 años. Sí No

Nombre del participante	Firma	Fecha (Día/Mes/Año)
Nombre del Testigo 1/Relación o Parentesco	Firma	Fecha (Día/Mes/Año)
Nombre del Testigo 2/Relación o Parentesco	Firma	Fecha (Día/Mes/Año)

Compromiso del Investigador:

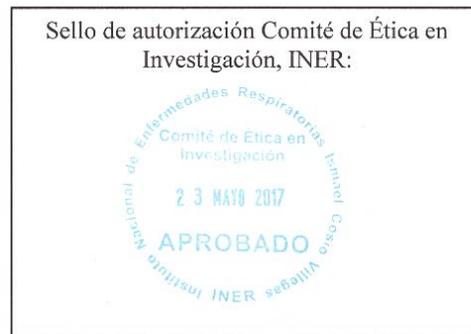
Certifico que, de acuerdo a lo mejor de mis conocimientos, el paciente (o su representante legal) que firmó este consentimiento recibió una explicación clara, completa y detallada de las investigaciones que se realizarán con su muestra de sangre y ha tenido oportunidad de preguntar libremente cualquier duda respecto a la naturaleza, riesgos y beneficios de su participación en este protocolo de investigación, así como una copia del presente consentimiento informado.

Este formato de consentimiento informado cumple con los lineamientos establecidos en el reglamento de la Ley General de Salud en materia de Investigación para la salud. Todos los conceptos plasmados en la declaración de Helsinki se han cumplido.

Investigador responsable: **Dr. Ramcés Falfán-Valencia**
Contacto: Laboratorio de HLA, 5487 1700 ext. 5152; Para preguntas, dudas y problemas relacionados a su participación con la Dra. Rocío Chapela M. Presidenta del Comité de Ciencia y Bioética en Investigación del INER: 5487 1700 exts. 5254 y 5110.

Fecha: _____
Día Mes Año

Firma del Investigador _____





INER

Anexo VII Cuestionario, parte 2/2

**Instituto Nacional de Enfermedades Respiratorias
Ismael Cosío Villegas**



Antecedentes hereditarios:

Enfermedad	Usted	Familiar	Parentesco
EPOC, Asma, CA de pulmón, fibrosis: _____			
Hipertensión arterial (HTA)			
Diabetes <i>mellitus</i> tipo II			
Enfermedades autoinmunes (LES, AR, DM)			
Cardiopatías (Enf. del corazón): _____			
Alergias (alimentarias, medicamentosas, ambientales).			

En caso de padecer EPOC, ¿hace cuánto tiempo lo diagnosticaron?: _____ y ¿Cuánto tiempo antes inicio con síntomas relacionados (dificultad para respirar, tos crónica, etc.)? _____

Tratamiento médico (S/N): Ventolin (Salbutamol) (___) Foradil (Formoterol) (___) Onbrize (Indacaterol) (___) Seretide (Fluticasona/Salmeterol) (___) Atrovent (Bromuro de Ipratropio) (___) Combivent (Salbutamol/Bromuro de Ipratropio) (___) Miflonide (Budesonida) (___) Spiriva (Bromuro de Tiotropio) (___) Otro: _____

Actualmente ¿es usted fumador? (S/N) _____ o ¿exfumador? (>1 año) (S/N)? _____

¿A qué edad empezó a fumar? _____ y si dejó de fumar ¿a qué edad lo hizo? _____

¿Cuántos años ha fumado de manera continua? _____ ¿Cuántos cigarrillos por día? _____

Calcule el índice tabáquico [(Cigarros al día) x (años fumando)]/20: _____

Las personas que lo rodean ¿fuman? _____, ¿Con qué frecuencia? _____

En el lugar donde vive ¿utilizan leña para cocinar u otros usos? _____

¿Con que frecuencia? (hrs. /día, años) _____ ¿Se expone directamente? _____

Calcule el índice de exposición al humo de leña [(Horas día) x (años exposición)]: _____

¿A qué se dedica usted? _____ ¿Es una actividad al aire libre? _____

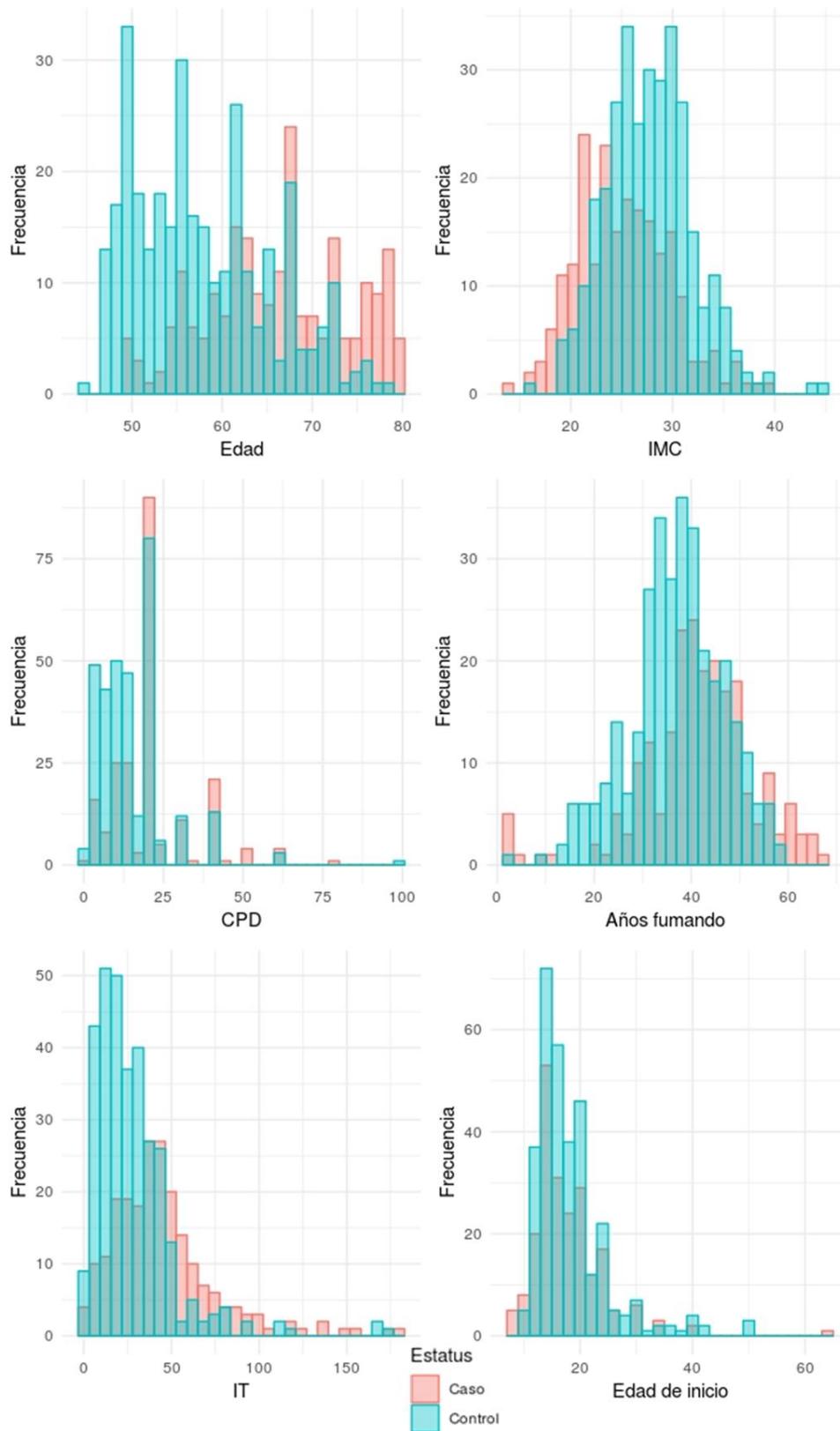
¿Se expone a humos (quema de pastos, sembradíos, caña, etc.), vapores, gases y/o solventes (industria textil, pinturas, calzado, etc.), polvos (asbesto, cemento, productos de construcción en general)? _____
Sí a ¿Cuál? _____

