

1 Novel risk loci for COVID-19 hospitalization among 2 admixed American populations

3 Silvia Diz-de Almeida,^{1,2,13,7} Raquel Cruz,^{1,2,13,7} Andre D. Luchessi,³ José M. Lorenzo-Salazar,⁴
4 Miguel López de Heredia,² Inés Quintela,⁵ Rafaela González-Montelongo,⁴ Vivian N. Silbiger,³
5 Marta Sevilla Porras,^{2,6} Jair Antonio Tenorio Castaño,^{2,6,7} Julian Nevado,^{2,6,7} Jose María
6 Aguado,^{8,9,10,11} Carlos Aguilar,¹² Sergio Aguilera-Albesa,^{13,14} Virginia Almadana,¹⁵ Berta
7 Almoguera,^{16,2} Nuria Alvarez,¹⁷ Álvaro Andreu-Bernabeu,^{18,10} Eunáte Arana-Arri,^{19,20} Celso
8 Arango,^{18,21,10} María J. Arranz,²² Maria-Jesus Artiga,²³ Raúl C. Baptista-Rosas,^{24,25,26} María
9 Barreda- Sánchez,^{27,28} Moncef Belhassen-Garcia,^{29,30} Joao F. Bezerra,³¹ Marcos A.C. Bezerra,³²
10 Lucía Boix-Palop,³³ María Brion,^{34,35} Ramón Brugada,^{36,37,35,38} Matilde Bustos,³⁹ Enrique J.
11 Calderón,^{39,40,41} Cristina Carbonell,^{42,30} Luis Castano,^{19,43,2,44,45} Jose E. Castelao,⁴⁶ Rosa Conde-
12 Vicente,⁴⁷ M. Lourdes Cordero-Lorenzana,⁴⁸ Jose L. Cortes-Sanchez,^{49,50} Marta Corton,^{16,2} M.
13 Teresa Darnaude,⁵¹ Alba De Martino-Rodríguez,^{52,53} Victor del Campo-Pérez,⁵⁴ Aranzazu Diaz
14 de Bustamante,⁵¹ Elena Domínguez-Garrido,⁵⁵ Rocío Eirós,⁵⁶ María Carmen Fariñas,^{57,58,59}
15 María J. Fernandez-Nestosa,⁶⁰ Uxía Fernández-Robelo,⁶¹ Amanda Fernández-Rodríguez,^{62,11}
16 Tania Fernández-Villa,^{63,41} Manuela Gago-Domínguez,^{5,64} Belén Gil-Fournier,⁶⁵ Javier Gómez-
17 Arrue,^{52,53} Beatriz González Álvarez,^{52,53} Fernan Gonzalez Bernaldo de Quirós,⁶⁶ Anna
18 González-Neira,¹⁷ Javier González-Peñas,^{18,10,21} Juan F. Gutiérrez-Bautista,⁶⁷ María José
19 Herrero,^{68,69} Antonio Herrero-Gonzalez,⁷⁰ María A. Jimenez-Sousa,^{62,11} María Claudia Lattig,^{71,72}
20 Anabel Liger Borja,⁷³ Rosario Lopez-Rodriguez,^{16,2,74} Esther Mancebo,^{75,76} Caridad Martín-
21 López,⁷³ Vicente Martín,^{63,41} Oscar Martinez-Nieto,^{77,72} Iciar Martinez-Lopez,^{78,79} Michel F.
22 Martinez-Resendez,⁴⁹ Ángel Martinez-Perez,⁸⁰ Juliana F. Mazzeu,^{81,82,83} Eleuterio Merayo
23 Macías,⁸⁴ Pablo Minguez,^{16,2} Victor Moreno Cuerda,^{85,86} Silviene F. Oliveira,^{87,88,83,89} Eva
24 Ortega-Paino,²³ Mara Parellada,^{18,21,10} Estela Paz-Artal,^{75,76,90} Ney P.C. Santos,⁹¹ Patricia Pérez-
25 Matute,⁹² Patricia Perez,⁹³ M. Elena Pérez-Tomás,²⁷ Teresa Perucho,⁹⁴ Mel-lina Pinsach-
26 Abuin,^{36,35} Guillermo Pita,¹⁷ Ericka N. Pompa-Mera,^{95,96} Gloria L. Porras-Hurtado,⁹⁷ Aurora

27 Pujol,^{98,2,99} Soraya Ramiro León,⁶⁵ Salvador Resino,^{62,11} Marianne R. Fernandes,^{91,100} Emilio
28 Rodríguez-Ruiz,^{101,64} Fernando Rodriguez-Artalejo,^{102,103,41,104} José A. Rodriguez-Garcia,¹⁰⁵
29 Francisco Ruiz-Cabello,^{67,106,107} Javier Ruiz-Hornillos,^{108,109,110} Pablo Ryan,^{111,112,113,11} José
30 Manuel Soria,⁸⁰ Juan Carlos Souto,¹¹⁴ Eduardo Tamayo,^{115,116,11} Alvaro Tamayo-Velasco,¹¹⁷
31 Juan Carlos Taracido-Fernandez,⁷⁰ Alejandro Teper,¹¹⁸ Lilian Torres-Tobar,¹¹⁹ Miguel
32 Urioste,¹²⁰ Juan Valencia-Ramos,¹²¹ Zuleima Yáñez,¹²² Ruth Zarate,¹²³ Itziar de Rojas,^{124,125}
33 Agustín Ruiz,^{124,125} Pascual Sánchez,¹²⁶ Luis Miguel Real,¹²⁷ SCOURGE Cohort Group, Encarna
34 Guillen-Navarro,^{27,128,129,130} Carmen Ayuso,^{16,2} Esteban Parra,¹³¹ José A. Riancho,^{57,58,59,2}
35 Augusto Rojas-Martinez,¹³² Carlos Flores,^{4,133,134,135,136,*} Pablo Lapunzina,^{2,6,7,136} Ángel
36 Carracedo,^{1,2,5,64,136,**}

37 **Affiliations**

38 ¹, Centro Singular de Investigación en Medicina Molecular y Enfermedades Crónicas (CIMUS),
39 Universidade de Santiago de Compostela, Santiago de Compostela, Spain
40 ², CIBERER, ISCIII, Madrid, Spain
41 ³, Universidade Federal do Rio Grande do Norte, Departamento de Analises Clinicas e
42 Toxicologicas, Natal, Brazil
43 ⁴, Genomics Division, Instituto Tecnológico y de Energías Renovables, Santa Cruz de Tenerife,
44 Spain
45 ⁵, Fundación Pública Galega de Medicina Xenómica, Sistema Galego de Saúde (SERGAS)
46 Santiago de Compostela, Spain
47 ⁶, Instituto de Genética Médica y Molecular (INGEMM), Hospital Universitario La Paz-
48 IDIPAZ, Madrid, Spain
49 ⁷, ERN-ITHACA-European Reference Network
50 ⁸, Unit of Infectious Diseases, Hospital Universitario 12 de Octubre, Instituto de Investigación
51 Sanitaria Hospital 12 de Octubre (imas12), Madrid, Spain
52 ⁹, Spanish Network for Research in Infectious Diseases (REIPI RD16/0016/0002), Instituto de
53 Salud Carlos III, Madrid, Spain

- 54 ¹⁰, School of Medicine, Universidad Complutense, Madrid, Spain
- 55 ¹¹, CIBERINFEC, ISCIII, Madrid, Spain
- 56 ¹², Hospital General Santa Bárbara de Soria, Soria, Spain
- 57 ¹³, Pediatric Neurology Unit, Department of Pediatrics, Navarra Health Service Hospital,
58 Pamplona, Spain
- 59 ¹⁴, Navarra Health Service, NavarraBioMed Research Group, Pamplona, Spain
- 60 ¹⁵, Hospital Universitario Virgen Macarena, Neumología, Seville, Spain
- 61 ¹⁶, Department of Genetics & Genomics, Instituto de Investigación Sanitaria-Fundación Jiménez
62 Díaz University Hospital - Universidad Autónoma de Madrid (IIS-FJD, UAM), Madrid, Spain
- 63 ¹⁷, Spanish National Cancer Research Centre, Human Genotyping-CEGEN Unit, Madrid, Spain
- 64 ¹⁸, Department of Child and Adolescent Psychiatry, Institute of Psychiatry and Mental Health,
65 Hospital General Universitario Gregorio Marañón (IiSGM), Madrid, Spain
- 66 ¹⁹, Biocruces Bizkai HRI, Bizkaia, Spain
- 67 ²⁰, Cruces University Hospital, Osakidetza, Bizkaia, Spain
- 68 ²¹, Centre for Biomedical Network Research on Mental Health (CIBERSAM), Instituto de Salud
69 Carlos III, Madrid, Spain
- 70 ²², Fundació Docència I Recerca Mutua Terrassa, Barcelona, Spain
- 71 ²³, Spanish National Cancer Research Center, CNIO Biobank, Madrid, Spain
- 72 ²⁴, Hospital General de Occidente, Zapopan Jalisco, Mexico
- 73 ²⁵, Centro Universitario de Tonalá, Universidad de Guadalajara, Tonalá Jalisco, Mexico
- 74 ²⁶, Centro de Investigación Multidisciplinario en Salud, Universidad de Guadalajara, Tonalá
75 Jalisco, Mexico
- 76 ²⁷, Instituto Murciano de Investigación Biosanitaria (IMIB-Arrixaca), Murcia, Spain
- 77 ²⁸, Universidad Católica San Antonio de Murcia (UCAM), Murcia, Spain
- 78 ²⁹, Hospital Universitario de Salamanca-IBSAL, Servicio de Medicina Interna-Unidad de
79 Enfermedades Infecciosas, Salamanca, Spain
- 80 ³⁰, Universidad de Salamanca, Salamanca, Spain
- 81 ³¹, Escola Tecnica de Saúde, Laboratório de Vigilância Molecular Aplicada, Brazil

