



ESCUELA DE DOCTORADO
INTERNACIONAL DE LA USC

Lilian Andrea
Torres Tobar

Tesis doctoral

Determinantes genéticos de
riesgo y comorbilidades en
pacientes colombianos con
infección por coronavirus
SARS-CoV-2

Santiago de Compostela, 2025

Programa de doctorado en Medicina Molecular

TESIS DOCTORAL

DETERMINANTES GENÉTICOS DE RIESGO Y COMORBILIDADES EN PACIENTES COLOMBIANOS CON INFECCIÓN POR CORONAVIRUS SARS-COV-2

Autor

Lilian Andrea Torres Tobar

Director: Ángel Carracedo

Tutor/a: Ángel Carracedo

DECLARACIÓN ÉTICA DE LA TESIS

La autora declara no tener ningún conflicto de interés relacionado con la tesis.

Las figuras incluidas aquí, o bien son obra propia de la doctoranda, o producto de los análisis propios de los datos.

La presente tesis doctoral fue llevada a cabo bajo la aprobación del Comité de Ética en Seres Humanos CEISH del Hospital de San José y Fundación Universitaria de Ciencias de la Salud.

Anexo 2.

AGRADECIMIENTOS

Esta tesis es el resultado de aprendizajes, experiencias, trabajo y de superar muchas dificultades, para su culminación he recibido el apoyo de personas maravillosas a las cuales agradezco su apoyo incondicional.

Un reconocimiento especial a las personas que voluntariamente decidieron dar sus muestras e información de historias clínicas para ser parte del proyecto SCOURGE, gracias a cada uno de ellos hoy, tenemos hallazgos interesantes y podremos seguir trabajando en la genómica humana, relacionada con diferentes fenotipos, interacción con el ambiente y la presencia del SARS-Cov-2 entre otros. Así mismo un agradecimiento a los investigadores de SCOURGE especialmente a Liliana Porras, Zuleima Yañez y María Claudia Lattig y cada una de las personas que hicieron parte de sus equipos de trabajo para vincular a SCOURGE muestras e historias clínicas de personas residentes en Pereira, Barranquilla y Bogotá.

Agradecimiento a los investigadores, residentes, enfermeras, auxiliares de laboratorio y estudiantes de la Fundación Universitaria de Ciencias de la Salud, del Instituto de Ciencias Básicas FUCS, del Hospital Infantil Universitario de San José y del Hospital de San José que participaron en el proceso de entrevistas a pacientes, toma de muestras y registro de historias clínicas, un agradecimiento especial a Luz Dary Gutiérrez y Carlos Maya, investigadores y amigos por sus aportes, dedicación a este proyecto y por su apoyo incondicional en todo momento.

A la Fundación Universitaria de Ciencias de la Salud, cada uno de los miembros del consejo superior especialmente al Dr. Sergio Parra Duarte por brindarme el apoyo para poder realizar el doctorado y culminar este proyecto.

A Raquel Cruz y Alicia de Coó Diz, un inmenso gracias porque sin su apoyo en la fase de análisis de datos y en la fase final de esta tesis, esto no habría sido posible.

A mis amigos y amigas, gracias por su compañía, apoyo, por sus comentarios siempre constructivos, su amistad y presión, me ayudaron a sobrevivir en los momentos difíciles.

A mi familia gallega, Montse, Mar, Guille, Marina, Ángel, gracias por recibirme siempre con amor y por su apoyo en todo momento.

A mi familia Torres, Tobar y Paredes, especialmente a mis primos Natalia, Catalina, Andrés Torres, Andrés Tovar y Rubén, por su apoyo incondicional, sus aportes fueron fundamentales para culminar este documento.

A mi abuelita, madre, hermano y cuñada les agradezco su apoyo constante, el estímulo permanente, su amor incondicional y por no permitir que yo abandonara este proyecto en los momentos más difíciles de mi vida.

A mi querido Ángel, director de tesis, maestro, amigo mi más sincero agradecimiento porque sin su conocimiento, enseñanzas, apoyo y cariño permanente yo no habría podido terminar esta tesis.

A Manuel el amor de mi vida, le agradezco, sus enseñanzas, su apoyo incondicional, sus contribuciones expertas e indispensables a este trabajo y por amarme hasta el final.

ABREVIATURAS

1KGP	1000 <i>Genomes</i> populations
ACE2	enzima convertidora de angiotensina 2
ACEI	enzima conversora de angiotensina I
AMR	poblaciones americanas mixtas
ARA	antagonistas de los receptores de angiotensina
AUC	<i>area under the ROC curve</i>
BOADICEA	<i>breast & Ovarian Analysis of Disease Incidence & Carrier Estimation algorithm</i>
CAC	Centro de Alto Costo
CADD	<i>Combined Annotation Dependent Depletion</i>
CEGEN	Centro Nacional de Genotipado
DBP	presión arterial diastólica
EPOC	enfermedad pulmonar obstructiva crónica
eQTL	<i>expression quantitative trait loci</i> , loci asociado a la expresión
FDR	<i>false discovery rate</i> , tasa de falsos positivos
FUMA	<i>Functional Mapping and Annotation of Genome-Wide Association Studies</i>
GenOMICC	<i>Genetics of Mortality in Critical Care</i>
GISAID	<i>Global initiative on sharing all influenza data</i>
GWAS	<i>Genome wide association studies</i>
HGI	<i>COVID-19 Host Genetics Initiative</i>
HTA	Hipertensión Arterial
HWE	Equilibrio de Hardy-Weinberg
IBD	<i>Identity by descent</i> / idénticos por descendencia
IC	Intervalo de confianza
ICBP	<i>International consortium of blood pressure</i>

IMC	índice de Masa Corporal
IMPUTE2	<i>Genotype imputation and haplotype phasing program software</i>
ISARICC	<i>International Severe Acute Respiratory & emerging Infection Consortium</i>
LASO	Consorcio Latinoamericano de Estudios en Obesidad
LD	<i>linkage disequilibrium</i> / desequilibrio de ligamiento
MA-GWAMA	<i>Genome-Wide Association Meta-Analysis software</i>
MAE	<i>mean absolute error</i> , error medio absoluto
MAF	<i>minor allele frequency</i> , frecuencia del alelo menor
METAL	<i>Meta-analysis genomewide association scans software</i>
MH	Hipertensión monogénica
MHC	<i>major histocompatibility complex</i>
MVP	<i>Million veteran program</i>
N-GWAMA	<i>Genome-Wide Association Meta-Analysis software</i>
NHS	<i>National Health Service</i>
OFH	<i>Our Future Health</i>
OMS	Organización Mundial de la Salud
OR	<i>Odds - Ratio</i>
PC	<i>Principal component</i> / componentes principales
PCA	<i>Principal component analysis</i> / Análisis de componentes principales
PGC	<i>Psychiatric Genome Consortium</i>
PH	Hipertensión poligénica
PP	presión de pulso
PRS	<i>Polygenic risk score</i> / Puntuación de riesgo poligénico
PSC	Catálogo de Score Poligénico
QC	<i>Quality Control</i> / Control de calidad
QQ	Gráfico Cuantil-Cuantil
QTL	<i>Quantitative trait loci</i>
SCOURGE	<i>Spanish Coalition to Unlock Research on host genetics</i>
TAG-VE	<i>Technical Advisory Group on SARS-CoV-2 Virus Evolution</i>

WES	<i>whole exome sequencing</i>
WGS	<i>whole genoma sequencing</i>
WHO	<i>World Health Organization</i>

INDICE

1. INTRODUCCIÓN 1

1.1 COVID-19 EPIDEMIOLOGÍA GLOBAL	1
1.2 EPIDEMIOLOGÍA EN COLOMBIA.....	2
1.3 EVOLUCIÓN MOLECULAR DEL GENOMA VIRAL	3
1.4 GWAS (GENOME-WIDE ASSOCIATION STUDIES)	5
1.4.1 GWAS flujo de trabajo	7
1.4.2 Imputación de genotipos.....	9
1.4.3 Valoración estadística de las asociaciones detectadas en GWAS.....	10
1.4.4 Representación gráfica de los resultados de un GWAS	11
1.4.5 Significado funcional de las asociaciones detectadas por estudios de GWAS.....	13
1.5 PRS (POLYGENIC RISK SCORE)	13
1.5.1 Control de calidad de un PRS	15
1.5.2 Cálculo del riesgo poligénico	17
1.5.3 Interpretación y presentación de resultados	18
1.5.4 Métricas usadas en la prueba de asociación.....	19
1.5.5 Representaciones gráficas de resultados	20
1.5.6 Utilidad clínica de las PRSs	20
1.5.7 Debate bioético sobre el uso o mal uso de las PRSs	22
1.6 HIPERTENSIÓN ARTERIAL Y GENÉTICA.....	23
1.6.1 Definición y diagnóstico.....	23
1.6.2 Mortalidad.....	23
1.6.3 Prevalencia	24
1.6.4 Fisiopatología de la HTA	25
1.6.5 Genética de la hipertensión	26
1.6.6 Hipertensión arterial y COVID	28

2. OBJETIVOS

29

3. METODOLOGÍA.....

30

3.1 CONSORCIO SCOURGE

30

3.2 PARTICIPANTES DEL ESTUDIO SCOURGE-COLOMBIA

30

3.2.1 Criterios de inclusión

30

3.2.2 Criterio de exclusión	31
3.2.3 Recolección de información clínica.....	31
3.2.4 Consideraciones éticas	32
3.3 EXTRACCIÓN DE ADN Y GENOTIPADO	34
3.4 CONTROL DE CALIDAD E IMPUTACIÓN DE VARIANTES	35
3.5 ANÁLISIS DE ASOCIACIÓN EN POBLACIÓN COLOMBIANA.....	35
3.6 ANOTACIÓN Y MAPEO INICIAL	36
3.7 ANÁLISIS DE SCORE DE RIESGO POLIGÉNICO	36
3.7.1 Puntuación de riesgo poligénico (PRS).....	36
3.7.2 Muestra diana	36
3.7.3 Muestra descubrimiento.....	36
3.7.4 Análisis estadísticos	37
4. RESULTADOS Y DISCUSION	37
4.1 RESULTADOS GWAS PARA EL RIESGO DE HOSPITALIZACIÓN AJUSTANDO PARA COMORBILIDADES CARDIOVASCULARES	44
4.2. ANÁLISIS A NIVEL DE GENES EN EL GWAS HOSPITALIZACIÓN AJUSTANDO PARA ANTECEDENTES CARDIOVASCULARES	47
4.3 GWAS PARA RIESGO DE SEVERIDAD 4 AJUSTANDO POR ANTECEDENTES CARDIOVASCULARES	50
4.4 ANÁLISIS A NIVEL DE GENES EN EL GWAS SEVERIDAD 4 AJUSTANDO PARA ANTECEDENTES CARDIOVASCULARES.....	54
4.4.1 PRS hipertensión - Asociación con riesgo de hospitalización por COVID-19	56
4.4.2 PRS hipertensión - Asociación con riesgo de gravedad crítica por COVID-19	60
4.4.3 PRS hipertensión - Asociación con antecedentes vasculares e hipertensión arterial	65
5. CONCLUSIONES... ..	73
6. BIBLIOGRAFÍA... ..	77
7.ÍNDICES COMPLEMENTARIOS.....	88
8. ANEXOS.....	92
ANEXO 1.....	92
ANEXO 2.....	93

RESUMEN

La severidad de la infección por el SARS Cox-2 y sus comorbilidades tiene un componente genético y un componente ambiental con una heredabilidad estimada de alrededor del 50% .

El descubrimiento de ese componente genético se está realizando con estudios de asociación de genomas completos (GWAS) para lo que se constituyeron muchos consorcios agrupados en una iniciativa global a nivel mundial (COVID-19 Host Genetics Initiative, HGI). Uno de los consorcios más importantes es SCOURGE que es uno de los que posee más información de variables clínicas de los pacientes y que contó con la participación de 46 grupos clínicos, 14 grupos de investigación, 9 biobancos de España y 22 grupos de Latinoamérica, de los cuales 4 grupos son colombianos.

Esta tesis es el resultado de la contribución de Colombia a este consorcio y parte de 1354 pacientes COVID con diferentes grados de afección y tiene como objetivo determinar variantes genéticas de severidad para la COVID y analizar el papel del riesgo cardiovascular en la severidad y hospitalización por COVID-19, aspecto que ha sido insuficientemente estudiado hasta la actualidad.

Se encontró una clara correlación entre las comorbilidades cardiovasculares y la hospitalización y severidad. Se construyó un GWAS con el requerido control de calidad y análisis funcional de resultados ajustando para comorbilidades cardiovasculares. Para el GWAS de riesgo de gravedad crítica ajustando para antecedentes cardiovasculares se observan dos SNPs, rs143861937; rs73078305, ubicados en el cromosoma 3 que pasan el umbral de referencia ($p=6.4 \times 10^{-9}$), cercanos a los genes PLXNB1 y al gen PFKFB4 que no se reportan en los estudios de SCOURGE, HGI ni en el metaanálisis de GenOMICC. PLXNB1 codifica para una proteína neurológica desregulada en el plasma de pacientes con COVID severa que se considera una diana prometedora para intervenciones terapéuticas, al igual que PKFB4 que actúa como un factor impulsor en la mediación de la liberación de citoquinas y se ha relacionado con infecciones virales y la COVID. La gran mayoría de los SNPs y genes asociados a hospitalización y severidad con valores de p inferiores a 10^{-6} se encuentran en genes que se han relacionado con mecanismos de patogenicidad del SARS-CoV2 y severidad para COVID. Las estimas de riesgo poligénico de hipertensión (presión arterial sistólica -SBP-, presión arterial diastólica -DBP- y presión de pulso -PP-) para predecir la hospitalización por COVID-19 (HOSP) y la gravedad de la COVID 19 (SEV4) arrojaron resultados no significativos.

1. INTRODUCCIÓN

1.1 COVID-19 EPIDEMIOLOGÍA GLOBAL

Desde que las autoridades de salud en China reportaron el 31 de diciembre de 2019 la aparición del virus SARS-CoV-2 en Wuhan, este se diseminó rápidamente en 5 países del sureste asiático y 7 días después ya se encontraba en 23 países, 6 de ellos europeos. En el transcurso de 10 meses se habían confirmado más de 30 millones de casos en todo el mundo (B. Hu et al., 2021). La Organización Mundial de la Salud (OMS) el 30 de enero de 2020 declaró la Emergencia de salud pública de interés internacional (PHEIC-MOS) y posteriormente el 11 de marzo de 2020 la pandemia global COVID-19. La epidemiología ha evolucionado con una tendencia a la disminución de muertes y hospitalizaciones y a altos niveles de inmunidad frente al virus. Ante esta situación la OMS declaró el 5 de mayo de 2023 que la COVID es un tema de salud en desarrollo pero que ya no constituía una emergencia internacional, aunque no está claro aún el criterio para establecer el nivel de “equilibrio endémico” de la infección a nivel global.

Entre tanto, la pandemia SARS-CoV-2 infectó a más de 560 millones de personas en el 2019 y de ellos fallecieron 6,4 millones. Ya en septiembre del 2022 se registraban más de 614 millones de infectados y 6,5 millones de fallecidos (Wolf et al., 2023). Finalmente, el registro a 31 de diciembre del 2023 es de más de 773 millones de casos de COVID-19 reportados por la Organización Mundial de la Salud y más de 7 millones de muertes (OMS reporte: <https://data.who.int/dashboards/covid19/cases>). Al mismo tiempo, la pandemia generó un desarrollo científico y tecnológico sin precedentes y facilitó el desarrollo de estrategias de vacunación junto con medidas de salud pública adecuadas en muchos países. En noviembre del 2023 cerca de 13.6 billones de dosis de vacunas COVID-19 habían sido administradas salvando más de 10 millones de vidas globalmente alterando así significativamente el curso de la pandemia (Ulrichs et al., 2024). En la distribución por sexo, se observó una mayor incidencia en hombres, quienes representaron aproximadamente el 60% de los casos e igualmente, la letalidad fue mayor en estos según el reporte de la OMS antes citado.

Se identificó una asociación entre enfermedades de base de los pacientes con la gravedad y el exitus. Así, múltiples investigaciones reportaron la hipertensión arterial, la diabetes mellitus, la enfermedad arterial coronaria y la obesidad como comorbilidades asociadas a un mayor riesgo de mortalidad en pacientes con COVID-19 (Yomayusa-Gonzalez et al., 2021). Casi todas las series indicaban que la hipertensión aumenta el riesgo de infección grave por COVID-19 o de la mortalidad, así como el ingreso en la UCI, aunque a veces no era claro si era independiente de otros factores de riesgo. Los pacientes con hipertensión presentan un riesgo 2.98 veces mayor de padecer una enfermedad grave, un riesgo 1.82 veces mayor de padecer

una enfermedad crítica y un riesgo de mortalidad entre 2.17 y 2.88 veces mayor, en comparación con los pacientes sin hipertensión (Razai et al., 2020)

Por otra parte, existía la preocupación de que las personas a las que se administraba inhibidores de la ECA o antagonistas de los receptores de la angiotensina II pudieran presentar un mayor riesgo de infección o enfermedad grave debido al aumento de la expresión del receptor de la ECA2, sin embargo, los datos de las series sugieren que el uso de estos fármacos no está asociado con la enfermedad grave, y no hay ninguna asociación entre el uso de estos medicamentos y un resultado positivo de la prueba del SARS-CoV-2 entre los pacientes sintomáticos. (García García, 2020)

1.2 EPIDEMIOLOGÍA EN COLOMBIA

El primer caso de enfermedad por coronavirus (COVID-19) en Colombia fue registrado en febrero de 2020 en un viajero proveniente de Italia. La confirmación de este primer caso se dio el 6 de marzo y por esta época ya se habían reportado casos en Brasil, Ecuador y Panamá. Ya el 26 de marzo en Colombia se habían registrado 470 casos, y más del 10% de los casos no tenían fuente de exposición conocida. En abril se habían reportado más de 2.500 casos, con 214 de pacientes recuperados y 100 fallecidos (Visscher & Goddard, 2019a). El análisis de las características clínicas y demográficas de los primeros 100 casos fatales por COVID-19 en Colombia, mostraron que 63 fueron hombres, con un promedio de edad de 65,75+/- 18,11, las comorbilidades relacionadas fueron hipertensión arterial, diabetes mellitus, enfermedades cardiovasculares y cerebrovasculares, enfermedad pulmonar obstructiva crónica, obesidad, tabaquismo y patología tiroidea (Marín-Sánchez, 2020a).

El 10 de junio del 2020, la letalidad reportada por el Instituto Nacional de Salud era del 3.3%, sin embargo, estos datos no reflejaban la información completa dado que todo dependía del número de pruebas diarias realizadas, la posibilidad de acceder a los exámenes, así como el registro adecuado de cada uno de los casos (Rosselli, 2021). La mayoría de los pacientes colombianos que fueron aislados presentaron síntomas (tos, fiebre) en el 59,6% y el resto de los pacientes presentaron más de 10 síntomas diferentes, 10 pacientes fallecieron, de los cuales la mayoría tenía enfermedades subyacentes (Laiton-Donato et al., 2020). Desde el 29 de marzo de 2020 hasta el 8 de diciembre de 2022, el Ministerio de Salud colombiano reportó 6.323.357 casos confirmados de COVID-19 y 141.943 defunciones, lo que representa una incidencia nacional de 12.553,2 casos por cada 100.000 habitantes y una tasa de mortalidad de 2.817,9 por cada millón de habitantes (INS Instituto nacional de salud, 2024). Para el 8 de diciembre de 2022, el país reportaba 5.887 casos activos, de los cuales el 10% (589 pacientes) se encontraban hospitalizados y el 2,2% requerían cuidados intensivos (UCI). Esta tendencia se mantuvo a lo largo de la pandemia, con el porcentaje de hospitalizados oscilando entre el 10% y el 27,7% del total de casos activos, y entre el 0,2% al 6,1% de los pacientes que requirieron ingreso en UCI (Instituto nacional de salud, 2024; Organización mundial de la salud) (<https://www.who.int/publications/m/item/covid-19-epidemiological-update-24-december-2024>).

La hipertensión fue la comorbilidad más frecuentemente reportada con alrededor de un 40% de los casos, seguido por la diabetes y la obesidad, no obstante, algunos estudios reportan la enfermedad arterial coronaria como una comorbilidad frecuente asociada a mayor riesgo de

mortalidad en pacientes diagnosticados con COVID-19 (Henríquez et al., 2022) (Reyes et al., 2022) (Marín-Sánchez, 2020b). Un metaanálisis publicado en 2021 encontró que la hipertensión, junto con la diabetes y las enfermedades cardiovasculares, aumentaba significativamente el riesgo de mortalidad en pacientes hospitalizados por COVID-19. Así mismo, la obesidad también fue identificada como un factor independiente de peor pronóstico, aumentando el riesgo de complicaciones graves y necesidad de ventilación mecánica (Cáceres Ramírez et al., 2023). En un estudio realizado en Bogotá utilizando datos de un esquema de vigilancia epidemiológica se concluye que alrededor del 53% (IC 95%: 45-62%) de la población había tenido infección por SARS-CoV-2 para el 2021, en ese mismo año se analizaron diferentes dinámicas de la infección según grupos de trabajadores sus funciones y lugares de trabajo, encontrando que aquellos trabajadores a los que les fue posible realizar sus actividades a distancia mantuvieron bajas tasas de infección, mientras que otros trabajadores que debían desplazarse y además tener contacto con posibles casos contribuyeron a los principales picos de la infección: cajeros, comerciantes, guardias de seguridad y todo el personal que participaba en la atención en salud. En el resto del país la tasa de infección era más alta que en Bogotá. Es posible que esto se deba a que las otras regiones diferentes a Bogotá tienen una menor capacidad institucional y de laboratorios y por ello no se registran ni se estudian todos los casos. Los resultados encontrados en Bogotá no implican que el 50% de la población tengan inmunidad dado que es posible que, aunque pueda observarse disminución de la transmisión también dicha inmunidad puede verse afectada por la presencia de nuevas variantes (Laajaj et al., 2021).

Durante el 2022 se publicaron diversos trabajos que tenían como objeto describir las características clínicas y mortalidad de los pacientes con COVID-19 en diferentes centros hospitalarios, uno de ellos en Bogotá en donde adicionalmente observaron como variable la altitud (2640 m) sobre el nivel del mar. Fueron analizados 5161 pacientes, de los cuales 1.305 fallecieron (25%), 1.223 pacientes fueron ingresados a la UCI (24%) concluyeron que la edad avanzada, la baja relación S/F (menor 259) y lactato deshidrogenasa (LDH) mayor 617U/L al ingreso, fueron asociadas a predictores de mortalidad (Rodríguez Lima et al., 2022).

1.3 EVOLUCIÓN MOLECULAR DEL GENOMA VIRAL

En su reporte del 16 de marzo de 2023 la OMS informó que el virus ha continuado evolucionando desde su identificación. La entidad mundial de salud ha definido Variantes de Preocupación (VOCs) y Variantes de Interés (VOIs) según su potencial de expansión y reemplazo de variantes previas y por causar nuevas oleadas de circulación y creó en junio de 2020 el Grupo técnico para el estudio de la evolución del virus o *Technical Advisory Group on SARS-CoV-2 Virus Evolution* (TAG-VE). (<https://www.who.int/groups/technical-advisory-group-on-virus-evolution/about>)

La cepa ancestral Wuhan-1 ha seguido evolucionando, las principales variantes que han aparecido hasta la fecha han incluido el linaje Alfa (principalmente B.1.1.7), el linaje Delta (B.1.617.2 y todos los sublinajes AY) y el linaje Ómicron más reciente (B.1.1.519 y todos los sublinajes BA), cada uno de los cuales ha tenido una ventaja selectiva en comparación con la variante predominante anterior. Inicialmente, las cepas iniciales tuvieron una sola sustitución en la secuencia para la proteína de la espícula, lo que resultó en una ventaja de crecimiento de aproximadamente el 20% en relación con las variantes anteriores. Las variantes posteriores

aumentaron la transmisibilidad del virus en los humanos. A partir de octubre de 2020, las nuevas variantes se distinguieron por un mayor número de mutaciones en la proteína de la espícula (S) del virus, en particular para Ómicron, alterando así la transmisibilidad y la antigenicidad. La variante Ómicron predominante ha representado más del 98% de las secuencias disponibles públicamente desde febrero de 2022 (Ulrichs et al., 2024). Aunque en la mayoría de informes se ha considerado a Ómicron como más infectiva pero menos letal, hay reportes de la OMS, sobre sublinajes que son capaces de evadir la vacuna (<https://www.who.int/news/item/16-03-2023-statement-on-the-update-of-who-s-working-definitions-and-tracking-system-for-sars-cov-2-variants-of-concern-and-variants-of-interest>).

Ómicron representa el VOC más divergente observado hasta la fecha. Desde su aparición, Ómicron ha seguido evolucionando genética y antigénicamente generando cada vez más sublinajes caracterizados por algunas mutaciones que comprometen la eficacia de neutralización de las vacunas existentes contra la enfermedad y la consecuente evasión de la inmunidad poblacional existente. Pero, por otra parte, Ómicron tiene una preferencia por infectar el tracto respiratorio superior (frente al tracto respiratorio inferior), en comparación con los VOC anteriores a Ómicron.

A medida que el SARS-CoV-2 ha evolucionado, ha surgido la necesidad de desarrollar nuevas formulaciones de vacunas contra la COVID-19 para atacar nuevas mutaciones y la necesidad de proponer nuevas recomendaciones de calendarios de vacunación. Es importante destacar que los calendarios de vacunación dependerán en gran medida de si la COVID-19 sigue un patrón estacional similar al de la gripe. Es de destacar que el conocimiento de la evolución del virus ha permitido el desarrollo de modelos experimentales del proceso evolutivo a partir de los cuales es posible predecir el diseño de vacunas más específicas para enfrentar las variantes nuevas.

Sobre las variantes detectadas en Colombia, a finales del 2020 Laiton-Donato y colaboradores analizaron 122 secuencias del genoma del SARS-CoV-2 a partir de muestras obtenidas en 20 departamentos colombianos y otras muestras también colombianas almacenadas en GISAID. En una reconstrucción filogenética las secuencias se clasificaron en 12 linajes: A.1.2, A.2, B, B.1.1, B.1.3, B.1.5, B.1.8, B.1.11, B.2 y B.2.5. mostrando que la mayor contribución de la fuente geográfica es europea, en segundo lugar, norte y suramericana y en menor proporción, africana y asiática. Esto coincide con el número de personas y vuelos internacionales que llegaron al país entre enero y marzo del 2020.

Se identificaron en el país múltiples linajes y variantes del SARS-CoV-2 durante la pandemia. El linaje B.621, determinado como variante Mu fue el primero detectado en el país, y fue predominante durante una gran parte de la pandemia (Jimenez-Silva et al., 2023). Posteriormente, diversos estudios reportaron cerca de 188 linajes diferentes entre los cuales se destacaron 6 grupos: B.1, B.1.111, B.1.420, B.1.1.237, B.1.1.161 y B.1.1.2555.

Las variantes de preocupación (VOCs) que circularon en Colombia incluyeron: Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), Ómicron (BA.1 y BA.2), Mu (B.621) y Lambda (C.37). Siendo Mu y Gamma las variantes que se correlacionan en mayor frecuencia con casos graves y fatales. Estas representaron alrededor del 69% y el 18% de los casos fatales respectivamente. Por tal motivo, estos linajes jugaron un papel clave en el curso

de la pandemia, afectando significativamente las tasas de hospitalización y mortalidad (Jimenez-Silva et al., 2023). Además, se identificaron un total de 14.653 secuencias SARS-CoV-2 en 6.035.143 de casos registrados hasta el 18 de febrero de 2022, donde se determinaron más de 1.716 mutaciones, de las cuales el 53,6% eran de tipo *missense*, el 37,4% sinónimas, el 1,51% *nonsense* y el 2,73% *frameshift*, motivo por el cual, el virus fue considerado de muy alta tasa de mutacional (Gálvez et al., 2022).

1.4 GWAS (GENOME-WIDE ASSOCIATION STUDIES)

Los estudios de asociación genómica son aplicados a la investigación de rasgos poblacionales comunes y a la investigación de la etiopatogenia enfermedades multifactoriales. Hace más de un siglo Ronald Fisher revolucionó el campo de la genética presentando una teoría unificadora que aportó un marco sólido para el estudio de la variación fenotípica de las poblaciones que no era posible explicar desde los conceptos mendelianos y que mantiene su validez hasta hoy (Nelson et al., 2013). Fisher desarrolló la teoría genética cuantitativa de forma simplificada y estadísticamente soportada, afirmando que la variación genética en una población era debida a un gran número de factores mendelianos que aportaban cada uno una pequeña contribución aditiva a un fenotipo particular lo que se conoció como el modelo infinitesimal (Fisher, 1919).

Hoy sabemos que muchos rasgos fenotípicos como la talla, el color de la piel, el índice de masa corporal (IMC), la presión arterial y muchas de las enfermedades comunes en la población humana como la hipertensión arterial (HTA), la diabetes, la enfermedad coronaria, la obesidad, la osteoporosis, el cáncer o la depresión se comportan en general como rasgos complejos y son el resultado de influencias tanto ambientales, asociadas al estilo de vida de la persona afectada y su entorno, como genéticas, determinadas por la acción conjunta de múltiples genes donde, como sugirió Fisher, cada uno tiene una pequeña influencia en la variación fenotípica del rasgo pudiendo incluso interactuar entre ellos para modificar el fenotipo (Visscher & Goddard, 2019b). Así, los rasgos fenotípicos mencionados y las enfermedades complejas tienen una herencia multifactorial o poligénica y su varianza es susceptible de medición, por lo que también se llaman rasgos cuantitativos y las regiones genómicas que contienen genes que controlan un carácter de esta naturaleza son conocidas como *loci* de caracteres cuantitativos o QTLs (*Quantitative trait loci*) (Ye et al., 2020).

Existe gran interés en la medicina actual por establecer modelos que permitan estimar el riesgo de padecer una de estas patologías y es que, dada la alta prevalencia en todas las poblaciones del mundo de las enfermedades poligénicas es necesario establecer herramientas que permitan detectar los genes causantes de las enfermedades y estimar el riesgo de expresión. Los estudios de ligamiento tradicionales realizados sobre familias portadoras han sido de gran utilidad para descifrar los genes relacionados con enfermedades de herencia mendeliana, pero han sido insuficientes para encontrar genes relacionados con enfermedades poligénicas.

A diferencia de las enfermedades monogénicas que pueden ser el resultado de una mutación en un solo gen y, por lo tanto seguir modelos determinísticos, las patologías multifactoriales pueden explicarse a través de modelos probabilísticos donde se asume que el genotipo complejo afecta la probabilidad de desarrollar la enfermedad; si bien, la presencia de una variante genética no es necesariamente la causa de esta, si puede modificar el riesgo de padecerla y sumarse al

de otras muchas variantes frecuentes en la población cuyos efectos sobre el riesgo pueden ser de diferente magnitud (López de Subijana 2016).

La variación común de rasgos en la población como la talla o la pigmentación de la piel se distribuyen según un modelo de normalidad probabilística donde los más altos y los más bajos, por ejemplo, se ubican en los extremos de la curva y la mayor parte de la población en la mitad de una distribución normal. Por su parte, para las enfermedades poligénicas o multifactoriales la distribución normal se refiere a la distribución del riesgo, la susceptibilidad o la predisposición a sufrir la enfermedad, así en los extremos de la distribución estarán los riesgos más altos y los más bajos y los valores intermedios en la mitad de la distribución. Abordar todas las posibles fuentes de variación que explican una patología compleja no es sencillo y se requiere de tecnologías que puedan evaluar el genoma completo; quizás, una de las estrategias de mayor impacto con que cuenta actualmente la medicina genómica en los últimos 20 años para estimar el riesgo de una enfermedad poligénica son los estudios de asociación de genoma completo o GWAS (*Genome-wide association studies*) que se han desarrollado con el objetivo de detectar asociaciones entre variantes genéticas comunes y fenotipos particulares.

Los GWAS no requieren de estudios familiares o de hipótesis previas, ni genes candidatos de una enfermedad; solo comparan poblaciones de individuos afectados (casos) contra individuos sanos (controles) (Uffelmann et al., 2021). Las posibles asociaciones detectadas en un GWAS entre variantes genéticas comunes y los fenotipos de patologías complejas de diversas poblaciones se basan en la diferencia de las distribuciones alélicas de las variantes entre los casos y los controles y en que el número de variantes genéticas asociadas al riesgo es mayor en los pacientes que en los controles lo cual habla de una mayor predisposición a la enfermedad.

Son numerosos los estudios que han utilizado los GWAS para el estudio de la variación genética común responsable de la enfermedad. Prácticamente en todas las enfermedades la contribución de la genética a la varianza, esto es, la heredabilidad depende mucho más de la variación común que de la variación rara funcional. Esta variación común es capturada mediante estudios de GWAS, normalmente de la mano de consorcios pues un número elevado casos y controles es esencial para capturar asociaciones con riesgos relativos muy bajos.

Los GWAS han representado una revolución en el descubrimiento de genes implicados en la enfermedad común, en el conocimiento de los mecanismos de las enfermedades, en la búsqueda de dianas para fármacos y están representando una revolución en la predicción del riesgo de enfermedad. Ya los primeros GWAS produjeron una revolución en el conocimiento de la etiopatogenia de las enfermedades, en ocasiones totalmente disruptivas como la asociación de genes de autoinmunidad en la esquizofrenia que fue un cambio conceptual que ha implicado nuevos enfoques de tratamiento (Stefansson et al., 2009).

Hoy día existen consorcios para la mayoría de las enfermedades humanas y no son infrecuentes metaanálisis en más de un millón de casos y controles que tienen potencia estadística. Son miles los loci de riesgo que se han asociado con enfermedades y rasgos poblacionales, muchos de ellos constituyendo dianas para fármacos; por ejemplo, *FTO2* se ha asociado a la obesidad, *PTPN22* a enfermedades autoinmunes y ocasionalmente han aportado pistas sobre la biología de estas enfermedades y nuevas terapias. Un GWAS involucró la vía

IL-12 / IL-23 con el desarrollo de la enfermedad de Crohn lo que sirvió de evidencia y soporte para la realización de ensayos clínicos con medicamentos dirigidos a esta vía metabólica. Rasgos psiquiátricos incluyendo esquizofrenia, trastorno del espectro autista, déficit de atención, trastorno depresivo mayor y trastorno bipolar han sido objeto de varios GWAS (Gelernter et al., 2014). El último GWAS sobre esquizofrenia asoció 287 loci en 120 genes con procesos sinápticos. Recientemente, Keaton, J. y Cols. realizaron un metaanálisis con datos de 12 diferentes GWAS que buscaban variantes genéticas asociadas a la presión sanguínea y a la hipertensión arterial en más de 1 millón de personas de ancestría europea, identificando 113 loci a partir de un total de 2103 SNPs con efecto significativo que explican más del 60% de la heredabilidad del rasgo en humanos. Para 2022 se habían desarrollado más de 5700 GWAS en la investigación de más de 3300 rasgos involucrando a millones de individuos (Keaton et al., 2024; Uffelmann et al., 2021).

Concretamente, los GWAS detectan variación común asociada a rasgos de interés a través del genotipado de polimorfismos de una sola base o SNPs aprovechando que existen más de 5 millones en el genoma humano comprendiendo un inmenso número de variantes (<https://www.ncbi.nlm.nih.gov/snp/>) que explican más del 90% de la variación del genoma humano y seguramente, muchos de ellos están asociados a genes responsables de fenotipos de interés en la medicina. Los estudios de asociación del genoma completo reportan bloques de SNPs correlacionados que muestran asociaciones estadísticamente significativas con el rasgo de interés y pueden ser usados para estimar la contribución relativa de los efectos del genotipo y de la interacción genotipo y ambiente a la varianza total de los caracteres fenotípicos.

Los estudios de asociación normalmente se realizan con arrays de SNPs de alta densidad, aunque en los últimos años algunos estudios han incorporado la secuenciación genómica masiva (WGS, *whole genome sequencing*) o de exoma total (WES, *whole exome sequencing*) lo cual permite detectar variación de baja frecuencia que por lo general tienen efectos mayores sobre el fenotipo y aportan información más funcional sobre el papel de los loci asociados.

1.4.1 GWAS flujo de trabajo

Un estudio de GWAS requiere, como se mencionó antes, de dos poblaciones de gran número de individuos ancestralmente similares, una de las cuales posee el rasgo fenotípico de interés (casos) y la otra no (controles). Habrá que definir con precisión los rasgos de interés o los componentes de varianza del rasgo para que sean evaluables posteriormente si resultan estar aparentemente asociados con variantes genotípicas (Marees et al., 2018). Las muestras son entonces procesadas a través de técnicas *arrays* o de secuenciación masiva (WGS o WES) y se obtiene un listado de genotipos para cada SNP estudiado en las dos poblaciones. Las plataformas actuales (Affymetrix o Illumina) detectan más de 700.000 SNPs a los cuales debe hacerse un control de calidad como se detalla más adelante. Las secuencias son ahora comparables y los genotipos obtenidos frente a los rasgos en estudio se prueban para verificar si las diferencias entre unos y otros se ajustan a un modelo estadístico de regresión que permita hacer predicciones del riesgo utilizando un *software* estadístico robusto como R o PLINK (Purcell et al., 2007a) el cual permite detectar y valorar estadísticamente las posibles asociaciones de las variantes SNPs en las poblaciones de casos y controles.

En general, el flujo de trabajo de un GWAS incluye las siguientes etapas:

1. Diseño adecuado del GWAS definiendo con precisión cómo seleccionar los participantes tanto pacientes como individuos control y hacer la caracterización detallada de los rasgos fenotípicos de interés. Es recomendable usar herramientas para cálculo del tamaño de las muestras como CaTS o GPC ya que se requieren grandes tamaños de muestras para tener poder estadístico suficiente para determinar asociaciones con un riesgo relativo bajo. Un GWAS puede diseñarse ya sea para el estudio de casos y controles si se trata de rasgos dicotómicos o medidas cuantitativas sobre la población si se trata de rasgos cuantitativos, en cada caso el diseño experimental dependerá de la disponibilidad de los datos preexistentes que puedan ser recogidos. Existen actualmente muchos recursos para lograr información de las poblaciones en estudio como son los biobancos y los repositorios de información médica a gran escala disponibles en grandes bases de datos en el mundo.
2. Toma de las muestras y obtención de la información de los participantes siguiendo un formato de registro estandarizado para recopilar la información fenotípica de los grupos de individuos en estudio (casos). Procesamiento de datos.