En caso afirmativo, ¿utiliza usted equipo de seguridad (mascarilla, cubrebocas, etc.)? _____

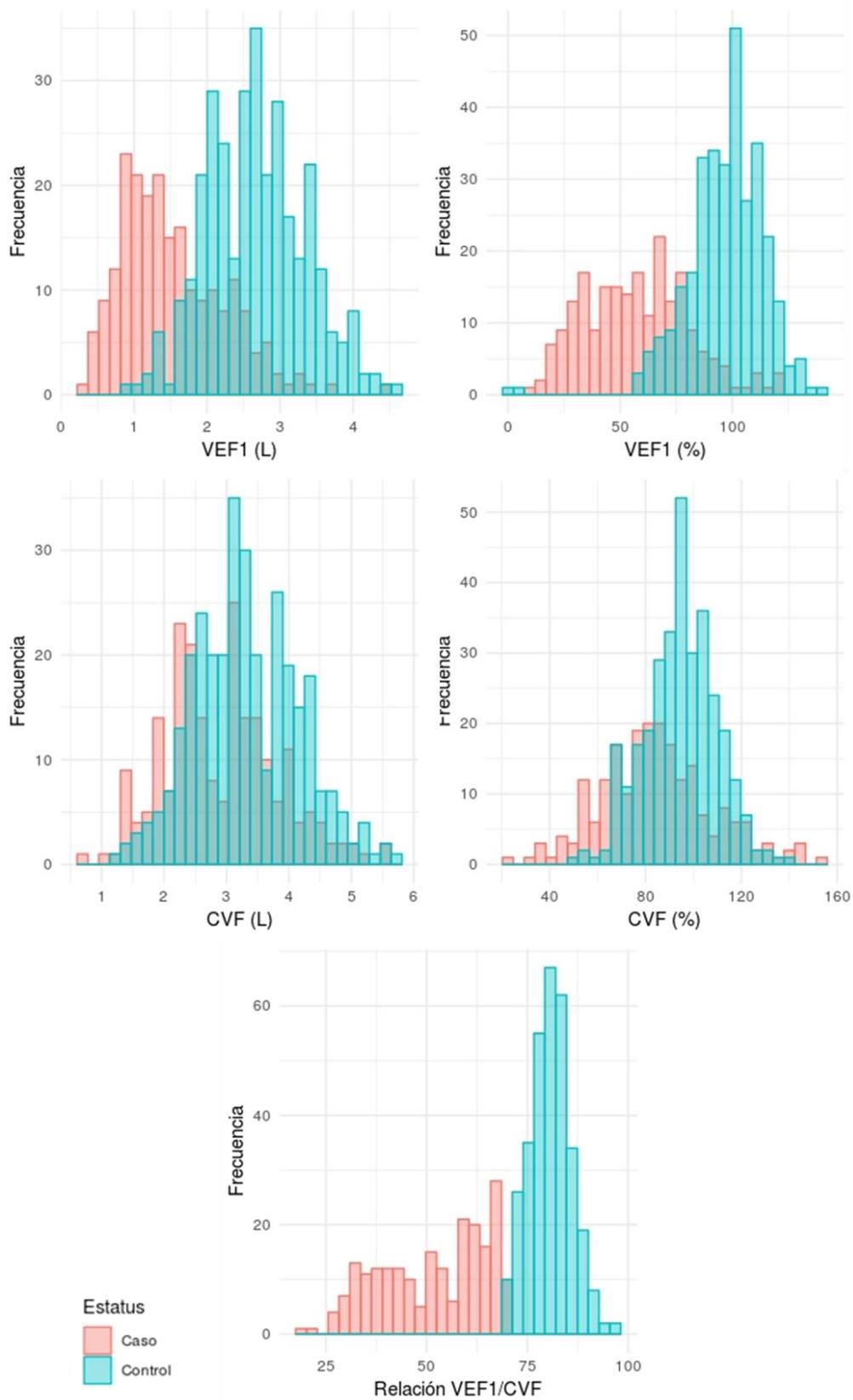
¿Convive con animales domésticos o de granja? _____ ¿Cuáles y cuántos? _____