- 82 ³², Federal University of Pernambuco, Genetics Postgraduate Program, Recife, PE, Brazil
- 83 ³³, Hospital Universitario Mutua Terrassa, Barcelona, Spain
- 84 ³⁴, Instituto de Investigación Sanitaria de Santiago (IDIS), Xenética Cardiovascular, Santiago de
- 85 Compostela, Spain
- 86 ³⁵, CIBERCV, ISCIII, Madrid, Spain
- 87 ³⁶, Cardiovascular Genetics Center, Institut d'Investigació Biomèdica Girona (IDIBGI), Girona,
- 88 Spain
- 89 ³⁷, Medical Science Department, School of Medicine, University of Girona, Girona, Spain
- 90 ³⁸, Hospital Josep Trueta, Cardiology Service, Girona, Spain
- 91 ³⁹, Institute of Biomedicine of Seville (IBiS), Consejo Superior de Investigaciones Científicas
- 92 (CSIC)- University of Seville- Virgen del Rocío University Hospital, Seville, Spain
- 93 ⁴⁰, Departamento de Medicina, Hospital Universitario Virgen del Rocío, Universidad de Sevilla,
- 94 Seville, Spain
- 95 ⁴¹, CIBERESP, ISCIII, Madrid, Spain
- 96 ⁴², Hospital Universitario de Salamanca-IBSAL, Servicio de Medicina Interna, Salamanca,
- 97 Spain
- 98 ⁴³, Osakidetza, Cruces University Hospital, Bizkaia, Spain
- 99 ⁴⁴, Centre for Biomedical Network Research on Diabetes and Metabolic Associated Diseases
- 100 (CIBERDEM), Instituto de Salud Carlos III, Madrid, Spain
- 101 ⁴⁵, University of Pais Vasco, UPV/EHU, Bizkaia, Spain
- 102 ⁴⁶, Oncology and Genetics Unit, Instituto de Investigación Sanitaria Galicia Sur, Xerencia de
- 103 Xestión Integrada de Vigo-Servizo Galego de Saúde, Vigo, Spain
- 104 ⁴⁷, Hospital Universitario Río Hortega, Valladolid, Spain
- 105 ⁴⁸, Servicio de Medicina intensiva, Complejo Hospitalario Universitario de A Coruña
- 106 (CHUAC), Sistema Galego de Saúde (SERGAS), A Coruña, Spain
- 107 ⁴⁹, Tecnológico de Monterrey, Monterrey, Mexico
- 108 ⁵⁰, Otto von Guericke University, Department of Microgravity and Translational Regenerative
- 109 Medicine, Magdeburg, Germany

- 110 ⁵¹, Hospital Universitario Mostoles, Unidad de Genética, Madrid, Spain
- 111 ⁵², Instituto Aragonés de Ciencias de la Salud (IACS), Zaragoza, Spain
- 112 ⁵³, Instituto Investigación Sanitaria Aragón (IIS-Aragon), Zaragoza, Spain
- 113 ⁵⁴, Preventive Medicine Department, Instituto de Investigacion Sanitaria Galicia Sur, Xerencia
- 114 de Xestion Integrada de Vigo-Servizo Galego de Saúde, Vigo, Spain
- 115 ⁵⁵, Unidad Diagnóstico Molecular. Fundación Rioja Salud, La Rioja, Spain
- 116 ⁵⁶, Hospital Universitario de Salamanca-IBSAL, Servicio de Cardiología, Salamanca, Spain
- 117 ⁵⁷, IDIVAL, Cantabria, Spain
- 118 ⁵⁸, Universidad de Cantabria, Cantabria, Spain
- 119 ⁵⁹, Hospital U M Valdecilla, Cantabria, Spain
- 120 ⁶⁰, Universidad Nacional de Asunción, Facultad de Politécnica, Paraguay
- 121 ⁶¹, Urgencias Hospitalarias, Complejo Hospitalario Universitario de A Coruña (CHUAC),
- 122 Sistema Galego de Saúde (SERGAS), A Coruña, Spain
- 123 ⁶², Unidad de Infección Viral e Inmunidad, Centro Nacional de Microbiología (CNM), Instituto
- 124 de Salud Carlos III (ISCIII), Madrid, Spain
- 125 ⁶³, Grupo de Investigación en Interacciones Gen-Ambiente y Salud (GIIGAS) - Instituto de
- 126 Biomedicina (IBIOMED), Universidad de León, León, Spain
- 127 ⁶⁴, IDIS
- 128 ⁶⁵, Hospital Universitario de Getafe, Servicio de Genética, Madrid, Spain
- 129 ⁶⁶, Ministerio de Salud Ciudad de Buenos Aires, Buenos Aires, Argentina
- 130 ⁶⁷, Hospital Universitario Virgen de las Nieves, Servicio de Análisis Clínicos e Inmunología,
- 131 Granada, Spain
- 132 ⁶⁸, IIS La Fe, Plataforma de Farmacogenética, Valencia, Spain
- 133 ⁶⁹, Universidad de Valencia, Departamento de Farmacología, Valencia, Spain
- 134 ⁷⁰, Data Analysis Department, Instituto de Investigación Sanitaria-Fundación Jiménez Díaz
- 135 University Hospital - Universidad Autónoma de Madrid (IIS-FJD, UAM), Madrid, Spain
- 136 ⁷¹, Universidad de los Andes, Facultad de Ciencias, Bogotá, Colombia

- 137 ⁷², SIGEN Alianza Universidad de los Andes - Fundación Santa Fe de Bogotá, Bogotá,
138 Colombia
- 139 ⁷³, Hospital General de Segovia, Medicina Intensiva, Segovia, Spain
- 140 ⁷⁴, Facultad de Farmacia, Universidad San Pablo-CEU, CEU Universities, Urbanización
141 Montepríncipe, 28660 Boadilla del Monte, España
- 142 ⁷⁵, Hospital Universitario 12 de Octubre, Department of Immunology, Madrid, Spain
- 143 ⁷⁶, Instituto de Investigación Sanitaria Hospital 12 de Octubre (imas12), Transplant
144 Immunology and Immunodeficiencies Group, Madrid, Spain
- 145 ⁷⁷, Fundación Santa Fe de Bogota, Departamento Patología y Laboratorios, Bogotá, Colombia
- 146 ⁷⁸, Unidad de Genética y Genómica Islas Baleares, Islas Baleares, Spain
- 147 ⁷⁹, Hospital Universitario Son Espases, Unidad de Diagnóstico Molecular y Genética Clínica,
148 Islas Baleares, Spain
- 149 ⁸⁰, Genomics of Complex Diseases Unit, Research Institute of Hospital de la Santa Creu i Sant
150 Pau, IIB Sant Pau, Barcelona, Spain
- 151 ⁸¹, Universidade de Brasília, Faculdade de Medicina, Brazil
- 152 ⁸², Programa de Pós-Graduação em Ciências Médicas (UnB), Brazil
- 153 ⁸³, Programa de Pós-Graduação em Ciências da Saude (UnB), Brazil
- 154 ⁸⁴, Hospital El Bierzo, Unidad Cuidados Intensivos, León, Spain
- 155 ⁸⁵, Hospital Universitario Mostoles, Medicina Interna, Madrid, Spain
- 156 ⁸⁶, Universidad Francisco de Vitoria, Madrid, Spain
- 157 ⁸⁷, Departamento de Genética e Morfologia, Instituto de Ciências Biológicas, Universidade de
158 Brasília, Brazil
- 159 ⁸⁸, Programa de Pós-Graduação em Biologia Animal (UnB), Brazil
- 160 ⁸⁹, Programa de Pós-Graduação Profissional em Ensino de Biologia (UnB), Brazil
- 161 ⁹⁰, Universidad Complutense de Madrid, Department of Immunology, Ophthalmology and ENT,
162 Madrid, Spain
- 163 ⁹¹, Universidade Federal do Pará, Núcleo de Pesquisas em Oncologia, Belém, Pará, Brazil