Los archivos de ingreso o *input files* incluyen el número de identificación anonimizado de las personas participantes, la codificación de las relaciones familiares posibles entre individuos, el sexo, información fenotípica, covariables y asignaciones de genotipos para todas las variantes estudiadas. Cada tipo de información requiere de un estricto control de calidad. Este proceso puede ayudarse con herramientas informáticas como PLINK, que ha sido específicamente diseñado para el análisis de datos genéticos incluyendo muchos controles de calidad.

3. Genotipado de las muestras: como hemos indicado se usan habitualmente *arrays de alta densidad* por su menor costo o estrategias de secuenciación masiva (WGS o WES). Cada etapa puede tener sesgos que deben ser vigilados desde el planteamiento del estudio. Estos controles se aplican tanto a GWAS que contienen variantes comunes como a los estudios que incluyen variantes raras, como el estudio de secuenciación del exoma total (WES) o los estudios de secuenciación del genoma total. Un estudio GWAS generalmente utiliza el genotipado de variantes preseleccionadas utilizando *arrays*, mientras que la secuenciación del exoma total y especialmente la secuenciación del genoma total ayuda a capturar toda la variación genética. Las variantes comunes son aquellas donde la frecuencia del alelo menor es mayor del 10%, aunque cuando el tamaño de las muestras es muy grande este umbral puede bajar al 1%. Existen diseños de arrays, como el utilizaremos en este estudio, que incorpora a los SNPs seleccionados para determinación de la variación (en general seleccionados a partir del proyecto HapMap), variantes raras funcionales que son muy específicas de población y que se suelen recoger de estudios de exomas o genomas completos en poblaciones, en nuestro caso de población española.

4. Control de calidad sobre los resultados del genotipado crudo que incluye los siguientes siete pasos:

- Verificación de datos perdidos: en el genotipado pueden perderse datos de alelos en varios individuos y será necesario eliminar dichos SNPs.

- Discordancias de sexo, por error en la transcripción de los datos o por confusión en la rotulación de las muestras.
- Frecuencia alélica menor (MAF) debe establecerse un umbral para aceptar frecuencias alélicas menores (0,01 o 0,05 o más restrictivas según el tamaño de la muestra).
- Equilibrio de Hardy-Weinberg (HWE). Deberán descartarse SNPs que se encuentren en desequilibrio genético (HWE).
- Estructura poblacional. Deberá determinarse si existe estructuración de la población que pueda sesgar las asociaciones detectadas. Una fuerza evolutiva que genere estructura poblacional podría sesgar el resultado porque la asociación no sería causal sino el resultado de una distribución asimétrica de las frecuencias alélicas por causas genético-poblacionales no relacionadas con la enfermedad.
- Heterocigosidad, deberán descartarse individuos con heterocigosidad media muy altas o bajas.
- Parentescos existentes entre los individuos de la muestra que deben detectarse mediante análisis de IBD (*Identity by descent*) y descartarse.

El análisis de componentes principales (PCA, por sus siglas en inglés) permite analizar también la ancestralidad de los individuos del estudio y reagruparlos haciendo asignaciones más precisas de las asociaciones detectadas. Con el PCA se calcula la proporción promedio de alelos compartidos en todo el genoma entre cualquier par de individuos dentro de la muestra para generar índices cuantitativos (componentes) de la variación genética para cada individuo. Se utilizan los genotipos de todos los individuos en un subgrupo de SNPs autosómicos independientes ((Price et al., 2006). Las puntuaciones de los componentes individuales se pueden representar gráficamente para explorar si hay grupos de individuos que son genéticamente más similares entre sí de lo esperado. Los individuos que son atípicos según el análisis de componentes principales deben eliminarse de los análisis y realizar un nuevo análisis PCA y sus componentes principales deben usarse como covariables en las pruebas de asociación para corregir cualquier estratificación poblacional remanente (Marees et al., 2018). La cantidad de componentes que deben incluirse depende de la estructura de la población y del tamaño de la muestra.

Los consorcios GWAS generalmente siguen diagramas de flujo o *pipelines* para realizar los pasos de control de calidad previo a la imputación, como RICOPILI y luego realizan la imputación, con frecuencia cargando sus datos a servidores de imputación, por ejemplo, el servidor de imputación Michigan o el TOPMed *Imputation Server*. Los análisis pueden ser corridos en paralelo en grupos o *clusters* de computadores o en ambientes en la nube que pueden distribuir trabajos a muchos computadores para conseguir abarcar el gran tamaño de muestra típico en estudios genéticos de asociación.

1.4.2 Imputación de genotipos

Las estrategias de secuenciación masiva favorecen la exploración de la diversidad genética de una población, pero generan matrices de genotipado con alto porcentaje de datos faltantes; para resolver esto se recurre a la predicción de los genotipos faltantes mediante la implementación de técnicas estadísticas de imputación. (Merino, 2018) . La imputación

también se aplica en el caso de datos de genotipado para inferir el genotipo de posiciones no incluidas en el array, aumentando así la densidad de marcadores analizables en el GWAS.

Esto se hace alineando las secuencias de estudio con genotipos de referencia, se compara el array de genotipo generalmente de 500 k a 1M de SNPs con el panel de referencia y se rellenan los marcadores que faltan en el array. Una vez realizado el control de calidad de las muestras y variantes en los datos genotípicos del array GWAS las variantes suelen alinearse en fase (*Phasing*) y los genotipos no tipados se imputan utilizando paneles de referencia de haplotipos secuenciados como el proyecto de 1000 genomas. A través de herramientas de *software* como IMPUTE2 o BEAGLE se realiza la imputación de genotipos contra un panel de referencia existente coincidente en ancestría (Howie et al., 2009). Uno de los más usados en la actualidad es el software Minimac4, implementado en el *TOPMed Imputation Server* usando su propio panel de referencia TOPMed.

1.4.3 Valoración estadística de las asociaciones detectadas en GWAS

Generalmente, se utilizan modelos de regresión lineal o logística para probar las asociaciones dependiendo de si el fenotipo es continuo o binario como la presencia o ausencia de una enfermedad lo que se puede analizar en estudios de casos y controles, por ejemplo. Se usan también covariables como el sexo, la edad y la ancestría para poder estratificar los resultados y evitar así la confusión por factores demográficos que pueden reducir el poder estadísticos de los datos. El modelo de regresión lineal para GWAS se describe de la siguiente forma:

$$Y = W\alpha + X_s\beta_s$$

Donde para cada individuo, Y es un vector de valores de fenotipo, W es un coeficiente o matriz de covariables que indica el punto donde la función corta el eje x (intercepto), α es un vector correspondiente de tamaños de efecto, X_s es un vector de valores de genotipo para todos los individuos en SNP s, β_s es el tamaño del efecto correspondiente de la variante genética en estudio o tamaño del efecto del SNP. En los modelos de regresión logística, se utiliza una función de enlace logit para fenotipos de casos y controles distribuidos binomialmente para modelar las probabilidades de resultados. Los resultados de un GWAS son inicialmente una serie de valores de p, el tamaño de sus efectos y sus direcciones en relación con las asociaciones detectadas. Determinar el valor correcto del umbral del valor de p para darle significancia estadística al análisis de asociación genómica es crítico para controlar la cantidad de estas posibles falsas asociaciones.

Al respecto, el valor de $P > 5 \times 10^{-8}$ se ha convertido en un estándar para estudios de GWAS, desde que, en 2005, el *International HapMap Consortium* utilizó pruebas de permutación de genotipos en 10 regiones genómicas densamente genotipadas de la *Encyclopedia of DNA elements* para estimar el número de variantes independientes comunes (alelo menor frecuencia (MAF) $\geq 5\%$) en 150 por cada 500 kilopares de bases (kb) en población europea. Generalizando a todo el genoma ($\sim 3,3$ Gb) sugirió un umbral de significancia de 5×10^{-8} . Desde entonces,

este valor no ha sido actualizado para incluir el espectro de frecuencias alélicas más bajas usadas en muchos estudios recientes de GWAS basados en *arrays* (Fadista et al., 2016a).

Se ha recurrido a diferentes métodos estadísticos para poder corregir para falsos positivos por comparaciones múltiples incluyendo la corrección de Bonferroni y se basa en el supuesto que cada variante genética es independiente. (Bonferroni, 1935) Aunque algunos autores consideran esta corrección demasiado conservadora es sin duda la más utilizada. También se utilizan enfoques basados en permutación, enfoques bayesianos, etc. Fadista y Cols, proponen usar enfoques de pruebas múltiples para evaluar variantes de baja frecuencia considerando que los estudios de secuenciación conllevan a analizar cada vez más variantes de baja frecuencia (MAF 5%) e incluso variantes raras (MAF 0,5%), lo que exige un umbral estadístico más estricto para las pruebas de asociación en los estudios realizados a partir de datos de secuencia (Fadista et al., 2016a). Concluyen que el valor tradicional es válido para variación genética común (MAF >5%) en población europea, pero para variantes de frecuencia más baja, el umbral debe ser más estricto para los estudios con individuos de ascendencia europea: 3×10^{-8} para $MAF \geq 1\%$, 2×10^{-8} para $MAF \geq 0,5\%$ y 1×10^{-8} para $MAF \geq 0,1\%$ en LD $r^2 < 0.8$ y afirman que estudios de otras ancestrías deben considerar el grado de variación genética esperada para determinar el umbral de significancia estadística apropiado. (Fadista et al., 2016b)

1.4.4 Representación gráfica de los resultados de un GWAS

Los resultados de un estudio de asociación se visualizan con dos tipos principales de gráficos: los Manhattan *plots* y los gráficos cuantil-cuantil (QQ *plots*), que se generan con herramientas como R o plataformas web como FUMA o Locus.

1.4.4.1 Manhattan *plot*

Este tipo de gráfico muestra la significancia de cada asociación detectada entre una variante y un fenotipo. Cada punto representa un SNP localizado sobre el eje X según su posición cromosómica. El eje Y representa la probabilidad de su asociación medida como $-\log_{10}(P \text{ valor})$ y la línea roja marca el umbral de significancia.

1.4.4.2 Gráfico Cuantil-Cuantil (QQ *plot*)

El QQ *plot* muestra la distribución de valores esperados de P en el eje X, bajo un modelo nulo (*null*) de no significancia versus los valores observados de P en el eje Y, permitiendo visualizar el enriquecimiento de la señal de asociación.

La desviación de los esperados bajo la H_0 (línea roja) indica la presencia de tanto efecto causal o de estratificación poblacional insuficientemente corregida. La diferencia más importante entre las poblaciones de estudio está en las frecuencias alélicas de cada SNP de modo que, en los individuos de control que no tienen la enfermedad, los alelos allí más representados se interpretarán como alelos protectores. Por su parte los SNPs que muestren frecuencias alélicas altas solo en los afectados se constituyen como marcadores de riesgo. Las asociaciones detectadas pueden no ser reales en muchos casos y aun siendo verdaderas, pueden no ser necesariamente los loci causales de la enfermedad, pero es muy probable que estén ligados al gen causante o en algunos casos, el mismo SNP puede ser el causante porque, aunque se encuentre en una región no expresiva (intrónica o intergénica) pueda ser parte de la secuencia reguladora de un gen.

Otras variantes sí que pueden estar ligadas a los genes causantes de la enfermedad, de todos modos, deben cumplirse las condiciones mencionadas antes para aceptar una variante SNP como asociada significativamente al rasgo fenotípico de interés. Puede decirse que los estudios de GWAS en general, informan de la correlación más que la causalidad.

Teniendo en cuenta que la significancia de un estudio de asociación depende de un tamaño de muestra suficientemente amplio los estudios de GWAS se realizan generalmente en el contexto de consorcios multicéntricos así por ejemplo, para la investigación genética de rasgos antropométricos existe el consorcio GIANT (https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium) o el consorcio de psiquiatría genómica (PGC) (<https://pgc.unc.edu/>) o SCOURGE (<https://www.scourge-covid.org>) para el estudio de los determinantes genéticos de la COVID 19, que en ocasiones se asocian en supra consorcios como en el caso de la COVID, el HGI (<https://www.covid19hg.org/>). Esta estrategia genera múltiples cohortes que pueden ser analizadas en conjunto usando herramientas de software como METAL, N-GWAMA o MA-GWAMA y pipelines de control de calidad como ya se han descrito (Willer et al., 2010).

Estos resultados pueden combinarse con otros estudios de GWAS con el mismo rasgo de interés (GWAS metaanálisis) para aumentar la potencia estadística de los hallazgos y ser objeto de estudios de mapeo funcional y anotaciones, es decir, anotar, priorizar, visualizar e interpretar los resultados de los GWAS. Los resultados de un GWAS son los estadísticos resumen (*Summary statistics*), este reporte incluye una lista de todos los SNPs probados y el tamaño de su efecto, la localización del SNP en el cromosoma, el número de identificación, la construcción genómica, los alelos genotipados, la hebra analizada, el error estándar, así como el valor de p y otros estadísticos de prueba, la frecuencia del alelo menor y el tamaño de la muestra. Estos archivos con los estadísticos resumen es lo único que hace falta para realizar un meta-análisis, con METAL por ejemplo, de donde se obtendrá un archivo de resultados semejante (tamaño y signo del efecto, error estándar, probabilidad...) incluyendo además información de los signos de los efectos en cada estudio que se meta analiza y una medida del grado de heterogeneidad entre los estudios que se combinan.

Otras herramientas estadísticas pueden utilizarse para comparar las diferentes cohortes y detectar errores de varianza entre ellas para probar la heterogeneidad en los resultados, igualmente, la combinación de las contribuciones de todas las cohortes aporta una estima más precisa del tamaño del efecto y la significancia de los efectos ya que puede ponderarse además cada resultado de una corte individual según el tamaño de la muestra usando métodos de varianza inversa (Willer et al., 2010). Adicionalmente, los hallazgos son sometidos a herramientas bioinformáticas que realizan anotaciones funcionales extensivas de los SNPs de importancia biológica.

Una crítica adicional que se hace a los estudios de GWAS es su falta de representatividad de poblaciones (Diz de Alemida, 2024); la mayoría de los estudios se han realizado sobre población europea o norteamericana y es muy conocido que las enfermedades poligénicas pueden variar según la población donde prevalecen hay y existen variantes causantes de patologías que tienen diferente frecuencia dependiendo de la población estudiada (Cohen et al., 2005). Esto es importante si se quiere llevar el estudio de la variación común como causa de riesgo de enfermedades a un nivel poblacional específico e incluso a un nivel personalizado.

No obstante, debe tenerse en cuenta que las poblaciones diferentes también están sometidas a influencias ambientales distintas que podrían afectar las estimas de riesgos.

1.4.5 Significado funcional de las asociaciones detectadas por estudios de GWAS

Una vez se conocen asociaciones entre un rasgo normal o patológico y un grupo de variantes genéticas a través de un GWAS y se seleccionan aquellas de mayor poder predictivo, es posible explorar el papel funcional de estas y comprender, por ejemplo, la patogénesis de una enfermedad y la evolución de las características de la enfermedad. Al respecto, en las últimas décadas se han adelantado una variedad de estudios a gran escala de loci de caracteres cuantitativos o QTLs (*Quantitative trait loci*) para indagar sobre el efecto genético en diversos fenotipos moleculares. Hasta el 2023, se han perfilado más de 20 rasgos moleculares a nivel de cohorte y se identificaron numerosos QTL de rasgos moleculares (xQTL) en diferentes tipos de tejidos/células. Estos esfuerzos han facilitado la interpretación de variantes asociadas a enfermedades/rasgos revelados y seguimientos funcionales que permiten conocer mejor los mecanismos moleculares subyacentes a estos fenotipos de interés para la medicina (Huang et al., 2023).

Un nivel de análisis superior que compensa la falta de potencia de los GWAS además de los metaanálisis es el estudio de estimadores o índices (*scores*) del riesgo poligénico o PRS que, además, permiten aplicar las asociaciones validadas a la práctica clínica de forma individual y personalizada (S. W. Choi et al., 2020) que pasamos a describir.

1.5 PRS (*POLYGENIC RISK SCORE*)

El índice o puntuación de riesgo poligénico (PRS) es un índice que resulta de la suma ponderada de los pesos estadísticos de cientos o millones de SNPs independientes y asociadas en un GWAS o un metaanálisis GWAS previo. Una PRS nos da información sobre el riesgo que puede tener un individuo en particular de tener un fenotipo/rasgo multifactorial o de sufrir una enfermedad poligénica, compensando así el bajo poder predictivo de los valores individuales de cada variante en los GWAS. Los PRS son, por tanto, predictores de susceptibilidad genética a enfermedades (Wray et al., 2021)

El peso asignado a cada variante está representando la fuerza de su asociación con la enfermedad en estudio detectada en un GWAS (Hingorani et al., 2023). El análisis de PRSs es el único estudio que provee una estima de la predisposición genética de un rasgo a nivel individual, es decir, que permite cuantificar el riesgo genético de un individuo a presentar una enfermedad. Además, la construcción de PRSs también permite estratificar la población en estudio según los niveles de riesgo.

Para poder generar PRSs son necesarios dos tipos de datos:

1. **Datos de descubrimiento** o estadísticos resumen (*summary statistics*): son los datos de asociación de cada variante analizada en un GWAS o metaanálisis GWAS -reúne la información de varios GWAS realizados sobre el mismo rasgo o enfermedad- en los cuales se hayan detectado asociaciones significativas de SNPs. Se trata de un archivo que debe incluir como mínimo el identificador del SNP analizado (rs), el cromosoma o ubicación del par de bases, los valores de significancia (P valor), el alelo de efecto y el

alelo de no efecto, la frecuencia del alelo de efecto, el efecto (*odds ratio* -OR- o beta) y el error estándar.

2. **Datos diana:** son los genotipos y fenotipos de los individuos que componen la muestra diana y en donde se quiere aplicar el PRS, es decir, grupos de pacientes con la patología de interés o incluso personas sanas con indicación para el estudio debido a antecedentes familiares, y para los cuales se construirán índices de riesgo individuales basados en los SNPs preseleccionados en los GWAS y metaanálisis de GWAS.

Los dos grupos de datos, descubrimiento y diana, deben proceder de cohortes independientes, es decir no debe existir solapamiento de individuos entre ellas. Los datos de descubrimiento se obtienen de asociaciones previas detectadas en GWAS con la mayor potencia estadística alcanzada y es el investigador quien decide cuántas variantes o SNPs incluir en la PRS. De hecho, el peso asignado en la PRS a cada variante está basado en la fuerza de su asociación con la enfermedad según el GWAS o meta-análisis previo. Uno de los desafíos más importantes en la construcción de PRSs es la selección de SNPs para ser incluidos en la puntuación y que umbral límite de probabilidad y ligamiento (R^2) establecer. Si estos parámetros ya se han tenido en cuenta las PRS puede ser aplicadas directamente sobre el individuo diana (S. W. Choi et al., 2020). Sin embargo, cuando los parámetros para generar una PRS óptima no se conocen, entonces la muestra diana puede ser usada para modelar el entrenamiento de los datos, permitiendo optimizar los parámetros del modelo (Diz de Alemida, 2024).

Una vez construido una PRS se aplica a los individuos de la cohorte diana a través de un análisis bioestadístico utilizando *software* especializado (por ejemplo, PLINK) obteniendo un valor de riesgo a cada uno. En el proceso se analiza la asociación y la predicción. En el primer caso se determina si los SNPs incluidos en el índice capturan correctamente la poligenicidad del fenotipo, y en el segundo se cuantifica en qué medida la PRS puede explicar la variabilidad del fenotipo y su utilidad como factor de riesgo. Además, a partir de las PRSs se puede estratificar a la población en riesgos altos, medios o bajos. La predicción de riesgo individual es la estima para un individuo del riesgo de llegar a ser afectado en un tiempo determinado. La estratificación del riesgo poblacional es la estima para un subgrupo poblacional del riesgo absoluto de llegar a ser afectado en ese mismo tiempo.

En resumen, de forma secuencial la creación de PRSs consta de las siguientes 4 etapas:

1. Selección de variantes SNPs a partir de estudios GWAS o metaanálisis en los que se encontró asociación significativa de dichos SNPs con la enfermedad o el rasgo estudiado.
2. Ponderación de las variantes seleccionadas según el tamaño de su efecto (beta o coeficiente de regresión) o impacto de asociación (asignación de pesos).
3. Agrupación (*Clumping*) en función de si hay LD entre las variantes seleccionadas para escoger solamente las que sean independientes entre sí y puedan ser utilizadas en el cálculo de la puntuación de riesgo.
4. Fijación de un umbral (*Thresholding*) de P valor para decidir qué variantes se incluirán en el cálculo de la PRS.

Los pasos 3 y 4 se conocen como método *clumping + thresholding* (C+T) o umbral + agrupamiento, uno de los métodos más empleados y que se encuentra integrado, entre otros, en PRSice (Euesden et al., 2015) o LDpred2 (Privé et al., 2021). Este método permite seleccionar los SNPs más asociados con el rasgo a estudio y retiene múltiples SNPs localizados en la misma región genómica con efectos independientes. También existen los métodos bayesianos, como PRS-CS (Ge et al., 2019) o su versión para multi-población PRS-CSx (Ruan et al., 2022). Y se ha propuesto recientemente el método PROSPER (por las siglas en inglés de *Polygenic Risk scores based on enSemble of PEnalized Regression models*), que permite aplicar PRSs a poblaciones de diferentes ancestrías (J. Zhang et al., 2024).

Sin embargo, las puntuaciones de riesgo poligénico sólo pueden explicar parte del componente genético de una enfermedad. Además, los factores no genéticos también contribuyen al riesgo, por lo que la precisión máxima de los índices de riesgo genéticos está limitada, entre otros, por la heredabilidad del trastorno. La heredabilidad (h^2) es la proporción de la varianza existente en una población para un rasgo o enfermedad, que se atribuye exclusivamente a factores genéticos.

Por otra parte, en la construcción de una PRS se incluyen variantes de riesgo que tienen frecuencias de al menos 1% en la población o en algunos casos solo si su frecuencia es superior al 10%. Por esta razón, las PRSs sólo capturan la variación genética aportada por variantes comunes de un solo nucleótido (h_{snp}^2) h_{snp}^2 . A medida que aumenta el tamaño de las muestras de GWAS, la varianza explicada por PRSs también aumentará y se aproximará a la heredabilidad real (Wray et al., 2021).

En otras palabras, si los efectos de un SNP sobre el fenotipo se estimaran correctamente, la PRS podría predecir el fenotipo de los individuos del grupo de pacientes diana con una varianza explicada igual a la heredabilidad del SNP (h_{snp}^2) del rasgo. Sin embargo, debido al error en la estima del tamaño del efecto del SNP y a las diferencias que puedan existir entre las muestras de descubrimiento y diana, el poder predictivo de las PRSs es generalmente más bajo que el h_{snp}^2 , pero tenderá a ella en la medida en que las muestras del GWAS aumenten (J. Choi et al., 2020). Si estos parámetros ya se han tenido en cuenta los PRS puede ser aplicados directamente sobre el individuo diana. Sin embargo, cuando los parámetros para generar una PRS óptima no se conocen, entonces la muestra diana puede ser usada para modelar el entrenamiento de los datos, permitiendo optimizar los parámetros del modelo (Diz de Alemida, 2024).

1.5.1 Control de calidad de un PRS

Tanto los datos procedentes de los estadísticos resumen de los GWAS que se van a evaluar en una PRS, como la información de fenotipos y genotipos de los individuos diana deben ser objeto de un control de calidad con altos estándares, lo cual afectará el poder y la validez del análisis de PRS (J. Choi et al., 2020). A continuación, en la Tabla 1 se recogen los ítems que según Choi y colaboradores (2020) deben tenerse en cuenta en el control de calidad para asegurar la precisión y el poder predictivo de las PRSs.

Tabla 1 Criterios de control de calidad según Choi y colaboradores (2020)

	Item	Recomendación	Efecto
En relación con la base de datos de descubrimiento GWAS	Verificación de la heredabilidad	Generar PRSs solo a partir de datos GWAS con una heredabilidad estimada de $h^2_{SNP} > 0,05$	Sobreestimación del valor de la asociación
	Definición del efecto del alelo	Aclarar cuál es el alelo de “efecto” y cuál es el alelo de “no efecto” o alelo de riesgo y alelo protector respectivamente.	Se invierte el resultado
Para los datos diana únicamente	Tamaño efectivo de la muestra	Realizar análisis PRS en muestras de al menos 100 individuos, tanto para el grupo de casos como controles.	Pérdida de rigor en el control de calidad que puede darse en muestras pequeñas
Para las dos bases de datos	Estándares para base GWAS	Seguir las recomendaciones de guías de estándares básicos, ya que controlar los datos de los GWAS no siempre es posible: • Tasa de genotipado $> 0,99$ • Muestras perdidas $< 0,22$ • $Htz P > 1 \times 10^{-6}$ • $MAF > 1\%$ • Imputación <i>info score</i> $> 0,8$ • Eliminar SNPs con frecuencias menores al 1%	
	Transferencia de archivos	Verificar que los estadísticos resumen de GWAS y los datos diana que son descargados <i>online</i> se mantengan sin alteración durante el proceso	Se afecta la calidad del dato
	Versión del genoma humano utilizado como referencia	Es necesario reconvertir las definiciones inconsistentes entre los <i>datasets</i> utilizados (Sulovari & Li, 2014) y asegurar que las dos bases de datos tengan las posiciones asignadas a partir del mismo panel genómico de referencia, ya que los archivos de datos crudos pueden haberse generado a partir de diferentes plataformas, <i>arrays</i> o procedimientos de tipado tomando como referencia distintas versiones del genoma humano.	Se afecta la definición de los alelos
	Parentescos encriptados en las muestras	Realizar búsqueda de parientes mediante IBD y descartar los que tengan un índice alto por consanguinidad de 1° y 2° grado	Sobreestimación de la asociación entre la PRS y el fenotipo de interés
	Solapamiento de muestras entre la cohorte descubrimiento y la cohorte diana	Detectar y eliminar las muestras redundantes	Sobreestimación de la asociación entre la PRS y el fenotipo de interés
	Discordancias en el sexo de las muestras	Eliminar individuos para los cuales hay diferencia entre el sexo reportado y el indicado por los cromosomas sexuales	Inclusión de muestras erróneas
	SNPs duplicados	Asegurarse de que no hay SNPs duplicados en ninguna de las bases de datos.	Puede causar colapso del software o generar errores a menos que el software este entrenado en detectarlos

	No coincidencia de genotipos (<i>strand mismatch</i>) o no coincidencia en la codificación de los alelos entre las dos bases de datos (Ej. A/C en la muestra descubrimiento y G/T en la muestra diana)	Puede resolverse cambiando los alelos de los datos diana a sus alelos complementarios (<i>flipping</i>). La mayoría de <i>softwares</i> de PRSs realizan el <i>flipping</i> automáticamente.	Se afecta la asignación alélica
	SNPs ambiguos Si las dos bases de datos fueron generadas usando diferentes <i>chips</i> de genotipado y no se conoce la hebra (+/-) para cada una. No es posible hacer coincidir SNPs ambiguos	Eliminar todos los SNPs ambiguos	No puede saberse si los datos de descubrimiento y la muestra diana se refieren al mismo alelo o a otro diferente. Pueden generarse sesgos
	Errores en la transferencia de los archivos	Verificar que el proceso no tenga archivos corruptos, ya que la mayoría de las bases de datos GWAS se descargan <i>online</i>	Se afecta la calidad del dato

Una vez completado el control de calidad de los datos del GWAS es necesario considerar si existe alguna estructura poblacional en la muestra estudiada, pues esta es la principal fuente de confusión en GWAS. Esto es debido a que los patrones de apareamiento y la dinámica de una población pueden generar distribuciones sesgadas de las frecuencias alélicas asociadas, como ocurre con la ubicación geográfica. Esto conduce a que se produzcan asociaciones entre muchas variantes genéticas y el rasgo fenotípico en estudio que se confunden con las variantes que se pretenden identificar mediante los estudios de asociación. De no corregirse esto, se genera sobreestimación de las predicciones de las PRSs. De hecho, la predicción de PRSs puede verse sobreestimada aún más por factores asociados al entorno doméstico del individuo (J. Choi et al., 2020) como la dieta y el ejercicio, compartidos por familiares, y no siendo estas fuentes de inflación de la PRS la causa del efecto fenotípico. Un ajuste estricto de los efectos a través del análisis de componentes principales (PC) o el uso de modelos mixtos dentro de las muestras descubrimiento y diana, pueden minimizar la inflación debida a la estructura de la población.

1.5.2 Cálculo del riesgo poligénico

Realizado el control de calidad y verificada la no existencia de estructura en las poblaciones estudiadas (muestra descubrimiento y muestra diana), puede procederse al cálculo de las puntuaciones del riesgo poligénico de los individuos de la muestra diana. Como consideraciones generales es importante tener en cuenta que los datos de GWAS se obtienen de poblaciones humanas particulares con lo cual las estimas del tamaño de efecto de los SNPs serán una combinación del verdadero efecto y de la variación estocástica, por lo que pueden no generalizarse bien para diferentes poblaciones. Además, el efecto de los SNPs también está afectado por el nivel de ligamiento entre ellos y su desequilibrio de ligamiento (LD). Por tanto, deben tenerse en cuenta los siguientes 3 factores para el cálculo del riesgo poligénico: el tamaño

del efecto estimado y su incorporación en su incertidumbre, la generalización de PRSs a las poblaciones diana y el LD entre los SNPs. Así pues, algunas estrategias propuestas por Choi y colaboradores en el 2020 para optimizar el cálculo de PRSs son:

1. Reducir la estima del tamaño del efecto de GWAS, ya que no todos los SNPs afectan el rasgo que se estudia y la estima inadecuada del tamaño del efecto puede generar PRSs pobremente estimadas.
2. Usar valores umbrales de P como criterio de selección de los SNPs en el estudio, como $P < 1 \times 10^{-8}$.
3. Reducción del número de SNPs según su estado de ligamiento (LD). Los SNPs se agrupan priorizando los asociados, de modo que los seleccionados sean independientes entre sí y sus efectos puedan sumarse. Esta selección se complementa con la del punto anterior (selección de aquellos que superan un umbral definido por el investigador para el valor de P) y se conoce, como ya se dijo anteriormente, como método C+T. Una crítica al método de agrupación es que el valor umbral de P es arbitrario para la eliminación de SNPs en LD.
4. Las unidades de los tamaños de efectos GWAS definen las unidades de la PRS, así, al calcular PRSs en fenotipos binarios, como casos/controles, los tamaños del efecto ponderados generalmente se informan como *log Odds Ratios (logOR)*, o mejor aún como una suma de genotipos ponderados.

Las PRS se calculan para cada individuo de la muestra diana como la suma de los alelos de riesgo que porta el individuo, ponderados por su correspondiente nivel de efecto. Los sujetos de la muestra diana presentarán 0 alelos de riesgo si son homocigotos para el alelo protector y 1 o 2 alelos de riesgo según sean heterocigotos u homocigotos para el alelo de riesgo. Las PRS se calculan en PLINK según la siguiente fórmula:

$$PRS_i = \sum_j g_{ij} \times \beta_j$$

Donde “i” representa al individuo para el que se realiza el cálculo de la PRS, “j” es un SNP de un conjunto de SNPs asociados con el fenotipo de interés, “g” es el número de alelos de riesgo para un determinado SNP que porta el individuo y “β” es el tamaño de efecto del SNP en cuestión (*lnOR*). Los tamaños del efecto para cada SNP se obtienen, como se dijo en apartados previos, de los estadísticos resumen del GWAS o metaanálisis de los datos de descubrimiento y se utilizan para construir las PRSs, las cuales se generan para cada uno de los individuos de la muestra diana.

1.5.3 Interpretación y presentación de resultados

Una vez que se ha calculado la PRS se comprueba si existe evidencia de asociación entre las PRSs y el rasgo de interés en los individuos diana realizando, para ello, un análisis de regresión en la muestra diana, utilizando la PRS como predictor del rasgo en estudio e incluyendo las covariables según corresponda. El estudio de PRSs implica encontrar evidencia de asociación entre la PRS y un rasgo en los datos diana y evaluar su efecto potencial. De igual forma, se busca comprobar si existe suficiente evidencia para diferenciar los casos de los controles en la muestra de pacientes diana. La asociación entre la PRS y el rasgo diana se prueba, generalmente, a través de una regresión lineal si los rasgos estudiados son continuos o logística si los rasgos son binarios, y ajustando para covariables como PCA, sexo y edad. La inclusión de covariables que predicen el resultado puede aumentar el poder estadístico y lograr

estimaciones más precisas de los efectos de la PRS en entornos de regresión lineal, pero en ciertos escenarios puede reducir el poder en entornos de regresión logística.

Para la construcción de una PRS es mejor seleccionar los hiperparámetros con mejores resultados, aunque esto puede generar un sesgo o inflación o el sobreajuste del modelo si los datos no son independientes (J. Choi et al., 2020). Así pues, se puede partir de una muestra de descubrimiento, que es el GWAS, y otra de validación. En la muestra de descubrimiento se seleccionan los parámetros que maximicen la precisión de la predicción, para luego probar el modelo en la muestra de validación o test.

En resumen, el procedimiento constaría de:

- Construcción/generación de las PRSs.
- Selección de la PRS con mejor predicción basándose en las variantes de R^2 .
- A partir de las PRSs seleccionadas ajuste de modelos de regresión logística usando como covariables la edad, el sexo y las 10 primeras PC y la PRS estandarizada.
- Análisis de sensibilidad para descartar sesgos por ancestralidad.

1.5.4 Métricas usadas en la prueba de asociación

P valor para probar una hipótesis nula de no asociación.

AUC: área bajo la curva (usado para regresión logística) como medida de discriminación en predicción de enfermedades.

AUC-ROC o área bajo la curva característica operativa del receptor. El área bajo la curva permite determinar si las PRSs calculadas (a partir de los puntajes de los genotipos SNPs) sobre la enfermedad pueden predecir el estado de gravedad. Mide la capacidad de distinguir con precisión a los pacientes con y sin la patología (especificidad y sensibilidad).

Varianza fenotípica explicada (R^2 empírica): la medida de la varianza explicada (R^2) está bien definida para rasgos continuos, pero no totalmente para rasgos binarios. Una aproximación de esta estima aplicable a estos casos se denomina “pseudo- R^2 ” de Nagelkerke o Lee, comúnmente usada en epidemiología.

Estimación del tamaño de efecto (beta u OR). por unidad de PRS o entre estratos específicos como individuos de alto versus bajo riesgo. En el cálculo de PRS sobre fenotipos binarios el tamaño del efecto usado es reportado como *log Odds Ratios* ($\log(ORs)$) (J. Choi et al., 2020).

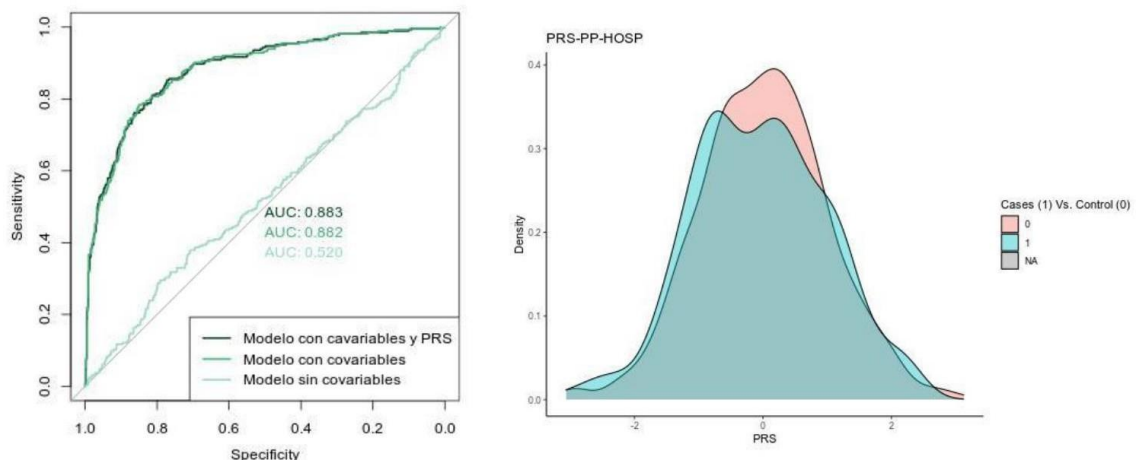
PCA: los componentes principales se relacionan con la estructura poblacional y ancestría de la población de estudio.

Prueba de Wald – Z score es una generalización multivariada que permite probar un conjunto de parámetros simultáneamente para ver si son lo suficientemente poco importantes como para poder eliminarlos (Wald, 1943).

1.5.5 Representaciones gráficas de resultados

En el enfoque C+T, los resultados de las pruebas de asociación de PRS pueden presentarse con un gráfico de barras, donde cada barra corresponde al resultado de probar una PRS que comprende SNPs con un P valor de GWAS que supera un umbral definido. Similar a un gráfico de barras es el gráfico de densidad que muestra la distribución de los datos en un intervalo continuo y permite comparar los dos grupos analizados, casos y controles, observando las diferencias existentes entre sus distribuciones en cuanto a su simetría o sesgos, así como la ubicación de la media y la mediana (Figura 1).

Figura 1. Representación gráfica de resultados de PRS A. Gráfico AUC área bajo la curva (sensibilidad/especificidad)-izquierda B. Gráfico de densidades-derecha



Si bien métricas como el AUC y R^2 pueden proporcionar resúmenes del poder predictivo de una PRS, también puede ser útil inspeccionar cómo los valores de los rasgos varían con el aumento de la PRS o medir el riesgo elevado de enfermedad entre los individuos con la PRS más alta. Esto se puede visualizar fácilmente utilizando un gráfico de cuantiles en el cual la muestra diana se separa primero en niveles crecientes de valores de PRS de 20 cuantiles de igual tamaño, cada uno de los cuales comprende el 5% de la distribución de la muestra PRS. Luego, los valores fenotípicos de cada nivel se trazan directamente como medias o prevalencias o se comparan con los de un estrato de referencia. Realizar una regresión permite ajustar las covariables y significará que el eje Y toma valores de beta (para rasgos continuos) u *odds ratio* (para rasgos binarios).