Entrevistador: _____ Firma: _____

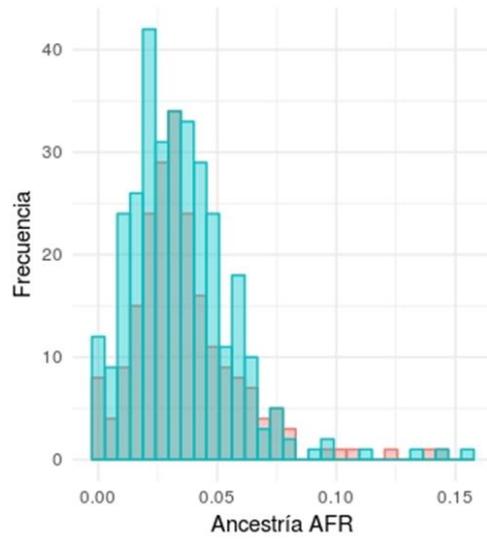
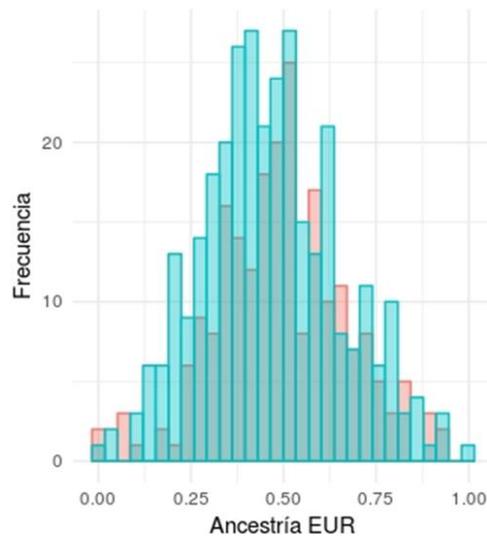
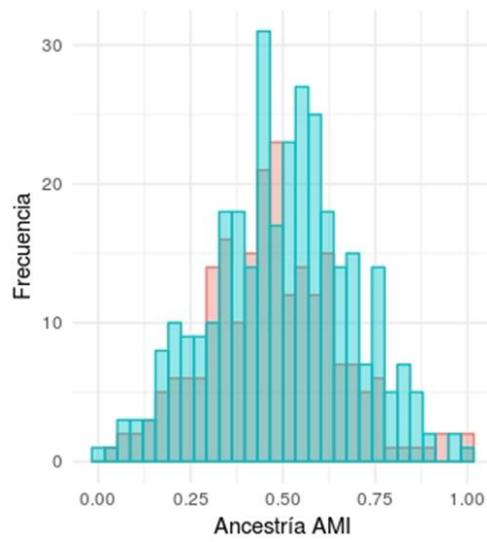
Anexo X Histogramas de variables demográficas y de tabaquismo.



Anexo XI Histogramas de variables de función pulmonar.

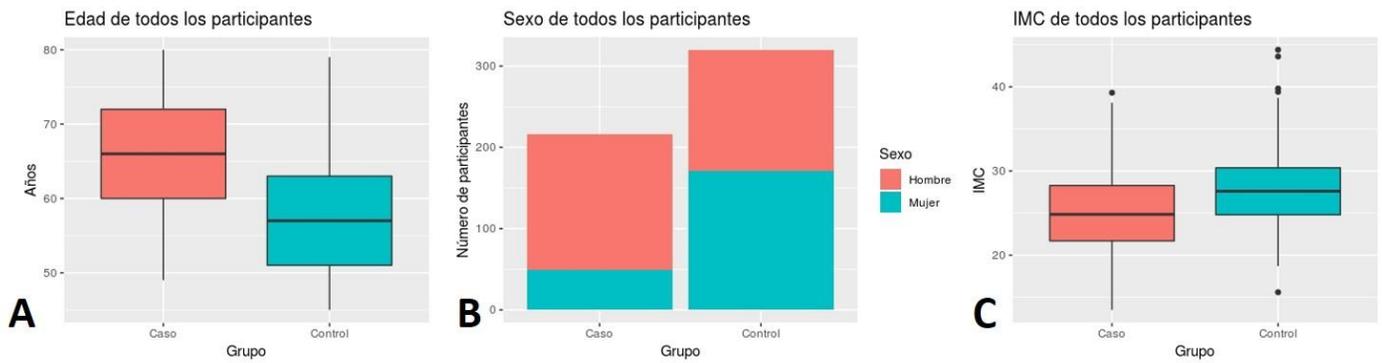


Anexo XII. Histogramas de componentes ancestrales.

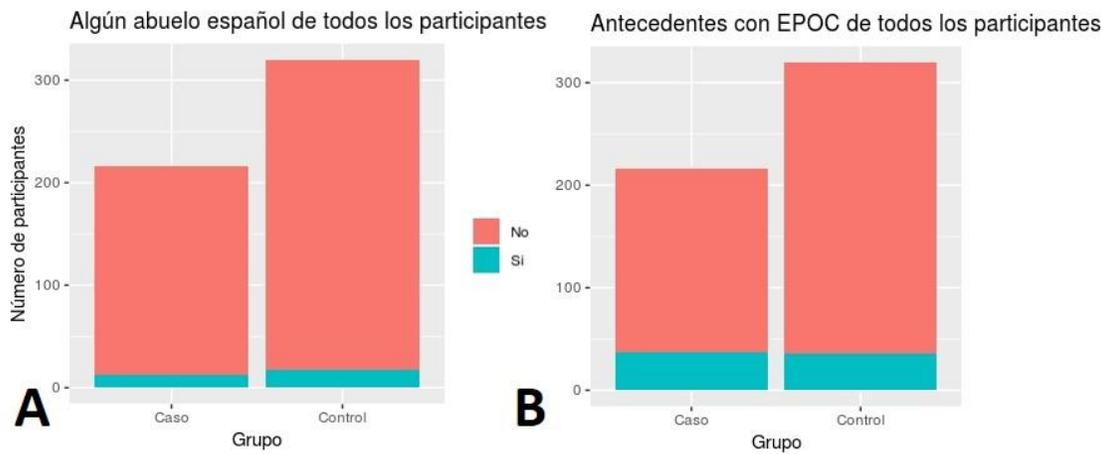


Estatus
Caso
Control

Anexo XIII Variables demográficas y de antecedentes de los grupos de estudio.