164 ⁹², Infectious Diseases, Microbiota and Metabolism Unit, CSIC Associated Unit, Center for
165 Biomedical Research of La Rioja (CIBIR), Logroño, Spain
166 ⁹³, Inditex, A Coruña, Spain
167 ⁹⁴, GENYCA, Madrid, Spain
168 ⁹⁵, Instituto Mexicano del Seguro Social (IMSS), Centro Médico Nacional Siglo XXI, Unidad
169 de Investigación Médica en Enfermedades Infecciosas y Parasitarias, Mexico City, Mexico
170 ⁹⁶, Instituto Mexicano del Seguro Social (IMSS), Centro Médico Nacional La Raza. Hospital de
171 Infectología, Mexico City, Mexico
172 ⁹⁷, Clínica Comfamiliar Risaralda, Pereira, Colombia
173 ⁹⁸, Bellvitge Biomedical Research Institute (IDIBELL), Neurometabolic Diseases Laboratory,
174 L'Hospitalet de Llobregat, Spain
175 ⁹⁹, Catalan Institution of Research and Advanced Studies (ICREA), Barcelona, Spain
176 ¹⁰⁰, Hospital Ophir Loyola, Departamento de Ensino e Pesquisa, Belém, Pará, Brazil
177 ¹⁰¹, Unidad de Cuidados Intensivos, Hospital Clínico Universitario de Santiago (CHUS),
178 Sistema Galego de Saúde (SERGAS), Santiago de Compostela, Spain
179 ¹⁰², Department of Preventive Medicine and Public Health, School of Medicine, Universidad
180 Autónoma de Madrid, Madrid, Spain
181 ¹⁰³, IdiPaz (Instituto de Investigación Sanitaria Hospital Universitario La Paz), Madrid, Spain
182 ¹⁰⁴, IMDEA-Food Institute, CEI UAM+CSIC, Madrid, Spain
183 ¹⁰⁵, Complejo Asistencial Universitario de León, León, Spain
184 ¹⁰⁶, Instituto de Investigación Biosanitaria de Granada (ibs GRANADA), Granada, Spain
185 ¹⁰⁷, Universidad de Granada, Departamento Bioquímica, Biología Molecular e Inmunología III,
186 Granada, Spain
187 ¹⁰⁸, Hospital Infanta Elena, Allergy Unit, Valdemoro, Madrid, Spain
188 ¹⁰⁹, Instituto de Investigación Sanitaria-Fundación Jiménez Díaz University Hospital -
189 Universidad Autónoma de Madrid (IIS-FJD, UAM), Madrid, Spain
190 ¹¹⁰, Faculty of Medicine, Universidad Francisco de Vitoria, Madrid, Spain
191 ¹¹¹, Hospital Universitario Infanta Leonor, Madrid, Spain

- 192 ¹¹², Complutense University of Madrid, Madrid, Spain
- 193 ¹¹³, Gregorio Marañón Health Research Institute (IiSGM), Madrid, Spain
- 194 ¹¹⁴, Haemostasis and Thrombosis Unit, Hospital de la Santa Creu i Sant Pau, IIB Sant Pau,
195 Barcelona, Spain
- 196 ¹¹⁵, Hospital Clinico Universitario de Valladolid, Servicio de Anestesiología y Reanimación,
197 Valladolid, Spain
- 198 ¹¹⁶, Universidad de Valladolid, Departamento de Cirugía, Valladolid, Spain
- 199 ¹¹⁷, Hospital Clinico Universitario de Valladolid, Servicio de Hematología y Hemoterapia,
200 Valladolid, Spain
- 201 ¹¹⁸, Hospital de Niños Ricardo Gutierrez, Buenos Aires, Argentina
- 202 ¹¹⁹, Fundación Universitaria de Ciencias de la Salud
- 203 ¹²⁰, Spanish National Cancer Research Centre, Familial Cancer Clinical Unit, Madrid, Spain
- 204 ¹²¹, University Hospital of Burgos, Burgos, Spain
- 205 ¹²², Universidad Simón Bolívar, Facultad de Ciencias de la Salud, Barranquilla, Colombia
- 206 ¹²³, Centro para el Desarrollo de la Investigación Científica, Paraguay
- 207 ¹²⁴, Research Center and Memory clinic, ACE Alzheimer Center Barcelona, Universitat
208 Internacional de Catalunya, Spain
- 209 ¹²⁵, Centre for Biomedical Network Research on Neurodegenerative Diseases (CIBERNED),
210 Instituto de Salud Carlos III, Madrid, Spain
- 211 ¹²⁶, CIEN Foundation/Queen Sofia Foundation Alzheimer Center, Madrid, Spain
- 212 ¹²⁷, Hospital Universitario de Valme, Unidad Clínica de Enfermedades Infecciosas y
213 Microbiología, Sevilla, Spain
- 214 ¹²⁸, Sección Genética Médica - Servicio de Pediatría, Hospital Clínico Universitario Virgen de la
215 Arrixaca, Servicio Murciano de Salud, Murcia, Spain
- 216 ¹²⁹, Departamento Cirugía, Pediatría, Obstetricia y Ginecología, Facultad de Medicina,
217 Universidad de Murcia (UMU), Murcia, Spain
- 218 ¹³⁰, Grupo Clínico Vinculado, Centre for Biomedical Network Research on Rare Diseases
219 (CIBERER), Instituto de Salud Carlos III, Madrid, Spain

220 ¹³¹ Department of Anthropology, University of Toronto at Mississauga, Mississauga, Ontario,
221 Canada.
222 ¹³², Tecnológico de Monterrey, Escuela de Medicina y Ciencias de la Salud, Monterrey, Mexico
223 ¹³³, Research Unit, Hospital Universitario Nuestra Señora de Candelaria, Instituto de
224 Investigación Sanitaria de Canarias, Santa Cruz de Tenerife, Spain
225 ¹³⁴, Centre for Biomedical Network Research on Respiratory Diseases (CIBERES), Instituto de
226 Salud Carlos III, Madrid, Spain
227 ¹³⁵, Department of Clinical Sciences, University Fernando Pessoa Canarias, Las Palmas de Gran
228 Canaria, Spain
229 ¹³⁶, These authors contributed equally: Carlos Flores, Pablo Lapunzina, Ángel Carracedo
230 ¹³⁷, These authors contributed equally: Silvia Diz-de Almeida, Raquel Cruz
231 *Correspondence:
232 **Correspondence: angel.carracedo@usc.es

233

234 **Abstract**

235 The genetic basis of severe COVID-19 has been thoroughly studied, and many genetic
236 risk factors shared between populations have been identified. However, reduced sample
237 sizes from non-European groups have limited the discovery of population-specific
238 common risk loci. In this second study nested in the SCOURGE consortium, we
239 conducted a GWAS for COVID-19 hospitalization in admixed Americans, comprising a
240 total of 4,702 hospitalized cases recruited by SCOURGE and seven other participating
241 studies in the COVID-19 Host Genetic Initiative. We identified four genome-wide
242 significant associations, two of which constitute novel loci and were first discovered in
243 Latin American populations (*BAZ2B* and *DDIAS*). A trans-ethnic meta-analysis revealed
244 another novel cross-population risk locus in *CREBBP*. Finally, we assessed the

245 performance of a cross-ancestry polygenic risk score in the SCOURGE admixed
246 American cohort. This study constitutes the largest GWAS for COVID-19
247 hospitalization in admixed Latin Americans conducted to date. This allowed to reveal
248 novel risk loci and emphasize the need of considering the diversity of populations in
249 genomic research.

250 **Introduction**

251 To date, more than 50 loci associated with COVID-19 susceptibility, hospitalization,
252 and severity have been identified using genome-wide association studies (GWAS)^{1,2}.
253 The COVID-19 Host Genetics Initiative (HGI) has made significant efforts³ to augment
254 the power to identify disease loci by recruiting individuals from diverse populations and
255 conducting a trans-ancestry meta-analysis. Despite this, the lack of genetic diversity and
256 a focus on cases of European ancestries still predominate in the studies^{4,5}. In addition,
257 while trans-ancestry meta-analyses are a powerful approach for discovering shared
258 genetic risk variants with similar effects across populations⁶, they may fail to identify
259 risk variants that have larger effects on particular underrepresented populations. Genetic
260 disease risk has been shaped by the particular evolutionary history of populations and
261 environmental exposures⁷. Their action is particularly important for infectious diseases
262 due to the selective constraints that are imposed by host-pathogen interactions^{8,9}.
263 Literature examples of this in COVID-19 severity include a *DOCK2* gene variant in
264 East Asians¹⁰ and frequent loss-of-function variants in *IFNAR1* and *IFNAR2* genes in
265 Polynesian and Inuit populations, respectively^{11,12}.

266 Including diverse populations in case-control GWAS studies with unrelated
267 participants usually requires a prior classification of individuals in genetically
268 homogeneous groups, which are typically analyzed separately to control the population
269 stratification effects¹³. Populations with recent admixture impose an additional

270 challenge to GWASs due to their complex genetic diversity and linkage disequilibrium
271 (LD) patterns, requiring the development of alternative approaches and a careful
272 inspection of results to reduce false positives due to population structure⁷. In fact, there
273 are benefits in study power from modeling the admixed ancestries either locally, at the
274 regional scale in the chromosomes, or globally, across the genome, depending on
275 factors such as the heterogeneity of the risk variant in frequencies or the effects among
276 the ancestry strata¹⁴. Despite the development of novel methods specifically tailored for
277 the analysis of admixed populations¹⁵, the lack of a standardized analysis framework
278 and the difficulties in confidently clustering admixed individuals into particular genetic
279 groups often leads to their exclusion from GWAS.