1.5.6 Utilidad clínica de las PRSs

Las puntuaciones de riesgo poligénico o PRSs tienen la capacidad de estudiar enfermedades comunes cuyo riesgo genético antes era imposible de calcular, pero sobre todo mejorar la predicción clínica del riesgo de enfermedad. Las PRSs han demostrado ser útiles como predictores de la susceptibilidad probabilística de una enfermedad y permiten establecer grupos de pacientes que podrían beneficiarse de un determinado tratamiento, o ser objeto de seguimientos prospectivos e incluso prevenir una patología compleja varios años antes de que aparezca. Esto facilita el cribado de los pacientes y permite tomar decisiones o conductas sobre

posibles intervenciones; también pueden ayudar a definir cómo disminuir el riesgo de una enfermedad realizando modificaciones en el estilo de vida del paciente. Es importante aclarar que, aunque el genotipado suele realizarse una vez en la vida de un individuo, la PRS puede recalcularse a medida que mejora la identificación de las variantes de riesgo y aportar información más precisa para calcular el riesgo de una enfermedad (Torkamani et al., 2018).

Wray y colaboradores (2021) mencionan que las PRSs por si solas, nunca podrán predecir de forma inequívoca diagnósticos de enfermedades complejas comunes, ya que los factores genéticos como se ha dicho anteriormente sólo contribuyen en parte al riesgo de sufrir la patología, por lo que las PRSs solo captarán una parte de dichas contribuciones genéticas. No obstante, una PRS puede mejorar la predicción de marcadores convencionales en medicina clínica (Inouye et al., 2018) o aportar un efecto aditivo de mayor utilidad, aunque sin alcanzar de todos modos la evaluación de todos los componentes de la varianza total alrededor de un rasgo complejo. Si bien es cierto que portar una mayor carga de variantes de riesgo significa mayor riesgo de sufrir la enfermedad, cada persona puede tener una combinación única de variantes de riesgo y una combinación única de experiencias en su vida y un estilo propio que afecten también al cálculo del riesgo genético (Wray et al., 2021).

Un trabajo clave en la posibilidad de traslación de las PRSs a la práctica clínica fue la demostración de que el cálculo de predicción de riesgo de algunas enfermedades comunes, en las que se han realizado grandes estudios de GWAS, era equivalente al de la enfermedad monogénica (Khera et al., 2018).

En el año 2018, Khera y colaboradores estudiaron a más de 250.000 personas para calcular el riesgo de padecer diferentes patologías, encontrando que el 8% de la población analizada tenía un riesgo 3 veces mayor que la población general a presentar enfermedad coronaria. También se identificaron 20 veces más personas con un riesgo alto de hipercolesterolemia familiar que las que se identificaban con paneles de mutaciones. Así mismo, observaron que el 3% de las personas estudiadas tenía 3 veces más riesgo a presentar diabetes que la población general (Khera et al., 2018). Hoy se conoce que 1 de cada 4 personas que presentan enfermedad coronaria no son detectadas en la evaluación de factores clínicos de riesgo convencionales, personas con un nivel de colesterol LDL dentro de los rangos normales pueden presentar una PRS de alto riesgo similar al de los pacientes con altos niveles de LDL.

Las áreas donde las PRSs ya están incluidas en la práctica clínica en algunos países y que están más validadas son la predicción del riesgo tanto de enfermedad coronaria, como de cáncer de mama. Parece claro que las PRSs para enfermedad coronaria, fibrilación auricular y diabetes tipo 2 mejoran la predicción cuando se incorporan a las estimas de riesgo clínicos. En cambio, para accidente cerebro vascular o hipertensión los datos no son tan claros (O'Sullivan et al., 2023), por lo que la Sociedad Americana de Cardiología no aconseja su uso. Por su parte, las PRSs de cáncer de mama están validadas y se están incorporando a la práctica clínica en algunos países. Se utilizan 313 SNPs que se han acordado en el seno de grandes proyectos internacionales dentro del consorcio BCAC (Breast Cancer Association Consortium) (<https://www.nature.com/nature-index/institution-outputs/united-kingdom-uk/breast-cancer-association-consortium-bcac/537af007140ba0a733000004>) y gozan de la ventaja del desarrollo de programas, también validados, para su cálculo, especialmente BOADICEA (X. Yang et al., 2022).

En el proyecto piloto más importante realizado en cáncer de mama, la historia familiar ajustada para *PRS313* fue asociada significativamente con dicho cáncer (OR=1.97, 95% CI 1.84-2.11). Pues bien, si el *PRS313* hubiese sido incluido en la predicción de riesgo familiar de BOADICEA, los criterios de cribado hubiesen cambiado en más de un 30% de los casos y en población control sin información de historia familiar en más de un 40% (Lakeman et al., 2023).

Hasta la fecha se han realizado estudios de GWAS para más de 2500 enfermedades o rasgos, y es cada vez más frecuente y de gran interés en el campo médico el uso de las PRSs en sus diversas aplicaciones en la salud humana. En la actualidad ya se dispone del servicio de perfiles PRS en varios centros de diagnóstico genético en el mundo, lo cual ha generado la elaboración de informes de muy variada composición (Hingorani et al., 2023). Es por esto por lo que hoy en día se cuenta con estándares reportados de presentación de informes de puntuación de riesgo poligénico (PRS-RS) para pacientes individuales, los cuales han sido publicados por el Grupo de trabajo sobre enfermedades complejas de *Clinical Genome Resource* (ClinGen), así como el Catálogo de puntuaciones poligénicas (PGS) que especifica la mínima información necesaria para interpretar y evaluar las PRSs (Wand et al., 2021).

Igualmente, en relación con rasgos poligénicos o cuantitativos, Lambert y colaboradores generaron el Catálogo de *Score Poligénico PSC* de acceso abierto de puntuaciones publicadas que incluye variantes SNPs, genotipado y pesos estadísticos, así como metadatos seleccionados para la reproducibilidad y aplicaciones independientes (Lambert et al., 2021). También proporciona plataformas para la difusión, investigación y traducción de PGS de estudios de *score* poligénicos de rasgos cuantitativos, como la talla y la presión sanguínea. Adicionalmente, incluye estimas de riesgo poligénico de enfermedades de interés en salud pública como el cáncer de mama (<https://www.PGSCatalog.org>).

Un estudio reciente con datos del UK Biobank (Thompson et al., 2024) calcula como el uso de PRSs mejoraría la predicción de 28 enfermedades en individuos de ancestralidad europea. Y el nuevo programa del Reino Unido, una extensión del UK Biobank denominado *Our Future Health* (OFH) (<https://ourfuturehealth.org.uk/>), está reclutando hasta 5 millones del *National Health Service* (NHS) y planea dar un resultado a los participantes en el riesgo de varias enfermedades comunes.

1.5.7 Debate bioético sobre el uso o mal uso de las PRSs

Si bien, contar con una herramienta molecular que aporte información fiable sobre el riesgo o la predisposición de una persona o un grupo de personas a sufrir una enfermedad es un ejemplo de beneficio para las poblaciones en general y con un gran impacto sobre la salud pública y la medicina preventiva, se abre nuevamente el debate sobre la protección de esta información individual. La información relativa las PRSs puede ser objeto de interés para aseguradoras de salud, aseguradoras de vida o incluso para empleadores, por lo que los países deben legislar al respecto para proteger dicha información (Knowles & Ashley, 2018). Otro aspecto en el campo bioético que merece discusión es el de los “factores no medidos”, es decir, no contar con información completa sobre una patología ya que la PRS evalúa muchas variantes comunes que aportan un nivel importante de probabilidad sobre el riesgo, pero no mide variantes raras que pueden ser determinantes para el riesgo (Knowles, 2018). Es cierto, no obstante, que muchas de las PRSs actuales también incluyen variantes raras funcionales o datos

de secuencia, y que en todo caso su inclusión es una mejora de la predicción, pero no impide la utilidad de la predicción basada solo en variantes comunes.

Otro tema es la accionabilidad. Las enfermedades cuyo riesgo se quiere predecir tienen que ser accionables, es decir tienen que poderse implementar medidas para disminuir su riesgo. Qué tan ético es informar del riesgo de desarrollar una enfermedad psiquiátrica a una persona sana, creando una peligrosa ansiedad, o no asesorar adecuadamente sobre posibles cambios en el estilo de vida. Estas conductas pueden ser peligrosas.

Otro potencial peligro es el de la equidad ¿Qué tan justo será el uso de PRS en las políticas de salud para la comunidad en general o solo podrá ser accesible para quienes puedan pagar los estudios genómicos? ¿Como garantizar un acceso equitativo y justo a las PRSs? (Khoury et al., 2022).

1.6 HIPERTENSIÓN ARTERIAL Y GENÉTICA

1.6.1 Definición y diagnóstico

La hipertensión arterial (HTA) es una condición patológica en la cual el individuo presenta una presión arterial sistémica persistentemente elevada, documentada a partir de varias mediciones. En general se considera hipertensión si la presión arterial sistólica persiste en valores iguales o superiores a 140 mm de Hg o la presión diastólica mantiene valores iguales o superiores a 90 mm de Hg. La *American Heart Association* ha definido los criterios diagnósticos para la hipertensión en las siguientes etapas: Grado 1, cuando la presión sistólica está entre 130 y 139 o la diastólica entre 80 y 89 en el grado 2, la presión sistólica es de al menos 140 mm o la diastólica al menos de 90 mm, la elevación que persiste más allá de estos umbrales es indicativa de hipertensión (Williams et al., 2018).

De acuerdo con las guías de la Sociedad Europea de cardiología de 2018 y la Sociedad Europea de hipertensión arterial se define como una presión sistólica de 140 o más y una diastólica de 90 o más (Zappa et al., 2024). La HTA está muchas veces asociada a las condiciones de vida de la sociedad actual es una enfermedad asintomática caracterizada por el incremento del volumen sanguíneo circulante, una vasoconstricción sostenida y un aumento en la resistencia vascular periférica que genera el aumento de la presión sanguínea sistémica.

1.6.2 Mortalidad

Las enfermedades cardiovasculares son la primera causa de mortalidad en el mundo y entre ellas la HTA es el principal factor de riesgo de muerte a nivel global constituyéndose como la principal causa subyacente de infartos de miocardio, accidentes cerebrovasculares, insuficiencia renal, degeneración retiniana, vasculopatía periférica e insuficiencia cardíaca según el ministerio de salud colombiano y aumenta además la severidad de otras condiciones patológicas como la demencia multi-infarto, la degeneración de las paredes de los vasos sanguíneos que conlleva disecciones aórticas o hemorragia vascular.

La HTA es la causa de muerte de aproximadamente 8.5 millones de personas en el mundo ya sea por enfermedad cardíaca isquémica, enfermedad renal y vascular y según un estudio reciente de carga global de la enfermedad estimó que 10.3 millones de muertes en el mundo en 2013 fueron causadas o asociadas directa o indirectamente a la hipertensión (Olsen et al., 2016).

En Colombia, para el período 2019 a 2020 la tasa de mortalidad por hipertensión se calculó en 190 casos por 100.000 habitantes (<https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/ED/GCFI/mortalidad-colombia-periodo-2020-2021.pdf>)

1.6.3 Prevalencia

Más del 30% de los adultos del mundo sufren de HTA en otras palabras, la prevalencia de la hipertensión arterial según la OMS a nivel global es de cerca de 1.3 billones de enfermos en todo el mundo (Adua, 2022) y se incrementa exponencialmente en la población adulta afectando su vida personal, social y financiera; además, varía según la región geográfica analizada. En Estados Unidos la hipertensión es más prevalente entre hombres adultos afro-americanos y aquellos con obesidad, enfermedad renal crónica y diabetes y su incidencia es más alta en áreas rurales; en general estos pacientes suman más de 80 millones de afectados solo en los Estados Unidos (Bress et al., 2017) (*World Health Association. Cardiovascular diseases 2021*). <https://www.who.int/health-topics/hypertension>).

La prevalencia es comparable globalmente, aunque varía entre países; en 1990 el 33.32% de los adultos en edades de 30 a 79 años tenían hipertensión en los Estados Unidos, estudios epidemiológicos han documentado una prevalencia del 55% entre pacientes de Europa mayor que en otras regiones. En 2019 cerca de 1.3 billones de adultos en el mundo estaban afectados por HTA como se mencionó antes y el 82% vivían en países de bajos y medianos ingresos (Zappa et al., 2024).

De acuerdo con recientes estudios el número de personas entre 30 y 79 años con hipertensión ha aumentado dramáticamente desde los años 90 al 2020 al partir de 331 millones de mujeres y 317 millones de hombres en el 90 a 626 de mujeres y 652 de hombres en 2020. En 2019 a pesar de la terapia antihipertensiva la hipertensión no controlada está ligada a 7.6 a 9 millones de muertes prematuras por año contribuyendo a cerca del 70% de la mortalidad global.

La naturaleza multifactorial y asintomática de la hipertensión complica su detección en aquellas personas que más probablemente pueden desarrollarla. Se calcula que entre 50 y 58 millones de personas no están protegidas y por lo tanto no están recibiendo tratamiento teniendo hipertensión. A nivel de Latinoamérica la información de prevalencia es limitada, se estima que varía entre el 12 y el 58%.

El estudio “CARMELA” de Evaluación de múltiples factores de riesgo cardiovascular en Latinoamérica realizado en 7 ciudades latinoamericanas (Barquisimeto, Bogotá, Buenos Aires, Lima, México, Quito y Santiago) con 11.550 pacientes de ambos sexos reportó una prevalencia de 18% (9- 29%) y el 16.3% eran hipertensos diagnosticados o con valores de presión elevados de *novo* (Pramparo et al., 2011).

El estudio del Consorcio Latinoamericano de Estudios en Obesidad LASO, (<https://laso.pophealth.wisc.edu>) encontró un promedio de prevalencia del 20.2% para la HTA en América Latina. Por su parte, el estudio de epidemiología rural y urbana prospectiva PURE en población latinoamericana documentó que al menos la mitad de la población latinoamericana

entre 35 y 70 años de edad sufre de hipertensión y su control es deficiente; Brasil para el 2008 mostró una prevalencia mayor del 50% en pacientes de mayores de 60 años (Lamelas et al., 2019). El mismo estudio PURE determinó que en Colombia había una prevalencia de 40.3 en el ámbito urbano y 34.9% en zonas rurales y a nivel de la ciudad de Bogotá, en mayores de 60 años la prevalencia fue de 56.9%, en el Eje Cafetero se encontró una prevalencia del 68% según el estudio *International Mobility in Aging Study* (MIAS). Según el Sistema de información de la protección social SISPRO en Colombia evaluando los años 2013 a 2017 se encontró un promedio de prevalencia nacional en pacientes mayores de 60 años del 28,14%, de nuevo, mayor en mujeres en proporción 1.8 a 1 (García-Peña et al., 2022)

En Colombia para el año 2015 según el Ministerio de salud la prevalencia de la HTA en personas entre 18 y 69 años se encontraba en el 7,23. Los departamentos de Antioquia y el Eje Cafetero presentaron las prevalencias más altas llegando al 10.75 comparado con el promedio del país que fue 7.23. Para el 2023, el Centro de Alto Costo (CAC) del Ministerio de salud de Colombia (<https://cuentadealtocosto.org>) reportó que entre el 1° de julio de 2020 y 30 de junio de 2021, se informaron 4.890.174 personas con HTA, lo que corresponde a una prevalencia del 9,64%. De estas, 413.091 fueron casos nuevos en la cohorte. Las mujeres seguían siendo las más afectadas reportando el 60.7% de los casos existentes y la mediana de edad fue de 65 años. El 28 de febrero de 2023 se reportaron 5.116.795 de personas diagnosticadas con hipertensión arterial, de nuevo, las mujeres tuvieron la mayor prevalencia (60,86%). La mayoría de los pacientes residía en las regiones Central y Caribe, con un 28,35% y un 20,98% de la totalidad de los casos, respectivamente.

1.6.4 Fisiopatología de la HTA

Los receptores de tensión del seno carotídeo y la aorta proporcionan un sistema de retroalimentación negativa para la regulación de la presión arterial. Cuando los vasos sanguíneos se dilatan debido al aumento de presión arterial se activan los baroreceptores que envían señales al cerebro (Núcleo del Tracto Solitario) en el tallo encefálico a través del nervio glosofaríngeo y el nervio vago. En consecuencia, se activa el sistema nervioso parasimpático que libera acetilcolina que actúa sobre las células marcapaso en el nódulo sinoauricular del corazón. Esto genera vasodilatación, reducción de la frecuencia cardíaca normalizándose la presión sanguínea. Por otra parte, si los varios receptores detectan una caída de la presión sanguínea envían señales al núcleo del tracto solitario desde donde se desactiva el sistema nervioso parasimpático y se activa en su lugar el sistema simpático (Adua, 2022).

La HTA se caracteriza básicamente por la existencia de una disfunción endotelial con ruptura del equilibrio entre los factores relajantes del vaso sanguíneo y los factores vasoconstrictores. A través de la acción de catecolaminas fibras eferentes del sistema incrementan la frecuencia cardíaca, la eyección de sangre del corazón y la contracción de vasos sanguíneos aumentando la resistencia vascular periférica y disminuyendo el flujo sanguíneo en ciertas regiones del cuerpo lo cual genera finalmente un incremento de la presión arterial. Vías eferentes del sistema simpático también inducen la secreción de renina en el riñón que activa el Sistema Renina Angiotensina Aldosterona (RAAS).

En condiciones fisiológicas el eje RAAS se activa en respuesta a la caída de la presión sanguínea renal y restaura el volumen y la presión sanguínea a nivel sistémico. Particularmente, las células yuxtaglomerulares del riñón liberan renina a la sangre que actúa sobre el angiotensinógeno del hígado convirtiéndolo en su forma activa, la angiotensina. La angiotensina I es catalizada por la enzima convertidora de angiotensina en el pulmón convirtiéndola en Angiotensina II. Esta se une a los receptores tipo 1 y 2 de Angiotensina II en la corteza adrenal, el riñón y las arteriolas. Esta unión dispara múltiples funciones efectoras incluyendo vasoconstricción e incremento de la resorción de sodio en el riñón por la Aldosterona. (Adua, 2022)

1.6.5 Genética de la hipertensión

1.6.5.1 Heredabilidad

Conocemos hoy que la hipertensión es consecuencia de la interacción de factores ambientales y genéticos; en cuanto a su origen genético se ha asociado a mutaciones de genes simples que codifican para proteínas relacionadas con la resorción de agua y sodio. Hoy sabemos que los factores genéticos pueden ser responsables del 30 al 60% del riesgo individual y conocemos también que la hipertensión es una enfermedad con una clara agregación familiar.

Varios estudios han documentado analizando familias con más de tres generaciones, que hijos cuyos padres y abuelos eran hipertensos tenían un riesgo mayor de desarrollar la enfermedad aún en ausencia de factores ambientales de riesgo. Otros estudios realizados en gemelos han demostrado una alta correlación entre gemelos monocigóticos e hipertensión al comparar con gemelos dicigóticos. La hipertensión es altamente heredable, con una heredabilidad media en los diferentes estudios de alrededor del 45% (Zappa et al., 2024). Aunque las variantes genéticas funcionales conocidas hasta hoy explican solo el 3% de la variación individual el resto se explica por variación común. (Bress et al., 2017). Los grupos familiares se ven influenciados por los mismos factores ambientales o genéticos, pero en realidad estos componentes tienen estrecha relación, por ejemplo, la obesidad que es el riesgo más importante para la hipertensión muestra una tendencia familiar muy fuerte igual que la ingesta de sodio y el estatus socioeconómico de los padres, pero, aun así, los estudios de gemelos han identificado influencias genéticas más que factores ambientales compartidos por los niños. Un estudio en 2008 sobre 1160 individuos encontró que los participantes cuyos padres ambos eran hipertensos podrían tener 2.4 veces más riesgo de desarrollar la hipertensión. (Zappa et al., 2024)

1.6.5.2 Hipertensión monogénica (MH)

La hipertensión monogénica es una forma rara de HTA causada por una sola mutación genética de herencia mendeliana dominante o recesiva. Se han descrito más de 35 genes afectados por mutaciones simples codificando enzimas o receptores que incrementan la reabsorción de sodio o aumentan la producción de aldosterona o la regulación de mecanismos del sistema metabólico de esteroides adrenales. Cada uno de estos efectos puede llevar a expansión del volumen plasmático y aumento de la presión sanguínea entre 20 y 50 mmHg (Khandelwal & Deinum, 2022) Síndromes de HTA monogénicos: El *Síndrome de Liddle* (LS) es uno de los más comunes, autosómico dominante, relacionado con mutaciones en las subunidades alfa, beta y gamma del canal de sodio epitelial1 (SCNN1A, SCNN1B

SCNN1G).LS es a menudo asociado a historia familiar de muerte súbita e HTA temprana (Schild, 1996).

El Síndrome de Gordon o pseudo hipoaldosteronismo tipo 2 resulta de mutaciones en 4 genes que regulan la actividad co-transportadora de sodio, cloro y potasio en los túbulos distales del riñón, Síndrome FH-I (GRA) o aldosteronismo respondedor a glucocorticoides, Síndrome AME o exceso aparente de mineralocorticoides y Síndrome de Geller, todos AD. Además, síndromes AR como la hiperplasia adrenal congénita (CAH). (Zappa et al., 2024)

1.6.5.3 Hipertensión poligénica (PH)

La HTA es una enfermedad poligénica compleja que involucra múltiples genes que contribuyen individualmente poco al riesgo, pero todos juntos explican la mayor parte de la heredabilidad de la HTA. Esta variación común es la que puede ser capturada por estudios de GWAS.

En el primer GWAS realizado en HTA en 2007 (The Wellcome Trust Case Control Consortium et al., 2007) se encontró solamente una débil asociación de variantes SNPs en el eje renina-angiotensina-aldosterona con presión sanguínea o rigidez arterial. En 2008 un segundo GWAS realizado sobre 1484 individuos de población japonesa detectó un locus con asociación significativa con la presión sanguínea vinculado al gen de la ATPasa ATP2B1 de transporte transmembrana de Ca^{2+} isoforma 1 que juega un papel importante en la regulación de la presión sanguínea (Kohara et al., 2008). En el GWAS liderado por Evangelou y Cols en 2008 se identificaron 535 nuevos loci asociados al control de la presión sanguínea el estudio se hizo sobre 1 millón de personas de ancestría europea dirigido a 3 rasgos particulares asociados a la presión sanguínea: presión sistólica, presión diastólica y presión de pulso o diferencial (the Million Veteran Program et al., 2018).

Recientemente, Keaton y Cols en 2024 reportaron el metaanálisis más grande realizado hasta la fecha sobre la presión sanguínea reuniendo cálculos estadísticos de al menos cuatro grandes bases de datos: UK Biobank, *International consortium of blood pressure* (ICBP), *Million veteran program* (MVP) y *Vanderbilt University Repository* (BioVU) además de los datos reportados por Evangelou 2018. El estudio se realizó sobre más de un millón de personas usando 7,5 millones de SNPs con una frecuencia alélica mínima (MAF) $>1\%$. Reportaron el descubrimiento de 113 nuevas variantes ($p > 1 \times 10^{-8}$) sobre las más de 500 previamente reportadas, incrementando la heredabilidad basada en SNPs ($h^2\text{SNP}$) a más del 60% por lo que los autores resaltan el alto potencial de este macro estudio para fabricar PRS dirigidos a estimar riesgos en HTA, de hecho, desarrollaron PRSs para predicción de la patología en poblaciones de ancestría europea y afroamericana. Este estudio se ha utilizado en la presente tesis como fuente de datos para calcular PRS en individuos de una muestra de pacientes colombianos que tuvieron COVID 19 grave.

Más del 90% de los loci novel se ubicaron en regiones no codificantes y 6 son variantes no sinónimas en los genes *TMPRSS6*, *GLRX2*, *RLF*, *HELQ*, *ZNF235* y *UNC13B*. Destacan el locus *TMPRSS6* situado en cromosoma 22, como un candidato prometedor como potencial diana farmacológica ya que codifica una proteína transmembrana serin proteasa 6 implicada en la atenuación de la carga dietaria de hierro en tejido cardíaco que tendría efectos protectores.

Las perspectivas en el estudio de la genética de la HTA son muy prometedoras. Aún están por realizarse estudios de asociación en poblaciones de ancestrías no europeas que permitan construir PRS más específicos para estas comunidades y, adicionalmente, el conocimiento de la biología de la presión sanguínea como rasgo complejo se espera que se apoye notablemente con el entendimiento de la función de cada uno de los cientos de genes detectados en los múltiples GWAS realizados hasta la fecha y en el futuro próximo. Finalmente, hay una gran esperanza para el avance de la medicina de precisión en la medida que una enfermedad de tanto impacto en la salud pública como la HTA cuente con evidencia científica sobre su origen genético y la interacción con factores asociados al estilo de vida de las sociedades contemporáneas.

1.6.6 Hipertensión arterial y COVID

Durante la experiencia médica vivida en la pandemia COVID-16 se conoció que la HTA, la diabetes, la enfermedad cardiovascular y el EPOC fueron prevalentes en pacientes ingresados con COVID-19 (Shi et al., 2020) (Guan et al., 2020). La HTA se consideró uno de los principales factores de riesgo para COVID-19. La prevalencia de HTA en COVID varía entre el 15% y el 20% hasta el 35% dependiendo de la edad de los pacientes. La infección por SARS-CoV-2 se activa cuando la proteína S del virus se une a la enzima convertidora de angiotensina 2 (ACE2) como receptora, la cual es altamente expresada en corazón, pulmones, riñón y tracto digestivo y tiene un papel importante en varias vías del sistema inmune y cardiovascular (Tadic et al., 2020). La pérdida de ACE2 puede desviar el sistema hacia unos niveles más altos de angiotensina II. El papel de la angiotensina II en los pacientes hipertensos con COVID-19 parece ser crucial debido a su papel en el funcionamiento del sistema renina-angiotensina-aldosterona (SRAA) que promueve la vasoconstricción, retención de sodio, estrés oxidativo, inflamación y fibrosis y aumenta el péptido bioactivo angiotensina.

Fang y colaboradores resaltan que la expresión de ACE2 es mayor en pacientes en tratamiento con inhibidores de la enzima conversora de angiotensina (ACEI) y/o antagonistas de los receptores de angiotensina (ARA), ya que esta expresión aumentada proporciona más receptores potenciales al SARS-CoV 2. Los autores también plantean la hipótesis de que los pacientes hipertensos o con otras afecciones cardiovasculares que requieren tratamiento con IECA/ARA tienen un mayor riesgo de desarrollar formas severas de COVID-19. Quedan muchos temas por definir sobre el papel del virus en la fisiopatología de la hipertensión de los pacientes infectados, comenzando por la relación de este con la enzima ACE. Además, es importante correlacionar estos efectos también con la predisposición genética del individuo infectado, tanto a desarrollar formas más severas de COVID, como a desarrollar HTA post COVID, o determinar el riesgo de este a desarrollar una infección más severa si es hipertenso y porta variantes genéticas definidas de susceptibilidad a la HTA (Fang et al., 2020).

2. OBJETIVOS

Los objetivos de esta tesis se derivan de la hipótesis de que la gravedad y las comorbilidades de la infección por SARS-CoV2 tienen, además de un componente ambiental, un componente genético que puede ser capturado por estudios de asociación de genoma completo, GWAS (*Genome Wide Association Studies*).

Para la captura de este componente el número de individuos es esencial por lo que es necesario establecer consorcios internacionales, y los esfuerzos de esta tesis se enmarcan en el consorcio SCOURGE (*Spanish COalition to Unlock Research on host GENetics on COVID-19*), a su vez integrado en Covid-19 Host Genetics Initiative (HGI).

Los resultados de los estudios de asociación GWAS de este consorcio fueron publicados en revistas internacionales de prestigio, primero en con datos de población europea (Cruz et al., 2022) después en población latinoamericana (Diz-de Almeida et al., 2024) y metaanalizados con otros estudios en el entorno de los consorcios HGI (The COVID-19 Host Genetics Initiative et al., 2023) y GenOMICC, ISARICC y 23andMe Covid19. (Pairo-Castineira et al., 2023). Contribuimos a estos últimos coordinando la recogida, en población colombiana, de casos y controles que han sufrido la infección por SARS-CoV-2 con distintos niveles de gravedad, y recogiendo muestras con todas las garantías éticas y legales, así como todas las variables fenotípicas acordadas dentro del consorcio SCOURGE. Este fue un objetivo general del proyecto.

Dado que las variables de riesgo cardiovascular no han sido estudiadas en el contexto de los GWAS realizados, planteamos como objetivos específicos:

- Realizar un GWAS en población colombiana para buscar de asociaciones específicas en esta población que puedan complementar lo que se está haciendo dentro del consorcio SCOURGE y los consorcios internacionales. En concreto analizaremos la asociación con gravedad y hospitalización, pero ajustando para comorbilidades cardiovasculares
- Analizar la capacidad de las puntuaciones de riesgo genético de hipertensión para predecir la hospitalización por COVID-19 (HOSP) y la gravedad de la COVID-19 (SEV4). Este último objetivo ha sido poco explorado y nunca en población colombiana. Mediante la generación de PRSs trataremos de ver si el componente genético que predispone a la hipertensión predispone también a la gravedad a la COVID-19, dado que la hipertensión sigue siendo un factor de riesgo controvertido desde el punto de vista de su contribución a la severidad en la COVID.

3. METODOLOGÍA

3.1 CONSORCIO SCOURGE

En los primeros meses del 2020, se consolidó el consorcio SCOURGE (*Spanish Coalition to Unlock Research on host Genetics on COVID-19*, <https://www.scourge-covid.org/>) con el liderazgo de los doctores Angel Carracedo (Universidad de Santiago y Fundación Galega de Medicina Xenómica, CIBERER) y Pablo Lapunzina (Instituto de Genética médica y molecular, INGEMM, Hospital La Paz, CIBERER) y con la participación de 46 grupos clínicos, 14 grupos de investigación, 9 biobancos de España y 22 grupos de Latinoamérica de los cuales 4 grupos son colombianos. El objetivo general del consorcio fue desarrollar un proyecto que permitiera buscar biomarcadores genómicos de riesgo que sean indicadores de predisposición a infección grave de COVID-19, para ello se plantea la realización de un estudio de Asociación de Genoma Completo (GWAS) de una muestra de más de 15.000 pacientes españoles y latinoamericanos que tuvieron COVID-19 y de quienes se tiene información clínica detallada de la gravedad del COVID-19 y la posible respuesta al tratamiento según el caso. La presente investigación forma parte del proyecto de este consorcio, específicamente del análisis de los datos de GWAS para la población colombiana y el análisis de PRS e hipertensión arterial.

3.2 PARTICIPANTES DEL ESTUDIO SCOURGE-COLOMBIA

En el proyecto SCOURGE fueron incluidos 1354 pacientes colombianos con prueba positiva para COVID-19 que asistieron a la convocatoria de participación en el proyecto. La mayoría de los participantes eran estudiantes, profesores, administrativos de las universidades o de los hospitales y/o pacientes, los cuales aceptaron participar en el proyecto en el momento de consulta u hospitalización en alguno de los servicios médicos de las Clínicas y Hospitales vinculados con el consorcio SCOURGE. Fueron cuatro los grupos de investigación líderes del proyecto en Colombia y adicionalmente se incluyeron pacientes colombianos residentes en España. Los grupos encargados de la recolección de muestras y del registro de información clínica se presentan a continuación:

1. Grupo Barranquilla: Universidad Simón Bolívar, IPS Camino Universitario Distrital de Adelita de Char y la Alcaldía de Barranquilla
2. Grupo Pereira: Clínica Comfamiliar
3. Grupo Bogotá-FSFB: Universidad de los Andes y la Fundación Santafé de Bogotá
4. Grupo Bogotá-FUCS: Sociedad de Cirugía de Bogotá Hospital de San José, Fundación Hospital Infantil Universitario de San José y Fundación Universitaria de Ciencias de la Salud
5. Grupo España: colombianos residentes en España, SCOURGE

3.2.1 Criterios de inclusión

Todos los pacientes con prueba molecular y/o serológica positiva para COVID-19

3.2.2 Criterio de exclusión

- Pacientes con prueba positiva para COVID-19 que durante la hospitalización o un mes antes de la toma de la muestra se le realizaran transfusiones de glóbulos rojos.
- Pacientes que hayan sido sometidos a trasplante de médula ósea.

Los menores de edad entre 12 y 18 años con resultados positivos para la prueba de COVID-19, también fueron invitados a participar, la aceptación de participación en el estudio se registró mediante la firma del asentimiento y el consentimiento de sus padres y/o representante legal. En el caso de las personas con autonomía disminuida como es el caso de los pacientes hospitalizados en UCI, la autorización la dieron los familiares quienes recibieron la información y firmaron el consentimiento autorizando la inclusión en el estudio, en los casos de pacientes recuperados cada uno firmó un nuevo consentimiento.

3.2.3 Recolección de información clínica

Para los casos leves o asintomáticos fue el personal de los Hospitales, Centros de investigación y de las Universidades los que realizaron el registro de información clínica mediante la realización de una entrevista y con la documentación que contaba el paciente, fueron registrados como datos mínimos. El registro de los datos clínicos (variables) de pacientes hospitalizados respetando los protocolos de consulta de registro e historia clínica de cada uno de los centros participantes se obtuvo la información general, y datos clínicos a partir de la Historia Clínica Institucional, toda la información obtenida de los pacientes fue registrada en RedCap SCOURGE <https://redcap.ciberisciii.es/>. La caracterización o definición operativa de las variables independientes está descrita en el documento Variables_ScourgeCRD_20200707 REDCap, localizado en el Centro de Investigación Biomédica en Red (CIBER) (Harris et al., 2019). Para la recolección de información clínica y paraclínica fueron incluidas las siguientes categorías:

1. Información general: Identificación de la entidad participante, procedimiento diagnóstico y verificación de consentimiento informado. A cada caso se le asigna un código que incluye el código del Centro o Institución y un número.
2. Datos demográficos: Edad, sexo, talla, peso, etnia, lugar de nacimiento y residencia y nivel educativo.
3. Antecedentes patológicos: factores de riesgo cardiovascular: hipertensión arterial, hipercolesterolemia, diabetes mellitus no complicada, diabetes mellitus con repercusión visceral, obesidad, cardiopatía isquémica, insuficiencia cardíaca, arritmia cardíaca, vasculopatía periférica; factores de riesgo neurológico enfermedad cerebrovascular infarto o hemorragia sin secuelas o con secuelas mínimas; factores de riesgo respiratorio EPOC, otra enfermedad respiratoria crónica; factores de riesgo digestivo úlcera péptica, hepatopatía crónica sin hipertensión vascular, hepatopatía crónica con hipertensión vascular, hematológico, tratamientos previos: tumor sólido localizado, tumor sólido metastásico, leucemia, linfoma, trasplante de médula/precursores hematopoyéticos.

4. Síntomas y signos: fiebre, cansancio, rinitis, tos, anorexia, mialgia, disnea, expectoración, odinofagia, cefalea, náuseas, diarrea, dolor abdominal, dolor tóxico, anosmia y manifestaciones dermatológicas.
5. Manejo clínico: duración de la hospitalización, tratamiento no farmacológico, variables de ingreso, pruebas complementarias y escalas clínicas.
6. Complicaciones
7. Tratamiento COVID-19
8. Desenlace clínico

Además, todos los pacientes fueron clasificados en 5 niveles de gravedad, como se muestra en la Tabla 2.

Tabla 2 . Clasificación de pacientes con COVID-19 según gravedad

Nivel	Criterio clínico
Gravedad 0 (asintomática)	Asintomático
Gravedad 1 (leve) ó severidad 1	Con síntomas, pero sin infiltración pulmonar o necesidad de terapia de oxígeno
Gravedad 2 (moderada) ó severidad 2	Con infiltración pulmonar afectando a <50% de los pulmones o necesidad de terapia de oxígeno
Gravedad 3 (grave) ó severidad 3	Hospitalización con algún de los siguientes criterios: PaO ₂ < 65 mmHg or SaO ₂ < 90% PaO ₂ /FiO ₂ < 300 SaO ₂ /FiO ₂ < 440 Disnea Frecuencia respiratoria ≥ 22 rpm Infiltración que afecta a > 50% de los pulmones
Gravedad 4 (crítica) ó Severidad 4	Ingreso en UCI o necesidad de ventilación mecánica (invasiva o no invasiva)

Modificada de criterios de SCOURGE PaO₂: Presión parcial de oxígeno en sangre arterial;
SaO₂: Saturación de oxígeno en sangre arterial; FiO₂: Fracción de oxígeno inspirado.

3.2.4 Consideraciones éticas

La presente investigación y los investigadores nos acogemos a los principios establecidos en el informe Belmont:

1. Respeto: Se aceptará la decisión de cada persona de participar o no en el proyecto, se manejará la información guardando la confidencialidad del nombre e identificación de los pacientes, las muestras se procesarán y se almacenarán bajo normas de bioseguridad y se marcarán con un número, y siempre serán procesadas por personal entrenado y capacitado para ello.
2. Beneficencia: La información obtenida de este estudio esperamos que sea útil para profundizar en el conocimiento de esta pandemia y beneficie así a toda la humanidad.
3. No maleficencia: durante la realización de este trabajo no se dañará ni se pondrá en riesgo a ninguna persona.

4. Justicia: Todos los participantes del estudio serán tratados de la misma manera durante el desarrollo del proyecto y los resultados serán de bien común.