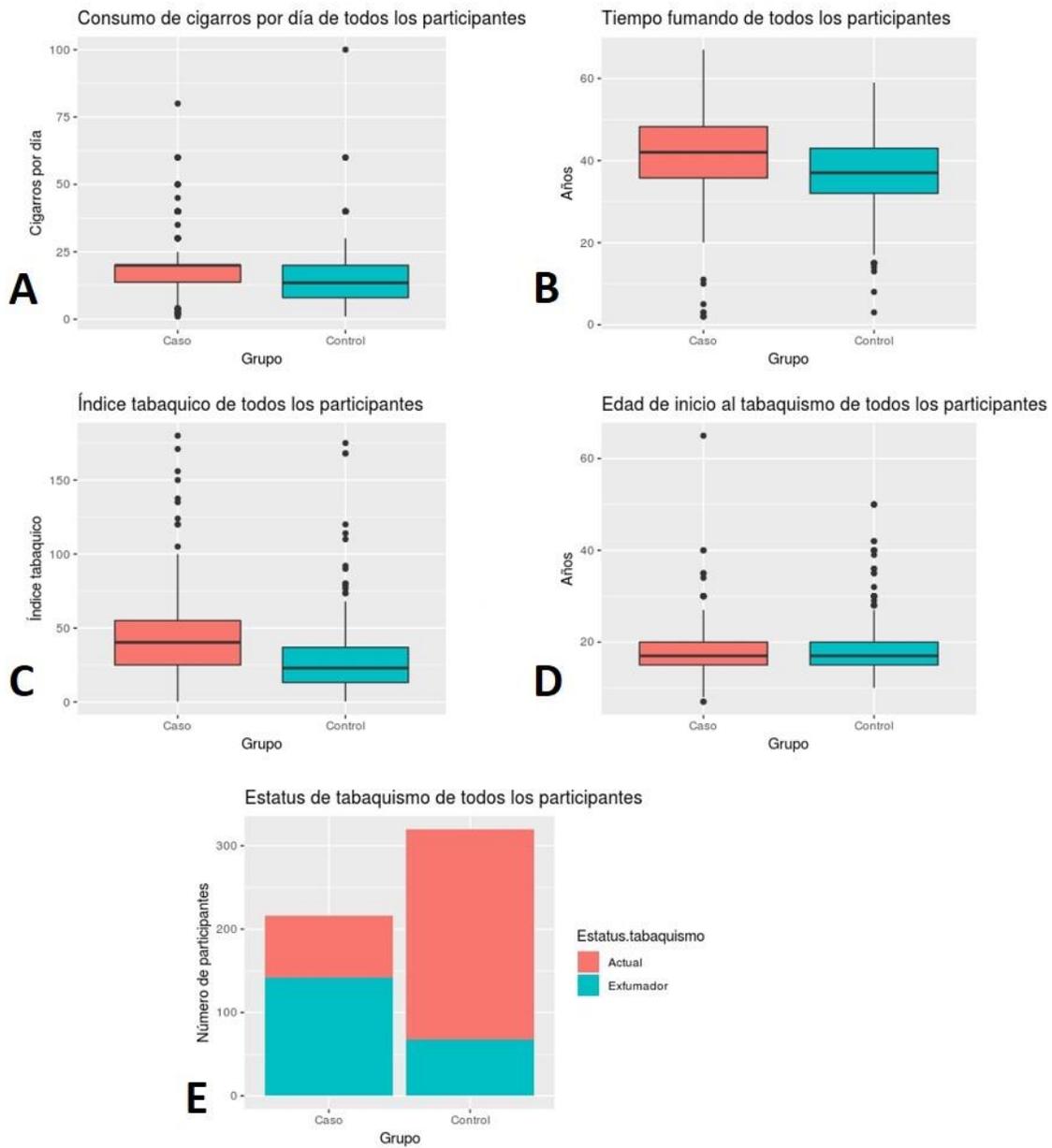


A. Edad en años, **B.** Sexo, **C.** IMC.



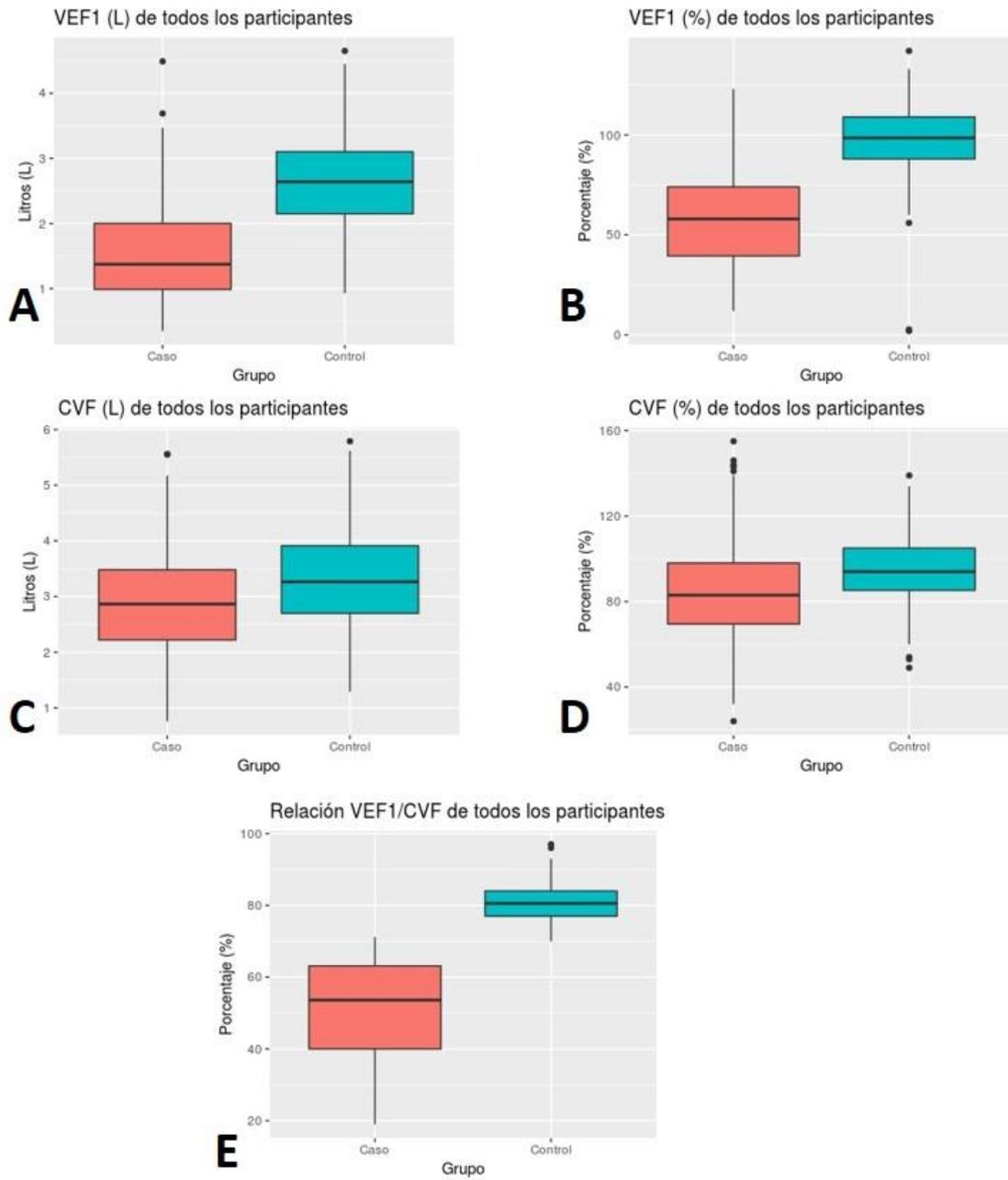
A. Participantes con algún abuelo español, **B.** Participantes con antecedentes de EPOC

Anexo XIV Variables de tabaquismo en los grupos de estudio.