280 The Spanish Coalition to Unlock Research on Host Genetics on COVID-19
281 (SCOURGE) recruited COVID-19 patients between March and December 2020 from
282 hospitals across Spain and from March 2020 to July 2021 in Latin America
283 (<https://www.scourge-covid.org>). A first GWAS of COVID-19 severity among Spanish
284 patients of European descent revealed novel disease loci and explored age- and sex-
285 varying effects of genetic factors¹⁶. Here, we present the findings of a GWAS meta-
286 analysis in admixed Latin American (AMR) populations, comprising individuals from
287 the SCOURGE Latin-American cohort and the HGI studies, which allowed us to
288 identify two novel severe COVID-19 loci, *BAZ2B* and *DDIAS*. Further analyses
289 modeling the admixture from three genetic ancestral components and performing a
290 trans-ethnic meta-analysis led to the identification of an additional risk locus near
291 *CREBBP*. We finally assessed a cross-ancestry polygenic risk score model with variants
292 associated with critical COVID-19.

293 **Results**

294 **Meta-analysis of COVID-19 hospitalization in admixed Americans**

295 *Study cohorts*

296 Within the SCOURGE consortium, we included 1,608 hospitalized cases and 1,887
297 controls (not hospitalized COVID-19 patients) from Latin American countries and from
298 recruitments of individuals of Latin American descent conducted in Spain
299 (Supplementary Table 1). Quality control details and estimation of global genetic
300 inferred ancestry (GIA) (Supplementary Figure 1) are described in Methods, whereas
301 clinical and demographic characteristics of patients included in the analysis are shown
302 in Table 1. Summary statistics from the SCOURGE cohort were obtained under a
303 logistic mixed model with the SAIGE model (Methods). Another seven studies
304 participating in the COVID-19 HGI consortium were included in the meta-analysis of
305 COVID-19 hospitalization in admixed Americans (Figure 1).

306 *GWAS meta-analysis*

307 We performed a fixed-effects GWAS meta-analysis using the inverse of the variance as
308 weights for the overlapping markers. The combined GWAS sample size consisted of
309 4,702 admixed AMR hospitalized cases and 68,573 controls.

310 This GWAS meta-analysis revealed genome-wide significant associations at four risk
311 loci (Table 2, Figure 2), two of which (*BAZ2B* and *DDIAS*) were novel discoveries.
312 Four lead variants were identified, linked to other 310 variants (Supplementary Tables
313 2-3). A gene-based association test revealed a significant association in *BAZ2B* and in
314 previously known COVID-19 risk loci: *LZTFL1*, *XCRI*, *FYCO1*, *CCR9*, and *IFNAR2*
315 (Supplementary Table 4).

316 Located within the *BAZ2B* gene, the sentinel variant rs13003835 is an intronic variant
317 associated with an increased risk of COVID-19 hospitalization (Odds Ratio [OR]=1.20,
318 95% Confidence Interval [CI]=1.12-1.27, $p=3.66 \times 10^{-8}$). This association was not

319 previously reported in any GWAS of COVID-19 published to date. Interestingly,
320 rs13003835 did not reach significance ($p=0.972$) in the COVID-19 HGI trans-ancestry
321 meta-analysis including the five population groups¹.

322 The other novel risk locus is led by the sentinel variant rs77599934, a rare intronic
323 variant located in chromosome 11 within *DDIAS* and associated with the risk of
324 COVID-19 hospitalization (OR=2.27, 95% CI=1.70-3.04, $p=2.26 \times 10^{-8}$).

325 We also observed a suggestive association with rs2601183 in chromosome 15, which is
326 located between *ZNF774* and *IQGAPI* (allele-G OR=1.20, 95% CI=1.12-1.29,
327 $p=6.11 \times 10^{-8}$, see Supplementary Table 2), which has not yet been reported in any other
328 GWAS of COVID-19 to date.

329 The GWAS meta-analysis also pinpointed two significant variants at known loci,
330 *LZTFL1* and *FOXP4*. The SNP rs35731912 was previously associated with COVID-19
331 severity in EUR populations¹⁷, and it was mapped to *LZTFL1*. While rs2477820 is a
332 novel risk variant within the *FOXP4* gene, it has a moderate LD ($r^2=0.295$) with
333 rs2496644, which has been linked to COVID-19 hospitalization¹⁸. This is consistent
334 with the effects of LD in tag-SNPs when conducting GWAS in diverse populations.

335 None of the lead variants was associated with the comorbidities included in Table 1.

336 **Functional mapping of novel risk variants**

337 Variants belonging to the lead loci were prioritized by positional and expression
338 quantitative trait loci (eQTL) mapping with FUMA, resulting in 31 mapped genes
339 (Supplementary Table 5). Within the region surrounding the lead variant in
340 chromosome 2, FUMA prioritized four genes in addition to *BAZ2B* (*PLA2R1*, *LY75*,
341 *WDSUB1*, and *CD302*). rs13003835 (allele C) is an eQTL of *LY75* in the esophagus
342 mucosa (NES=0.27) and of *BAZ2B-AS* in whole blood (NES=0.27), while rs2884110

343 ($R^2=0.85$) is an eQTL of *LY75* in lung (NES=0.22). As for the chromosome 11,
344 rs77599934 (allele G) is in moderate-to-strong LD ($r^2=0.776$) with rs60606421 (G
345 deletion, allele -) which is an eQTL associated with a reduced expression of *DDIAS* in
346 the lungs (NES=-0.49, allele -). Associations with expression are shown in the
347 supplementary Figure 4. The sentinel variant for the region in chromosome 16 is in
348 perfect LD ($r^2=1$) with rs601183, an eQTL of *ZNF774* in the lung.

349 *Bayesian fine mapping*

350 We performed different approaches to narrow down the prioritized loci to a set of most
351 likely genes driving the associations. First, we computed credible sets at the 95%
352 confidence level for causal variants and annotated them with VEP (and V2G aggregate
353 scoring. The 95% confidence credible set from the region of chromosome 2 around
354 rs13003835 included 76 variants, which can be found in the Supplementary Table 6
355 (VEP and V2G annotations are included in the Supplementary Tables 7 and 8). The V2G
356 score prioritized *BAZ2B* as the most likely gene driving the association. However, the
357 approach was unable to converge allocating variants in a 95% confidence credible set
358 for the region in chromosome 11.

359 *Transcriptome-wide association study (TWAS)*

360 Five novel genes, namely, *SLC25A37*, *SMARCC1*, *CAMP*, *TYW3*, and *SI00A12*
361 (supplementary Table 9), were found to be significantly associated in the cross-tissue
362 TWAS. To our knowledge, these genes have not been reported previously in any
363 COVID-19 TWAS or GWAS analyses published to date. In the single tissue analyses,
364 *ATP5O* and *CXCR6* were significantly associated in the lungs, *CCR9* was significantly
365 associated in whole blood, and *IFNAR2* and *SLC25A37* were associated in lymphocytes.

366 Likewise, we carried out TWAS analyses using the models trained in the admixed
367 populations. However, no significant gene pairs were detected in this case. The top 10
368 genes with the lowest p-values for each of the datasets (Puerto-Ricans, Mexicans,
369 African-Americans and pooled cohorts) are shown in the Supplementary Table 10.
370 Although not significant, *KCNC3* was repeated in the four analyses, whereas
371 *MAPKAPK3*, *NAPSA* and *THAP5* were repeated in 3 out of 4. Both *NAPSA* and *KCNC3*
372 are located in the chromosome 19 and were reported in the latest HGI meta-analysis¹⁹.
373 All mapped genes from analyses conducted in AMR populations are shown in Figure 5.

374

375 **Genetic architecture of COVID-19 hospitalization in AMR populations**

376 *Allele frequencies of rs13003835 and rs77599934 across ancestries*

377 Neither rs13003835 (*BAZ2B*) nor rs77599934 (*DDIAS*) were significantly associated in
378 the COVID-19 HGI B2 cross-population or population-specific meta-analyses. Thus,
379 we investigated their allele frequencies (AF) across populations and compared their
380 effect sizes.

381 According to gnomAD v3.1.2, the T allele at rs13003835 (*BAZ2B*) has an AF of 43% in
382 admixed AMR groups, while AF is lower in the EUR populations (16%) and in the
383 global sample (29%). Local ancestry inference (LAI) reported by gnomAD shows that
384 within the Native-American component, the risk allele T is the major allele, whereas it
385 is the minor allele within the African and European LAI components. These large
386 differences in AF might be the reason underlying the association found in AMR
387 populations. However, when comparing effect sizes between populations, we found that
388 they were in opposite directions between SAS-AMR and EUR-AFR-EAS and that there
389 was large heterogeneity among them (Figure 3). We queried SNPs within 50kb

390 windows of the lead variant in each of the other populations that had p-value<0.01. The
391 variant with the lowest p-value in the EUR population was rs559179177 ($p=1.72 \times 10^{-4}$),
392 which is in perfect LD ($r^2=1$) in the 1KGP EUR population with our sentinel variant
393 (rs13003835), and in moderate LD $r^2=0.4$ in AMR populations. Since this variant was
394 absent from the AMR analysis, probably due to its low frequency, it could not be meta-
395 analyzed. Power calculations revealed that the EUR analysis was underpowered for this
396 variant to achieve genome-wide significance (77.6%, assuming an effect size of 0.46,
397 EAF= 0.0027, and number of cases/controls as shown in the HGI website for B2-EUR).
398 In the cross-population meta-analysis (B2-ALL), rs559179177 obtained a p-value of
399 5.9×10^{-4} .