Así mismo esta investigación sigue los “Principios éticos para las investigaciones médicas en seres humanos” de la Declaración de Helsinki de la Asamblea Médica Mundial 1964-2013 y se rige por las normas nacionales consignadas en la resolución 8430 del 4 de octubre de 1993. Y debido a que esta investigación se clasifica como “investigación de riesgo mínimo” (Artículo 10, Resolución 8430 de 1993, “Investigación de riesgo mínimo: Son estudios prospectivos que emplean el registro de datos a través de procedimientos comunes consistentes en: Extracción de sangre por punción venosa en adultos en buen estado de salud, con frecuencia máxima de dos veces por semana y volumen máximo de 450 ml en dos meses, excepto durante el embarazo...” y en cumplimiento con los aspectos mencionados con el Artículo 6 de la presente Resolución, este estudio se desarrollara conforme a los siguientes criterios:

- Se ajusta a los principios científicos y éticos, dado que: Se respetará a cada una de las personas, permitiéndoles ingresar voluntariamente a la investigación.
- La población humana se beneficiará del conocimiento y de futuros desarrollos con la información que proporcionarán los resultados de este proyecto.
- Se fundamenta en la experimentación previa realizada para otras patologías, que permitieron establecer relaciones de causalidad entre variantes poblacionales y desarrollo algunas patologías.
- No hay otro medio que permita obtener el conocimiento que se busca, dado que es la única forma de acceder a información genética que permita asociar los diferentes fenotipos clínicos a las variantes en el genoma.
- Prevalecerá la seguridad de los participantes, quienes conocerán de antemano los riesgos mínimos a los cuales serán sometidos. Se aclara que el análisis del ADN que será realizado no se usará para la identificación genética de una persona. No se analizará información que permita realizar diagnósticos genéticos. Solo será utilizado para el análisis de variantes como marcadores de asociación poblacional. Los riesgos de extraer ADN y enviarlo a España para su análisis estará dado por la metodología de extracción, que no se obtenga muestra, para ello se utilizarán kits comerciales, registro de insumos para investigación registro INVIMA reactivo para docencia e investigación, se realizará la estandarización en el laboratorio y lo realizará un experto. El siguiente riesgo es que se deteriore la muestra en el envío, y para ello se contará con una empresa con certificación para transporte de muestras biológicas (IATA).
- La investigación será realizada por profesionales con conocimiento y experiencia en el área de urgencias, medicina interna, cuidados intensivos y genética, bajo la responsabilidad de la Fundación Universitaria de Ciencias de la Salud, el Hospital de San José, el Hospital Infantil Universitario de San José y el Centro dermatológico Federico Lleras Acosta, así como genetistas, biólogos moleculares ingenieros y demás personal del CEGEN y del CIBERER en España.

- Se realizará el Consentimiento informado por escrito del sujeto de investigación o su representante legal con las excepciones dispuestas en la Resolución 008430/93.

Nos acogemos a la Declaración Internacional sobre los Datos Genéticos Humanos y Declaración Universal sobre el Genoma Humano y los derechos Humanos 29ª sesión de la UNESCO (1997) particularmente en el apartado: Artículo 5; “Finalidades: Los datos genéticos humanos podrán ser recolectados, tratados y conservados solamente con los siguientes fines: específicamente este trabajo... ii) investigación médica y otras formas de investigación científica comprendiendo los estudios epidemiológicos en especial los de genética de poblaciones...”.

Investigaciones sobre el Genoma Humano artículo 12 b. “La libertad de investigación, que es necesaria para el progreso del saber, procede de la libertad de pensamiento. Las aplicaciones de la investigación sobre el genoma humano, sobre todo en el campo de la biología, la genética y la medicina, deben orientarse a aliviar el sufrimiento y a mejorar la salud del individuo y de toda la humanidad”.

En el apartado: Solidaridad y Cooperación Internacional Artículo 19 a) ii) “*desarrollar y fortalecer la capacidad de los países en desarrollo...iii) permitir a los países en desarrollo sacar provecho de las investigaciones científicas y tecnológicas a fin de que su utilización en pro del progreso económico y social pueda redundar en beneficio de todos. vi) fomentar el libre intercambio de conocimientos e información científicos en los campos de la biología, la genética y la medicina*”.

También cumple los requisitos establecidos en la legislación española en el ámbito de la investigación biomédica, la protección de datos de carácter personal y la bioética. Dado que el proyecto implica la utilización de muestras biológicas de origen humano, se respetará lo establecido en la Ley 14/2007, de 3 de julio, de Investigación Biomédica y demás legislación vigente sobre la materia.

Los investigadores declaramos que no tenemos ningún conflicto de interés y manifestamos que el fin de este proyecto en todas sus fases es académico con implicaciones clínicas.

Como Anexo 1. Carta de Consorcio SCOURGE. Anexo 2. Comité de ética en investigaciones en seres humanos del Hospital de San José.

3.3 EXTRACCIÓN DE ADN Y GENOTIPADO

El ADN genómico se obtuvo de sangre periférica y se aisló utilizando el kit *Chemagic DNA Blood 100 (PerkinElmer Chemagen Technologies GmbH)* y *PureLink Genomic DNA Mini Kit INVITROGEN®* para muestras tomadas en España y en Colombia, respectivamente, según las recomendaciones del fabricante. El ADN obtenido fue cuantificado mediante Protocolo *QUBIT Qubit 1X dsDNA BR Assay Kit Thermo Fisher Scientific®*.

Las muestras tomadas en Colombia fueron enviadas a Santiago de Compostela, España, cumpliendo con las normas IATA y con las autorizaciones correspondientes del ministerio de salud colombiano y el ministerio de sanidad español. La genotipificación de las muestras se realizó en el Nodo de Santiago de Compostela del Centro Nacional de Genotipado (CeGen-

ISCI; <http://www.usc.es/cegen>) con el *Axiom Spain Biobank Array* (Thermo Fisher Scientific) según recomendaciones del fabricante. El array utilizado contiene 757.836 marcadores , incluidas variantes raras seleccionadas para población española. La agrupación y la determinación del genotipo se realizó utilizando el software *Axiom Analysis Suite* v4.0.3.3.

3.4 CONTROL DE CALIDAD E IMPUTACIÓN DE VARIANTES

Se aplicó un procedimiento de control de calidad (QC) utilizando PLINK 1.9 (Purcell et al., 2007b) siguiendo el protocolo realizado en el análisis del conjunto de población latina (Diz-de Almeida et al., 2024). Respecto a la calidad de los marcadores, fueron excluidas las variantes con una frecuencia de alelos menores (MAF) $<1\%$, una tasa de llamada $<98\%$ y aquellos marcadores que se desviaban fuertemente de las expectativas de equilibrio de Hardy-Weinberg ($p < 1 \times 10^{-6}$). En lo que concierne a la calidad de las muestras, fueron excluidas aquellas a las que les faltaba $>2\%$ de las variantes y también se exploró el exceso de heterocigosidad para descartar una posible contaminación entre muestras. Posteriormente, para analizar el parentesco entre individuos y la existencia de estratificación, se mantuvieron los SNP autosómicos, fueron eliminadas las regiones en alto desequilibrio de ligamiento y se realizó una poda de LD (ventanas de 1000 SNP, con un tamaño de paso de 80 y un umbral r^2 de 0,1) para obtener un grupo de SNPs independientes con los que evaluar el parentesco y estimar las proporciones ancestrales globales con el software ADMIXTURE. El parentesco se evaluó en PLINK, con base en los valores de IBD, eliminando un individuo de cada par con $PI_HAT > 0,25$ que mostrara un patrón coherente Z0, Z1 y Z2 (según los valores teóricos esperados para cada nivel de parentesco). Los componentes principales (PC) genéticos se calcularon también con PLINK con el mismo subconjunto de variantes independientes.

Los genotipos se imputaron con el panel de referencia TOPMed versión r2 (GRCh38) utilizando el servidor de imputación TOPMed, y se filtraron las variantes con $Rsq < 0,8$ o con $MAF < 1\%$. Tras todo este proceso de control de calidad de genotipado e imputación se incluyeron un total de 4348 individuos y 10 671 028 variantes genéticas en los análisis de la cohorte Latinoamérica de SCOURGE.

3.5 ANÁLISIS DE ASOCIACIÓN EN POBLACIÓN COLOMBIANA

En el periodo comprendido entre marzo del 2020 y febrero del 2021 se realizó la recolección de muestras y de información clínica de los pacientes con prueba positiva para COVID y que decidieron participar voluntariamente. De los 1354 pacientes registrados inicialmente en RedCap se trabajó un conjunto de datos imputados, tras pasar el proceso de control de calidad se seleccionó el subconjunto de 1001 pacientes reclutados en Colombia. De los 1001 pacientes seleccionados 186 (19%) formaban parte del Grupo de Barranquilla, 178 pacientes (18%) procedían del grupo de Pereira, 137 pacientes (14%) pertenecían al grupo Bogotá-FSFB y 404 pacientes equivalente al 40% provenían del Grupo Bogotá-FUCS. Del total de pacientes seleccionados el 57% fueron mujeres y el promedio de edad fue de 41 años, pacientes desde 7 hasta 90 años.

En esta población se realizó el estudio de asociación de genoma completo (GWAS) para dos variables indicativas de la gravedad de COVID:



Hospitalización: se compararon 265 casos hospitalizados frente a 737 casos no hospitalizados que servían como controles

- Enfermedad crítica: 127 casos con severidad 4 según la escala de gravedad definida en SCOURGE y 875 casos con severidad menor, que serían el grupo control.

El estudio de asociación en cada caso implicó el ajuste de un modelo de regresión logística ajustando como covariables: sexo, edad y los primeros 4 PC. Adicionalmente, y para evaluar específicamente la importancia de los antecedentes vasculares (AV) y cardíacos (AC) se añadió la presencia de dichas comorbilidades a las anteriormente mencionadas. Los antecedentes vasculares (AV) se definieron como AV=1 si había presencia de cualquier tipo de comorbilidad vascular (av1 a av5) y AV=0 si había ausencia total de antecedentes vasculares.

3.6 ANOTACIÓN Y MAPEO INICIAL

La anotación funcional se realizó con FUMA para aquellas variantes con un valor $p < 5 \times 10^{-5}$ o en LD moderada a fuerte ($r^2 > 0,6$) con las variantes principales, donde la LD se calculó a partir del panel 1KGP AMR. Los loci de riesgo genético se definieron colapsando bloques LD dentro de 250 kb. Luego, los genes, las puntuaciones escaladas de CADD v1.4 y las puntuaciones de *Regulome* DB v1.1 se anotaron para las variantes resultantes con ANNOVAR en FUMA (Watanabe et al., 2017). El análisis basado en genes también se realizó utilizando MAGMA (De Leeuw et al., 2015) que está implementado en FUMA. La significancia se estableció en un umbral $p < 2,66 \times 10^{-6}$ (que supone que las variantes pueden asignarse a un total de 18.817 genes). FUMA nos permitió realizar un mapeo genético inicial mediante dos enfoques: (1) mapeo posicional, que asigna variantes a genes por distancia física utilizando ventanas de 10 kb; y (2) mapeo eQTL basado en datos GTEx v.8 de sangre completa, pulmones, linfocitos y tejidos de la mucosa esofágica, estableciendo una tasa de falsos descubrimientos (FDR) de 0,05 para declarar la significancia de pares variante-gen.

3.7 ANÁLISIS DE SCORE DE RIESGO POLIGÉNICO

3.7.1 Puntuación de riesgo poligénico (PRS)

Se construyeron PRSs para 3 fenotipos de hipertensión: hipertensión sistólica (SBP), hipertensión diastólica (DBP) y presión de pulso (PP), utilizando la lista de SNPs asociados con la hipertensión en el mayor metaanálisis de GWAS de hipertensión realizado hasta la fecha por Keaton y colaboradores en el 2024 (muestra descubrimiento) [ver tabla ST4 de Keaton et al., 2024 (Keaton et al., 2024)]. Después las PRSs se aplicaron a la cohorte de Colombia (muestra diana) para valorar el poder de predicción del estatus caso/control (definidos en función de hospitalización y gravedad) de COVID-19 en dicha cohorte.

3.7.2 Muestra diana

La muestra diana es la misma cohorte de Colombia que se empleó para la realización del GWAS de COVID-19. Esta muestra diana está constituida por un total de 1001 individuos, de los cuales 264 son casos de COVID-19 y 737 son controles.

3.7.3 Muestra descubrimiento

Los estadísticos resumen del último gran metaanálisis de hipertensión de Keaton y colaboradores (Keaton et al., 2024) fueron utilizados como muestra de descubrimiento para

calcular las PRSs de riesgo de hipertensión. La cohorte de Keaton y colaboradores está constituida por 1,028,980 individuos europeos. Los estadísticos resumen fueron recuperados de la tabla suplementaria 4 (ST4). Se emplearon los 3 fenotipos de hipertensión: SBP (hipertensión sistólica), DBP (hipertensión diastólica) y PP (presión de pulso).

3.7.4 Análisis estadísticos

Se realizaron modelos de regresión logística para evaluar la capacidad de las PRSs de hipertensión (SBP, DBP y PP) para predecir la hospitalización por COVID-19 (HOSP) y la gravedad de la COVID-19 (SEV4). Las puntuaciones se calcularon con PLINK v1.90b5 y utilizando todos los marcadores de la tabla ST4 (2,068 SNPs) y, por otro lado, solo aquellos con significación GWAS ($p\text{-valor} < 5 \times 10^{-8}$). Se analizaron, por tanto, un total de 6 PRS (SBP, DBP y PP, con dos selecciones de SNPs) en relación con dos variables dependientes (12 análisis). Los modelos de regresión logística incluyeron el estatus caso/control -HOSP y SEV4- de COVID-19 como variable dependiente y se ajustaron para las covariables: sexo, edad, 4 primeros componentes principales (PC), con y sin PRS. Y se calculó el R² de Nagelkerke (pseudo-R²). Por último, se realizaron análisis del área bajo la curva característica operativa del receptor (AUC-ROC) utilizando el paquete de R pROC para determinar si las PRSs de hipertensión podían predecir el estado COVID-19 HOSP y SEV4 en la cohorte. Se comparó el AUC al incluir y al excluir las PRSs en el modelo de covariables.

4. RESULTADOS Y DISCUSION

El proyecto se inició en marzo del 2020 y se vincularon pacientes hasta marzo del 2021, durante este periodo de tiempo 1354 pacientes se registraron en RedCap. Habiendo firmado todos ellos un consentimiento informado, se tomó una muestra de sangre para extraer el ADN y se registró la información clínica. Al realizar una evaluación de la primera parte del proceso se identificaron algunos inconvenientes que no permitieron al final analizar los datos de todos los pacientes. Así, en cuanto a los pacientes hospitalizados durante el periodo de cuarentena con aislamientos estrictos en los hospitales, fue complejo acceder a ellos para realizar simultáneamente al menos las tres actividades iniciales importantes: la explicación y firma del consentimiento y la toma de la muestra. Cuando esto no se pudo realizar simultáneamente en algunos casos se recibían muestras que no coincidían con los datos de registro en la hoja de recolección de datos o en el consentimiento, por lo tanto, todas estas muestras se excluyeron del estudio. Finalmente, solo 1001 muestras superaron los criterios de calidad, 186 pacientes -19%- formaban parte del Grupo de Barranquilla, 178 pacientes -el 18%- provenían del grupo de Pereira, 137 pacientes -correspondiente al 14%- pertenecían al grupo Bogotá-FSFB y 404 pacientes -equivalente al 40%- al Grupo Bogotá-FUCS. De los 1001 casos analizados el 57% fueron mujeres y la edad promedio fue de 41 años (entre 7 y 90 años). Para el primer análisis 265 pacientes, el 22% de la muestra total, fueron pacientes hospitalizados y de ellos 148 tenían antecedentes vasculares, para el segundo análisis 127, el 10.5% de pacientes de la muestra total, fueron clasificados con gravedad 4 equivalente a SEV4 de los cuales 81 tenían antecedentes cardiovasculares.

Como hemos indicado en el capítulo anterior en el GWAS realizado en población colombiana hemos utilizado como covariables los antecedentes vasculares (AV) y los antecedentes cardiacos (AC).

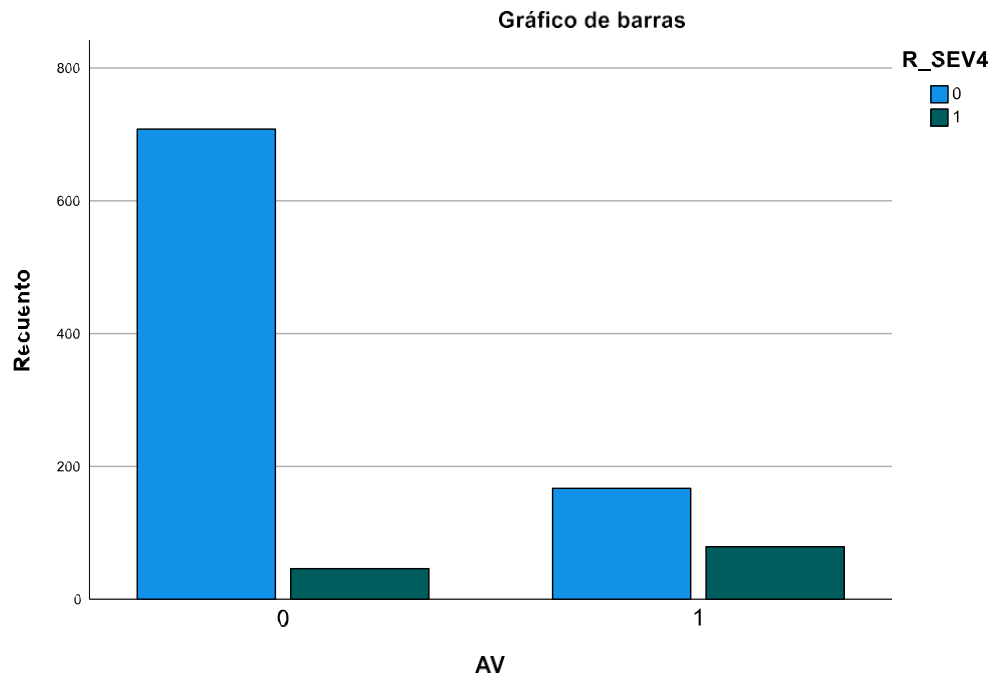
En la Tabla que sigue se puede ver el recuento de casos de severidad 4 (SEV4) en relación con los AV que se ilustran también en la Figura 3.

Tabla 3 Recuentos y frecuencias de antecedentes vasculares en casos graves (SEV4=1) y casos más leves (SEV4=0)

			SEV4		Total
			0	1	
AV	0	Recuento	708	46	754
		% dentro de AV	93,9%	6,1%	100,0%
		% dentro de R_SEV4	80,9%	36,8%	75,4%
	1	Recuento	167	79	246
		% dentro de AV	67,9%	32,1%	100,0%
		% dentro de R_SEV4	19,1%	63,2%	24,6%

Total	Recuento	875	125	1000
	% dentro de AV	87,5%	12,5%	100,0%
	% dentro de R_SEV4	100,0%	100,0%	100,0%

Figura 2 Distribución de casos AV*_SEV4



Como es fácilmente deducible de pruebas de chi-cuadrado, hay una asociación significativa entre la severidad del COVID-19 y los antecedentes vasculares

Tabla 4 Pruebas de chi-cuadrado AV*R_SEV4

	Valor	Gl	Significación asintótica (bilateral)	Significación exacta (bilateral)	Significación exacta (unilateral)
Chi-cuadrado de Pearson	114,755 ^a	1	<.001		
Corrección de continuidad ^b	112,389	1	<.001		
Razón de verosimilitud	98,264	1	<.001		
Prueba exacta de Fisher				<.001	<.001
Asociación lineal por lineal	114,640	1	<.001		
N de casos válidos	1000				

Del mismo modo en la Tabla 5 se puede ver una tabla cruzada de casos de AV con hospitalización (HOSP), los cuales también se ilustran en la Figura 4.

Tabla 5 Tabla cruzada AV*R_HOSP

		R_HOSP		Total
		0	1	
AV	0	Recuento	634	755
		% dentro de AV	84,0%	100,0%
		% dentro de R_HOSP	86,0%	75,4%
	1	Recuento	103	246
		% dentro de AV	41,9%	100,0%
		% dentro de R_HOSP	14,0%	24,6%
Total	Recuento		737	1001
	% dentro de AV		73,6%	100,0%
	% dentro de R_HOSP		100,0%	100,0%

Figura 3 Distribución de casos AV*R_HOSP

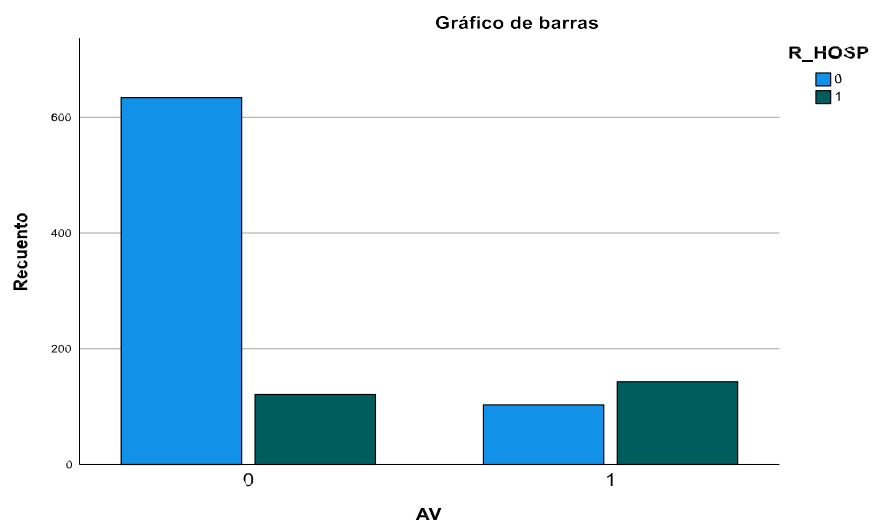


Tabla 6 AV*R_HOSP: Pruebas de chi-cuadrado

	Valor	gl	Significación asintótica (bilateral)	Significación exacta (bilateral)	Significación exacta (unilateral)
Chi-cuadrado de Pearson	169,388 ^a	1	<.001		
Corrección de continuidad ^b	167,227	1	<.001		
Razón de verosimilitud	155,952	1	<.001		
Prueba exacta de Fisher				<.001	<.001
Asociación lineal por lineal	169,219	1	<.001		
N de casos válidos	1001				

- a. 0 casillas (0,0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 64,88.
- b. Sólo se ha calculado para una tabla 2x2

La Figura 5 muestra la distribución de los AV en los diferentes grupos de gravedad, que también se pueden ver en forma de suma (SAV) en la figura 6, que ilustran muy bien el peso de los antecedentes de problemas vasculares en la severidad.

Figura 4 Distribución de AV (alguno/ninguno) en los diferentes grupos de gravedad

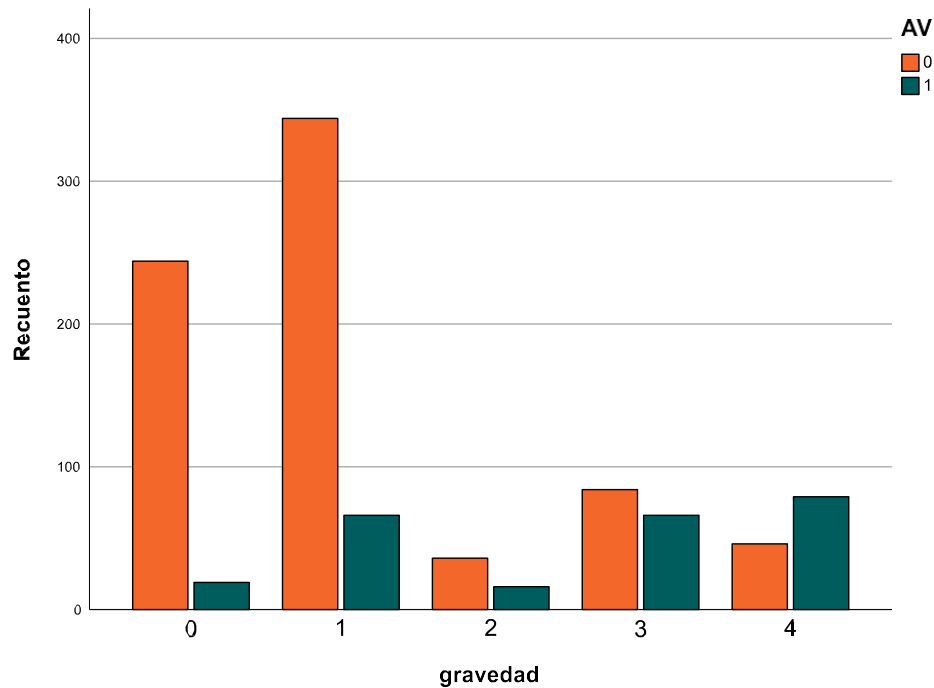
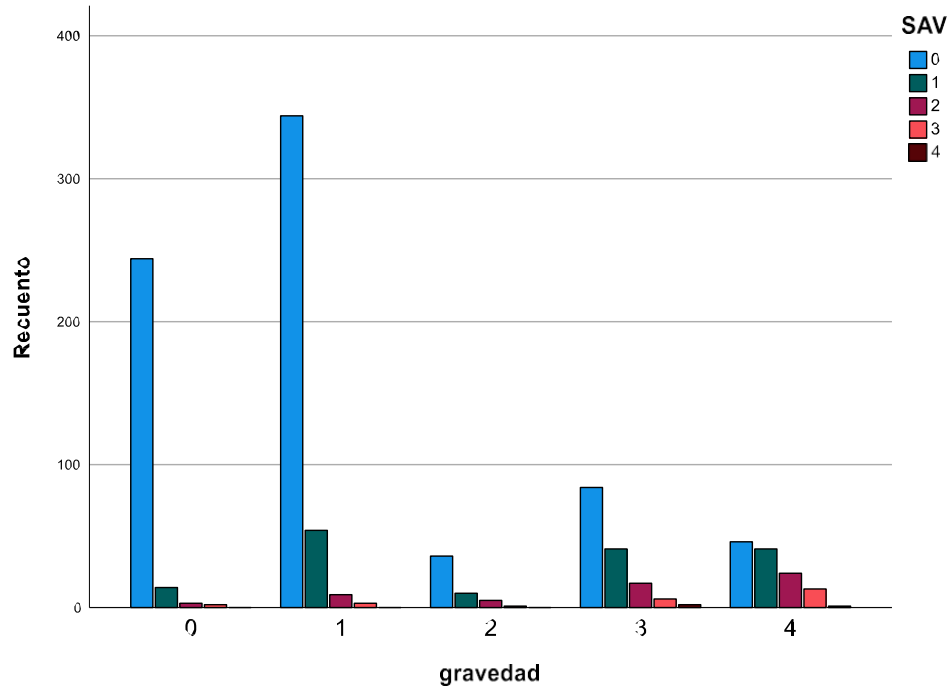
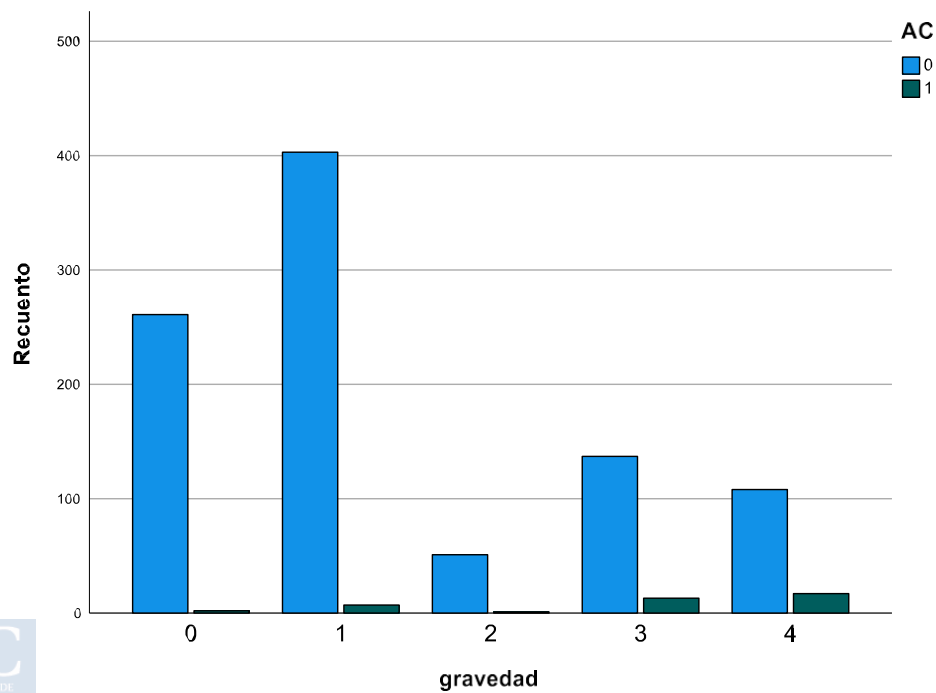


Figura 5 Distribución de SAV (suma de AVs) en los diferentes grupos de gravedad



La distribución de antecedentes de problemas cardíacos (AC) en los diferentes grupos de gravedad se ven en la Figura 7. Como se puede observar, son mucho menos frecuentes que los AV y tienen un número insuficiente como para alcanzar significancia estadística

Figura 6 Distribución de antecedentes cardiacos (AC) (alguno/ninguno) en los diferentes grupos de gravedad



En las siguientes figuras se ve la distribución de cada AV en los diferentes grupos de gravedad, incluyendo la hipertensión arterial (HTA), que es la comorbilidad más asociada (Figura 8); hipercolesterolemia (Figura 9), diabetes no complicada (Figura 10), diabetes complicada (Figura 11) y obesidad, el segundo AV en frecuencia (Figura 12).

Figura 7 Distribución de AV1 (HTA) en los diferentes grupos de gravedad

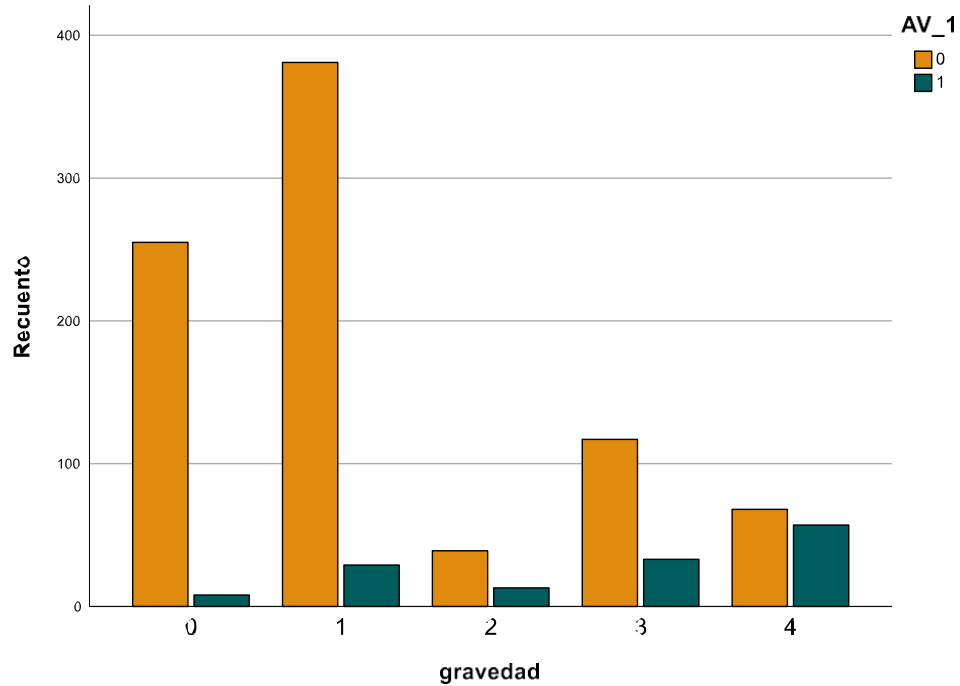


Figura 8 Distribución de AV2 (hipercolesterolemia) en los diferentes grupos de gravedad

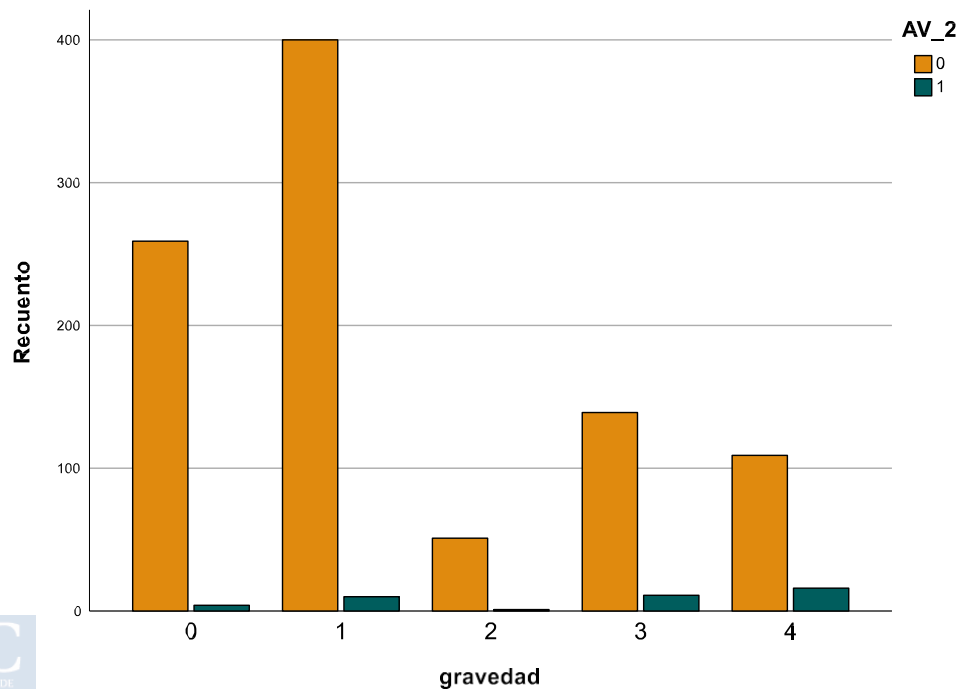


Figura 9 Distribución de AV3 (diabetes no complicada) en los diferentes grupos de gravedad

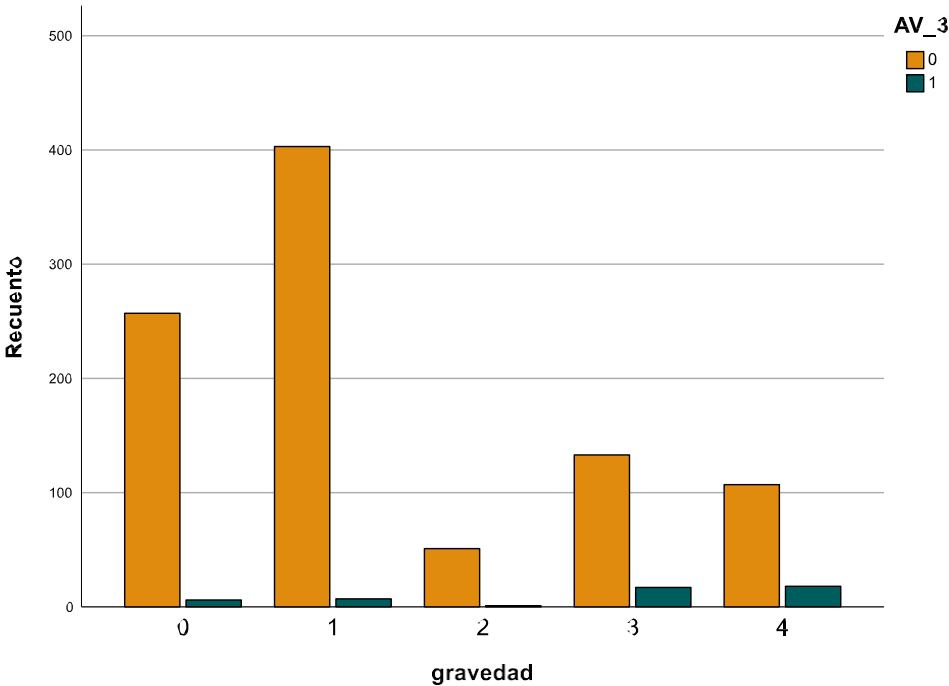


Figura 10 Distribución de AV4 (diabetes complicada) en los diferentes grupos de gravedad

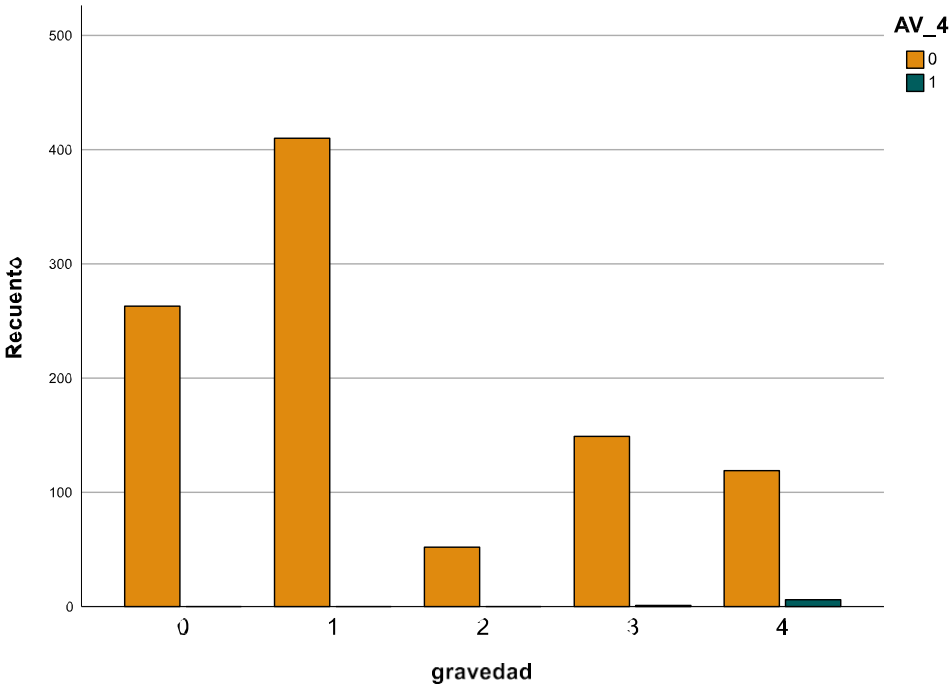
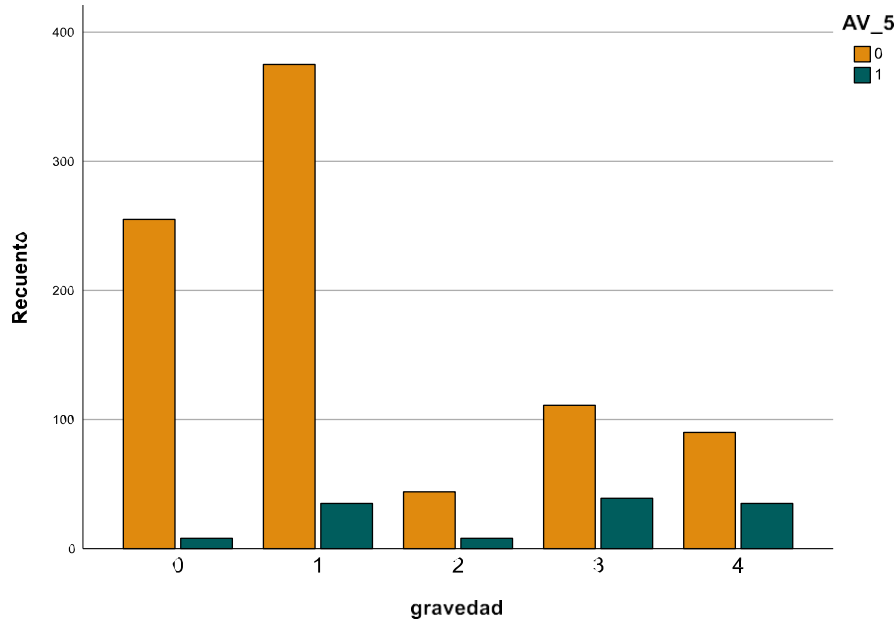
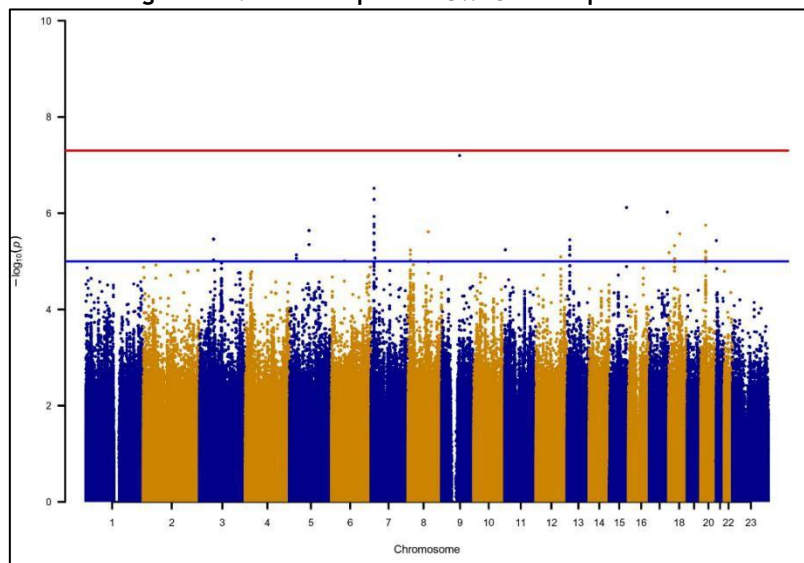


Figura 11 Distribución de 10 (obesidad) en los diferentes grupos de gravedad

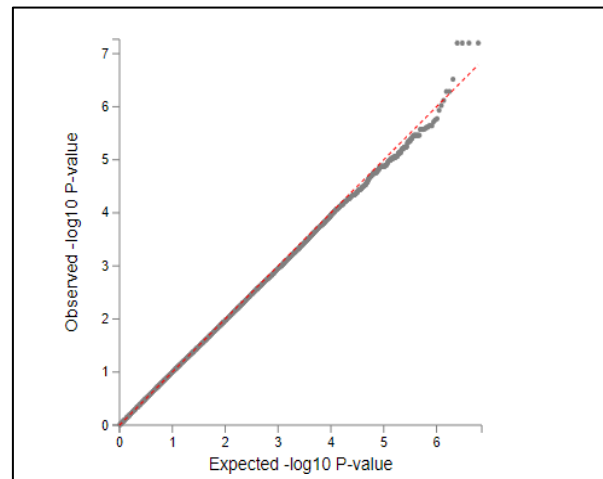
4.1 RESULTADOS GWAS PARA EL RIESGO DE HOSPITALIZACIÓN AJUSTANDO PARA COMORBILIDADES CARDIOVASCULARES

La pregunta a resolver era si existe una asociación de variantes genéticas y el riesgo de hospitalización por COVID-19 de los pacientes de la cohorte Colombia-SCOURGE, ajustando para las comorbilidades cardiovasculares (AV+AC) además de las covariables usuales (sexo, edad y PCs). En la Figura 2 se observa el Manhattan Plot en donde se puede ver que si bien ninguna señal alcanza la significación GWAS ($p=1 \times 10^{-8}$) sí que se observan señales interesantes si se considera un umbral más modesto ($p=1 \times 10^{-5}$). Así, se observan cuatro señales relevantes para el análisis, un grupo de señales en el cromosoma 7, una señal en el cromosoma 3 y otra en el 9 y un grupo de señales en el cromosoma 13, y otras señales menores en los cromosomas 18 y 20 (Figura 2).