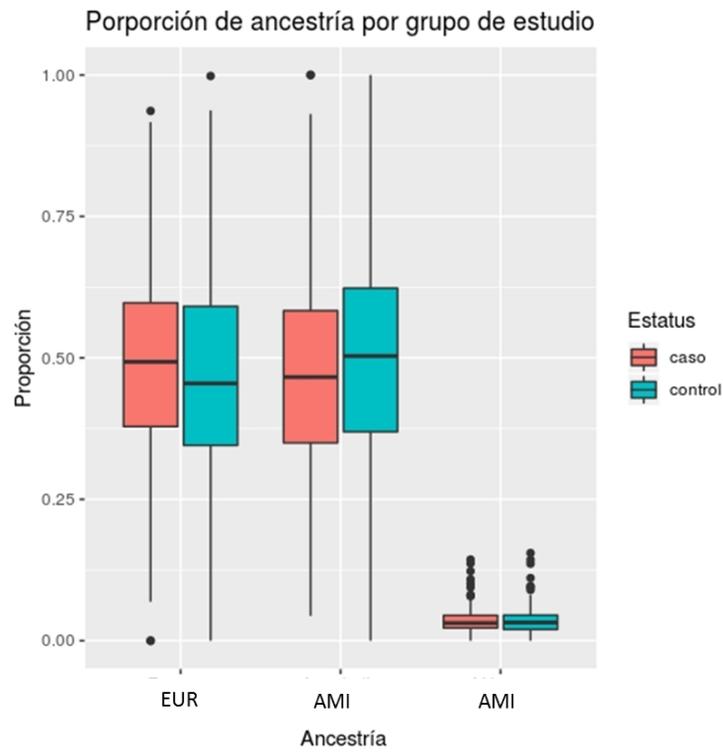
A. Consumo de cigarras por día, **B.** Tiempo fumando en años, **C.** Índice tabáquico en paquetes-año, **D.** Estatus actual de tabaquismo.

Anexo XV Variables espirométricas de los grupos de estudio.

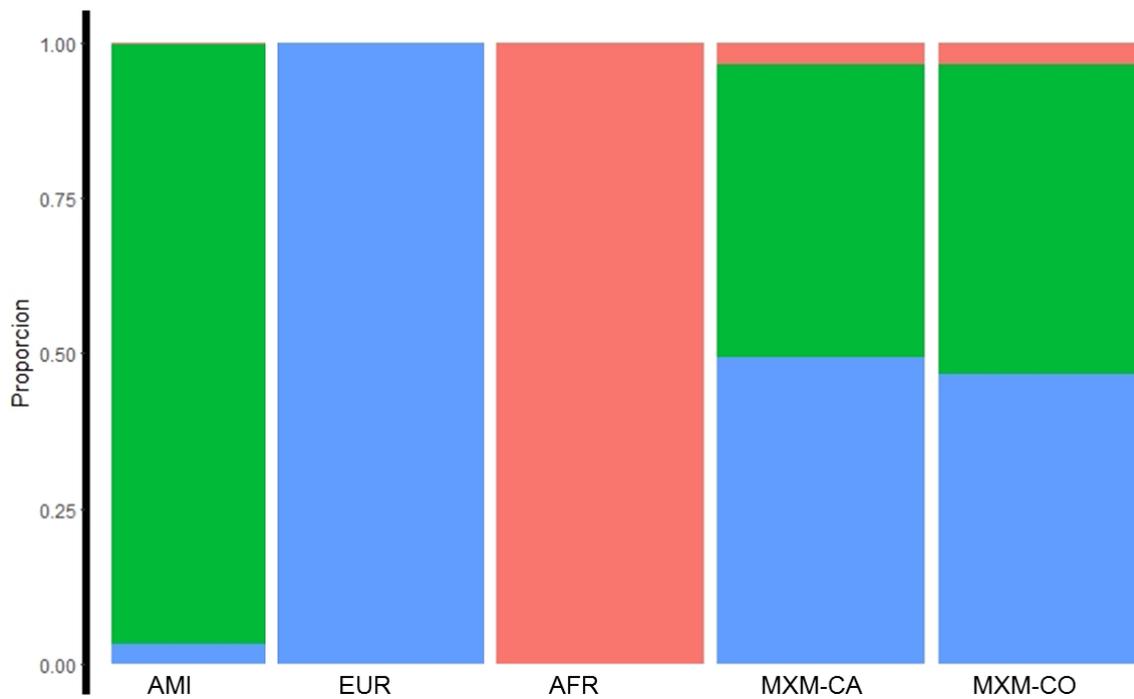


A. VEF₁ en Litros, **B.** VEF₁ en porcentaje, **C.** CVF en Litros, **D.** CVF en porcentaje. **E.** Relación VEF₁/CVF en porcentaje.

Anexo XVI Variables de ancestría en los grupos de estudio.