400 rs77599934 (*DDIAS*) had an AF of 1.1% for the G allele in the nonhospitalized
401 controls (Table 2), in line with the recorded gnomAD AF of 1% in admixed AMR
402 groups. This variant has the potential to be a population-specific variant, given the allele
403 frequencies in other population groups, such as EUR (0% in Finnish, 0.025% in non-
404 Finnish), EAS (0%) and SAS (0.042%), and its greater effect size over AFR populations
405 (Figure 3). Examining the LAI, the G allele occurs at a 10.8% frequency in the African
406 component, while it is almost absent in the Native-American and European. Due to its
407 low MAF, rs77599934 was not analyzed in the COVID-19 HGI B2 cross-population
408 meta-analysis and was only present in the HGI B2 AFR population-specific meta-
409 analysis, precluding the comparison (Figure 3). For this reason, we retrieved the variant
410 with the lowest p-value within a 50 kb region around rs77599934 in the COVID-19
411 HGI cross-population analysis to investigate whether it was in moderate-to-strong LD
412 with our sentinel variant. The variant with the smallest p-value was rs75684040
413 (OR=1.07, 95% CI=1.03-1.12, $p=1.84 \times 10^{-3}$). However, LD calculations using the 1KGP
414 phase 3 dataset indicated that rs77599934 and rs75684040 were poorly correlated

415 ($r^2=0.11$). As for AFR populations, the variant with the lowest p-value was rs138860115
416 ($p=8.3 \times 10^{-3}$), but it was not correlated with the lead SNP of this locus.

417

418 *Cross-population meta-analyses*

419 We carried out two cross-ancestry inverse variance-weighted fixed-effects meta-
420 analyses with the admixed AMR GWAS meta-analysis results to evaluate whether the
421 discovered risk loci replicated when considering other population groups. In doing so,
422 we also identified novel cross-population COVID-19 hospitalization risk loci.

423 First, we combined the SCOURGE Latin American GWAS results with the HGI B2
424 ALL analysis (supplementary Table 11). We refer to this analysis as the SC-HGI_{ALL}
425 meta-analysis. Out of the 40 genome-wide significant loci associated with COVID-19
426 hospitalization in the last HGI release¹, this study replicated 39, and the association was
427 stronger than in the original study in 29 of those (supplementary Table 12). However,
428 the variant rs13003835 located in *BAZ2B* did not replicate (OR=1.00, 95% CI=0.98-
429 1.03, $p=0.644$).

430 In this cross-ancestry meta-analysis, we replicated two associations that were not found
431 in HGIv7, albeit they were sentinel variants in the latest GenOMICC meta-analysis².
432 We found an association at the *CASC20* locus led by the variant rs2876034 (OR=0.95,
433 95% CI=0.93-0.97, $p=2.83 \times 10^{-8}$). This variant is in strong LD with the sentinel variant
434 of that study (rs2326788, $r^2=0.92$), which was associated with critical COVID-19². In
435 addition, this meta-analysis identified the variant rs66833742 near *ZBTB7A* associated
436 with COVID-19 hospitalization (OR=0.94, 95% CI=0.92-0.96, $p=2.50 \times 10^{-8}$). Notably,
437 rs66833742 or its perfect proxy rs67602344 ($r^2=1$) are also associated with upregulation

438 of *ZBTB7A* in whole blood and in esophageal mucosa. This variant was previously
439 associated with COVID-19 hospitalization².

440 In a second analysis, we also explored the associations across the defined admixed
441 AMR, EUR, and AFR ancestral sources by combining through meta-analysis the
442 SCOURGE Latin American GWAS results with the HGI studies in EUR, AFR, and
443 admixed AMR and excluding those from EAS and SAS (supplementary Table 13). We
444 refer to this as the SC-HGI_{3POP} meta-analysis. The association at rs13003835 (*BAZ2B*,
445 OR=1.01, 95% CI=0.98-1.03, p=0.605) was not replicated, and rs77599934 near *DDIAS*
446 could not be assessed, although the association at the *ZBTB7A* locus was confirmed
447 (rs66833742, OR=0.94, 95% CI=0.92-0.96, p=1.89x10⁻⁸). The variant rs76564172
448 located near *CREBBP* also reached statistical significance (OR=1.31, 95% CI=1.25-
449 1.38, p=9.64x10⁻⁹). The sentinel variant of the region linked to *CREBBP* (in the trans-
450 ancestry meta-analysis) was also subjected a Bayesian fine mapping (supplementary
451 Table 6). Eight variants were included in the credible set for the region in chromosome
452 16 (meta-analysis SC-HGI_{3POP}).

453 **Polygenic risk score models**

454 Using the 49 variants associated with disease severity that are shared across populations
455 according to the HGIv7, we constructed a polygenic risk score (PGS) model to assess its
456 generalizability in the admixed AMR (supplementary Table 14). First, we calculated the
457 PGS for the SCOURGE Latin Americans and explored the association with COVID-19
458 hospitalization under a logistic regression model. The PGS model was associated with a
459 1.48-fold increase in COVID-19 hospitalization risk per every PGS standard deviation.
460 It also contributed to explaining a slightly larger variance ($\Delta R^2=1.07\%$) than the
461 baseline model.

462 Subsequently, we divided the individuals into PGS deciles and percentiles to assess
463 their risk stratification. The median percentile among controls was 40, while in cases, it
464 was 63. Those in the top PGS decile exhibited a 2.89-fold (95% CI=2.37-3.54,
465 $p=1.29 \times 10^{-7}$) greater risk compared to individuals in the deciles between 4 and 6
466 (corresponding to a score of the median distribution).

467 We also examined the distribution of PGS scores across a 5-level severity scale to
468 further determine if there was any correspondence between clinical severity and genetic
469 risk. Median PGS scores were lower in the asymptomatic and mild groups, whereas
470 higher median scores were observed in the moderate, severe, and critical patients
471 (Figure 4). We fitted a multinomial model using the asymptomatic class as a reference
472 and calculated the OR for each category (supplementary Table 15), observing that the
473 disease genetic risk was similar among asymptomatic, mild, and moderate patients.
474 Given that the PGS was built with variants associated with critical disease and/or
475 hospitalization and that the categories severe and critical correspond to hospitalized
476 patients, these results underscore the ability of cross-ancestry PGS for risk stratification
477 even in an admixed population.

478

479 **DISCUSSION**

480 We have conducted the largest GWAS meta-analysis of COVID-19 hospitalization in
481 admixed AMR to date. While the genetic risk basis discovered for COVID-19 is largely
482 shared among populations, trans-ancestry meta-analyses on this disease have primarily
483 included EUR samples. This dominance of GWAS in Europeans and the subsequent
484 bias in sample sizes can mask population-specific genetic risks (i.e., variants that are
485 monomorphic in some populations) or be less powered to detect risk variants having

486 higher allele frequencies in population groups other than Europeans. In this sense, after
487 combining data from admixed AMR patients, we found two risk loci that were first
488 discovered in a GWAS of Latin American populations. Interestingly, the sentinel
489 variant rs77599934 in the *DDIAS* gene is a rare coding variant (~1% for allele G) with a
490 large effect on COVID-19 hospitalization that is nearly monomorphic in most of the
491 other populations. This has likely led to its exclusion from the cross-population meta-
492 analyses conducted to date, remaining undetectable.

493 Fine mapping of the region harboring *DDIAS* did not reveal further information about
494 which gene could be the more prone to be causal or about the functional consequences
495 of the risk variant, but our sentinel variant was in strong LD with an eQTL that
496 associated with reduced gene expression of *DDIAS* in the lung. *DDIAS*, known as
497 damage-induced apoptosis suppressor gene, is itself a plausible candidate gene. It has
498 been linked to DNA damage repair mechanisms: research has shown that depletion of
499 *DDIAS* leads to an increase in ATM phosphorylation and the formation of p53-binding
500 protein (53BP1) foci, a known biomarker of DNA double-strand breaks, suggesting a
501 potential role in double-strand break repair²⁰. Interestingly, a study found that infection
502 by SARS-CoV-2 also triggered the phosphorylation of ATM kinase and inhibited repair
503 mechanisms, causing the accumulation of DNA damage²¹. This gene has also been
504 proposed as a potential biomarker for lung cancer after finding that it interacts with
505 STAT3 in lung cancer cells, regulating IL-6^{22,23} and thus mediating inflammatory
506 processes, while another study determined that its blockade inhibited lung cancer cell
507 growth²⁴. Another prioritized gene from this region was *PRCP*, an angiotensinase which
508 shares substrate specificity with ACE2 receptor. It has been positively linked to
509 hypertension and some studies have raised hypotheses on its role in COVID-19
510 progression, particularly in relation to the development of pro-thrombotic events^{25,26}.