Figura 12 Manhattan plot de GWAS de Hospitalizados

En el gráfico **QQ** se observa la distribución de valores esperados de P en el eje X, bajo un modelo nulo (*null*) de no significancia versus los valores observados de P en el eje Y. Este gráfico permite visualizar el enriquecimiento de la señal de asociación, esto indica la presencia de cuánto efecto causal existe en el análisis. En la Figura 14 se observa que solo hay efecto causal en el extremo de la gráfica.

Figura 13 Gráfico de calidad QQ de GWAS de pacientes hospitalizados con antecedentes cardiovasculares



La Tabla 7 muestra la selección de *LeadSNPs* realizada en FUMA, en donde se seleccionan los rs líderes o los que muestran mejor asociación según el modelo, y se complementa con información del mapeo por posición (tabla “snps” de FUMA) que aporta la localización de la región de interés ubicando el o los genes más cercanos según la posición de los rs líderes. Para el GWAS de riesgo de hospitalización ajustando para antecedentes cardiovasculares se observa el rs563581300 ubicado en el cromosoma 9 ($p=6.3 \times 10^{-8}$) que es el único que se acerca al umbral de referencia y que está cercano al gen *VPS13A*, ubicado en 9q21.2. Este gen no ha sido reportado en relación con la severidad ni con las manifestaciones clínicas de la COVID-19. Se trata de una señal aislada, no hay señales previas que indiquen desequilibrio de ligamiento de una región, y esto disminuye la importancia de esta asociación. Tampoco la función del gen, relacionado con el transporte de lípidos entre orgánulos, mayoritariamente expresado en cerebro y relacionado con enfermedades del SNC, apunta a una relación causal.

Tabla 7 Lead SNPs del GWAS de riesgo de hospitalización

ID	rsID	chr	P	Gen más cercano
9:79942549:A:ATAT	rs563581300	9	6.34E-08	VPS13A
7:12712310:C:T	rs17166276	7	3.04E-07	ARL4A
15:94554370:C:T	rs117269204	15	7.63E-07	CTD-3049M7.1:CTD-2643K12.3
17:75731550:C:T	rs61342344	17	9.50E-07	RP11-316M20.1
20:21344648:C:T	rs6137329	20	1.78E-06	ZNF877P
8:88380376:C:T	rs116990669	8	2.44E-06	CNBD1
18:45670076:A:C	rs68145908	18	2.68E-06	ZBTB7C

13:31325643:A:C	rs11147439	13	3.57E-06	ALOX5AP
21:14855186:G:T	rs455290	21	3.71E-06	POTE
18:23221222:A:AT	rs149699267	18	4.72E-06	MRPL23
11:2897883:C:G	rs4929953	11	5.76E-06	CDKN1C
8:9857604:A:T	rs1996733	8	5.85E-06	SLC22A18AS
18:1798171:A:G	rs112362481	18	6.59E-06	SLC22A18
5:29691415:C:G	rs146134227	5	7.38E-06	NKX2-4
12:106969216:C:T	rs1991297	12	8.06E-06	NA
7:17022854:A:G	rs62445133	7	8.55E-06	RFX4
6:54980902:A:G	rs11961656	6	9.66E-06	RGMA

Aunque los SNPs más relevantes para el análisis son rs17166276 ($p=3.4 \times 10^{-7}$) en el cromosoma 7 que es cercano al gen *ARL4A*, y el rs11147439 del cromosoma 13 ($p=3.5 \times 10^{-6}$) próximo al gen *ALOX5AP* (Tabla 3) por los resultados que después comentaremos obtenidos en MAGMA, la mayoría de los SNPs encontrados están en genes que se han relacionado con la COVID y su severidad, no siempre en estudios de GWAS si no con otras aproximaciones y que son relevantes para entender la etiopatogenia de la COVID, sus complicaciones, y para potenciales terapias.

Así sucede con CNBD1 (*cyclic nucleotide binding domain containing 1*) que se ha encontrado en varios estudios y está incluido en la “The COVID-19 Drug and Gene Set Library (<https://maayanlab.cloud/covid19/genes/CNBD1>) y que se encuentra up-regulado en pacientes con COVID-19 (<https://maayanlab.cloud/covid19/genesets/95>).

ZBTP7C, es una proteína de encendido y apagado de metaloproteasas de la matriz que son endopeptidasas que contienen zinc y juegan un papel importante en la proliferación celular, migración, diferenciación, angiogénesis y apoptosis (Jeon et al., 2016), encontrada asociada en varios GWAS y en el transcriptoma pulmonar de COVID e implicado en el papel de la modificación N6-metiladenosina del ARN en la regulación de la infección por SARS-CoV-2 e interacciones virus-huésped. Es una de las dianas más prometedoras para el tratamiento de la COVID (J. Liu et al., 2024).

POTE, la familia de POTEF y POTED contribuye a la respuesta antiviral asociada a macrófagos a (POTEF y POTED) y ha sido identificada entre los genes que juegan un papel crítico en la infección por SARS-CoV-2 (Y. Zhang et al., 2020).

MRPL23 codifica la proteína ribosómica 39S L23, la cual es de interés particular en este estudio no solo porque se ha asociado con la severidad por COVID sino porque parece que juega un papel en la tromboinflamación conducida por la fibrina en la COVID-19 (Ryu et al., 2024).

Las proteínas ribosómicas mitocondriales de mamíferos están codificadas por genes nucleares y ayudan en la síntesis de proteínas dentro de la mitocondria y hemos demostrado recientemente dentro del consorcio SCOURGE el papel que juegan en la severidad del COVID las variantes de ADN mitocondrial (Cabrera-Alarcon et al., 2025).

El gen *CDKN1C* codifica la proteína p57Kip2, que ha sido identificada como el tercer miembro de la familia CIP/Kip, que también incluye p27Kip1 y p21Cip1. En analogía con estas proteínas, p57Kip2 es capaz de unirse firmemente e inhibir los complejos ciclina/quinasa dependiente de ciclina y, a su vez, modular la progresión del ciclo de división celular. Miembros de esta familia han sido también relacionados con la severidad en la COVID (Carreras-Gallo et al., 2022).

SLC22A18AS y *SLC22A18*, están *down* regulados en la infección por COVID y sobreexpresados en cáncer de pulmón pueden ser un link entre la infección por COVID y el cáncer de pulmón según apuntan algunos estudios (Amara et al., 2024).

Como los anteriores, el gen *NKX2-4* está en el repositorio del Biogrid COVID-19 Coronavirus Curation Project (<https://thebiogrid.org/569674/summary/homo-sapiens/nkx2-4.html>) y está entre las dianas para el tratamiento de infecciones virales como las del coronavirus (Mao et al., 2024). Igualmente, está ligado a enfermedades cardiovasculares pues es un gen esencial para el desarrollo del corazón (Cao et al., 2023).

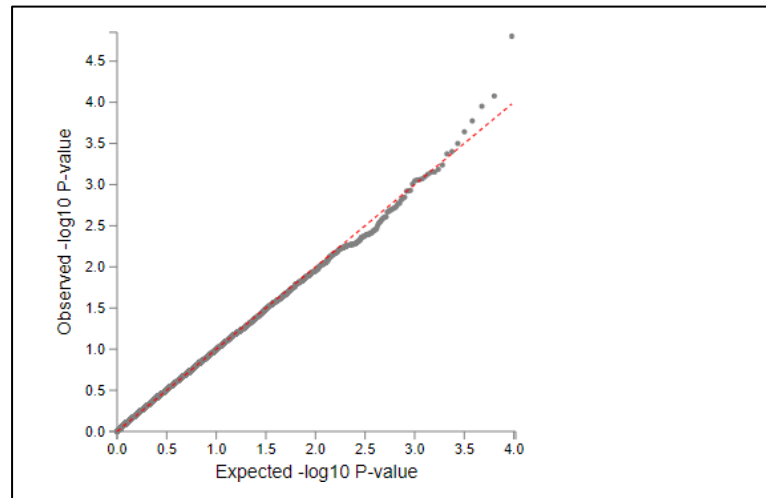
RGMA codifica un miembro de la familia de moléculas guía repulsivas. La proteína codificada es una glicoproteína anclada a glicosilfosfatidilinositol que funciona como una proteína guía de axones en el sistema nervioso central en desarrollo y en el adulto. Se encuentra asociado a severidad para COVID en varios estudios y se ha revisado de forma específica su papel en la infección por COVID. Al mismo tiempo está asociado a aterosclerosis y riesgo de accidentes cerebro-vasculares (Q. Hu et al., 2021).

Es muy significativo que la mayoría de los SNPs asociados con p valores inferiores a 10^{-6} jueguen un papel en la severidad para COVID y que varios se encuentran ligados también a problemas cardiovasculares.

4.2. ANÁLISIS A NIVEL DE GENES EN EL GWAS HOSPITALIZACIÓN AJUSTANDO PARA ANTECEDENTES CARDIOVASCULARES

Las Figuras 15 y 16 muestran los resultados del análisis a nivel de gen, MAGMA (análisis a nivel de gen a partir de los estadísticos resumen del GWAS) realizado en FUMA

En el gráfico **QQ** de genes se observa la distribución de valores esperados de P en el eje x, bajo un modelo nulo (*null*) de no significancia versus los valores observados de P en el eje y. permite visualizar el enriquecimiento de la señal de asociación. Esto indica la presencia de un efecto causal. (Figura 15). En la figura 16 se muestra el Manhattan *plot*, destacando los dos genes con mayor probabilidad de asociación: el gen *ALOX5AP* y el gen *ARL4A*, dos genes que ya aparecían en la tabla 7 del mapeo por posición de los *lead* SNP rs11147439 y rs17166276 respectivamente.

Figura 14 Gráfico de calidad Gráfico de calidad QQ de Genes de GWAS hospitalización

Las Figuras 16 y 17 muestran el análisis regional para estos *lead* SNPs. El análisis regional corresponde a una ampliación de la zona de la gráfica del Manhattan, cada SNP está codificado por colores en función del r^2 más alto de uno de los SNP, si es mayor o igual a $p=1 \times 10^{-5}$, los SNP coloreados de gris corresponden a los que están por debajo del umbral. Los SNP principales superiores en loci de riesgo genómico, SNP principales y SNP ind. Sig. están rodeados por un círculo negro y coloreados con violeta oscuro, violeta y rojo. En la Figura 16A se observa la región ampliada del cromosoma 7 en donde se identifican los SNP principales, en este caso rs17166276 en desequilibrio de ligamiento. En la figura 16B se observa el gráfico regional identificando adicionalmente la posición del gen más cercano, en este caso para el rs17166276 del cromosoma 7 es el gen *ARL4A*. En la Figura 17A se observa la región ampliada del cromosoma 13 en donde se identifican los SNP principales, en este caso rs11147439. Por último, en la Figura 17B se registra en el gráfico regional la posición del rs11147439 dentro el gen *ALOX5AP*.

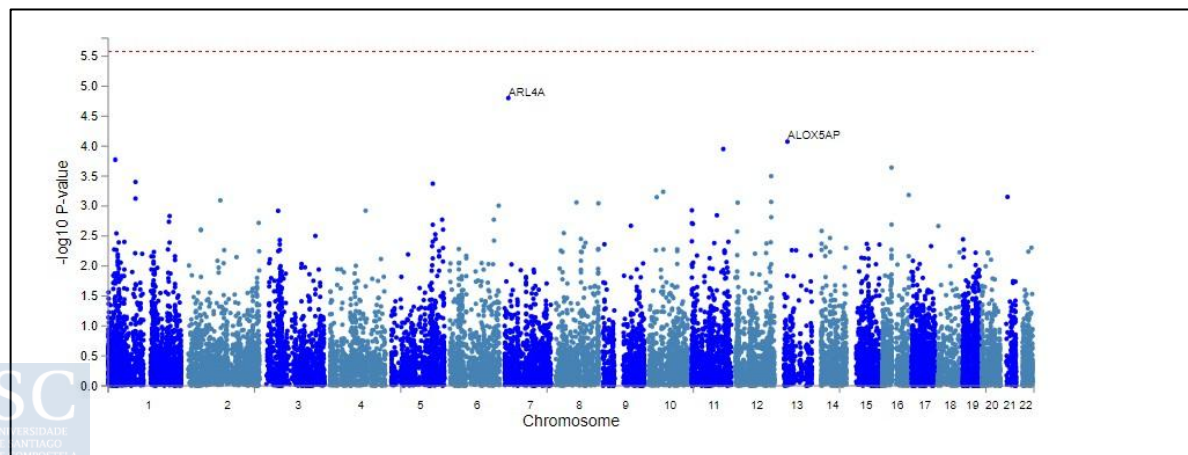
Figura 15 Manhattan plot de resultados de MAGMA (análisis a nivel de gen) para el riesgo de hospitalización

Figura 16 Gráficos regionales de la señal en ARL4A del cromosoma 7: A. Región de SBP asociados

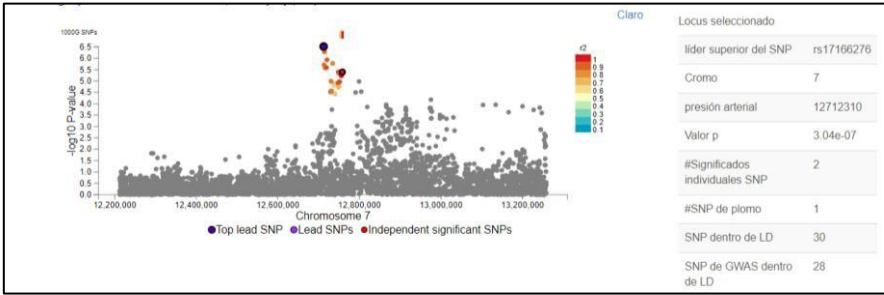


Figura 17 Gráficos regionales de la señal en ARL4A del cromosoma 7: B. Relación con el gen más cercano

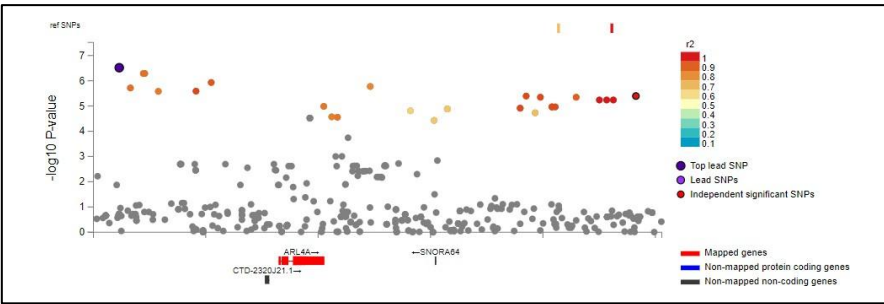


Figura 18 . Gráficos regionales de la señal en ALOX5AP del cromosoma 13: A. Región de SBP asociados

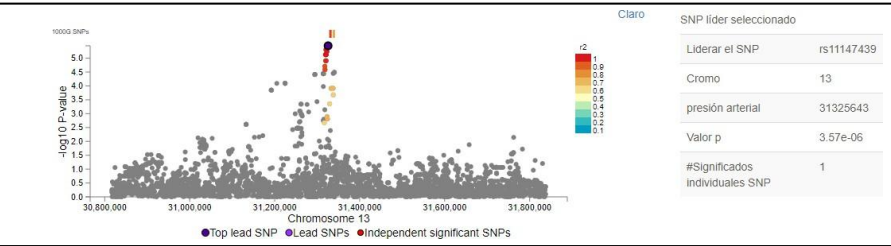
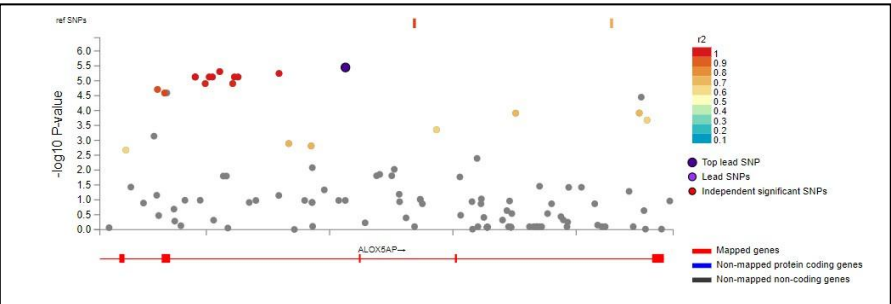


Figura 19 . Gráficos regionales de la señal en ALOX5AP del cromosoma 13: B. Relación con el gen



Los dos genes priorizados en FUMA y más relevantes en MAGMA son *ARL4A* y *ALOX5AP*.

El gen *ARL4A* [GTPasa 4ª similar al factor de ADP-ribosilación 7p21.3 (7:12,686,903-12,690,958)] es el encargado de generar una proteína reguladora de funciones celulares como el transporte de vesículas, la endocitosis y la actividad de fosforilasa D. Esta reportado en

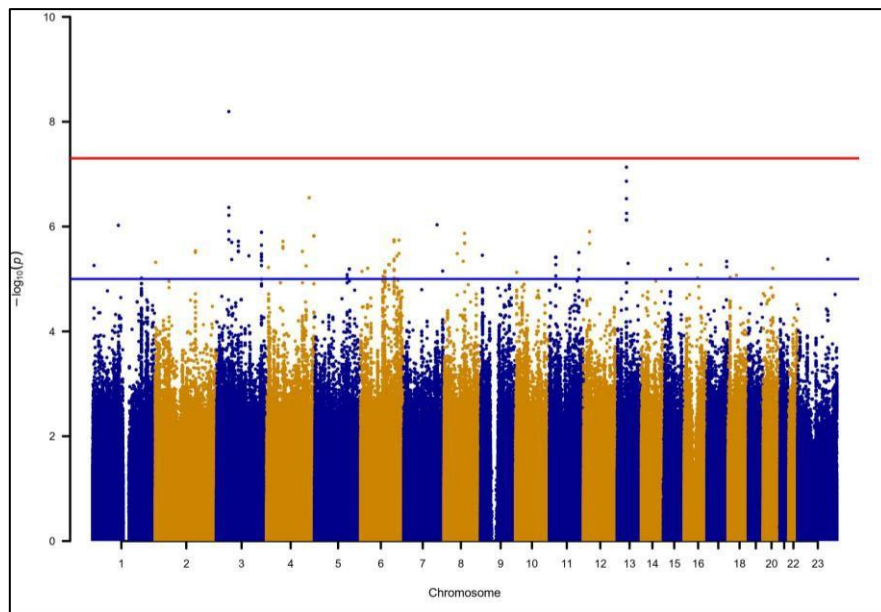
estudios de GWAS, uno de ellos en COVID 19, y está identificado como un gen de interés para posibles dianas farmacológicas para COVID 19 (Selvaraj et al., 2021).

El gen *ALOX5AP* codifica para la proteína activadora de la araquidonato 5-lipoxinasa (13q12.3 13:30,713,485-30,764,426), la cual fue reportada en asociación con COVID-19 como proteína relacionada con la inmunidad innata y mediadores lipídicos bioactivos (Sahanic et al., 2021) y en un análisis celular para identificar rutas hiperinflamatorias e inmunosupresoras en monocitos (N. Liu et al., 2021). Así mismo, también se observó asociación de esta proteína en un estudio relacionado con severidad de la COVID-19 en pacientes con cáncer (Camps-Vilaró et al., 2024) y ha sido relacionada con rasgos de presión arterial (Gouveia et al., 2021).

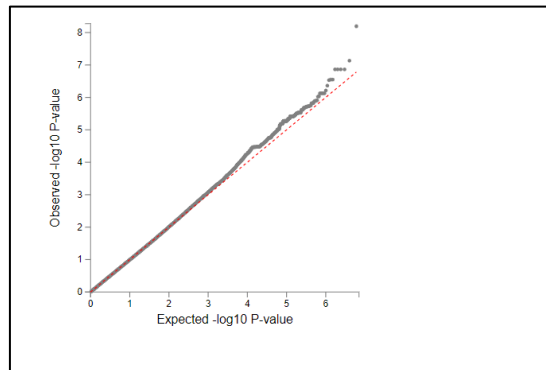
4.3 GWAS PARA RIESGO DE SEVERIDAD 4 AJUSTANDO POR ANTECEDENTES CARDIOVASCULARES

En la Figura 21 se observa el *Manhattan Plot* con los resultados del GWAS para la cohorte SCOURGE-Colombia para el riesgo de enfermedad crítica (gravedad 4) ajustando para antecedentes cardiovasculares, en donde se observa una señal que supera el umbral de significación GWAS ($p=5 \times 10^{-8}$) pero también se observan varias señales relevantes ($p < 10^{-5}$) para el análisis, un grupo de señales en el cromosoma 3, y un grupo de señales en el cromosoma 13 y unas señales menores en los cromosomas 4, 6 y 12.

Figura 20 Manhattan plot de GWAS Severidad 4



En el gráfico QQ se observa la distribución de valores esperados de P en el eje X, bajo un modelo nulo (*null*) de no significancia versus los valores observados de P en el eje Y. permite visualizar el enriquecimiento de la señal de asociación, esto indica la presencia de que tanto efecto causal existe en el análisis. En la figura 22 se observa que solo hay efecto causal en el extremo de la gráfica.

Figura 21 Gráfico de calidad QQ de GWAS de pacientes severidad4 con antecedentes cardiovasculares

Para la anotación de los SNPs representativos se combinaron los resultados de las tablas que arroja FUMA específicamente *LeadSNPs*, en donde se seleccionan los rs líderes o los que muestran mejor asociación según el modelo, y se complementa con información de la tabla “snps” que aporta la localización de la región de interés ubicando el o los genes más cercanos según la posición de los rs líderes (Tabla 8). Para el GWAS de riesgo de gravedad crítica ajustando para antecedentes cardiovasculares se observan dos SNPs, rs143861937; rs73078305, ubicados en el cromosoma 3 que pasan el umbral de referencia ($p=6.4 \times 10^{-9}$), cercanos a los genes *PLXNB1* y al gen *PFHB4*. Otros SNPs relevantes para el análisis son rs73151085 ($p=9.2 \times 10^{-7}$) del cromosoma 7 que es cercano al gen *PLXNA4*, y el rs116849267 del cromosoma 12 ($p=1.25 \times 10^{-6}$) cercano al gen *SOX5*. Otro SNP de interés en la tabla 8 es el rs74083314 cercano al gen *PCDH20* porque es el SNP seleccionado de un grupo de puntos que se ven en la imagen del Manhattan (Figura 21).

Tabla 8 Lead SNPs del GWAS gravedad 4 ajustando para antecedentes cardiovasculares

uniqID	Chr	p	IndSigSNPs	GEN más cercano
3:48451307:A:G	3	6.40E-09	rs143861937;rs73078305	PLXNB1;PFKB4
7:131721534:A:G	7	9.26E-07	rs73151085	PLXNA4
12:24073260:C:T	12	1.25E-06	rs116849267	SOX5
3:177301894:A:G	3	1.29E-06	rs6767036	LINC00578
8:82671111:C:T	8	1.35E-06	rs150715017	CHMP4C
6:132888583:G:T	6	1.80E-06	rs9493380;rs4085406	TAAR6
6:152386628:A:G	6	1.82E-06	rs3778081	ESR1
3:86215425:C:T	3	1.92E-06	rs4133179;rs145832288	snoZ40
4:65227230:A:T	4	1.92E-06	rs78476998	TECRL
3:59780948:C:T	3	2.01E-06	rs146617297	FHIT
2:159072217:A:G	2	2.92E-06	rs4664934	CCDC148-AS1:CCDC148
13:61923572:A:C	13	5.04E-06	rs74083314	PCDH20

El gen *PLXNB1* ubicado en 3p21.31, genera una proteína denominada plexina B1. Este gen no aparece como resultado significativo en los GWAS realizados para severidad para COVID. Sin embargo, un estudio reciente lo encuentra entre las proteínas neurológicas desreguladas en el plasma de pacientes con COVID severa y la considera una diana prometedora para intervenciones terapéuticas (El-Agnaf et al., 2023).

En la vecindad de *PLXNB1* se encuentra *PFKFB4*, también ubicado en 3p21.31, traduce para una fructuosa 2,6 bifosfatasa 4, que es un activador de una la fosfofructoquinasa, enzima reguladora clave de la glucólisis. Posee varias isoenzimas que difieren en las regiones y son importantes para la respuesta diferencial a los efectores alostéricos y las señales hormonales en diferentes tejidos. La expresión de este gen se encontró regulado positivamente en las células CAR-T CD22 administradas a pacientes con infecciones virales, estos resultados fueron validados en cohortes de pacientes con COVID-19, Influenza, enfermedades autoinmunes y en tejidos tumorales. Se sugiere que *PFKFB4* actúa como un factor impulsor en la mediación de la liberación de citoquinas in vivo al regular la actividad glicolítica, es posible pensar entonces en que sería beneficioso desarrollar fármacos dirigidos a *PFKFB4* y la vía glucolítica para el tratamiento del Síndrome de liberación de citoquinas (Shao et al., 2022).

PLXNA4 codifica para la plexina A4. Esta proteína se une a la neuropilina 1 (Nrp1) y la neuropilina 2 (Nrp2) y transduce señales de Sema3A, Sema6A y Sema6B. Estos complejos Nrp-plexina y semaforina inician cascadas que regulan diversos procesos como la poda y repulsión de axones, la atracción y ramificación dendrítica, la regulación de la migración celular y la remodelación vascular. Concretamente, *PLXNA4* es un gen que se ha encontrado asociado con el COVID persistente particularmente en los casos de mayor severidad de síntomas (Udomsinprasert et al., 2023).

SOX5 codifica un miembro de la familia de factores de transcripción SOX (SRY- related HMG-box) implicados en la regulación del desarrollo embrionario y muerte celular. La proteína codificada puede actuar como regulador transcripcional después de formar un complejo proteico con otras proteínas. Fue incluida en un estudio reciente que identificó dos poblaciones de memoria (CD80+/ISG15+ y CD11c+/SOX5+/T-bet+/-) con signos inmunogenéticos y transcripcionales de autorreactividad que pueden ser la fuente celular de autoanticuerpos en COVID-19 y que pueden persistir más allá de la recuperación, por lo que puede estar involucrada en la severidad y persistencia de la COVID (Schultheiß et al., 2021).

LINC00578, es un *long non-coding RNA* está incluido como un actor relevante en la COVID en el repositorio “*The COVID-19 Drug and Gene Set Library*” (“<https://maayanlab.cloud/covid19/genes/LINC00578>”), donde aparece como un gen *down* regulado en el epitelio de las vías respiratorias en la infección por coronavirus. También muy asociado con infecciones virales es un gen que aparece como el más significativamente asociado en las mismas (Silva et al., 2022)

CHMP4C pertenece a la familia de proteínas modificadoras de la cromatina -proteínas corporales multivesiculares cargadas (CHMP). Está también referido en la *The COVID-19 Drug and Gene Set Library* (<https://maayanlab.cloud/covid19/genes/LINC00578>) y se ha asociado a la respuesta inmune en pacientes con COVID-19 (Bi et al., 2022).

ERS1 es un gen que codifica para un receptor de estrógenos, el cual está asociado con la severidad a COVID y parece que puede estar implicado en las diferencias que se observan en la gravedad por el sexo (Stilhan et al., 2020), y que también hemos visto en nuestro artículo principal del consorcio SCOURGE (Cruz et al., 2022).



TECRL codifica para una proteína de retículo endoplásmico que se expresa preferencialmente en el corazón. Es un gen muy ligado a arritmias y a muerte súbita de origen

cardíaco ligado a que algunas variantes producen una función mitocondrial aberrante en los cardiomiocitos (Hou et al., 2022). Está incluido también en estudios para valorar su papel, no solo en la fase aguda de COVID, sino en el COVID persistente (Vallejo Camazón et al., 2022).

PCDH20 es uno de los genes de mayor interés. Es un gen localizado en 13q21.2 que codifica para la protocadherina 20. Las protocadherinas son proteínas transmembrana que participan en la adhesión celular. Este gen está registrado en el catálogo GWAS en un análisis de interacción entre genes y ambiente que incorpora el sexo, las enfermedades cardiometabólicas y el Índice de privaciones múltiples en el que muestra una asociación con la gravedad de la COVID-19 (Westerman et al., 2022).

El gen de la tríada frágil de histidina diadenosina trifosfatasa (*FHIT*) se encuentra en el sitio frágil común FRA3B y codifica una enzima que hidroliza el dinucleótido Ap3A. *FHIT* es un gen que ha sido asociado con la severidad para la COVID a nivel de metilación y parece que podría tener implicaciones en los síntomas para COVID a largo plazo (Z. Zhang, 2024).

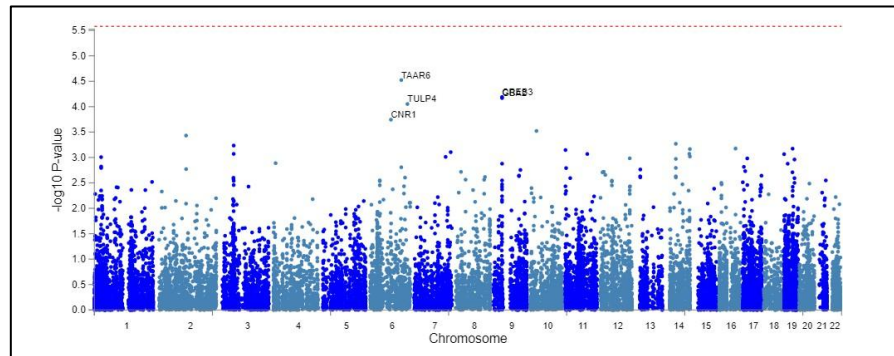
Como en el caso del GWAS de hospitalización ajustando para antecedentes de riesgo cardiovascular es relevante el número de genes en los SNPs priorizados que pueden tener un papel relevante en la severidad para COVID, también en relación con este tipo de problemas y en la COVID persistente.

4.4 ANÁLISIS A NIVEL DE GENES EN EL GWAS SEVERIDAD 4 AJUSTANDO PARA ANTECEDENTES CARDIOVASCULARES

El análisis a nivel de genes es producto del proceso de los datos en FUMA, en donde obtiene una probabilidad de asociación por gen calculado por MAGMA en función de las estadísticas de resumen de los GWAS, estos resultados se observan en un gráfico de Manhattan de genes que tienen su correspondiente análisis de calidad. Así, en el gráfico QQ de genes se observa la distribución de valores esperados de P en el eje x, bajo un modelo nulo (null) de no significancia versus los valores observados de P en el eje y. permite visualizar el enriquecimiento de la señal de asociación, esto indica la presencia de un efecto causal. (Figura 22.)

Los resultados de MAGMA se observan en el Manhattan plot de la figura 23. Allí se resaltan los genes que tienen mayor significancia o impacto, en este caso son los genes *TAAR6*, *TULP4*, *CNR1*, del cromosoma 6 y el *CREB*, *GBA2* del cromosoma 9.

Figura 22 Manhattan plot de genes para análisis de gravedad crítica ajustando para antecedentes cardiovasculares



TAAR6 es un gen sin intrones que codifica una proteína de choque térmico de 70 kDa que es miembro de la familia de proteínas de choque térmico 70. Junto con otras proteínas de choque térmico, esta proteína estabiliza las proteínas existentes contra la agregación y media en el plegamiento de proteínas recién traducidas en el citosol y en los orgánulos. También participa en la vía ubiquitina-proteasoma a través de la interacción con la proteína 1 de unión a ARN del elemento rico en AU. El gen está ubicado en la región de clase III del complejo mayor de histocompatibilidad, en un grupo con dos genes estrechamente relacionados que codifican proteínas similares. Se sabe que los TAAR están ampliamente representados en las células del sistema inmunológico humano y aunque hay solo evidencias indirectas de que puede estar involucrado en la severidad para la COVID si puede estar relacionado con la respuesta al tratamiento por corticoides (Moiseenko et al., 2024). *TULP4* (TUB Like Protein 4) es un factor de transcripción que en genética clínica es conocido por estar asociado con el síndrome Coffin-Siris 1 y la displasia craneoectodérmica. Aunque hay muy poca evidencia de su papel en COVID, entre sus vías relacionadas se encuentran el metabolismo de proteínas y el procesamiento y presentación de antígenos mediados por MHC de clase I. Un parálogo importante de este gen es *WDR35* que, si aparece en algún estudio de GWAS de COVID, pero no ha sido replicado.

CNR1 es un gen que codifica uno de los dos receptores de cannabinoides y su expresión está aumentada en los pulmones de los pacientes con COVID-19 y está considerado como una potencial diana terapéutica para los pacientes con severidad (Cinar et al., 2022).

CREB (*cAMP response element-binding*) es una proteína que actúa como factor de transcripción. Se une a ciertas secuencias de ADN llamadas "elementos de respuesta a AMPc" (*cAMP Response Element*), mediante los cuales incrementa o reduce la transcripción "corriente abajo" (*downstream*) regulada por estos genes. Varios estudios han demostrado la importancia de *CREB/CBP* inducida por la vía AMPc-PKA durante la infección por SARS-CoV-2 y, además, proporcionaron nuevos objetivos farmacológicos terapéuticos de interacción *CREB/CBP* para COVID-19. En concreto, la vía de la proteína quinasa AMPc (PKA) tiene un papel esencial en la infección por SARS-CoV-2, debido a la interacción entre la proteína de

unión al elemento de respuesta del AMP cíclico (CREB) y la proteína de unión a CREB (CBP) que podría inducirse y conducir para la mejora de la actividad transcripcional CREB/CBP (Q. Yang et al., 2023).

GBA2 es el gen que codifica la glucosilceramidasa no lisosómica en humanos. Tiene actividad glucosilceramidasa. Se trata de una beta-glucosidasa microsomal que cataliza a hidrólisis de los 3-O-glucósidos de ácidos biliares como compuestos endógenos. Se cree que esta supuesta proteína transmembrana desempeña un papel en el transporte y el metabolismo de los carbohidratos. Ha sido relacionado en un estudio GWAS reciente de epítasis con la severidad para COVID (Lin et al., 2023).