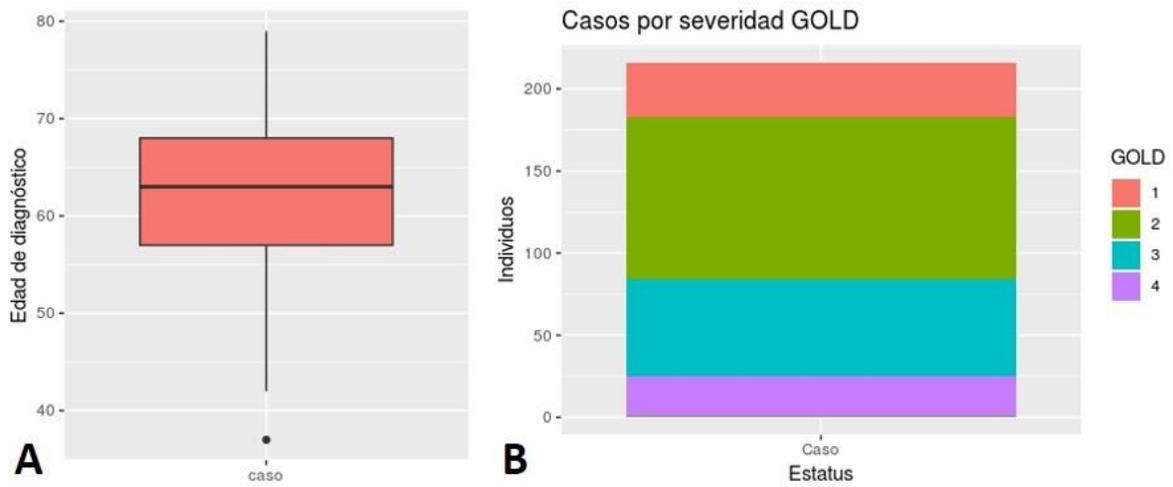


Componentes ancestrales en el grupo de estudio



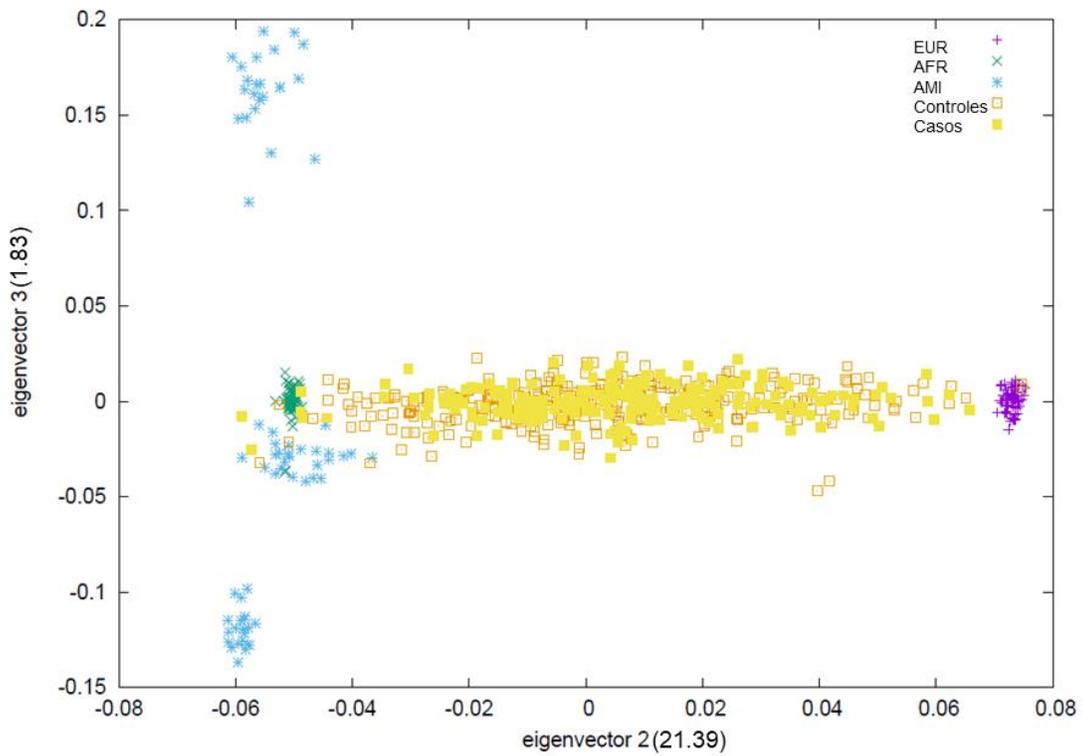
Proporción de los componentes ancestrales promedio de grupo de estudio. MXM-CA: Casos mexicano mestizos, MXM-CO: Controles mexicanos mestizos

Anexo XVII Variables clínicas en los pacientes con EPOC

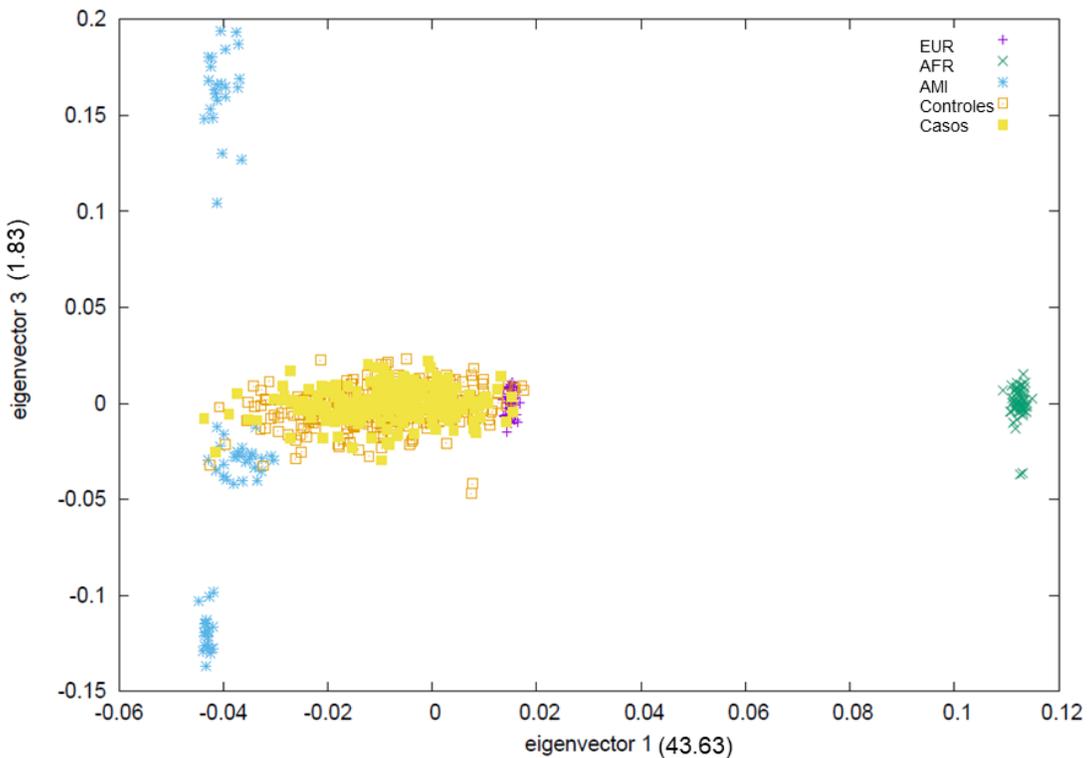


A. Edad de diagnóstico en años, **B.** Proporción de pacientes con EPOC según estadio GOLD.

Anexo XVIII Análisis de componentes principales entre PC1-PC2 y PC3 en los grupos de estudio



Análisis de componentes principales entre PC2 y PC3 en los grupos de estudio



Análisis de componentes principales entre PC1 y PC3 en los grupos de estudio

Anexo XIX Frecuencias alélicas de las variantes asociadas de acuerdo al modelo resultante.