511 The risk region found in chromosome 2 harbors more than one gene. The lead variant
512 rs13003835 is located within *BAZ2B*, and it increases the expression of the antisense
513 *BAZ2B* gene in whole blood. *BAZ2B* encodes one of the regulatory subunits of the
514 Imitation switch (ISWI) chromatin remodelers²⁷ constituting the BRF-1/BRF-5
515 complexes with SMARCA1 and SMARCA5, respectively. Interestingly, it was
516 discovered that *lnc-BAZ2B* promotes macrophage activation through regulation of
517 *BAZ2B* expression. Its over-expression resulted in pulmonary inflammation and
518 elevated levels of *MUC5AC* in mice with asthma²⁸. This variant was also an eQTL for
519 *LY75* (encoding lymphocyte antigen 75) in the esophageal mucosa tissue. Lymphocyte
520 antigen 75 is involved in immune processes through antigen presentation in dendritic
521 cells and endocytosis²⁹ and has been associated with inflammatory diseases,
522 representing a compelling candidate for the region. Increased expression of *LY75* has
523 been detected within hours after infection by SARS-CoV-2^{30,31}. It is worth noting that
524 differences in AF for this variant suggest that analyses in AMR populations might be
525 more powered to detect the association, supporting the necessity of population-specific
526 studies.

527 A third novel risk region was observed on chromosome 15 between the *IQGAPI* and
528 *ZNF774* genes, although it did not reach genome-wide significance.

529 Secondary analyses revealed five TWAS-associated genes, some of which have already
530 been linked to severe COVID-19. In a comprehensive multitissue gene expression
531 profiling study³², decreased expression of *CAMP* and *S100A8/S100A9* genes in patients
532 with severe COVID-19 was observed, while another study detected the upregulation of
533 *SCL25A37* among patients with severe COVID-19³³. *SMARCC1* is a subunit of the
534 SWI/SNF chromatin remodeling complex that has been identified as proviral for SARS-
535 CoV-2 and other coronavirus strains through a genome-wide screen³⁴. This complex is

536 crucial for *ACE2* expression and viral entry into the cell³⁵. However, it should be noted
537 that using eQTL mostly from European populations such as those in GTEx could result
538 in reduced power to detect associations.

539 To explore the genetic architecture of the trait among admixed AMR populations, we
540 performed two cross-ancestry meta-analyses including the SCOURGE Latin American
541 cohort GWAS findings. We found that the two novel risk variants were not associated
542 with COVID-19 hospitalization outside the population-specific meta-analysis,
543 highlighting the importance of complementing trans-ancestry meta-analyses with group-
544 specific analyses. Notably, this analysis did not replicate the association at the *DSTYK*
545 locus, which was associated with severe COVID-19 in Brazilian individuals with higher
546 European admixture³⁶. This lack of replication aligns with the initial hypothesis of that
547 study suggesting that the risk haplotype was derived from European populations, as we
548 reduced the weight of this ancestral contribution in our study by excluding those
549 individuals.

550 Moreover, these cross-ancestry meta-analyses pointed to three loci that were not
551 genome-wide significant in the HGIV7 ALL meta-analysis: a novel locus at *CREBBP*
552 and two loci at *ZBTB7A* and *CASC20* that were reported in another meta-analysis.
553 *CREBBP* and *ZBTB7A* achieved a stronger significance when considering only the
554 EUR, AFR, and admixed AMR GIA groups. According to a recent study, elevated
555 levels of the *ZBTB7A* gene promote a quasihomeostatic state between coronaviruses and
556 host cells, preventing cell death by regulating oxidative stress pathways³⁷. This gene is
557 involved in several signaling pathways, such as B and T-cell differentiation³⁸. On a
558 separate note, *CREBBP* encodes the CREB binding protein (CBP), which is involved in
559 transcriptional activation and is known to positively regulate the type I interferon
560 response through virus-induced phosphorylation of IRF-3³⁹. In addition, the

561 CREBP/CBP interaction has been implicated in SARS-CoV-2 infection⁴⁰ via the
562 cAMP/PKA pathway. In fact, cells with suppressed *CREBBP* gene expression exhibit
563 reduced replication of the so-called Delta and Omicron SARS-CoV-2 variants⁴⁰.

564 We developed a cross-population PGS model, which effectively stratified individuals
565 based on their genetic risk and demonstrated consistency with the clinical severity
566 classification of the patients. Only a few polygenic scores were derived from COVID-
567 19 GWAS data. Horowitz et al. (2022)⁴¹ developed a score using 6 and 12 associated
568 variants (PGS ID: PGP000302) and reported an associated OR (top 10% vs rest) of 1.38
569 for risk of hospitalization in European populations, whereas the OR for Latin-American
570 populations was 1.56. Since their sample size and the number of variants included in the
571 PGS were lower, direct comparisons are not straightforward. Nevertheless, our analysis
572 provides the first results for a PRS applied to a relatively large AMR cohort, being of
573 value for future analyses regarding PRS transferability.

574 This study is subject to limitations, mostly concerning sample recruitment and
575 composition. The SCOURGE Latino-American sample size is small, and the GWAS is
576 likely underpowered. Another limitation is the difference in case-control recruitment
577 across sampling regions that, yet controlled for, may reduce the ability to observe
578 significant associations driven by different compositions of the populations. In this
579 sense, the identified risk loci might not replicate in a cohort lacking any of the parental
580 population sources from the three-way admixture. Likewise, we could not explicitly
581 control for socioenvironmental factors that could have affected COVID-19 spread and
582 hospitalization rates, although genetic principal components are known to capture
583 nongenetic factors. Finally, we must acknowledge the lack of a replication cohort. We
584 used all the available GWAS data for COVID-19 hospitalization in admixed AMR in
585 this meta-analysis due to the low number of studies conducted. Therefore, we had no

586 studies to replicate or validate the results. These concerns may be addressed in the
587 future by including more AMR GWAS in the meta-analysis, both by involving diverse
588 populations in study designs and by supporting research from countries in Latin
589 America.

590 This study provides novel insights into the genetic basis of COVID-19 severity,
591 emphasizing the importance of considering host genetic factors by using non-European
592 populations, especially of admixed sources. Such complementary efforts can pin down
593 new variants and increase our knowledge on the host genetic factors of severe COVID-
594 19.

595 **Materials and methods**

596 **GWAS in Latin Americans from SCOURGE**

597 *The SCOURGE Latin American cohort*

598 A total of 3,729 COVID-19-positive cases were recruited across five countries from
599 Latin America (Mexico, Brazil, Colombia, Paraguay, and Ecuador) by 13 participating
600 centers (supplementary Table 1) from March 2020 to July 2021. In addition, we
601 included 1,082 COVID-19-positive individuals recruited between March and December
602 2020 in Spain who either had evidence of origin from a Latin American country or
603 showed inferred genetic admixture between AMR, EUR, and AFR (with < 0.05%
604 SAS/EAS). These individuals were excluded from a previous SCOURGE study that
605 focused on participants with European genetic ancestries¹⁶. We used hospitalization as a
606 proxy for disease severity and defined COVID-19-positive patients who underwent
607 hospitalization as a consequence of the infection as cases and those who did not need
608 hospitalization due to COVID-19 as controls.

609 Samples and data were collected with informed consent after the approval of the Ethics
610 and Scientific Committees from the participating centers and by the Galician Ethics
611 Committee Ref 2020/197. Recruitment of patients from IMSS (in Mexico, City) was
612 approved by the National Committee of Clinical Research from Instituto Mexicano del
613 Seguro Social, Mexico (protocol R-2020-785-082).

614 Samples and data were processed following normalized procedures. The REDCap
615 electronic data capture tool^{42,43}, hosted at Centro de Investigación Biomédica en Red
616 (CIBER) from the Instituto de Salud Carlos III (ISCIII), was used to collect and manage
617 demographic, epidemiological, and clinical variables. Subjects were diagnosed with
618 COVID-19 based on quantitative PCR tests (79.3%) or according to clinical (2.2%) or
619 laboratory procedures (antibody tests: 16.3%; other microbiological tests: 2.2%).

620 *SNP array genotyping*

621 Genomic DNA was obtained from peripheral blood and isolated using the Chemagic
622 DNA Blood 100 kit (PerkinElmer Chemagen Technologies GmbH), following the
623 manufacturer's recommendations.

624 Samples were genotyped with the Axiom Spain Biobank Array (Thermo Fisher
625 Scientific) following the manufacturer's instructions in the Santiago de Compostela
626 Node of the National Genotyping Center (CeGen-ISCIII; <http://www.usc.es/cegen>).
627 This array contains probes for genotyping a total of 757,836 SNPs. Clustering and
628 genotype calling were performed using Axiom Analysis Suite v4.0.3.3 software.

629 *Quality control steps and variant imputation*

630 A quality control (QC) procedure using PLINK 1.9⁴⁴ was applied to both samples and
631 the genotyped SNPs. We excluded variants with a minor allele frequency (MAF) <1%,
632 a call rate <98%, and markers strongly deviating from Hardy-Weinberg equilibrium

633 expectations ($p < 1 \times 10^{-6}$) with mid-p adjustment. We also explored the excess of
634 heterozygosity to discard potential cross-sample contamination. Samples missing >2%
635 of the variants were filtered out. Subsequently, we kept the autosomal SNPs, removed
636 high-LD regions and conducted LD pruning (windows of 1,000 SNPs, with a step size
637 of 80 and a r^2 threshold of 0.1) to assess kinship and estimate the global ancestral
638 proportions. Kinship was evaluated based on IBD values, removing one individual from
639 each pair with $PI_HAT > 0.25$ that showed a Z0, Z1, and Z2 coherent pattern (according
640 to the theoretical expected values for each relatedness level). Genetic principal
641 components (PCs) were calculated with PLINK with the subset of LD pruned variants.