4.5 RESULTADOS DE PRS

A continuación, se muestran los resultados obtenidos para los distintos modelos. Los modelos de regresión logística fueron utilizados para evaluar la capacidad de las PRSs de hipertensión (SBP, DBP y PP) para predecir la hospitalización por COVID-19 (HOSP) y la gravedad de la COVID-19 (SEV4). Estos modelos incluyeron el estatus caso/control -HOSP y SEV4- de COVID-19 como variable dependiente y se ajustaron para las covariables: sexo, edad y 4 primeros componentes principales (PC), con y sin el PRS a evaluar. Se generaron PRSs teniendo en cuenta dos selecciones de SNPs diferentes:

- los marcadores de hipertensión de la tabla ST4 de Keaton et al., 2024 (2,068 SNPs).
- los marcadores con asociación significativa a nivel GWAS en el mismo artículo ($p < 5 \times 10^{-8}$) -hit SNPs-.

Se analizaron, por tanto, un total de 6 PRS (PRS_{SBP}, PRS_{DBP}, PRS_{PP}, PRS_{Hits SBP}, PRS_{Hits DBP}, PRS_{Hits PP}) con dos variables dependientes (HOSP y SEV4).

4.4.1 PRS hipertensión - Asociación con riesgo de hospitalización por COVID-19

La tabla 9 muestra los resultados del análisis de regresión logística sobre el riesgo de hospitalización para las tres PRS de hipertensión generadas a partir de los 2068 SNPs de la tabla suplementaria de Keaton et al., 2024. En todos los casos los resultados fueron muy semejantes: el PRS fue no significativo y solo mostraron asociación significativa las covariables sexo (pendiente negativa, mayor riesgo en hombres), edad (pendiente positiva, mayor riesgo a mayor edad) y el primer componente principal.

Las figuras 24, 25 y 26 muestran las gráficas de área bajo la curva (AUC) para el modelo de covariables con y sin la PRS y adicionalmente se muestra, a título orientativo, el AUC del modelo que incluye únicamente la PRS (sin covariables). Las figuras de las tres PRS fueron muy similares, mostrando muy poca diferencia al incorporar la PRS con respecto al modelo basal con covariables.

Tabla 9 Resultados del análisis de regresión logística del modelo completo (covariables y PRS) sobre el riesgo de hospitalización por COVID-19

Modelo	Variables	Estimado	Error estándar	Valor Z	Pr(> z)
PRS _{SBP}	R ² Nagelkerke	0.0763			

	(Intercept)	-4.255837	0.434193	-9.802	< 2e-16 ***
	SCORE_st	-0.079841	0.090873	-0.879	0.380
	coPC1	-14.644519	3.295127	-4.444	8.82e-06 ***
	coPC2	3.893208	2.860757	1.361	0.174
	coPC3	-2.897491	2.783653	-1.041	0.298
	coPC4	3.750787	2.945949	1.273	0.203
	SEX	-0.857089	0.184947	-4.634	3.58e-06 ***
	EDAD	0.099674	0.007074	14.090	< 2e-16 ***
<u>PRSDBP</u>	<i>R2 Nagelkerke</i>	<i>0.0392</i>			
	(Intercept)	-4.272217	0.433883	-9.846	< 2e-16 ***
	SCORE_st	-0.057221	0.090877	-0.630	0.529
	coPC1	-14.706154	3.289404	-4.471	7.79e-06 ***
	coPC2	4.065569	2.860478	1.421	0.155
	coPC3	-2.848767	2.783920	-1.023	0.306
	coPC4	3.707408	2.947458	1.258	0.208
	SEX	-0.845802	0.184143	-4.593	4.37e-06 ***
	EDAD	0.099666	0.007073	14.091	< 2e-16 ***
<u>PRSPp</u>	<i>R2 Nagelkerke</i>	<i>0.0773</i>			
	(Intercept)	-4.252585	0.434688	-9.783	< 2e-16 ***
	SCORE_st	-0.081323	0.092004	-0.884	0.377
	coPC1	-14.709502	3.294872	-4.464	8.03e-06 ***
	coPC2	3.761835	2.870992	1.310	0.190
	coPC3	-2.873522	2.782804	-1.033	0.302
	coPC4	3.868394	2.949762	1.311	0.190
	SEX	-0.861016	0.185436	-4.643	3.43e-06 ***
	EDAD	0.099749	0.007076	14.097	< 2e-16 ***

(***) < 0.001

Figura 23 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión: sistólica (SBP) para la hospitalización por COVID-19

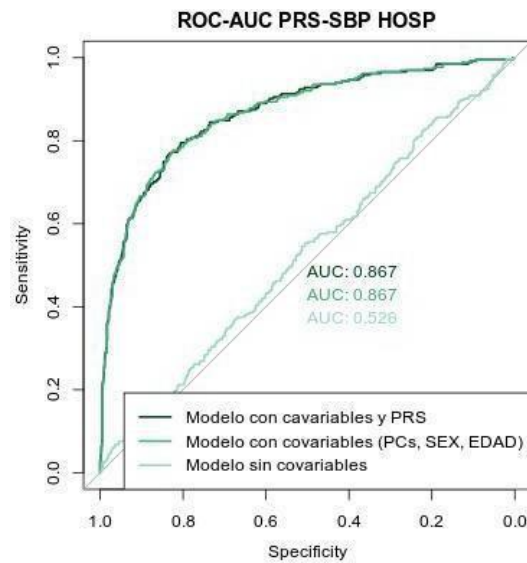


Figura 24 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión: diastólica (DBP) para la hospitalización por COVID-19

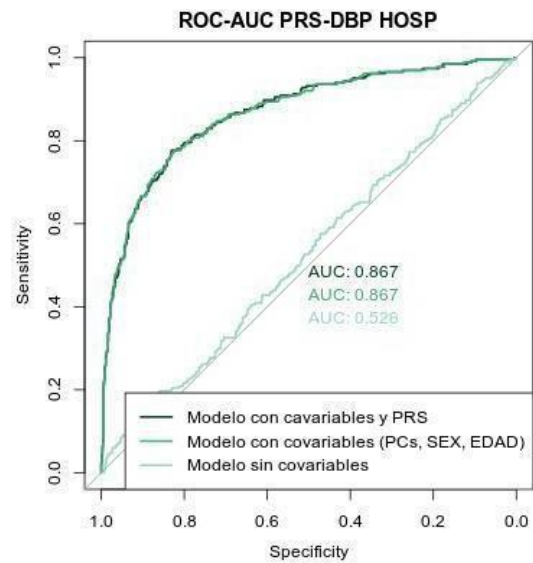
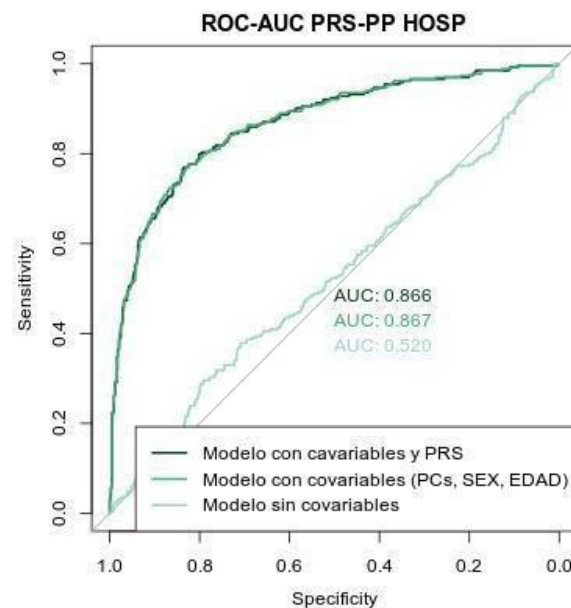


Figura 25 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión: presión de pulso (PP) para la hospitalización por COVID-19



Por otra parte, la tabla 10 y las figura 27, 28 y 29 muestran los resultados de los modelos de regresión logística sobre hospitalización para las PRS de hipertensión basadas en los hits ($p < 5 \times 10^{-8}$) del GWAS de Keaton y colaboradores (2024). Los resultados son muy similares a los encontrados con las PRS basadas en 2068 SNPs; ninguna PRS estuvo significativamente

asociada con el riesgo de hospitalización y el incremento en AUC del modelo con PRS fue mínimo con respecto al modelo de covariables.

Tabla 10 Asociación hospitalización y PRS basadas en hits. Resultados del análisis de regresión logística del modelo completo (covariables y PRS) sobre el riesgo de hospitalización

Modelo	Variables	Estimado	Error estándar	Valor Z	Pr(> z)
<u>PRS_{Hits SBP}</u>	<i>R2 Nagelkerke</i>	<i>0.1976</i>			
	(Intercept)	-4.2532	0.4336	-9.809	< 2e-16 ***
	SCORE_st	-0.1284	0.0909	-1.414	0.157
	coPC1	-14.5027	3.3020	-4.392	1.12e-05 ***
	coPC2	3.7287	2.8629	1.302	0.193
	coPC3	-3.0022	2.7835	-1.079	0.281
	coPC4	3.5267	2.9522	1.195	0.232
	SEX	-0.8654	0.1851	-4.675	2.94e-06 ***
	EDAD	0.0998	0.0071	14.099	< 2e-16 ***
<u>PRS_{Hits DBP}</u>	<i>R2 Nagelkerke</i>	<i>0.01955</i>			
	(Intercept)	-4.2749	0.4341	-9.847	< 2e-16 ***
	SCORE_st	-0.0401	0.0903	-0.445	0.656
	coPC1	-14.7095	3.2932	-4.467	7.95e-06 ***
	coPC2	3.9447	2.8599	1.379	0.168
	coPC3	-2.8302	2.7852	-1.016	0.310
	coPC4	3.7125	2.9554	1.256	0.209
	SEX	-0.8450	0.1841	-4.589	4.45e-06 ***
	EDAD	0.0997	0.0070	14.098	< 2e-16 ***
<u>PRS_{Hits PP}</u>	<i>R2 Nagelkerke</i>	<i>0.1044</i>			
	(Intercept)	-4.255561	0.4344	-9.796	< 2e-16 ***
	SCORE_st	-0.095531	0.0930	-1.027	0.304
	coPC1	-14.927924	3.2967	-4.528	5.95e-06 ***
	coPC2	3.611197	2.8823	1.253	0.210
	coPC3	-2.900530	2.7828	-1.042	0.297
	coPC4	3.901992	2.9514	1.322	0.186
	SEX	-0.863478	0.1854	-4.656	3.22e-06 ***
	EDAD	0.099893	0.0071	14.115	< 2e-16 ***

(***) < 0.001

Figura 26 Área bajo la curva (AUC) de la predicción de los PRS de hipertensión basados en hits: sistólica (hits SBP, arriba izquierda) para la hospitalización por COVID-19. Las líneas corresponden a la precisión obtenida por los modelos.

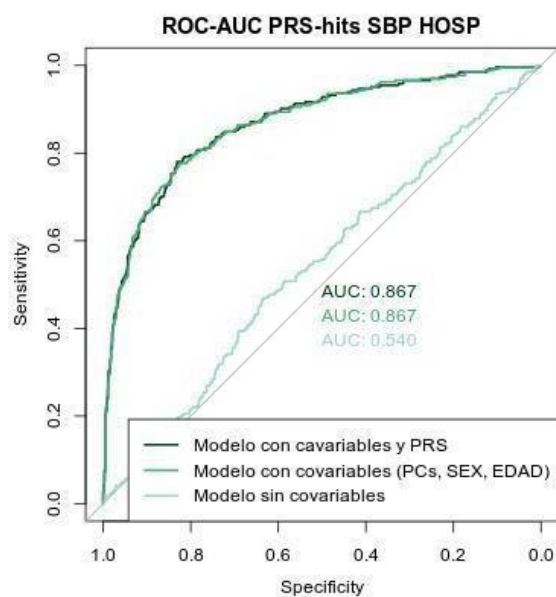


Figura 27 Área bajo la curva (AUC) de la predicción de los PRS de hipertensión basados en hits: diastólica (hits DBP). Las líneas corresponden a la precisión obtenida por los modelos.

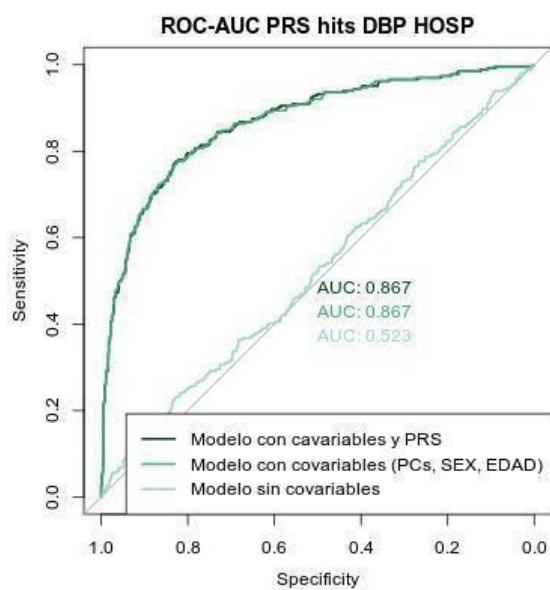
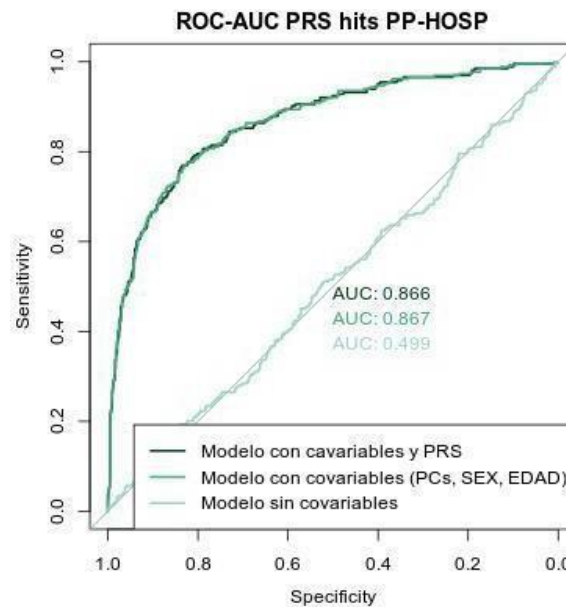


Figura 28 Área bajo la curva (AUC) de la predicción de los PRS de hipertensión basados en hits: sistólica (hits SBP, arriba izquierda), diastólica (hits DBP, arriba derecha) y presión de pulso (hits PP, abajo) para la hospitalización por COVID-19. Las líneas corresponden a la precisión obtenida por los modelos.



4.4.2 PRS hipertensión - Asociación con riesgo de gravedad crítica por COVID-19

Los resultados de los modelos de regresión logística usando como variable dependiente la gravedad crítica (SEV4) e incluyendo en el modelo las covariables y cada uno de las PRS de hipertensión probadas se muestran en la tabla 11 (PRS con los 2068 SNPs de la tabla de Keaton) y en la tabla 11 (PRS con el subgrupo de *hits* del GWAS de Keaton). Las figuras siguientes muestran los cambios en AUC al añadir el PRS al modelo de covariables, considerando los 2068 SNPs o los *hits*.

Tabla 11 Asociación gravedad crítica (SEV4) y PRS de hipertensión. Resultados del análisis de regresión logística del modelo completo (covariables y PRS) sobre el riesgo de COVID con gravedad crítica

Modelo	Variables	Estimado	Error estándar	Valor Z	Pr(> z)
PRS _{SBP}	R2 Nagelkerke	0.2674			
	(Intercept)	-5.794878	0.582701	-9.945	< 2e-16 ***
	SCORE_st	-0.153056	0.114937	-1.332	0.1830
	coPC1	-8.635960	3.930489	-2.197	0.0280 *
	coPC2	7.906141	3.592715	2.201	0.0278 *
	coPC3	5.386388	3.830576	1.406	0.1597
	coPC4	-0.193537	3.721025	-0.052	0.9585
	SEX	-0.476022	0.232369	-2.049	0.0405 *
	EDAD	0.091687	0.007958	11.521	< 2e-16 ***
PRS _{DBP}	R2 Nagelkerke	0.1304			
	(Intercept)	-5.812899	0.581949	-9.989	< 2e-16 ***
	SCORE_st	-0.107571	0.115554	-0.931	0.3519

	coPC1	-8.701796	3.913999	-2.223	0.0262 *
	coPC2	8.322855	3.580725	2.324	0.0201 *
	coPC3	5.540117	3.830259	1.446	0.1481
	coPC4	-0.198631	3.722670	-0.053	0.9574
	SEX	-0.460191	0.231650	-1.987	0.0470 *
	EDAD	0.091622	0.007939	11.541	< 2e-16 ***
PRS_{pp}	R² Nagelkerke	0.2717			
	(Intercept)	-5.796099	0.583720	-9.930	< 2e-16 ***
	SCORE _{st}	-0.154310	0.115112	-1.341	0.1801
	coPC1	-8.779554	3.933398	-2.232	0.0256 *
	coPC2	7.634701	3.623753	2.107	0.0351 *
	coPC3	5.491419	3.824543	1.436	0.1510
	coPC4	0.007781	3.723418	0.002	0.9983
	SEX	-0.478365	0.232584	-2.057	0.0397 *
	EDAD	0.091811	0.007963	11.529	< 2e-16 ***

(*) < 0.05; (***) < 0.001

Figura 29 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión: sistólica (SBP)

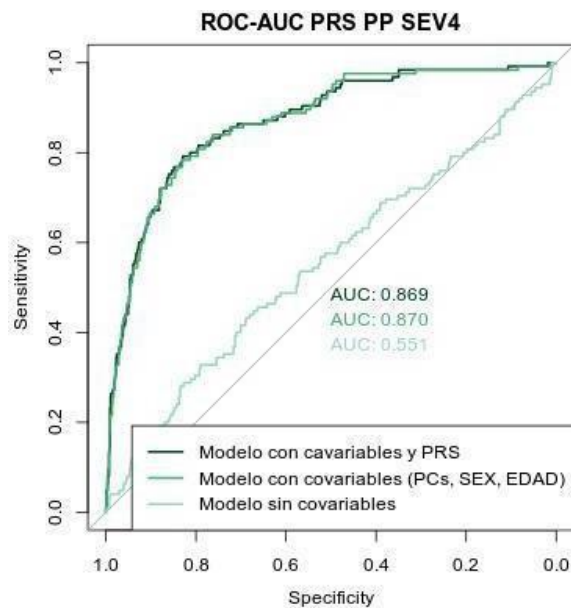


Figura 30 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión: sistólica (SBP). Para la gravedad (SEV4) por COVID-19. Las líneas corresponden a la precisión obtenida para los modelos

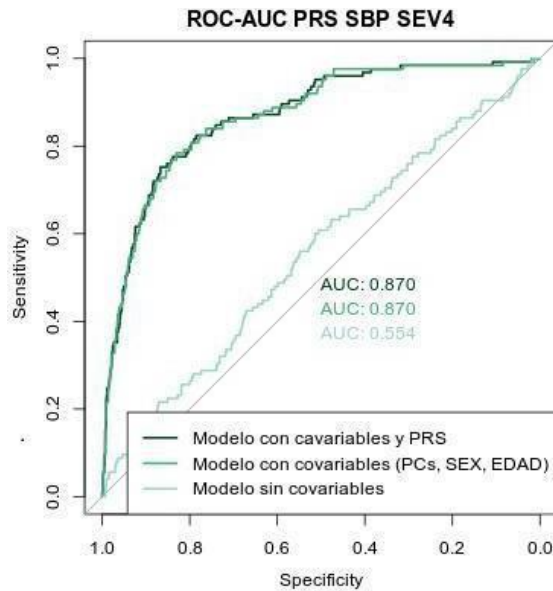


Figura 32 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión: diastólica (DBP) para la gravedad (SEV4) por COVID-19. Las líneas corresponden a la precisión obtenida para los modelos

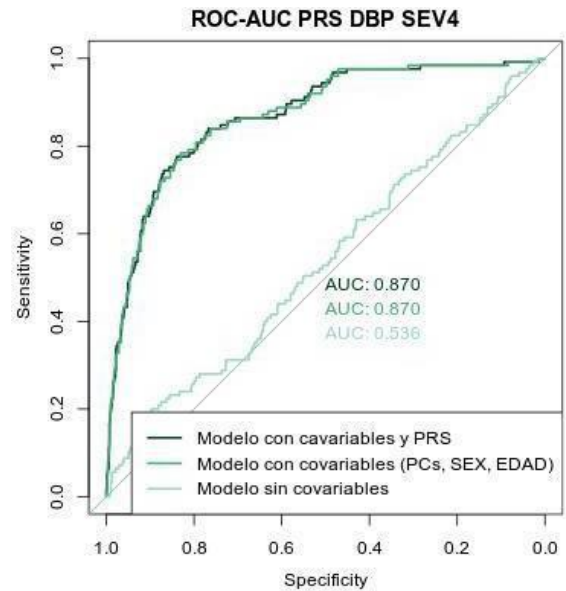


Figura 31 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión: presión de pulso (PP) para la gravedad (SEV4) por COVID-19. Las líneas corresponden a la precisión obtenida para los modelos

Tabla 12 Asociación gravedad crítica (SEV4) y PRS basadas en hits. Resultados del análisis de regresión logística del modelo completo (covariables y PRS) sobre el riesgo de gravedad crítica

Modelo	Variabes	Estimado	Error estándar	Valor Z	Pr(> z)
<u>PRS_{SBP}</u>	<i>R² Nagelkerke</i>	<i>0.2803</i>			
	(Intercept)	-5.801822	0.582370	-9.962	< 2e-16 ***
	SCORE_st	-0.157888	0.115601	-1.366	0.1720
	coPC1	-8.615334	3.935065	-2.189	0.0286 *
	coPC2	7.737762	3.597305	2.151	0.0315 *
	coPC3	5.319305	3.837384	1.386	0.1657
	coPC4	-0.426237	3.733609	-0.114	0.9091
	SEX	-0.478800	0.232520	-2.059	0.0395 *
	EDAD	0.091906	0.007968	11.534	< 2e-16 ***
<u>PRS_{DBP}</u>	<i>R² Nagelkerke</i>	<i>0.1421</i>			
	(Intercept)	-5.813055	0.581563	-9.996	< 2e-16 ***
	SCORE_st	-0.111285	0.114460	-0.972	0.3309
	coPC1	-8.628722	3.916560	-2.203	0.0276 *
	coPC2	8.032627	3.584798	2.241	0.0250 *
	coPC3	5.510876	3.834247	1.437	0.1506
	coPC4	-0.304421	3.728285	-0.082	0.9349
	SEX	-0.461880	0.231719	-1.993	0.0462 *
	EDAD	0.091697	0.007938	11.552	< 2e-16 ***
<u>PRS_{PP}</u>	<i>R² Nagelkerke</i>	<i>0.3094</i>			
	(Intercept)	-5.80972	0.58374	-9.953	< 2e-16 ***
	SCORE_st	-0.16758	0.11711	-1.431	0.1524
	coPC1	-9.17805	3.94211	-2.328	0.0199 *
	coPC2	7.34725	3.65281	2.011	0.0443 *
	coPC3	5.54437	3.82221	1.451	0.1469
	coPC4	0.05956	3.71869	0.016	0.9872
	SEX	-0.48400	0.23291	-2.078	0.0377 *
	EDAD	0.09226	0.00799	11.548	< 2e-16 ***

(*) < 0.05; (***) < 0.001

Figura 33 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión basadas en hits: sistólica (hits SBP), para la gravedad (SEV4) por COVID-19. Las líneas corresponden a la precisión obtenida para los modelos.

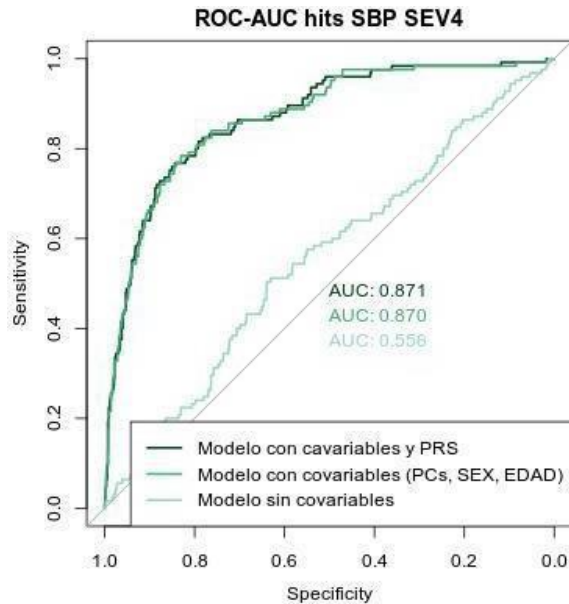


Figura 34 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión basadas en hits: diastólica (hits DBP), para la gravedad (SEV4) por COVID-19. Las líneas corresponden a la precisión obtenida para los modelos.

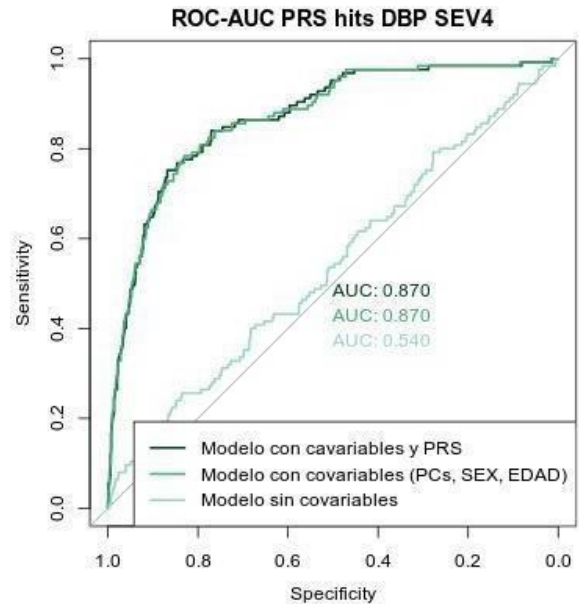
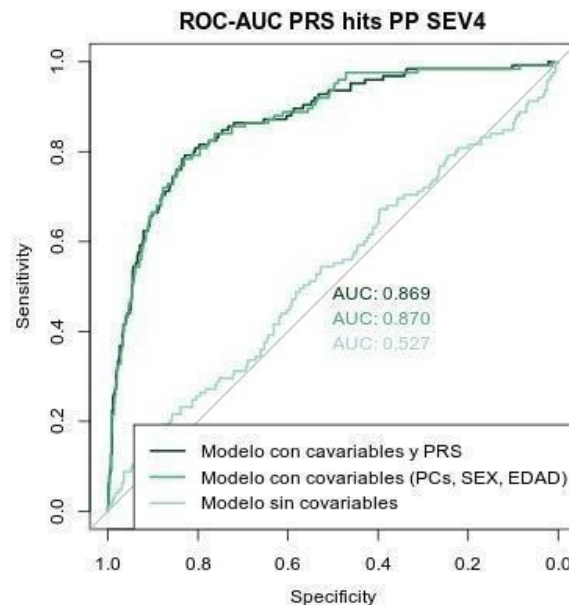


Figura 35 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión basadas en hits: Presión de Pulso (hits PP), para la gravedad (SEV4) por COVID-19. Las líneas corresponden a la precisión obtenida para los modelos.



Los resultados para el estudio de asociación con el riesgo de gravedad crítica nuevamente muestran ausencia de relación de cada uno de los PRS probados. Solo hay sutiles diferencias en la asociación mostrada por las covariables: se observa una relación significativa ($p < 0.05$) de los dos primeros PC y el sexo (no tan importante como en el análisis de hospitalización) y la covariable más asociada es la edad. Nuevamente, al igual que en el caso de la hospitalización y

coherente con la ausencia de significación, apenas hay incremento en AUC al añadir cada PRS al modelo de covariables.

4.4.3 PRS hipertensión - Asociación con antecedentes vasculares e hipertensión arterial

Finalmente, para evaluar si la falta de relación entre las PRSs generadas y analizadas, y la gravedad o el riesgo de hospitalización por COVID puede estar relacionado con la falta de validez de dichas PRSs en la predicción del riesgo cardiovascular en población colombiana, se evaluó de forma complementaria su asociación con el riesgo de antecedentes vasculares (AV) en los datos de SCOURGE-Colombia. Para ello se aplicó un modelo de regresión logística empleando AV como variable dependiente e incluyendo en el modelo la PRS a analizar y las covariables: sexo, edad y los cuatro primeros PCs.

Tabla 13 Asociación de antecedentes vasculares (AV) y PRS de hipertensión. Resultados del análisis de regresión logística del modelo completo (covariables y PRS) sobre el riesgo de antecedentes vasculares

Modelo	Variables	Estimado	Error estándar	Valor Z	Pr(> z)
<u>PRS_{SBP}</u>					
	(Intercept)	-4.19522	0.41381	-10.138	< 2e-16 ***
	SCORE_st	0.15551	0.08389	1.854	0.06377
	coPC1	-8.13644	2.95303	-2.755	0.00586 **
	coPC2	-5.83127	2.61704	-2.228	0.02587 *
	coPC3	2.73124	2.56039	1.067	0.28609
	coPC4	-2.62653	2.70825	-0.970	0.33213
	SEX	-0.08466	0.16972	-0.499	0.61790
	EDAD	0.07102	0.00582	12.204	< 2e-16 ***
<u>PRS_{DBP}</u>					
	(Intercept)	-4.127418	0.409143	-10.088	< 2e-16 ***
	SCORE_st	0.005600	0.083575	0.067	0.94657
	coPC1	-7.800639	2.946740	-2.647	0.00812 **
	coPC2	-5.996338	2.610143	-2.297	0.02160 *
	coPC3	2.430848	2.549520	0.953	0.34036
	coPC4	-2.780255	2.702579	-1.029	0.30360
	SEX	-0.104999	0.168830	-0.622	0.53399
	EDAD	0.070309	0.005785	12.153	< 2e-16 ***
<u>PRS_{PP}</u>					
	(Intercept)	-4.247971	0.416234	-10.206	< 2e-16 ***
	SCORE_st	0.224085	0.084501	2.652	0.00801 **
	coPC1	-8.167983	2.953896	-2.765	0.00569 **
	coPC2	-5.369459	2.624333	-2.046	0.04075 *
	coPC3	2.744660	2.564703	1.070	0.28454
	coPC4	-2.822427	2.702603	-1.044	0.29633
	SEX	-0.063835	0.170438	-0.375	0.70801
	EDAD	0.071367	0.005839	12.223	< 2e-16 ***

(*) < 0.05; (**) < 0.01 ; (***) < 0.001

Tabla 14 Asociación de antecedentes vasculares (AV) y PRS basadas en hits. Resultados del análisis de regresión logística del modelo completo (covariables y PRS) sobre el riesgo de antecedentes vasculares

Modelo	Variables	Estimado	Error estándar	Valor Z	Pr(> z)
<u>PRS_{SBP}</u>					
	(Intercept)	-4.157908	0.411480	-10.105	< 2e-16 ***
	SCORE_st	0.089478	0.083693	1.069	0.28502
	coPC1	-8.005165	2.950185	-2.713	0.00666 **
	coPC2	-5.820500	2.615351	-2.226	0.02605 *
	coPC3	2.598625	2.553984	1.017	0.30893
	coPC4	-2.586039	2.708349	-0.955	0.33966
	SEX	-0.093403	0.169390	-0.551	0.58135
	EDAD	0.070564	0.005796	12.176	< 2e-16 ***
<u>PRS_{DBP}</u>					
	(Intercept)	-4.122351	0.408658	-10.088	< 2e-16 ***
	SCORE_st	-0.044911	0.083314	-0.539	0.58984
	coPC1	-7.700518	2.947347	-2.613	0.00898 **
	coPC2	-6.027280	2.605141	-2.314	0.02069 *
	coPC3	2.321600	2.549033	0.911	0.36241
	coPC4	-2.911117	2.702741	-1.077	0.28144
	SEX	-0.106682	0.168838	-0.632	0.52748
	EDAD	0.070240	0.005782	12.148	< 2e-16 ***
<u>PRS_{PP}</u>					
	(Intercept)	-4.199463	0.413334	-10.160	< 2e-16 ***
	SCORE_st	0.169368	0.084905	1.995	0.04607 *
	coPC1	-7.644535	2.945697	-2.595	0.00945 **
	coPC2	-5.301778	2.631506	-2.015	0.04393 *
	coPC3	2.622092	2.558521	1.025	0.30544
	coPC4	-2.873689	2.699960	-1.064	0.28717
	SEX	-0.073000	0.170090	-0.429	0.66779
	EDAD	0.070681	0.005809	12.168	< 2e-16 ***

(*) < 0.05; (**) < 0.01; (***) < 0.001

Las figuras 36 y 37 muestran los cambios en AUC al añadir la PRS al modelo de covariables, considerando los 2068 SNPs (Figura 36) o los *hits* (Figura 37).

Figura 36 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión: sistólica (SBP, arriba izquierda), diastólica (DBP, arriba derecha) y presión de pulso (PP, abajo) para los antecedentes vasculares. Las líneas corresponden a la precisión obtenida para los modelos

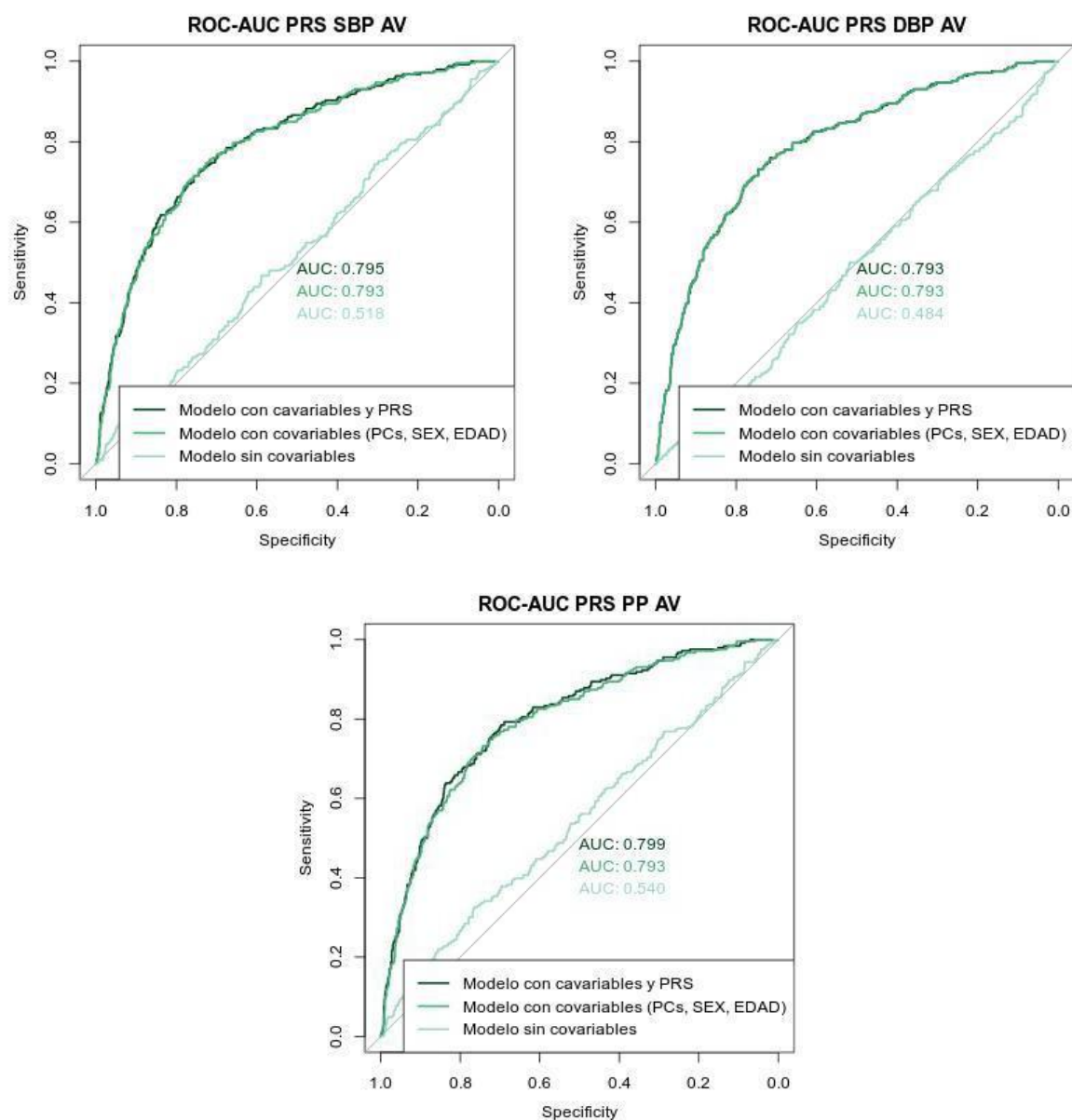
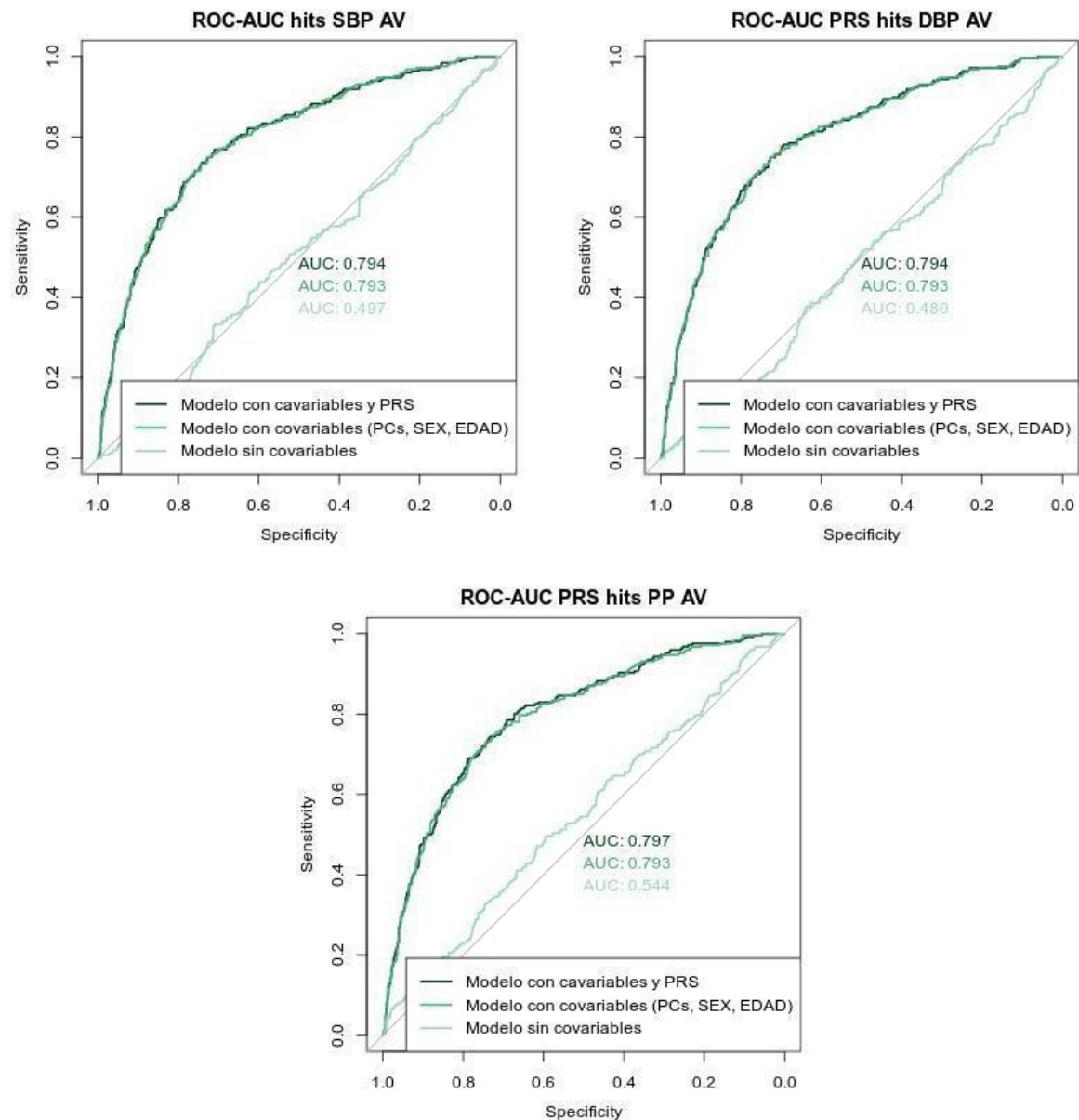


Figura 37 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión basadas en hits: sistólica (hits SBP, arriba izquierda), diastólica (hits DBP, arriba derecha) y presión de pulso (hits PP, abajo) para los antecedentes vasculares (AV). Las líneas corresponden a la precisión obtenida para los modelos



Como se puede observar en las figuras 36 y 37 el incremento en AUC del modelo con PRS fue mínimo con respecto al modelo de covariables, tanto para la PRS generada a partir de los 2068 SNPs, como a partir de los hits.