SNP	Alelo asociado	Frecuencia en estudio			Frecuencia en 1kGP		
		Global	Casos	Controles	Global	Europeos	Americanos
Modelo sin ajustar (modelo 0)							
rs6110505	C	0.48	0.57	0.42	0.45	0.63	0.49
rs12518621	A	0.44	0.35	0.5	0.34	0.26	0.43
16:67993643-T-C	C	0.21	0.14	0.26	ND	ND	ND
exm-rs255049	C	0.21	0.14	0.26	0.34	0.19	0.26
rs4711289	A	0.25	0.33	0.21	0.29	0.22	0.25
16:67977382-T-C	C	0.20	0.13	0.25	ND	ND	ND
16:67833171-T-C	C	0.19	0.12	0.23	ND	ND	ND
exm2264618	C	0.19	0.12	0.23	ND	ND	ND
Modelo ajustado							
rs4559136	A	0.12	0.07	0.015	0.19	0.06	0.10
JHU_15.94073449	G	0.03	0.06	0.01	ND	ND	ND
rs2901070	T	0.08	0.12	0.06	0.04	0.02	0.08
rs352523	T	0.23	0.30	0.19	0.33	0.17	0.22
rs4320447	C	0.12	0.08	0.15	0.21	0.06	0.10
rs9363933	T	0.21	0.27	0.16	0.25	0.12	0.19
rs17257107	G	0.11	0.16	0.08	0.12	0.18	0.11
1:26450101-A-G	G	0.11	0.16	0.08	ND	ND	ND
Estratificado por sexo							
rs2762924	A	0.3	0.26	0.32	0.37	0.32	0.28
rs10420331	A	0.37	0.43	0.33	0.53	0.46	0.41
rs11084307	C	0.27	0.32	0.23	0.3	0.36	0.24
Estratificado por estatus de tabaquismo							
rs3825365	C	0.43	0.38	0.47	0.35	0.13	0.37
rs7644995	C	0.13	0.11	0.14	0.09	0.23	0.12

Anexo XX Comparación de variantes asociadas a EPOC y función pulmonar previamente en la literatura y nuestro estudio.

Locus/loci	SNP	Alelo de efecto	Estudio				Reporte previo		Frecuencia en 1kGP			
			OR	IC 95%	p	MAF global	OR/ β ^{Ref}	Rasgo asociado	EUR	AMR	AFR	EAS
HHIP	rs13147758	G	0.96	0.69-1.34	0.845	0.278	0.25 ^{114 §}	VEF ₁ /CVF	0.420	0.320	0.042	0.313
	rs1828591	G	1.07	0.77-1.49	0.662	0.335	0.72 ^{108 §}	EPOC	0.426	0.330	0.482	0.313
	rs13118928	A	0.97 ⁺	0.69-1.35 ⁺	0.831	0.676	0.72 ^{108 §}	EPOC	0.578	0.700	0.915	0.686
	rs12504628	C	0.98	0.70-1.37	0.936	0.334	-0.07 ^{109 §}	VEF ₁ /CVF	0.421	0.320	0.092	0.315
	rs1980057	T	1.00	0.71-1.40	0.982	0.321	0.37 ^{123 μ}	VEF ₁ /CVF	0.422	0.300	0.042	0.314
	rs13141641	T	1.07 ⁺	0.77-1.51 ⁺	0.673	0.680	1.32 ^{122 §}	EPOC	0.575	0.710	0.951	0.674
	rs13118083	A	1.00	0.72-1.39	0.967	0.431	1.23 ^{113 §}	EPOC	0.494	0.500	0.896	0.514
CHRNA3/5	rs1051730	A	1.08	0.73-1.58	0.687	0.234	1.17 ^{127 §}	EPOC	0.369	0.220	0.087	0.027
	rs12914385	T	1.05	0.72-1.53	0.777	0.261	1.39 ^{117 §}	EPOC	0.405	0.230	0.165	0.289
	rs17486278	A	0.90 ⁺	0.62-1.31 ⁺	0.569	0.747	1.43 ^{122 §}	EPOC	0.634	0.780	0.703	0.711
FAM13A	rs1903003	C	0.70	0.51-0.97	0.033	0.465	0.76 ^{125 §}	EPOC	0.446	0.530	0.370	0.429
	rs2869967	C	1.44	1.04-1.99	0.025	0.468	1.29 ^{125 §}	EPOC	0.400	0.420	0.670	0.542
	rs6830970	A	0.89	0.64-1.23	0.510	0.478	0.30 ^{115 §}	VEF ₁ /CVF	0.659	0.530	0.575	0.489
	rs7671167	C	0.64	0.47-0.89	0.007	0.486	0.73 ^{116 §}	EPOC	0.506	0.550	0.419	0.447
	rs2609255	T	1.47 ⁺	1.02-2.12 ⁺	0.041	0.742	0.51 ^{126 *}	VEF ₁ /CVF	0.759	0.700	0.695	0.583
	rs2609264	T	1.51 ⁺	1.05-2.17 ⁺	0.028	0.748	0.51 ^{126 *}	VEF ₁ /CVF	0.764	0.700	0.722	0.579
GSTCD-INTS12-NPNT	rs10516526	G	1.23	0.61-2.48	0.556	0.044	0.09 ^{123 μ}	VEF ₁	0.064	0.060	0.027	0.000
	rs17331332	A	1.08	0.52-2.22	0.820	0.043	56.79 ^{115 §}	VEF ₁	0.065	0.060	0.005	0.000
	rs34712979	A	1.28	0.77-2.12	0.326	0.099	1.26 ^{129 §}	EPOC	0.231	0.130	0.013	0.000
HTR4	rs7733088	A	0.96 ⁺	0.68-1.35 ⁺	0.809	0.636	0.81 ^{127 §}	EPOC	0.367	0.550	0.289	0.636
	rs3995090	C	0.96 ⁺	0.68-1.35 ⁺	0.787	0.645	0.03 ^{123 μ}	VEF ₁	0.381	0.550	0.247	0.719
	rs11168048	T	1.10	0.78-1.53	0.572	0.352	-0.40 ^{115 §}	VEF ₁ /CVF	0.612	0.450	0.807	0.282
TGFB2	rs993925	T	0.87	0.63-1.21	0.418	0.298	0.04 ^{123 μ}	VEF ₁ /CVF	0.309	0.290	0.337	0.464
ADGRG6	rs3817928	A	1.09 ⁺	0.96-1.69 ⁺	0.682	0.834	-0.57 ^{123 μ}	VEF ₁ /CVF	0.767	0.800	0.792	0.910
ADAM19	rs2277027	A	0.66 ⁺	0.48-0.90 ⁺	0.012	0.539	0.33 ^{123 μ}	VEF ₁ /CVF	0.642	0.660	0.359	0.874
MMP12	rs615098	A	0.56	0.32-0.98	0.043	0.094	0.01 ^{112 §}	VEF ₁ /CVF	0.190	0.140	0.368	0.150
THSD4	rs12899618	G	0.77 ⁺	0.40-1.47 ⁺	0.427	0.933	0.08 ^{123 μ}	VEF ₁ /CVF	0.857	0.910	0.884	0.898
ARMC2	rs2798641	T	1.00	0.72-1.39	0.977	0.315	-0.04 ^{123 μ}	VEF ₁ /CVF	0.183	0.270	0.039	0.285
PTCH1	rs16909898	A	0.60 ⁺	0.32-1.11 ⁺	0.100	0.949	0.38 ^{123 μ}	VEF ₁ /CVF	0.903	0.930	0.869	0.916
MECOM/EVI1	rs1344555	T	1.15	0.82-1.61	0.390	0.362	-0.03 ^{123 μ}	VEF ₁ /CVF	0.201	0.310	0.202	0.683
HDAC4	rs12477314	T	0.68	0.40-1.14	0.150	0.106	0.05 ^{123 μ}	VEF ₁ /CVF	0.198	0.150	0.011	0.288
RARB	rs1529672	C	0.90 ⁺	0.63-1.28 ⁺	0.546	0.718	-0.05 ^{123 μ}	VEF ₁ /CVF	0.853	0.730	0.784	0.606
MMP15	rs12447804	T	0.91	0.66-1.25	0.570	0.430	-0.04 ^{123 μ}	VEF ₁ /CVF	0.183	0.390	0.031	0.354
KCNE2	rs9978142	T	0.84	0.46-4.53	0.576	0.073	-0.05 ^{123 μ}	VEF ₁ /CVF	0.138	0.060	0.168	0.002