642 Genotypes were imputed with the TOPMed version r2 reference panel (GRCh38) using
643 the TOPMed Imputation Server, and variants with $Rsq < 0.3$ or with $MAF < 1\%$ were
644 filtered out. A total of 4,348 individuals and 10,671,028 genetic variants were included
645 in the analyses.

646 *Genetic admixture estimation*

647 Global genetic inferred ancestry (GIA), referred to the genetic similarity to the used
648 reference individuals, was estimated with ADMIXTURE⁴⁵ v1.3 software following a
649 two-step procedure. First, we randomly sampled 79 European (EUR) and 79 African
650 (AFR) samples from The 1000 Genomes Project (1KGP)⁴⁶ and merged them with the
651 79 Native American (AMR) samples from Mao *et al.*⁴⁷ keeping the biallelic SNPs. LD-
652 pruned variants were selected from this merge using the same parameters as in the QC.
653 We then run an unsupervised analysis with $K=3$ to redefine and homogenize the clusters
654 and to compose a refined reference for the analyses by applying a threshold of $\geq 95\%$ of
655 belonging to a particular cluster. As a result, 20 AFR, 18 EUR, and 38 AMR individuals
656 were removed. The same LD-pruned variants data from the remaining individuals were

657 merged with the SCOURGE Latin American cohort to perform supervised clustering
658 and estimate admixture proportions. A total of 471 samples from the SCOURGE cohort
659 with >80% estimated European GIA were removed to reduce the weight of the
660 European ancestral component, leaving a total of 3,512 admixed Latin American
661 (AMR) subjects for downstream analyses.

662 *Association analysis*

663 The results for the SCOURGE Latin American GWAS were obtained by testing for
664 COVID-19 hospitalization as a surrogate of severity. To accommodate the continuum of
665 GIA in the cohort, we opted for a joint testing of all the individuals as a single study
666 using a mixed regression model, as this approach has demonstrated a greater power and
667 sufficient control of population structure⁴⁸. The SCOURGE cohort consisted of 3,512
668 COVID-19-positive patients: cases (n=1,625) were defined as hospitalized COVID-19
669 patients, and controls (n=1,887) were defined as nonhospitalized COVID-19-positive
670 patients.

671 Logistic mixed regression models were fitted using the SAIGEgds⁴⁹ package in R,
672 which implements the two-step mixed SAIGE⁵⁰ model methodology and the SPA test.
673 Baseline covariables included sex, age (continuous), and the first 10 PCs. To account
674 for potential heterogeneity in the recruitment and hospitalization criteria across the
675 participating countries, we adjusted the models by groups of the recruitment areas
676 classified into six categories: Brazil, Colombia, Ecuador, Mexico, Paraguay, and Spain.
677 This dataset has not been used in any previously published GWAS of COVID-19.

678 **Meta-analysis of Latin American populations**

679 The results of the SCOURGE Latin American cohort were meta-analyzed with the
680 AMR HGI-B2 data, conforming our primary analysis. Summary results from the HGI

681 freeze 7 B2 analysis corresponding to the admixed AMR population were obtained from
682 the public repository (April 8, 2022: <https://www.covid19hg.org/results/r7/>), summing
683 up 3,077 cases and 66,686 controls from seven contributing studies. We selected the B2
684 phenotype definition because it offered more power, and the presence of population
685 controls not ascertained for COVID-19 does not have a drastic impact on the association
686 results.

687 The meta-analysis was performed using an inverse-variance weighting method in
688 METAL⁵¹. The average allele frequency was calculated, and variants with low
689 imputation quality ($R_{sq} < 0.3$) were filtered out, leaving 10,121,172 variants for meta-
690 analysis.

691 Heterogeneity between studies was evaluated with Cochran's Q test. The inflation of
692 results was assessed based on a genomic control (λ).

693 *Replicability of associations*

694 The model-based method MAMBA⁵² was used to calculate the posterior probabilities
695 of replication for each of the lead variants. Variants with $p < 1 \times 10^{-05}$ were clumped and
696 combined with random pruned variants from the 1KGP AMR reference panel. Then,
697 MAMBA was applied to the set of significant and non-significant variants.

698 Each of the lead variants was also tested for association with the main comorbidities in
699 the SCOURGE cohort with logistic regression models (adjusted by the same base
700 covariables as the GWAS).

701 **Definition of the genetic risk loci and putative functional impact**

702 *Definition of lead variant and novel loci*

703 To define the lead variants in the loci that were genome-wide significant, LD-clumping
704 was performed on the meta-analysis data using a threshold $p\text{-value} < 5 \times 10^{-8}$, clump
705 distance=1500 kb, independence set at a threshold $r^2=0.1$ and the SCOURGE cohort
706 genotype data as the LD reference panel. Independent loci were deemed as a novel
707 finding if they met the following criteria: 1) $p\text{-value} < 5 \times 10^{-8}$ in the meta-analysis and $p\text{-value} > 5 \times 10^{-8}$
708 in the HGI B2 ALL meta-analysis or in the HGI B2 AMR and AFR and
709 EUR analyses when considered separately; 2) Cochran's Q-test for heterogeneity of
710 effects is $< 0.05/N_{\text{loci}}$, where N_{loci} is the number of independent variants with $p < 5 \times 10^{-8}$;
711 and 3) the nearest gene has not been previously described in the latest HGIv7 update.

712 *Annotation and initial mapping*

713 Functional annotation was performed with FUMA⁵³ for those variants with a $p\text{-value} < 5 \times 10^{-8}$
714 or in moderate-to-strong LD ($r^2 > 0.6$) with the lead variants, where the LD
715 was calculated from the 1KGP AMR panel. Genetic risk loci were defined by collapsing
716 LD blocks within 250 kb. Then, genes, scaled CADD v1.4 scores, and RegulomeDB
717 v1.1 scores were annotated for the resulting variants with ANNOVAR in FUMA⁵³.
718 Gene-based analysis was also performed using MAGMA⁵⁴ as implemented in FUMA
719 under the SNP-wide mean model using the 1KGP AMR reference panel. Significance
720 was set at a threshold $p < 2.66 \times 10^{-6}$ (which assumes that variants can be mapped to a total
721 of 18,817 genes).

722 FUMA allowed us to perform initial gene mapping by two approaches: (1) positional
723 mapping, which assigns variants to genes by physical distance using 10-kb windows;
724 and (2) eQTL mapping based on GTEx v.8 data from whole blood, lungs, lymphocytes,
725 and esophageal mucosa tissues, establishing a false discovery rate (FDR) of 0.05 to
726 declare significance for variant-gene pairs.

727 Subsequently, to assign the variants to the most likely gene driving the association, we
728 refined the candidate genes by fine mapping the discovered regions

729 *Bayesian fine-mapping*

730 To conduct a Bayesian fine mapping, credible sets for the genetic loci considered novel
731 findings were calculated on the results from each of the three meta-analyses to identify a
732 subset of variants most likely containing the causal variant at the 95% confidence level,
733 assuming that there is a single causal variant and that it has been tested. We used
734 *corrcoverage* (<https://cran.rstudio.com/web/packages/corrcoverage/index.html>) for R to
735 calculate the posterior probabilities of the variant being causal for all variants with a
736 $r^2 > 0.1$ with the leading SNP and within 1 Mb except for the novel variant in
737 chromosome 19, for which we used a window of 0.5 Mb. Variants were added to the
738 credible set until the sum of the posterior probabilities was ≥ 0.95 .

739 *VEP and V2G annotation*

740 We used the Variant-to-Gene (V2G) score to prioritize the genes that were most likely
741 affected by the functional evidence based on expression quantitative trait loci (eQTL),
742 chromatin interactions, in silico functional predictions, and distance between the
743 prioritized variants and transcription start site (TSS), based on data from the Open
744 Targets Genetics portal⁵⁵. Details of the data integration and the weighting of each of
745 the datasets are described with detail here: <https://genetics-docs.opentargets.org/our-approach/data-pipeline>. V2G is a score for ranking the functional genomics evidence
746 that supports the connection between variants and genes (the higher the score the more
747 likely the variant to be functionally implicated on the assigned gene). We used VEP
748 release 111 (URL: <https://www.ensembl.org/info/docs/tools/vep/index.html>; accessed
749

750 April 10, 2024)⁵⁶ to annotate the following: gene symbol, function (exonic, intronic,
751 intergenic, non-coding RNA, etc.), impact, feature type, feature, and biotype.

752 We queried the GWAS Catalog (date of accession: 01/07/2024) for evidence of
753 association of each of the prioritized genes with traits related to lung diseases or
754 phenotypes. Lastly, those which were linked to COVID-19, infection, or lung diseases
755 in the revised literature were classified as “literature evidence”.