Si bien la PRS de presión de pulso sí que mostró una asociación significativa y positiva con el riesgo de AV ($P < 0.05$) (Tablas 13 y 14), cuando se analizó específicamente su capacidad predictiva del riesgo de hipertensión arterial (comorbilidad AV_1) la asociación resultó no significativa (Tablas 15 y 16). Esto apunta a una cierta falta de aplicabilidad de estas PRS de hipertensión en población colombiana.

Tabla 11. Asociación hipertensión (comorbilidad AV_1) y PRS de hipertensión. Resultados del análisis de regresión logística del modelo completo (covariables y PRS) sobre el riesgo de hipertensión.

Tabla 15 Asociación hipertensión (comorbilidad AV_1) y PRS de hipertensión. Resultados del análisis de regresión logística del modelo completo (covariables y PRS) sobre el riesgo de hipertensión

Modelo	Variables	Estimado	Error estándar	Valor Z	Pr(> z)
<u>PRS_{SBP}</u>					
	(Intercept)	-6.767941	0.598199	-11.314	<2e-16 ***
	SCORE_st	0.200117	0.108628	1.842	0.0654
	coPC1	-4.861921	3.618051	-1.344	0.1790
	coPC2	-3.614523	3.267385	-1.106	0.2686
	coPC3	-0.347828	3.232926	-0.108	0.9143
	coPC4	3.027577	3.483344	0.869	0.3848
	SEX	0.216474	0.222328	0.974	0.3302
	EDAD	0.094609	0.007723	12.251	<2e-16 ***
<u>PRS_{DBP}</u>					
	(Intercept)	-6.702900	0.593536	-11.293	<2e-16 ***
	SCORE_st	0.142949	0.108708	1.315	0.189
	coPC1	-4.694092	3.624343	-1.295	0.195
	coPC2	-4.125471	3.268061	-1.262	0.207
	coPC3	-0.483746	3.224377	-0.150	0.881
	coPC4	3.068788	3.495047	0.878	0.380
	SEX	0.190517	0.221128	0.862	0.389
	EDAD	0.094195	0.007702	12.230	<2e-16 ***
<u>PRS_{PP}</u>					
	(Intercept)	-6.74988	0.59615	-11.322	<2e-16 ***
	SCORE_st	0.17847	0.10807	1.651	0.0987
	coPC1	-4.71046	3.60981	-1.305	0.1919
	coPC2	-3.32346	3.27124	-1.016	0.3096
	coPC3	-0.54293	3.22394	-0.168	0.8663
	coPC4	2.75578	3.46290	0.796	0.4261
	SEX	0.21826	0.22236	0.982	0.3263
	EDAD	0.09420	0.00769	12.249	<2e-16 ***

(***) < 0.001

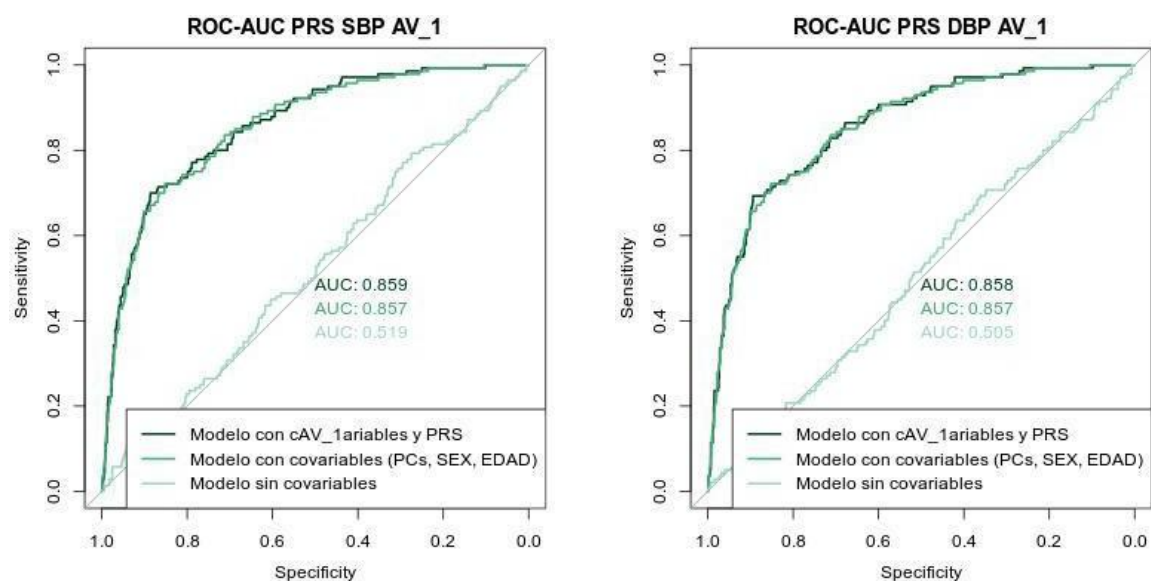
Tabla 16 Asociación hipertensión (comorbilidad AV_1) y PRS basada en hits. Resultados del análisis de regresión logística del modelo completo (covariables y PRS) sobre el riesgo de hipertensión

Modelo	Variables	Estimado	Error estándar	Valor Z	Pr(> z)
<u>PRS_{SBP}</u>					
	(Intercept)	-6.710183	0.594361	-11.290	<2e-16 ***
	SCORE_st	0.129413	0.109332	1.184	0.237
	coPC1	-4.726526	3.616146	-1.307	0.191
	coPC2	-3.599475	3.268163	-1.101	0.271
	coPC3	-0.493878	3.225908	-0.153	0.878
	coPC4	3.067966	3.484123	0.881	0.379
	SEX	0.204433	0.221880	0.921	0.357
	EDAD	0.093917	0.007678	12.233	<2e-16 ***
<u>PRS_{DBP}</u>					
	(Intercept)	-6.700825	0.593503	-11.290	<2e-16 ***
	SCORE_st	0.140803	0.108672	1.296	0.195
	coPC1	-4.745433	3.629310	-1.308	0.191
	coPC2	-3.762154	3.261955	-1.153	0.249
	coPC3	-0.459113	3.226860	-0.142	0.887
	coPC4	3.174709	3.501925	0.907	0.365
	SEX	0.192277	0.221152	0.869	0.385
	EDAD	0.094076	0.007693	12.228	<2e-16 ***
<u>PRS_{PP}</u>					
	(Intercept)	-6.678152	0.591280	-11.294	<2e-16 ***
	SCORE_st	0.079998	0.109546	0.730	0.465
	coPC1	-4.367163	3.608279	-1.210	0.226
	coPC2	-3.534508	3.289809	-1.074	0.283
	coPC3	-0.762543	3.211960	-0.237	0.812
	coPC4	2.688750	3.461840	0.777	0.437
	SEX	0.198566	0.222026	0.894	0.371
	EDAD	0.093467	0.007643	12.229	<2e-16 ***

(***) < 0.001

Y como se puede observar en las figuras 38 y 39, el incremento en AUC del modelo con PRS fue mínimo con respecto al modelo de covariables, tanto para la PRS generada a partir de los 2068 SNPs, como a partir de los *hits*.

Figura 38 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión: sistólica (SBP, arriba izquierda), diastólica (DBP, arriba derecha) y presión de pulso (PP, abajo) para la hipertensión (comorbilidad AV_1). Las líneas corresponden a la precisión obtenida para los modelos



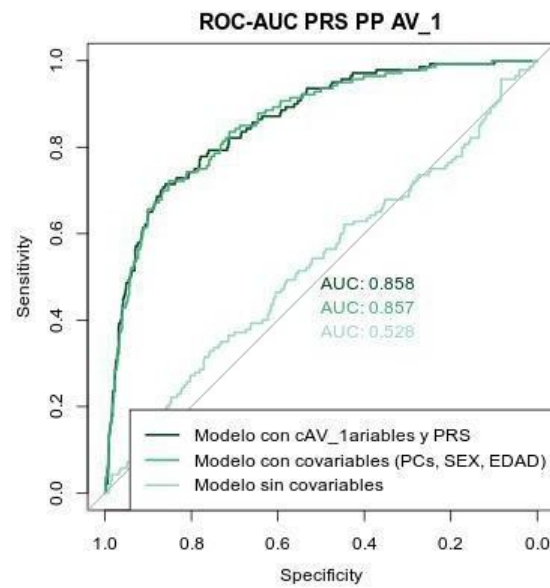
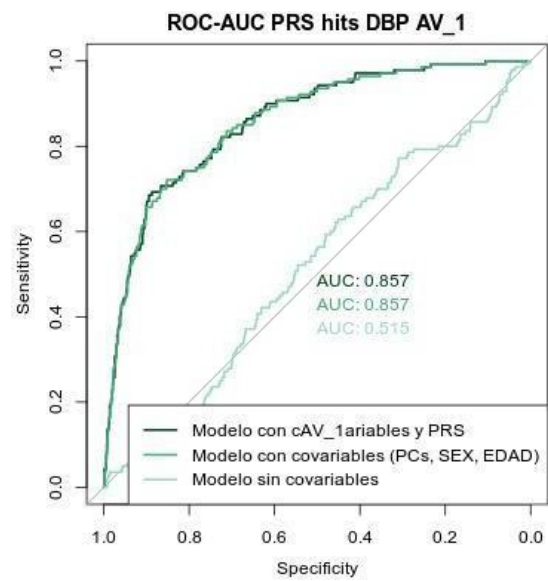
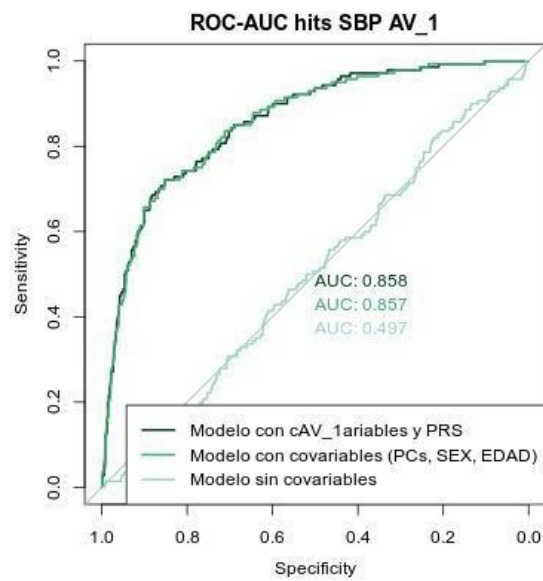


Figura 39 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión basadas en hits: sistólica (hits SBP, arriba izquierda), diastólica (hits DBP, arriba derecha) y presión de pulso (hits PP, abajo) para la hipertensión (comorbilidad AV_1). Las líneas corresponden a la precisión obtenida para los modelos.



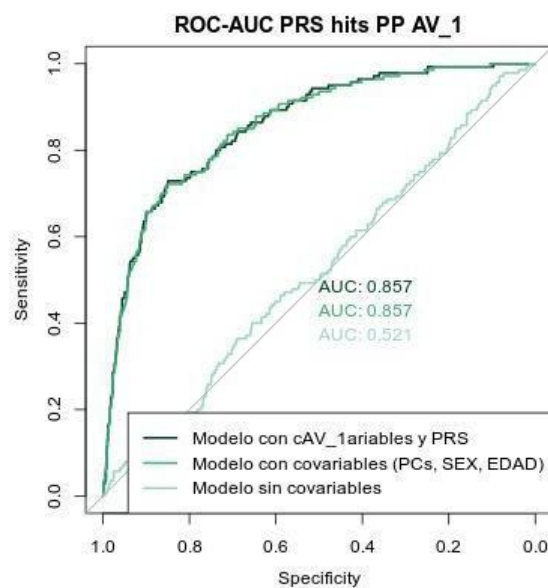


Figura 40. Área bajo la curva (AUC) de la predicción de las PRS de hipertensión basadas en hits: sistólica (hits SBP, arriba izquierda), diastólica (hits DBP, arriba derecha) y presión de pulso (hits PP, abajo) para la hipertensión (comorbilidad AV_1). Las líneas corresponden a la precisión obtenida para los modelos.

5. CONCLUSIONES

1. Se recogieron muestras de 1001 individuos de población colombiana, de ellas, 265 requirieron hospitalización durante la infección por COVID-19 y de este grupo 127 fueron clasificados con gravedad o severidad 4. Se obtuvo información de 64 variables clínicas que se recogieron en RedCap. Con este estudio se contribuyó al consorcio SCOURGE y a los estudios derivados del mismo. Se encontró una clara asociación entre los antecedentes de enfermedad vascular y la severidad para COVID.
2. El GWAS entre hospitalización y controles en población colombiana ajustado para riesgo cardiovascular no mostró ningún SNP significativo, seguramente por insuficiente tamaño muestral, pero es muy relevante que la mayoría de los SNPs asociados con p valores inferiores a 10^{-6} se encuentran en genes que juegan un papel en la severidad para COVID y que varios que encuentran ligados también a problemas cardiovasculares.
3. En el GWAS de hospitalización ajustado para riesgo cardiovascular, dos de los SNPs entre los que tenían los p valores más bajos en el GWAS, situados en los genes *ARL4A* ($p=1 \times 10^{-6}$) en cromosoma 7, y *ALOX5AD* ($p=3 \times 10^{-6}$) en cromosoma 13, fueron confirmados con MAGMA a nivel de gen sin llegar al umbral de significación. Ambos se han reportado asociados a severidad para COVID en algún estudio GWAS aunque aparecen en los resultados de los metaanálisis de los grandes consorcios HGI y GenOMICC. *ALOX5AD* se encuentra también asociada a hipertensión y *ARL4A* se ha reportado además como una diana de interés para fármacos de reposicionamiento para COVID.
4. Para el GWAS de riesgo de gravedad crítica ajustando para antecedentes cardiovasculares se observan dos SNPs, rs143861937; rs73078305, ubicados en el cromosoma 3 que pasan el umbral de referencia ($p=6.4 \times 10^{-9}$), cercanos a los genes *PLXNB1* y al gen *PFKFB4* que no se reportan en los estudios de SCOURGE, HGI ni en el metaanálisis de GenOMICC. *PLXNB1* codifica para una proteína neurológica desregulada en el plasma de pacientes con COVID severa que se considera una diana prometedora para intervenciones terapéuticas, al igual que

PKFB4 que actúa como un factor impulsor en la mediación de la liberación de citoquinas y se ha relacionado con infecciones virales y la COVID.

5. En el GWAS de severidad 4, la mayoría de los otros SNPs asociados con p valores inferiores a 10^{-6} se han relacionado con COVID y algunos muestran señales de asociación muy consistentes en el *Manhattan Plot*. Lo mismo sucede con los genes que se encuentran con mejores valores de asociación en MAGMA que se han relacionado con mecanismos de patogenicidad del SARS-CoV2 y severidad para COVID.
6. Las estimas de riesgo poligénico de hipertensión (presión arterial sistólica -SBP-, presión arterial diastólica -DBP- y presión de pulso -PP-) para predecir la hospitalización por COVID- 19 (HOSP) y la gravedad de la COVID-19 (SEV4), arrojaron resultados no significativos. Esto indica que, aunque la hipertensión contribuye a la severidad, seguramente asociada con otros factores, no comparte en nuestra serie base genética con el riesgo de hospitalización y la severidad para la COVID-19. Como quiera que se han encontrado asociadas en otras cohortes, una explicación puede estar en que las PRSs generadas a partir de SNPs identificados en estudios europeos pueden no ser adecuados para la generación de PRSs en poblaciones como la colombiana. La falta de asociación entre estos PRS y el riesgo de hipertensión arterial en población colombiana también apunta en esta dirección.

6. BIBLIOGRAFÍA

- Adua, E. (2022). Decoding the mechanism of hypertension through multiomics profiling. *Journal of Human Hypertension*, 37(4), 253-264. <https://doi.org/10.1038/s41371-022-00769-8>
- Amara, A., Trabelsi, S., Hai, A., Zaidi, S. H. H., Siddiqui, F., & Alsaeed, S. (2024). Equivocating and Deliberating on the Probability of COVID-19 Infection Serving as a Risk Factor for Lung Cancer and Common Molecular Pathways Serving as a Link. *Pathogens*, 13(12), 1070. <https://doi.org/10.3390/pathogens13121070>
- Bi, X., Liu, W., Ding, X., Liang, S., Zheng, Y., Zhu, X., Quan, S., Yi, X., Xiang, N., Du, J., Lyu, H., Yu, D., Zhang, C., Xu, L., Ge, W., Zhan, X., He, J., Xiong, Z., Zhang, S., ... Guo, T. (2022). Proteomic and metabolomic profiling of urine uncovers immune responses in patients with COVID-19. *Cell Reports*, 38(3), 110271. <https://doi.org/10.1016/j.celrep.2021.110271>
- Bonferroni, C. E. (1935). *Il calcolo delle assicurazioni su gruppi di teste*.
- Bress, A. P., Irvin, M. R., & Muntner, P. (2017). Genetics of Blood Pressure: New Insights Into a Complex Trait. *American Journal of Kidney Diseases*, 69(6), 723-725. <https://doi.org/10.1053/j.ajkd.2017.02.365>
- Cabrera-Alarcon, J. L., Cruz, R., Rosa-Moreno, M., Latorre-Pellicer, A., De Almeida, S. D., Scourge Cohort Group, Abellán, J., Acosta-Isaac, R., Aguado, J. M., Aguilar, C., Aguilera-Albesa, S., Sabbagh, A. A., Alba, J., Albu, S., Alcalá-Gallardo, K. A. M., Alcoba-Florez, J., Batres, S. A., Algarin-Lara, H. R., Almadana, V., ... Enriquez, J. A. (2025). Shaping current European mitochondrial haplogroup frequency in response to infection: The case of SARS-CoV-2 severity. *Communications Biology*, 8(1), 33. <https://doi.org/10.1038/s42003-024-07314-y>
- Cáceres Ramírez, C., Lora Mantilla, A. J., Parra Gómez, L. A., Ortegón Vargas, V., Posso Paz, M., Flórez Esparza, V., Gómez Lahitton, E., Villabona Flórez, S. J., Rocha Lezama, M. C., & Camacho López, P. A. (2023). General Hospitalization and Intensive Care Unit-Related Factors of COVID-19 Patients in Northeastern Colombia: Baseline Characteristics of a Cohort Study. *Cureus*. <https://doi.org/10.7759/cureus.43888>
- Camps-Vilaró, A., Pinsach-Abuin, M., Degano, I. R., Ramos, R., Martí-Lluch, R., Elosua, R., Subirana, I., Solà-Richarte, C., Puigmulé, M., Pérez, A., Vilaró, I., Cruz, R., Diz-de Almeida, S., Nogues, X., Masclans, J. R., Güerri-Fernández, R., Marin, J., Tizon-Marcos, H., Vaquerizo, B., ... Marrugat, J. (2024). Genetic characteristics involved in COVID-19 severity. The CARGENCORS case-control study and meta-analysis. *Journal of Medical Virology*, 96(2), e29404. <https://doi.org/10.1002/jmv.29404>
- Cao, C., Li, L., Zhang, Q., Li, H., Wang, Z., Wang, A., & Liu, J. (2023). Nkx2.5: A crucial regulator of cardiac development, regeneration and diseases. *Frontiers in Cardiovascular Medicine*, 10, 1270951. <https://doi.org/10.3389/fcvm.2023.1270951>
- Carreras-Gallo, N., Cáceres, A., Balagué-Dobón, L., Ruiz-Arenas, C., Andrusaityte, S., Carracedo, Á., Casas, M., Chatzi, L., Grazuleviciene, R., Gutzkow, K. B., Lepeule, J., Maitre, L., Nieuwenhuijsen, M., Slama, R., Stratakis, N., Thomsen, C., Urquiza, J., Wright, J., Yang, T., ... González, J. R. (2022). The early-life exposome modulates the

- effect of polymorphic inversions on DNA methylation. *Communications Biology*, 5(1), 455. <https://doi.org/10.1038/s42003-022-03380-2>
- Choi, J., Jia, G., Wen, W., Long, J., & Zheng, W. (2020). Evaluating polygenic risk scores in assessing risk of nine solid and hematologic cancers in European descendants. *International Journal of Cancer*, 147(12), 3416-3423. <https://doi.org/10.1002/ijc.33176>
- Choi, S. W., Mak, T. S.-H., & O'Reilly, P. F. (2020). Tutorial: A guide to performing polygenic risk score analyses. *Nature Protocols*, 15(9), 2759-2772. <https://doi.org/10.1038/s41596-020-0353-1>
- Cinar, R., Iyer, M. R., & Kunos, G. (2022). Dual inhibition of CB₁ receptors and iNOS, as a potential novel approach to the pharmacological management of acute and long COVID-19. *British Journal of Pharmacology*, 179(10), 2121-2127. <https://doi.org/10.1111/bph.15461>
- Cohen, J., Pertsemlidis, A., Kotowski, I. K., Graham, R., Garcia, C. K., & Hobbs, H. H. (2005). Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nature Genetics*, 37(2), 161-165. <https://doi.org/10.1038/ng1509>
- Cruz, R., Diz-de Almeida, S., López De Heredia, M., Quintela, I., Ceballos, F. C., Pita, G., Lorenzo-Salazar, J. M., González-Montelongo, R., Gago-Domínguez, M., Sevilla Porras, M., Tenorio Castaño, J. A., Nevado, J., Aguado, J. M., Aguilar, C., Aguilera-Albesa, S., Almadana, V., Almoguera, B., Alvarez, N., Andreu-Bernabeu, Á., ... Carracedo, A. (2022). Novel genes and sex differences in COVID-19 severity. *Human Molecular Genetics*, 31(22), 3789-3806. <https://doi.org/10.1093/hmg/ddac132>
- De Leeuw, C. A., Mooij, J. M., Heskes, T., & Posthuma, D. (2015). MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Computational Biology*, 11(4), e1004219. <https://doi.org/10.1371/journal.pcbi.1004219>
- Diz de Alemida, S. (2024). *Análisis Genético Poblacional de la Variabilidad en la gravedad de la COVID-19* [Tesis Doctoral]. Universidad de Santiago de Compostela.
- Diz-de Almeida, S., Cruz, R., Luchessi, A. D., Lorenzo-Salazar, J. M., De Heredia, M. L., Quintela, I., González-Montelongo, R., Nogueira Silbiger, V., Porras, M. S., Tenorio Castaño, J. A., Nevado, J., Aguado, J. M., Aguilar, C., Aguilera-Albesa, S., Almadana, V., Almoguera, B., Alvarez, N., Andreu-Bernabeu, Á., Arana-Arri, E., ... Carracedo, Á. (2024). Novel risk loci for COVID-19 hospitalization among admixed American populations. *eLife*, 13, RP93666. <https://doi.org/10.7554/eLife.93666.3>
- El-Agnaf, O., Bensmail, I., Al-Nesf, M. A. Y., Flynn, J., Taylor, M., Majbour, N. K., Abdi, I. Y., Vaikath, N. N., Farooq, A., Vemulapalli, P. B., Schmidt, F., Ouararhni, K., Al-Siddiqi, H. H., Arredouani, A., Wijten, P., Al-Maadheed, M., Mohamed-Ali, V., Decock, J., & Abdesslem, H. B. (2023). Uncovering a neurological protein signature for severe COVID-19. *Neurobiology of Disease*, 182, 106147. <https://doi.org/10.1016/j.nbd.2023.106147>
- Euesden, J., Lewis, C. M., & O'Reilly, P. F. (2015). PRSice: Polygenic Risk Score software. *Bioinformatics*, 31(9), 1466-1468. <https://doi.org/10.1093/bioinformatics/btu848>
- Fadista, J., Manning, A. K., Florez, J. C., & Groop, L. (2016a). The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *European Journal of Human Genetics*, 24(8), 1202-1205. <https://doi.org/10.1038/ejhg.2015.269>
- Fadista, J., Manning, A. K., Florez, J. C., & Groop, L. (2016b). The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *European Journal of Human Genetics*, 24(8), 1202-1205. <https://doi.org/10.1038/ejhg.2015.269>

- Fang, L., Karakiulakis, G., & Roth, M. (2020). Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *The Lancet Respiratory Medicine*, 8(4), e21. [https://doi.org/10.1016/S2213-2600\(20\)30116-8](https://doi.org/10.1016/S2213-2600(20)30116-8)
- Fisher, R. A. (1919). The causes of human variability. *The Eugenics Review*, 10(4), 213-220.
- Gálvez, J. M., Chaparro-Solano, H. M., Pinzón-Rondón, Á. M., Alborno, L. L., Pardo-Oviedo, J. M., Zapata-Gómez, F. A., Patiño-Aldana, A. F., Hernández-Rodríguez, A. D. P., Díaz-Quiroz, M., & Ruiz-Sternberg, Á. M. (2022). Mutation profile of SARS-CoV-2 genome in a sample from the first year of the pandemic in Colombia. *Infection, Genetics and Evolution*, 97, 105192. <https://doi.org/10.1016/j.meegid.2021.105192>
- García García, D. L. (2020). COVID19 y sistema renina angiotensina aldosterona (SRAA). *Revista Colombiana de Nefrología*, 7(1), 45-46. <https://doi.org/10.22265/acnef.7.Supl.2.424>
- García-Peña, Á. A., Ospina, D., Rico, J., Fernández-Ávila, D. G., Muñoz-Velandia, Ó., & Suárez-Obando, F. (2022). Prevalencia de hipertensión arterial en Colombia según información del Sistema Integral de Información de la Protección Social (SISPRO). *Revista Colombiana de Cardiología*, 29(1), 8216. <https://doi.org/10.24875/RCCAR.M22000114>
- Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A., & Smoller, J. W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*, 10(1), 1776. <https://doi.org/10.1038/s41467-019-09718-5>
- Gelernter, J., Sherva, R., Koesterer, R., Almasy, L., Zhao, H., Kranzler, H. R., & Farrer, L. (2014). Genome-wide association study of cocaine dependence and related traits: FAM53B identified as a risk gene. *Molecular Psychiatry*, 19(6), 717-723. <https://doi.org/10.1038/mp.2013.99>
- Gouveia, M. H., Bentley, A. R., Leonard, H., Meeks, K. A. C., Ekoru, K., Chen, G., Nalls, M. A., Simonsick, E. M., Tarazona-Santos, E., Lima-Costa, M. F., Adeyemo, A., Shriner, D., & Rotimi, C. N. (2021). Trans-ethnic meta-analysis identifies new loci associated with longitudinal blood pressure traits. *Scientific Reports*, 11(1), 4075. <https://doi.org/10.1038/s41598-021-83450-3>
- Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., Liu, L., Shan, H., Lei, C., Hui, D. S. C., Du, B., Li, L., Zeng, G., Yuen, K.-Y., Chen, R., Tang, C., Wang, T., Chen, P., Xiang, J., ... Zhong, N. (2020). Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine*, 382(18), 1708-1720. <https://doi.org/10.1056/NEJMoa2002032>
- Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., & Duda, S. N. (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics*, 95, 103208. <https://doi.org/10.1016/j.jbi.2019.103208>
- Henríquez, A., Accini, J., Baquero, H., Molina, F., Rey, A., Ángel, V. E., Barraza, E., Moreno-Bedoya, S., Maloof-Cuse, D. N., Martínez-Díaz, J. I., Cavadia, G., & Viasus, D. (2022). Clinical features and prognostic factors of adults with COVID-19 admitted to intensive care units in Colombia: A multicentre retrospective study during the first wave of the pandemic. *Acta Colombiana de Cuidado Intensivo*, 22(2), 95-99. <https://doi.org/10.1016/j.acci.2021.02.001>
- Hingorani, A. D., Gratton, J., Finan, C., Schmidt, A. F., Patel, R., Sofat, R., Kuan, V., Langenberg, C., Hemingway, H., Morris, J. K., & Wald, N. J. (2023). Performance of polygenic risk scores in screening, prediction, and risk stratification: Secondary

- analysis of data in the Polygenic Score Catalog. *BMJ Medicine*, 2(1), e000554. <https://doi.org/10.1136/bmjmed-2023-000554>
- Hou, C., Jiang, X., Zhang, H., Zheng, J., Qiu, Q., Zhang, Y., Sun, X., Xu, M., Chang, A. C. Y., Xie, L., & Xiao, T. (2022). TECRL deficiency results in aberrant mitochondrial function in cardiomyocytes. *Communications Biology*, 5(1), 470. <https://doi.org/10.1038/s42003-022-03414-9>
- Howie, B. N., Donnelly, P., & Marchini, J. (2009). A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genetics*, 5(6), e1000529. <https://doi.org/10.1371/journal.pgen.1000529>
- Hu, B., Guo, H., Zhou, P., & Shi, Z.-L. (2021). Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology*, 19(3), 141-154. <https://doi.org/10.1038/s41579-020-00459-7>
- Hu, Q., Chen, Z., Yuan, X., Li, S., Zhang, R., & Qin, X. (2021). Common Polymorphisms in the RGMa Promoter Are Associated With Cerebrovascular Atherosclerosis Burden in Chinese Han Patients With Acute Ischemic Cerebrovascular Accident. *Frontiers in Cardiovascular Medicine*, 8, 743868. <https://doi.org/10.3389/fcvm.2021.743868>
- Huang, D., Feng, X., Yang, H., Wang, J., Zhang, W., Fan, X., Dong, X., Chen, K., Yu, Y., Ma, X., Yi, X., & Li, M. J. (2023). QTLbase2: An enhanced catalog of human quantitative trait loci on extensive molecular phenotypes. *Nucleic Acids Research*, 51(D1), D1122-D1128. <https://doi.org/10.1093/nar/gkac1020>
- Inouye, M., Abraham, G., Nelson, C. P., Wood, A. M., Sweeting, M. J., Dudbridge, F., Lai, F. Y., Kaptoge, S., Brozynska, M., Wang, T., Ye, S., Webb, T. R., Rutter, M. K., Tzoulaki, I., Patel, R. S., Loos, R. J. F., Keavney, B., Hemingway, H., Thompson, J., ... UK Biobank CardioMetabolic Consortium CHD Working Group. (2018). Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. *Journal of the American College of Cardiology*, 72(16), 1883-1893. <https://doi.org/10.1016/j.jacc.2018.07.079>
- Jeon, B.-N., Yoon, J.-H., Kim, M.-K., Choi, W.-I., Koh, D.-I., Hur, B., Kim, K., Kim, K.-S., & Hur, M.-W. (2016). Zbtb7c is a molecular ‘off’ and ‘on’ switch of Mmp gene transcription. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1859(11), 1429-1439. <https://doi.org/10.1016/j.bbagr.2016.09.004>
- Jimenez-Silva, C., Rivero, R., Douglas, J., Bouckaert, R., Villabona-Arenas, Ch. J., Atkins, K. E., Gastelbondo, B., Calderon, A., Guzman, C., Echeverri-De La Hoz, D., Muñoz, M., Ballesteros, N., Castañeda, S., Patiño, L. H., Ramirez, A., Luna, N., Paniz-Mondolfi, A., Serrano-Coll, H., Ramirez, J. D., ... Drummond, A. J. (2023). Genomic epidemiology of SARS-CoV-2 variants during the first two years of the pandemic in Colombia. *Communications Medicine*, 3(1), 97. <https://doi.org/10.1038/s43856-023-00328-3>
- Keaton, J. M., Kamali, Z., Xie, T., Vaez, A., Williams, A., Goleva, S. B., Ani, A., Evangelou, E., Hellwege, J. N., Yengo, L., Young, W. J., Traylor, M., Giri, A., Zheng, Z., Zeng, J., Chasman, D. I., Morris, A. P., Caulfield, M. J., Hwang, S.-J., ... Warren, H. R. (2024). Genome-wide analysis in over 1 million individuals of European ancestry yields improved polygenic risk scores for blood pressure traits. *Nature Genetics*, 56(5), 778-791. <https://doi.org/10.1038/s41588-024-01714-w>
- Khandelwal, P., & Deinum, J. (2022). Monogenic forms of low-renin hypertension: Clinical and molecular insights. *Pediatric Nephrology*, 37(7), 1495-1509. <https://doi.org/10.1007/s00467-021-05246-x>

- Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., Natarajan, P., Lander, E. S., Lubitz, S. A., Ellinor, P. T., & Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*, 50(9), 1219-1224. <https://doi.org/10.1038/s41588-018-0183-z>
- Khoury, M. J., Bowen, S., Dotson, W. D., Drzymalla, E., Green, R. F., Goldstein, R., Kolor, K., Liburd, L. C., Sperling, L. S., & Bunnell, R. (2022). Health equity in the implementation of genomics and precision medicine: A public health imperative. *Genetics in Medicine*, 24(8), 1630-1639. <https://doi.org/10.1016/j.gim.2022.04.009>
- Knowles, J. W., & Ashley, E. A. (2018). Cardiovascular disease: The rise of the genetic risk score. *PLOS Medicine*, 15(3), e1002546. <https://doi.org/10.1371/journal.pmed.1002546>
- Kohara, K., Tabara, Y., Nakura, J., Imai, Y., Ohkubo, T., Hata, A., Soma, M., Nakayama, T., Umemura, S., Hirawa, N., Ueshima, H., Kita, Y., Ogihara, T., Katsuya, T., Takahashi, N., Tokunaga, K., & Miki, T. (2008). Identification of Hypertension-Susceptibility Genes and Pathways by a Systemic Multiple Candidate Gene Approach: The Millennium Genome Project for Hypertension. *Hypertension Research*, 31(2), 203-212. <https://doi.org/10.1291/hypres.31.203>
- Laajaj, R., De Los Rios, C., Sarmiento-Barbieri, I., Aristizabal, D., Behrentz, E., Bernal, R., Buitrago, G., Cucunubá, Z., De La Hoz, F., Gaviria, A., Hernández, L. J., León, L., Moyano, D., Osorio, E., Varela, A. R., Restrepo, S., Rodriguez, R., Schady, N., Vives, M., & Webb, D. (2021). COVID-19 spread, detection, and dynamics in Bogota, Colombia. *Nature Communications*, 12(1), 4726. <https://doi.org/10.1038/s41467-021-25038-z>
- Laiton-Donato, K., Villabona-Arenas, C. J., Usme-Ciro, J. A., Franco-Muñoz, C., Álvarez-Díaz, D. A., Villabona-Arenas, L. S., Echeverría-Londoño, S., Cucunubá, Z. M., Franco-Sierra, N. D., Flórez, A. C., Ferro, C., Ajami, N. J., Walteros, D. M., Prieto, F., Durán, C. A., Ospina-Martínez, M. L., & Mercado-Reyes, M. (2020). Genomic Epidemiology of Severe Acute Respiratory Syndrome Coronavirus 2, Colombia. *Emerging Infectious Diseases*, 26(12), 2854-2862. <https://doi.org/10.3201/eid2612.202969>
- Lakeman, I. M. M., Rodríguez-Girondo, M. D. M., Lee, A., Celosse, N., Braspenning, M. E., Van Engelen, K., Van De Beek, I., Van Der Hout, A. H., Gómez García, E. B., Mensenkamp, A. R., Ausems, M. G. E. M., Hooning, M. J., Adank, M. A., Hollestelle, A., Schmidt, M. K., Van Asperen, C. J., & Devilee, P. (2023). Clinical applicability of the Polygenic Risk Score for breast cancer risk prediction in familial cases. *Journal of Medical Genetics*, 60(4), 327-336. <https://doi.org/10.1136/jmg-2022-108502>
- Lambert, S. A., Gil, L., Jupp, S., Ritchie, S. C., Xu, Y., Buniello, A., McMahon, A., Abraham, G., Chapman, M., Parkinson, H., Danesh, J., MacArthur, J. A. L., & Inouye, M. (2021). The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nature Genetics*, 53(4), 420-425. <https://doi.org/10.1038/s41588-021-00783-5>
- Lamelas, P., Diaz, R., Orlandini, A., Avezum, A., Oliveira, G., Mattos, A., Lanás, F., Seron, P., Oliveros, M. J., Lopez-Jaramillo, P., Otero, J., Camacho, P., Miranda, J., Bernabe-Ortiz, A., Malaga, G., Irazola, V., Gutierrez, L., Rubinstein, A., Castellana, N., ... Yusuf, S. (2019). Prevalence, awareness, treatment and control of hypertension in rural and urban communities in Latin American countries. *Journal of Hypertension*, 37(9), 1813-1821. <https://doi.org/10.1097/HJH.0000000000002108>