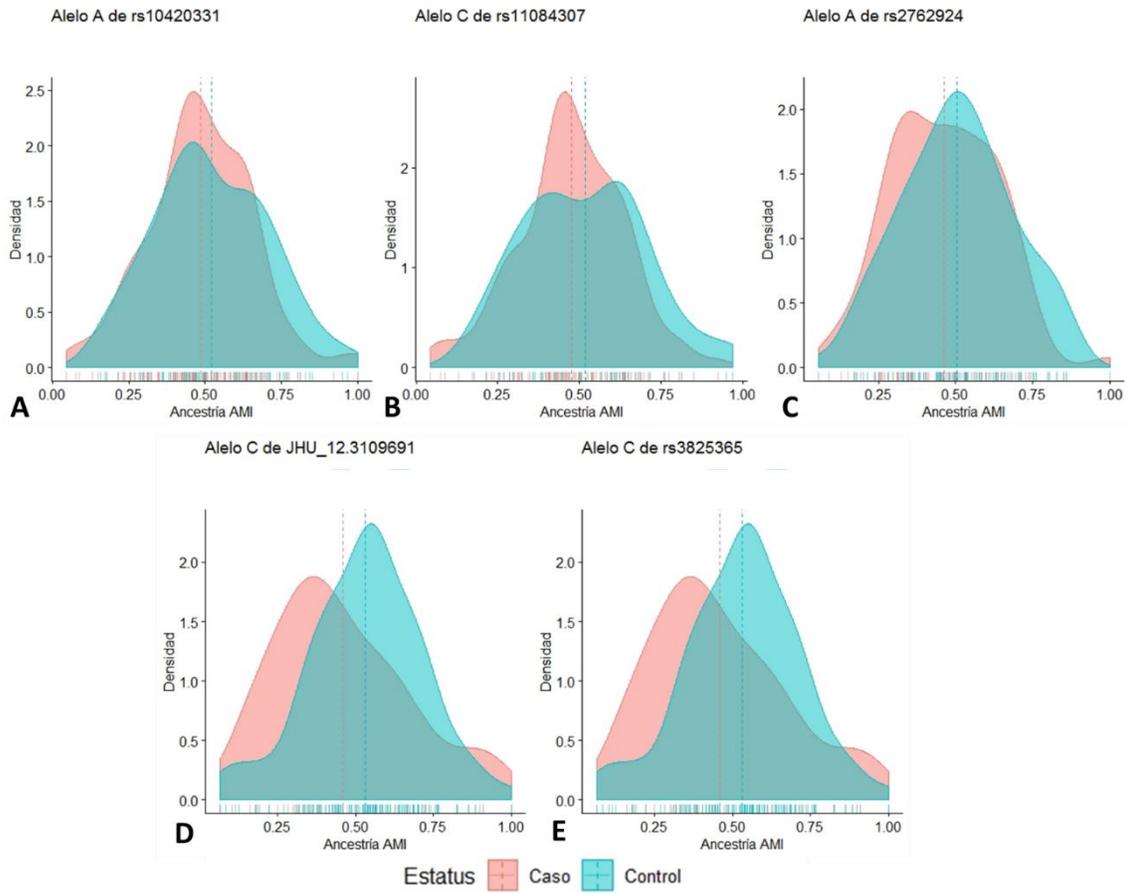
+: Valores inversos a los encontrados, §: Europeos, *: Asiáticos, μ: Hispanos.

Anexo XXI Variantes asociadas en el modelo ajustado estratificado por sexo y estatus de tabaquismo y su relación respecto al componente AMI en casos y controles.

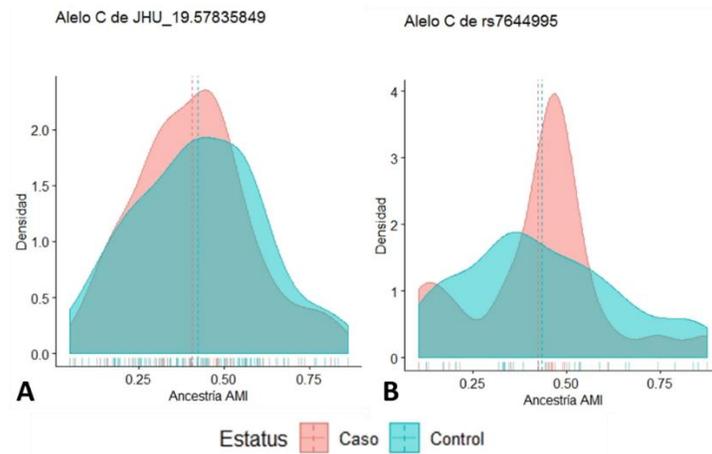
SNP	Alelo asociado	p de Normalidad*	Media AMI en alelo asociado (casos/controles)	p de diferencia entre medias [§]	Magnitud media del efecto
Hombres					
rs2762924	A	0.446	0.486/0.520	0.163	0.197
rs10420331	A	0.163	0.474/0.518	0.146	0.253
rs11084307	C	0.710	0.462/0.506	0.090	0.249
Mujeres					
JHU_12.3109691	C	0.482	0.458/0.531	0.139	0.392
rs3825365	C	0.482	0.458/0.531	0.139	0.392
Fumadores actuales					
JHU_19.57835849	C	0.233	0.407/0.424	0.555	0.092
Exfumadores					
rs7644995	C	0.069	0.423/0.433	0.859	0.003

*: Prueba de Shapiro-Wilk, §: T-Student para muestras independientes

Anexo XXII Distribución de los alelos asociados en el modelo ajustado estratificado por sexo y estatus de tabaquismo respecto a la ancestría amerindia.



Variantes asociadas en hombres A. rs10420331, B. rs11084307, C. rs2762924. Variantes asociadas en mujeres D. JHU_12.3109691, E. rs3825365



Variantes asociadas en fumadores actuales A. JHU_19.57835849. Variantes asociadas en exfumadores B. rs7644995.