- Lin, S., Gao, X., Degenhardt, F., Qian, Y., Liu, T., Ramon, X. F., Hadi, S. S., Romero-Gómez, M., Fernández, J., Albillos, A., Ferret, M. B., Bujanda, L., Julià, A., De Cid, R., Asselta, R., Franke, A., & Liu, F. (2023). Genome-wide epistasis study highlights genetic interactions influencing severity of COVID-19. *European Journal of Epidemiology*, 38(8), 883-889. <https://doi.org/10.1007/s10654-023-01020-5>
- Liu, J., Chen, L., Guo, X., Zhao, B., & Jiang, J. (2024). Emerging role of N6-methyladenosine RNA modification in regulation of SARS-CoV-2 infection and virus-host interactions. *Biomedicine & Pharmacotherapy*, 173, 116231. <https://doi.org/10.1016/j.biopha.2024.116231>
- Liu, N., Jiang, C., Cai, P., Shen, Z., Sun, W., Xu, H., Fang, M., Yao, X., Zhu, L., Gao, X., Fang, J., Lin, J., Guo, C., & Qu, K. (2021). Single-cell analysis of COVID-19, sepsis, and HIV infection reveals hyperinflammatory and immunosuppressive signatures in monocytes. *Cell Reports*, 37(1), 109793. <https://doi.org/10.1016/j.celrep.2021.109793>
- Mao, Y.-Q., Jahanshahi, S., Maly, R., Van Ommen, D. A. J., Wan, Y., Morey, T. M., Chuang, S. H. W., Pavlova, V., Ahmed, C., Dahal, S., Lin, F., Mangos, M., Nurtanto, J., Song, Y., Been, T., Christie-Holmes, N., Gray-Owen, S. D., Babu, M., Wong, A. P., ... Houry, W. A. (2024). Targeting protein homeostasis with small molecules as a strategy for the development of pan-coronavirus antiviral therapies. *Communications Biology*, 7(1), 1460. <https://doi.org/10.1038/s42003-024-07143-z>
- Marees, A. T., De Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., & Derks, E. M. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *International Journal of Methods in Psychiatric Research*, 27(2), e1608. <https://doi.org/10.1002/mpr.1608>
- Marín-Sánchez, A. (2020a). Características clínicas básicas en los primeros 100 casos fatales de COVID-19 en Colombia. *Revista Panamericana de Salud Pública*, 44, 1. <https://doi.org/10.26633/RPSP.2020.87>
- Marín-Sánchez, A. (2020b). Características clínicas básicas en los primeros 100 casos fatales de COVID-19 en Colombia. *Revista Panamericana de Salud Pública*, 44, 1. <https://doi.org/10.26633/RPSP.2020.87>
- Merino, G. A. (2018). *Imputación de genotipos faltantes en datos de secuenciación masiva* [Tesis de Maestría]. Universidad Nacional de Córdoba.
- Moiseenko, V. I., Aparyatina, V. A., Gainetdinov, R. R., & Aparyatin, S. A. (2024). Trace Amine-Associated Receptors' Role in Immune System Functions. *Biomedicine*, 12(4), 893. <https://doi.org/10.3390/biomedicine12040893>
- Nelson, R. M., Pettersson, M. E., & Carlborg, Ö. (2013). A century after Fisher: Time for a new paradigm in quantitative genetics. *Trends in Genetics*, 29(12), 669-676. <https://doi.org/10.1016/j.tig.2013.09.006>
- Olsen, M. H., Angell, S. Y., Asma, S., Boutouyrie, P., Burger, D., Chirinos, J. A., Damasceno, A., Delles, C., Gimenez-Roqueplo, A.-P., Hering, D., López-Jaramillo, P., Martinez, F., Perkovic, V., Rietzschel, E. R., Schillaci, G., Schutte, A. E., Scuteri, A., Sharman, J. E., Wachtell, K., & Wang, J. G. (2016). A call to action and a lifecourse strategy to address the global burden of raised blood pressure on current and future generations: The Lancet Commission on hypertension. *The Lancet*, 388(10060), 2665-2712. [https://doi.org/10.1016/S0140-6736\(16\)31134-5](https://doi.org/10.1016/S0140-6736(16)31134-5)
- O'Sullivan, J. W., Ashley, E. A., & Elliott, P. M. (2023). Polygenic risk scores for the prediction of cardiometabolic disease. *European Heart Journal*, 44(2), 89-99. <https://doi.org/10.1093/eurheartj/ehac648>

- Pairo-Castineira, E., Rawlik, K., Bretherick, A. D., Qi, T., Wu, Y., Nassiri, I., McConkey, G. A., Zechner, M., Klaric, L., Griffiths, F., Oosthuyzen, W., Kousathanas, A., Richmond, A., Millar, J., Russell, C. D., Malinauskas, T., Thwaites, R., Morrice, K., Keating, S., ... Baillie, J. K. (2023). Author Correction: GWAS and meta-analysis identifies 49 genetic variants underlying critical COVID-19. *Nature*, 619(7971), E61-E61. <https://doi.org/10.1038/s41586-023-06383-z>
- Pramparo, P., Boissonnet, C., & Schargrodsky, H. (2011). Evaluación del riesgo cardiovascular en siete ciudades latinoamericanas: Principales conclusiones del estudio y subestudios CARMELA. *Rdo. argén. cardiol*, 79(4), 377-382.
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8), 904-909. <https://doi.org/10.1038/ng1847>
- Privé, F., Arbel, J., & Vilhjálmsson, B. J. (2021). LDpred2: Better, faster, stronger. *Bioinformatics*, 36(22-23), 5424-5431. <https://doi.org/10.1093/bioinformatics/btaa1029>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007a). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, 81(3), 559-575. <https://doi.org/10.1086/519795>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007b). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, 81(3), 559-575. <https://doi.org/10.1086/519795>
- Razai, M. S., Doerholt, K., Ladhani, S., & Oakeshott, P. (2020). Coronavirus disease 2019 (covid-19): A guide for UK GPs. *BMJ*, m800. <https://doi.org/10.1136/bmj.m800>
- Reyes, L. F., Bastidas, A., Narváez, P. O., Parra-Tanoux, D., Fuentes, Y. V., Serrano-Mayorga, C. C., Ortíz, V., Caceres, E. L., Ospina-Tascon, G., Díaz, A. M., Jibaja, M., Vera, M., Silva, E., Gorordo-Delsol, L. A., Maraschin, F., Varón-Vega, F., Buitrago, R., Poveda, M., Saucedo, L. M., ... for the LIVEN-Covid-19 Investigators. (2022). Clinical characteristics, systemic complications, and in-hospital outcomes for patients with COVID-19 in Latin America. LIVEN-Covid-19 study: A prospective, multicenter, multinational, cohort study. *PLOS ONE*, 17(3), e0265529. <https://doi.org/10.1371/journal.pone.0265529>
- Rodriguez Lima, D. R., Pinzón Rondón, Á. M., Rubio Ramos, C., Pinilla Rojas, D. I., Niño Orrego, M. J., Díaz Quiroz, M. A., Molano-González, N., Ceballos Quintero, J. E., Arroyo Santos, A. F., & Ruiz Sternberg, Á. M. (2022). Clinical characteristics and mortality associated with COVID-19 at high altitude: A cohort of 5161 patients in Bogotá, Colombia. *International Journal of Emergency Medicine*, 15(1), 22. <https://doi.org/10.1186/s12245-022-00426-4>
- Rosselli, D. (2021). Covid-19 en Colombia: Los primeros 90 días. *Acta Neurológica Colombiana*, 36(2 Supl 1), 1-6. <https://doi.org/10.22379/24224022287>
- Ruan, Y., Lin, Y.-F., Feng, Y.-C. A., Chen, C.-Y., Lam, M., Guo, Z., Stanley Global Asia Initiatives, Ahn, Y. M., Akiyama, K., Arai, M., Baek, J. H., Chen, W. J., Chung, Y.-C., Feng, G., Fujii, K., Glatt, S. J., Ha, K., Hattori, K., Higuchi, T., ... Ge, T. (2022). Improving polygenic prediction in ancestrally diverse populations. *Nature Genetics*, 54(5), 573-580. <https://doi.org/10.1038/s41588-022-01054-7>
- Ryu, J. K., Yan, Z., Montano, M., Sozmen, E. G., Dixit, K., Suryawanshi, R. K., Matsui, Y., Helmy, E., Kaushal, P., Makanani, S. K., Deerinck, T. J., Meyer-Franke, A., Rios

- Coronado, P. E., Trevino, T. N., Shin, M.-G., Tognatta, R., Liu, Y., Schuck, R., Le, L., ... Akassoglou, K. (2024). Fibrin drives thromboinflammation and neuropathology in COVID-19. *Nature*, 633(8031), 905-913. <https://doi.org/10.1038/s41586-024-07873-4>
- Sahanic, S., Löffler-Ragg, J., Tymoszyk, P., Hilbe, R., Demetz, E., Masanetz, R. K., Theurl, M., Holfeld, J., Gollmann-Tepeköylü, C., Tzankov, A., Weiss, G., Giera, M., & Tancevski, I. (2021). The Role of Innate Immunity and Bioactive Lipid Mediators in COVID-19 and Influenza. *Frontiers in Physiology*, 12, 688946. <https://doi.org/10.3389/fphys.2021.688946>
- Schild, L. (1996). The ENaC channel as the primary determinant of two human diseases: Liddle syndrome and pseudohypoaldosteronism. *Nephrologie*, 17(7), 395-400.
- Schultheiß, C., Paschold, L., Willscher, E., Simnica, D., Wöstemeier, A., Muscate, F., Wass, M., Eisenmann, S., Dutzmann, J., Keyßer, G., Gagliani, N., & Binder, M. (2021). Maturation trajectories and transcriptional landscape of plasmablasts and autoreactive B cells in COVID-19. *iScience*, 24(11), 103325. <https://doi.org/10.1016/j.isci.2021.103325>
- Selvaraj, G., Kaliyandurthi, S., Peslherbe, G. H., & Wei, D.-Q. (2021). Identifying potential drug targets and candidate drugs for COVID-19: Biological networks and structural modeling approaches. *F1000Research*, 10, 127. <https://doi.org/10.12688/f1000research.50850.3>
- Shao, L., Pelayo, A., Shi, R., Ma, J., Liu, H., Cai, Y., Prochazkova, M., Somerville, R. P., Panch, S. R., Shah, N. N., Stroncek, D. F., & Jin, P. (2022). Identification of genomic determinants contributing to cytokine release in immunotherapies and human diseases. *Journal of Translational Medicine*, 20(1), 338. <https://doi.org/10.1186/s12967-022-03531-3>
- Shi, S., Qin, M., Shen, B., Cai, Y., Liu, T., Yang, F., Gong, W., Liu, X., Liang, J., Zhao, Q., Huang, H., Yang, B., & Huang, C. (2020). Association of Cardiac Injury With Mortality in Hospitalized Patients With COVID-19 in Wuhan, China. *JAMA Cardiology*, 5(7), 802. <https://doi.org/10.1001/jamacardio.2020.0950>
- Silva, P. P., Gaudillo, J. D., Vilela, J. A., Roxas-Villanueva, R. M. L., Tiangco, B. J., Domingo, M. R., & Albia, J. R. (2022). A machine learning-based SNP-set analysis approach for identifying disease-associated susceptibility loci. *Scientific Reports*, 12(1), 15817. <https://doi.org/10.1038/s41598-022-19708-1>
- Stefansson, H., Ophoff, R. A., Steinberg, S., Andreassen, O. A., Cichon, S., Rujescu, D., Werge, T., Pietiläinen, O. P. H., Mors, O., Mortensen, P. B., Sigurdsson, E., Gustafsson, O., Nyegaard, M., Tuulio-Henriksson, A., Ingason, A., Hansen, T., Suvisaari, J., Lonnqvist, J., Paunio, T., ... Collier, D. A. (2009). Common variants conferring risk of schizophrenia. *Nature*, 460(7256), 744-747. <https://doi.org/10.1038/nature08186>
- Stilhano, R. S., Costa, A. J., Nishino, M. S., Shams, S., Bartolomeo, C. S., Breithaupt-Faloppa, A. C., Silva, E. A., Ramirez, A. L., Prado, C. M., & Ureshino, R. P. (2020). SARS-CoV-2 and the possible connection to ERs, ACE2, and RAGE: Focus on susceptibility factors. *The FASEB Journal*, 34(11), 14103-14119. <https://doi.org/10.1096/fj.202001394RR>
- Sulovari, A., & Li, D. (2014). GACT: A Genome build and Allele definition Conversion Tool for SNP imputation and meta-analysis in genetic association studies. *BMC Genomics*, 15(1), 610. <https://doi.org/10.1186/1471-2164-15-610>

- Tadic, M., Cuspidi, C., Grassi, G., & Mancia, G. (2020). COVID-19 and arterial hypertension: Hypothesis or evidence? *The Journal of Clinical Hypertension*, 22(7), 1120-1126. <https://doi.org/10.1111/jch.13925>
- The COVID-19 Host Genetics Initiative, Leadership, Kanai, M., Andrews, S. J., Cordioli, M., Stevens, C., Neale, B. M., Daly, M., Ganna, A., Writing group, Writing group lead, Pathak, G. A., Writing group member, Iwasaki, A., Analysis group, Manuscript analysis team lead, Karjalainen, J., Mehtonen, J., Manuscript analysis team members: phenome-wide association study, ... Solomonson, M. (2023). A second update on mapping the human genetic architecture of COVID-19. *Nature*, 621(7977), E7-E26. <https://doi.org/10.1038/s41586-023-06355-3>
- the Million Veteran Program, Evangelou, E., Warren, H. R., Mosen-Ansorena, D., Mifsud, B., Pazoki, R., Gao, H., Ntritsos, G., Dimou, N., Cabrera, C. P., Karaman, I., Ng, F. L., Evangelou, M., Witkowska, K., Tzanis, E., Hellwege, J. N., Giri, A., Velez Edwards, D. R., Sun, Y. V., ... Caulfield, M. J. (2018). Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nature Genetics*, 50(10), 1412-1425. <https://doi.org/10.1038/s41588-018-0205-x>
- The Wellcome Trust Case Control Consortium, Management Committee, Burton, P. R., Clayton, D. G., Cardon, L. R., Craddock, N., Deloukas, P., Duncanson, A., Kwiatkowski, D. P., McCarthy, M. I., Ouwehand, W. H., Samani, N. J., Todd, J. A., Donnelly, P., Data and Analysis Committee, Barrett, J. C., Burton, P. R., Davison, D., Donnelly, P., ... Worthington, J. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145), 661-678. <https://doi.org/10.1038/nature05911>
- Thompson, D. J., Wells, D., Selzam, S., Peneva, I., Moore, R., Sharp, K., Tarran, W. A., Beard, E. J., Riveros-Mckay, F., Giner-Delgado, C., Palmer, D., Seth, P., Harrison, J., Futema, M., Genomics England Research Consortium, McVean, G., Plagnol, V., Donnelly, P., & Weale, M. E. (2024). A systematic evaluation of the performance and properties of the UK Biobank Polygenic Risk Score (PRS) Release. *PLOS ONE*, 19(9), e0307270. <https://doi.org/10.1371/journal.pone.0307270>
- Torkamani, A., Wineinger, N. E., & Topol, E. J. (2018). The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics*, 19(9), 581-590. <https://doi.org/10.1038/s41576-018-0018-x>
- Udomsinprasert, W., Nontawong, N., Saengsiwaritt, W., Panthan, B., Jiaranai, P., Thongchompoo, N., Santon, S., Runcharoen, C., Sensorn, I., Jittikoon, J., Chaikledkaew, U., & Chantratita, W. (2023). Host genetic polymorphisms involved in long-term symptoms of COVID-19. *Emerging Microbes & Infections*, 12(2), 2239952. <https://doi.org/10.1080/22221751.2023.2239952>
- Uffelmann, E., Huang, Q. Q., Munung, N. S., De Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, 1(1), 59. <https://doi.org/10.1038/s43586-021-00056-9>
- Ulrichs, T., Rolland, M., Wu, J., Nunes, M. C., El Guerche-Séblain, C., & Chit, A. (2024). Changing epidemiology of COVID-19: Potential future impact on vaccines and vaccination strategies. *Expert Review of Vaccines*, 23(1), 510-522. <https://doi.org/10.1080/14760584.2024.2346589>
- Vallejo Camazón, N., Teis, A., Martínez Membrive, M. J., Llibre, C., Bayés-Genís, A., & Mateu, L. (2022). COVID-19 persistente y angina microvascular. *Revista Española de Cardiología*, 75(5), 444-446. <https://doi.org/10.1016/j.recesp.2021.10.011>

- Visser, P. M., & Goddard, M. E. (2019a). From R.A. Fisher's 1918 Paper to GWAS a Century Later. *Genetics*, 211(4), 1125-1130. <https://doi.org/10.1534/genetics.118.301594>
- Visser, P. M., & Goddard, M. E. (2019b). From R.A. Fisher's 1918 Paper to GWAS a Century Later. *Genetics*, 211(4), 1125-1130. <https://doi.org/10.1534/genetics.118.301594>
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54(3), 426-482. <https://doi.org/10.1090/S0002-9947-1943-0012401-3>
- Wand, H., Lambert, S. A., Tamburro, C., Iacocca, M. A., O'Sullivan, J. W., Sillari, C., Kullo, I. J., Rowley, R., Dron, J. S., Brockman, D., Venner, E., McCarthy, M. I., Antoniou, A. C., Easton, D. F., Hegele, R. A., Khera, A. V., Chatterjee, N., Kooperberg, C., Edwards, K., ... Wojcik, G. L. (2021). Improving reporting standards for polygenic scores in risk prediction studies. *Nature*, 591(7849), 211-219. <https://doi.org/10.1038/s41586-021-03243-6>
- Watanabe, K., Taskesen, E., Van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, 8(1), 1826. <https://doi.org/10.1038/s41467-017-01261-5>
- Westerman, K. E., Lin, J., Sevilla-Gonzalez, M. D. R., Tadess, B., Marchek, C., & Manning, A. K. (2022). Gene-Environment Interaction Analysis Incorporating Sex, Cardiometabolic Diseases, and Multiple Deprivation Index Reveals Novel Genetic Associations With COVID-19 Severity. *Frontiers in Genetics*, 12, 782172. <https://doi.org/10.3389/fgene.2021.782172>
- Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, 26(17), 2190-2191. <https://doi.org/10.1093/bioinformatics/btq340>
- Williams, B., Mancina, G., Spiering, W., Agabiti Rosei, E., Azizi, M., Burnier, M., Clement, D. L., Coca, A., De Simone, G., Dominiczak, A., Kahan, T., Mahfoud, F., Redon, J., Ruilope, L., Zanchetti, A., Kerins, M., Kjeldsen, S. E., Kreutz, R., Laurent, S., ... Brady, A. (2018). 2018 ESC/ESH Guidelines for the management of arterial hypertension. *European Heart Journal*, 39(33), 3021-3104. <https://doi.org/10.1093/eurheartj/ehy339>
- Wolf, J. M., Wolf, L. M., Bello, G. L., Maccari, J. G., & Nasi, L. A. (2023). Molecular evolution of SARS-CoV-2 from December 2019 to August 2022. *Journal of Medical Virology*, 95(1), e28366. <https://doi.org/10.1002/jmv.28366>
- Wray, N. R., Lin, T., Austin, J., McGrath, J. J., Hickie, I. B., Murray, G. K., & Visscher, P. M. (2021). From Basic Science to Clinical Application of Polygenic Risk Scores: A Primer. *JAMA Psychiatry*, 78(1), 101-109. <https://doi.org/10.1001/jamapsychiatry.2020.3049>
- Yang, Q., Tang, J., Cao, J., Liu, F., Fu, M., Xue, B., Zhou, A., Chen, S., Liu, J., Zhou, Y., Shi, Y., Peng, W., & Chen, X. (2023). SARS-CoV-2 infection activates CREB/CBP in cellular cyclic AMP-dependent pathways. *Journal of Medical Virology*, 95(1), e28383. <https://doi.org/10.1002/jmv.28383>
- Yang, X., Eriksson, M., Czene, K., Lee, A., Leslie, G., Lush, M., Wang, J., Dennis, J., Dorling, L., Carvalho, S., Mavaddat, N., Simard, J., Schmidt, M. K., Easton, D. F., Hall, P., & Antoniou, A. C. (2022). Prospective validation of the BOADICEA multifactorial breast cancer risk prediction model in a large prospective cohort study.

- Journal of Medical Genetics*, 59(12), 1196-1205. <https://doi.org/10.1136/jmg-2022-108806>
- Ye, Y., Zhang, Z., Liu, Y., Diao, L., & Han, L. (2020). A Multi-Omics Perspective of Quantitative Trait Loci in Precision Medicine. *Trends in Genetics*, 36(5), 318-336. <https://doi.org/10.1016/j.tig.2020.01.009>
- Yomayusa-Gonzalez, N., Chacón-Acevedo, K. R., Ávila-Reina, A. J., Rincón Karen, L., Toloza- Parra, C. H., Combariza-Vallejo, J. F., & Álvarez-Moreno, C. A. (2021). Clinical course, biomarkers, management and outcomes of patients hospitalised due to COVID-19 in Colombia. *Infectio*, 25(4), 262. <https://doi.org/10.22354/in.v25i4.958>
- Zappa, M., Golino, M., Verdecchia, P., & Angeli, F. (2024). Genetics of Hypertension: From Monogenic Analysis to GETomics. *Journal of Cardiovascular Development and Disease*, 11(5), 154. <https://doi.org/10.3390/jcdd11050154>
- Zhang, J., Zhan, J., Jin, J., Ma, C., Zhao, R., O'Connell, J., Jiang, Y., 23andMe Research Team, Koelsch, B. L., Zhang, H., & Chatterjee, N. (2024). An ensemble penalized regression method for multi-ancestry polygenic risk prediction. *Nature Communications*, 15(1), 3238. <https://doi.org/10.1038/s41467-024-47357-7>
- Zhang, Y., Zeng, T., Chen, L., Ding, S., Huang, T., & Cai, Y.-D. (2020). Identification of COVID-19 Infection-Related Human Genes Based on a Random Walk Model in a Virus–Human Protein Interaction Network. *BioMed Research International*, 2020(1), 4256301. <https://doi.org/10.1155/2020/4256301>
- Zhang, Z. (2024). The Initial COVID-19 Reliable Interactive DNA Methylation Markers and Biological Implications. *Biology*, 13(4), 245. <https://doi.org/10.3390/biology13040245>

7. ÍNDICES COMPLEMENTARIOS

ÍNDICE DE FIGURAS

Figura 1. Representación gráfica de resultados de PRS A. Gráfico AUC área bajo la curva (sensibilidad/ especificidad)-izquierda B. Gráfico de densidades-derecha	20
Figura 2 Distribución de casos AV*_SEV4.....	39
Figura 3 Distribución de casos AV*R_HOSP	40
Figura 4 Distribución de AV (alguno/ninguno) en los diferentes grupos de gravedad.....	41
Figura 5 Distribución de SAV (suma de AVs) en los diferentes grupos de gravedad.....	42
Figura 6 Distribución de antecedentes cardiacos (AC) (alguno/ninguno) en los diferentes grupos de gravedad	42
Figura 7 Distribución de AV1 (HTA) en los diferentes grupos de gravedad.....	43
Figura 8 Distribución de AV2 (hipercolesterolemia) en los diferentes grupos de gravedad...43	
Figura 9 Distribución de AV3 (diabetes no complicada) en los diferentes grupos de gravedad	44
Figura 10 Distribución de AV4 (diabetes complicada) en los diferentes grupos de gravedad	44
Figura 11 Distribución de 10 (obesidad) en los diferentes grupos de gravedad.....	45
Figura 12 Manhattan plot de GWAS de Hospitalizados.....	45
Figura 13 Gráfico de calidad QQ de GWAS de pacientes hospitalizados con antecedentes cardiovasculares	46
Figura 14 Gráfico de calidad QQ de Genes de GWAS hospitalización.....	49
Figura 15 Manhattan plot de resultados de MAGMA (análisis a nivel de gen) para el riesgo de hospitalización	49
Figura 16 Gráficos regionales de la señal en ARL4A del cromosoma 7: A. Región de SBP asociados	50
Figura 17 Gráficos regionales de la señal en ARL4A del cromosoma 7: B. Relación con el gen más cercano.....	50
Figura 18 Gráficos regionales de a señal en ALOX5AP del cromosoma 13: A, Región de SBP asociados	50
Figura 19 Gráficos regionales de la señal en ALOX5AP del cromosoma 13: B. Relación con el gen.....	50
Figura 20 . Manhattan plot de GWAS Severidad 4	51
Figura 21 Gráfico de calidad QQ de GWAS de pacientes severidad4 con antecedentes cardiovasculares	52
Figura 22 Manhattan plot de genes para análisis de gravedad crítica ajustando para antecedentes cardiovasculares	55

Figura 23 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión: sistólica (SBP) para la hospitalización por COVID-19	58
Figura 24 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión: diastólica (DBP) para la hospitalización por COVID-19	58
Figura 25 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión: presión de pulso (PP) para la hospitalización por COVID-19	58
Figura 25 Área bajo la curva (AUC) de la predicción de los PRS de hipertensión basados en hits: sistólica (hits SBP, arriba izquierda) para la hospitalización por COVID-19. Las líneas corresponden a la precisión obtenida por los modelos	60
Figura 26 Área bajo la curva (AUC) de la predicción de los PRS de hipertensión basados en hits: sistólica (hits, SBP, arriba izquierda) para la hospitalización por COVID-19. Las líneas corresponden a la precisión obtenida por los modelos	60
Figura 27 Área bajo la curva (AUC) de la predicción de los PRS de hipertensión basados en hits: diastólica (hits DBP). Las líneas corresponden a la precisión obtenida por los modelos	60
Figura 28 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión: sistólica (hits SBP, arriba izquierda), diastólica (hits DBP, arriba derecha) y presión de pulso (hits PP, abajo) para la hospitalización por COVID-19. Las líneas corresponden a la precisión obtenida por los modelos	62
Figura 29 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión: sistólica (SBP)	62
Figura 40 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión: sistólica (SBP). Para la gravedad (SEV4) por COVID-19. Las líneas corresponden a la precisión obtenida para los modelos.....	63
Figura 31 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión: presión de pulso (PP) para la gravedad (SEV4) por COVID-19. Las líneas corresponden a la precisión obtenida para los modelos	63
Figura 32 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión: diastólica (DBP) para la gravedad (SEV4) por COVID-19. Las líneas corresponden a la precisión obtenida para los modelos	63
Figura 33 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión basadas en hits: sistólica (hits SBP), para la gravedad (SEV4) por COVID-19. Las líneas corresponden a la precisión obtenida para los modelos	64
Figura 34 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión basadas en hits: diastólica (hits DBP), para la gravedad (SEV4) por COVID-19. Las líneas corresponden a la precisión obtenida para los modelos	64
Figura 35 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión basadas en hits: Presión de Pulso (hits PP), para la gravedad (SEV4) por COVID-19. Las líneas corresponden a la precisión obtenida para los modelos	64
Figura 36 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión: sistólica (SBP, arriba izquierda), diastólica (DBP, arriba derecha) y presión de pulso (PP, abajo) para los antecedentes vasculares. Las líneas corresponden a la precisión obtenida para los modelos	68
Figura 37 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión basadas en hits: sistólica (hits SBP, arriba izquierda), diastólica (hits DBP, arriba derecha) y presión de	

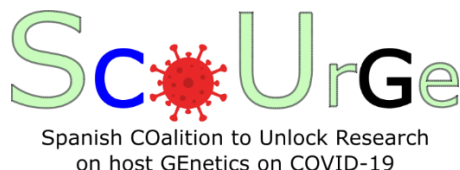
pulso (hits PP, abajo) para los antecedentes vasculares (AV). Las líneas corresponden a la precisión obtenida para los modelos	69
Figura 38 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión: sistólica (SBP, arriba izquierda), diastólica (DBP, arriba derecha) y presión de pulso (PP, abajo) para la hipertensión (comorbilidad AV_1). Las líneas corresponden a la precisión obtenida para los modelos	72
Figura 39 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión basadas en hits: sistólica (hits SBP, arriba izquierda), diastólica (hits DBP, arriba derecha) y presión de pulso (hits PP, abajo) para la hipertensión (comorbilidad AV_1). Las líneas corresponden a la precisión obtenida para los modelos	73
Figura 40 Área bajo la curva (AUC) de la predicción de las PRS de hipertensión basadas en hits: sistólica (hits SBP, arriba izquierda), diastólica (hits DBP, arriba derecha) y presión de pulso (hits PP, abajo) para la hipertensión (comorbilidad AV_1). Las líneas corresponden a la precisión obtenida para los modelos	74

ÍNDICE DE TABLAS

Tabla 1 Criterios de control de calidad según Choi y colaboradores (2020)	16
Tabla 2 . Clasificación de pacientes con COVID-19 según gravedad	32
Tabla 3 Recuentos y frecuencias de antecedentes vasculares en casos graves (SEV4=1) y casos más leves (SEV4=0)	38
Tabla 4 Pruebas de chi-cuadrado AV*R_SEV4.....	39
Tabla 5 Tabla cruzada AV*R_HOSP	40
Tabla 6 AV*R_HOSP: Pruebas de chi-cuadrado.....	40
Tabla 7 Lead SNPs del GWAS de riesgo de hospitalización.....	48
Tabla 8 Lead SNPs del GWAS gravedad 4 ajustando para antecedentes cardiovasculares	52
Tabla 9 Resultados del análisis de regresión logística del modelo completo (covariables y PRS) sobre el riesgo de hospitalización por COVID-19	56
Tabla 10 Asociación hospitalización y PRS basadas en hits. Resultados del análisis de regresión logística del modelo completo (covariables y PRS) sobre el riesgo de hospitalización.....	59
Tabla 11 Asociación gravedad crítica (SEV4) y PRS de hipertensión. Resultados del análisis de regresión logística del modelo completo (covariables y PRS) sobre el riesgo de COVID con gravedad crítica	61
Tabla 12 Asociación gravedad crítica (SEV4) y PRS basadas en hits. Resultados del análisis de regresión logística del modelo completo (covariables y PRS) sobre el riesgo de gravedad crítica.....	64
Tabla 13 Asociación de antecedentes vasculares (AV) y PRS de hipertensión. Resultados del análisis de regresión logística del modelo completo (covariables y PRS) sobre el riesgo de antecedentes vasculares	66
Tabla 14 Asociación de antecedentes vasculares (AV) y PRS basadas en hits. Resultados del análisis de regresión logística del modelo completo (covariables y PRS) sobre el riesgo de antecedentes vasculares	67
Tabla 15 Asociación hipertensión (comorbilidad AV_1) y PRS de hipertensión. Resultados del análisis de regresión logística del modelo completo (covariables y PRS) sobre el riesgo de hipertensión	70
Tabla 16 Asociación hipertensión (comorbilidad AV_1) y PRS basada en hits. Resultados del análisis de regresión logística del modelo completo (covariables y PRS) sobre el riesgo de hipertensión	71

8. ANEXOS

ANEXO 1.



Santiago de Compostela, 16 de julio 2020

Señores:

FUNDACIÓN UNIVERSITARIA DE CIENCIAS DE LA SALUD

Por este medio certifico la invitación a participar en el proyecto: **“DETERMINANTES GENÉTICOS Y BIOMARCADORES GENÓMICOS DE RIESGO EN PACIENTES CON INFECCIÓN POR**

CORONAVIRUS SARS-COV-2” a los centros: **Sociedad de Cirugía de Bogotá Hospital de San José, Hospital Infantil Universitario de San José, Centro dermatológico Federico Lleras Acosta y al Centro de Gestión líder: Fundación Universitaria de Ciencias de la Salud.**

Así mismo se aprueba el presupuesto para compra de materiales e insumos por: 5.726€ que serán enviados por transferencia bancaria a la Fundación Universitaria de Ciencias de la Salud, se informa que el presupuesto del proyecto del Consorcio para 16.000 muestras españolas y 8.000 latinoamericanas es de 650.000€.

Atentamente;

USC
UNIVERSIDAD
DE COMPOSTELA
Ángel Carracedo (IP)

Fundación Pública Galga de Medicina Xenómica-Universidad de Santiago de Compostela (España).
Investigador principal del Proyecto SCOURGE

ANEXO 2.



COMITÉ DE ÉTICA EN INVESTIGACIÓN EN SERES HUMANOS CEISH



Bogotá, 31 de julio de 2020
(CEISH) 0339-2020

Doctora
LILIAN TORRES TOBAR
Investigador Principal
Ciudad

REF.: "DETERMINANTES GENÉTICOS Y BIOMARCADORES GENÓMICOS DE RIESGO EN PACIENTES CON INFECCIÓN POR CORONAVIRUS SARS-COV-2"

Respetada doctora:

El Comité de Ética en Investigación con Seres Humanos HSJ-FUCS (CEISH) del Hospital de San José, constituido mediante acta No. 2 del 29 de septiembre de 2006 y conformado por once (11) miembros permanentes con un requisito de Quórum para sesionar de manera ordinaria de la mitad más uno (Estatutos vigentes), certificado por el INVIMA, Res. 2017007903 del 27 de febrero de 2017 IPS Sociedad de Cirugía de Bogotá, Hospital de San José, certificada por la [Secretaría](#) de Salud, para la realización de Investigación Clínica.

Consignando sus deliberaciones en acta 14 del 22 de julio de 2020,

APRUEBA:

1. Protocolo versión 2 del 02 de agosto de 2020.
2. Consentimiento informado versión 2 del 22jul2020
3. Asentimiento versión 2 del 22jul2020
4. Consentimiento menor de edad versión 2 22jul2020
5. Consentimiento pacientes UCI versión 1 22jul2020
6. Hoja de recolección de información de datos

Calle 10 No. 18-75- Teléfono 3538100 Extensión 3621
Emails: ceish@hospitaldesanjose.org.co
www.fucsalud.edu.co - www.hospitaldesanjose.org.co
Bogotá, Colombia



COMITÉ DE ÉTICA EN INVESTIGACIÓN EN SERES HUMANOS CEISH



7. Carta de aprobación del comité de ética del Centro dermatológico Federico Lleras Acosta
8. Certificado 016 de acreditación del Centro Dermatológico Federico Lleras
9. Informe de visita de la secretaria de Salud FUCS-norte
10. Directorio de Laboratorios inscritos a la Red de Laboratorios de la Secretaría de Salud de Bogotá
11. Variables ScourgeCRD
12. Manual Scourge
13. Listado de centros del consorcio
14. Carta de la contrapartida del consorcio
15. Plantilla COVID19 (proyecto internacional)
16. Análisis bioinformático
17. Presupuesto

Queda pendiente envío de respuesta Invima a procedimiento de exportación de muestras, favor hacer llegar a este comité en cuanto se obtenga respuesta.

Dando cumplimiento a la Resolución 2378 de 2008, se informa que los estatutos que rigen este comité, así como las guías operativas de su funcionamiento, están disponibles en la página WEB de la Fundación Universitaria de Ciencias de la Salud para su conocimiento y aplicación.

Se oíten los miembros participantes en esta sesión:

Pablo Vilasquez Hoyos. (Presidente) Médico, Especialista en Pediatría, Intensivista, Especialista en Epidemiología, **María Carolina Ayala Ramírez (Secretaria)** Licenciada en lenguas Modernas., Magister en Ciudadanía y Derechos Humanos, Ética y Política, Especialista en Bioética, Doctoranda en Ciudadanía en Derechos Humanos, **Michael A. Vallejo Urrego**, Médico Geriatra, Especialista en Epidemiología, Especialista en Gerencia de la Calidad en Salud, Magister en Genética Humana y Genética

Calle 10 No. 18-75- Teléfono 3538100 Extensión 3621
Email: ceish@hospitaldesenjo.org - ori@ceish.org
www.fuacsalud.edu.co - www.hospitaldesenjo.org - ori@ceish.org
Bogotá, Colombia



COMITÉ DE ÉTICA EN INVESTIGACIÓN EN SERES HUMANOS CEISH



Luis A. Castro Pineda, Miembro de la Comunidad, **Patricia Pérez Abo11da1110**, Psicólogo, **María Inés Sarmiento Medina**, Médico,, Especialista en Epidemiología, Especialista en Bioética, Magíster en Bioética, **Guillermo Sánchez Vanegas**, Médico, Especialista y MSc en Epidemiología, Doctor en Salud Pública, **Jean11Bier1re Ca margo**, Abogado, Diego Alberto Caioe.do, Químico Farmacéutico, Moisés H. Silvera, Enfermero, Especialista en Bioética, cl a111dia, Ul1iana López B., Médico,, Especialista y Magister en Bioética..

Atentamente,


PABLO VASQUEZ HOYOS MD. MSc.
 Presidente

Preparó y Elaboró: Patricia Perez A

cam" 10 f1lo.18-75- h'éfono 35:381'1JO Exten,i6n 3621
 Emails: ceish@hos,ita1de.anjos!.oq;.co
 www.fuC!aludedu.,o - www.hos,pita1de.anjose.oq;.co
 Bog,otá, C:ollormaia



La contribución de Colombia al consorcio SCOURGE parte de 1354 pacientes con diferentes grados de afección con el objetivo de determinar variantes genéticas de severidad y analizar el poder de las estimas de riesgo poligénico de hipertensión para predecir la severidad y hospitalización por COVID-19. Se encontró una clara correlación entre las comorbilidades cardiovasculares, hospitalización y severidad. Se construyó un GWAS ajustado para comorbilidades cardiovasculares. En el caso del GWAS de riesgo de gravedad crítica ajustando con antecedentes cardiovasculares se observan dos SNPs, rs143861937; rs73078305, ubicados en el cromosoma 3 que pasan umbral de referencia ($p=6.4 \times 10^{-9}$), cercanos a los genes PLXNB1 y al gen PFKFB4 no reportados hasta ahora en GWAS previos.