756 **Transcription-wide association studies**

757 Transcriptome-wide association studies (TWAS) were conducted using the pretrained
758 prediction models with MASHR-computed effect sizes on GTEx v8 datasets^{57,58}. The
759 results from the Latin American meta-analysis were harmonized and integrated with the
760 prediction models through S-PrediXcan⁵⁹ for lung, whole blood, lymphocyte and
761 esophageal mucosal tissues. Statistical significance was set at p-value<0.05 divided by
762 the number of genes that were tested for each tissue. Subsequently, we leveraged results
763 for all 49 tissues and ran a multitissue TWAS to improve the power for association, as
764 demonstrated recently⁶⁰. TWAS was also performed using recently published gene
765 expression datasets derived from a cohort of African Americans, Puerto Ricans, and
766 Mexican Americans (GALA II-SAGE)⁶¹.

767

768 **Cross-population meta-analyses**

769 We conducted two additional meta-analyses to investigate the ability of combining
770 populations to replicate our discovered risk loci. This methodology enabled the
771 comparison of effects and the significance of associations in the novel risk loci between
772 the results from analyses that included or excluded other population groups.

773 The first meta-analysis comprised the five populations analyzed within HGI (B2-ALL).
774 Additionally, to evaluate the three GIA components within the SCOURGE Latin
775 American cohort⁶², we conducted a meta-analysis of the admixed AMR, EUR, and AFR
776 cohorts (B2). All summary statistics were retrieved from the HGI repository. We
777 applied the same meta-analysis methodology and filters as in the admixed AMR meta-
778 analysis.

779 **Cross-population Polygenic Risk Score**

780 A polygenic risk score (PGS) for critical COVID-19 was derived by combining the
781 variants associated with hospitalization or disease severity that have been discovered to
782 date. We curated a list of lead variants that were 1) associated with either severe disease
783 or hospitalization in the latest HGIv7 release¹ (using the hospitalization weights) or 2)
784 associated with severe disease in the latest GenOMICC meta-analysis² that were not
785 reported in the latest HGI release. A total of 48 markers were used in the PGS model
786 (see supplementary Table 13) since two variants were absent from our study.

787 Scores were calculated and normalized for the SCOURGE Latin American cohort with
788 PLINK 1.9. This cross-ancestry PGS was used as a predictor for hospitalization
789 (COVID-19-positive patients who were hospitalized *vs.* COVID-19-positive patients
790 who did not necessitate hospital admission) by fitting a logistic regression model.
791 Prediction accuracy for the PGS was assessed by performing 500 bootstrap resamples of
792 the increase in the pseudo-R-squared. We also divided the sample into deciles and
793 percentiles to assess risk stratification. The models were fit for the dependent variable
794 adjusting for sex, age, the first 10 PCs, and the sampling region (in the Admixed AMR
795 cohort) with and without the PGS, and the partial pseudo-R² was computed and
796 averaged among the resamples.

797 A clinical severity scale was used in a multinomial regression model to further evaluate
798 the power of this cross-ancestry PGS for risk stratification. These severity strata were
799 defined as follows: 0) asymptomatic; 1) mild, that is, with symptoms, but without
800 pulmonary infiltrates or need of oxygen therapy; 2) moderate, that is, with pulmonary
801 infiltrates affecting <50% of the lungs or need of supplemental oxygen therapy; 3)
802 severe disease, that is, with hospital admission and $\text{PaO}_2 < 65$ mmHg or $\text{SaO}_2 < 90\%$,
803 $\text{PaO}_2/\text{FiO}_2 < 300$, $\text{SaO}_2/\text{FiO}_2 < 440$, dyspnea, respiratory frequency ≥ 22 bpm, and
804 infiltrates affecting >50% of the lungs; and 4) critical disease, that is, with an admission
805 to the ICU or need of mechanical ventilation (invasive or noninvasive).

806 **Data availability**

807 Summary statistics from the SCOURGE Latin American GWAS and the analysis scripts
808 are available from the public repository [https://github.com/CIBERER/Scourge-](https://github.com/CIBERER/Scourge-COVID19)
809 COVID19.

810

811 **Funding**

812 Instituto de Salud Carlos III (COV20_00622 to A.C., COV20/00792 to M.B.,
813 COV20_00181 to C.A., COV20_1144 to M.A.J.S. and A.F.R., PI20/00876 to C.F.);
814 European Union (ERDF) ‘A way of making Europe’. Fundación Amancio Ortega,
815 Banco de Santander (to A.C.), Estrella de Levante S.A. and Colabora Mujer Association
816 (to E.G.-N.) and Obra Social La Caixa (to R.B.); Agencia Estatal de Investigación
817 (RTC-2017-6471-1 to C.F.), Cabildo Insular de Tenerife (CGIEU0000219140
818 ‘Apuestas científicas del ITER para colaborar en la lucha contra la COVID-19’ to C.F.)
819 and Fundación Canaria Instituto de Investigación Sanitaria de Canarias (PIFIISC20/57
820 to C.F.).

821 SD-DA was supported by a Xunta de Galicia predoctoral fellowship.

822 **Author contributions**

823 Study design: RC, AC, CF. Data collection: SCOURGE cohort group. Data analysis:
824 SD-DA, RC, ADL, CF, JML-S. Interpretation: SD-DA, RC, ADL. Drafting of the
825 manuscript: SD-DA, RC, ADL, CF, AR-M, AC. Critical revision of the manuscript:
826 SD-DA, RC, ADL, AC, CF, JAR, AR-M, and PL. Approval of the final version of the
827 publication: all coauthors.

828 **Acknowledgments**

829 The contribution of the Centro Nacional de Genotipado (CEGEN) and Centro de
830 Supercomputación de Galicia (CESGA) for funding this project by providing
831 supercomputing infrastructures is also acknowledged. The authors are also particularly
832 grateful for the supply of material and the collaboration of patients, health professionals
833 from participating centers and biobanks. Namely, Biobanc-Mur, and biobanks of the
834 Complejo Hospitalario Universitario de A Coruña, Complejo Hospitalario
835 Universitario de Santiago, Hospital Clínico San Carlos, Hospital La Fe, Hospital
836 Universitario Puerta de Hierro Majadahonda—Instituto de Investigación Sanitaria
837 Puerta de Hierro—Segovia de Arana, Hospital Ramón y Cajal, IDIBGI, IdISBa, IIS
838 Biocruces Bizkaia, IIS Galicia Sur. Also biobanks of the Sistema de Salud de Aragón,
839 Sistema Sanitario Público de Andalucía, and Banco Nacional de ADN.

840

841 **References**

842 1. Initiative, T. C.-19 H. G. & Ganna, A. A second update on mapping the human genetic
843 architecture of COVID-19. 2022.12.24.22283874 Preprint at
844 <https://doi.org/10.1101/2022.12.24.22283874> (2023).

- 845 2. GWAS and meta-analysis identifies 49 genetic variants underlying critical COVID-19 |
846 Nature. <https://www.nature.com/articles/s41586-023-06034-3>.
- 847 3. Niemi, M. E. K. *et al.* Mapping the human genetic architecture of COVID-19. *Nature*
848 **600**, 472–477 (2021).
- 849 4. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161–164
850 (2016).
- 851 5. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The Missing Diversity in Human Genetic
852 Studies. *Cell* **177**, 26–31 (2019).
- 853 6. Li, Y. R. & Keating, B. J. Trans-ethnic genome-wide association studies: advantages and
854 challenges of mapping in diverse populations. *Genome Med.* **6**, 91 (2014).
- 855 7. Rosenberg, N. A. *et al.* Genome-wide association studies in diverse populations. *Nat.*
856 *Rev. Genet.* **11**, 356–366 (2010).
- 857 8. Kwok, A. J., Mentzer, A. & Knight, J. C. Host genetics and infectious disease: new tools,
858 insights and translational opportunities. *Nat. Rev. Genet.* **22**, 137–153 (2021).
- 859 9. Karlsson, E. K., Kwiatkowski, D. P. & Sabeti, P. C. Natural selection and infectious
860 disease in human populations. *Nat. Rev. Genet.* **15**, 379–393 (2014).
- 861 10. Namkoong, H. *et al.* DOCK2 is involved in the host genetics and biology of severe
862 COVID-19. *Nature* **609**, 754–760 (2022).
- 863 11. Bastard, P. *et al.* A loss-of-function IFNAR1 allele in Polynesia underlies severe viral
864 diseases in homozygotes. *J. Exp. Med.* **219**, e20220028 (2022).
- 865 12. Duncan, C. J. A. *et al.* Life-threatening viral disease in a novel form of autosomal
866 recessive IFNAR2 deficiency in the Arctic. *J. Exp. Med.* **219**, e20212427 (2022).
- 867 13. Peterson, R. E. *et al.* Genome-wide Association Studies in Ancestrally Diverse
868 Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell* **179**, 589–603
869 (2019).
- 870 14. Mester, R. *et al.* Impact of cross-ancestry genetic architecture on GWAS in admixed
871 populations. 2023.01.20.524946 Preprint at <https://doi.org/10.1101/2023.01.20.524946>
872 (2023).
- 873 15. Tractor uses local ancestry to enable the inclusion of admixed individuals in GWAS and
874 to boost power | Nature Genetics. <https://www.nature.com/articles/s41588-020-00766-y>.
- 875 16. Cruz, R. *et al.* Novel genes and sex differences in COVID-19 severity. *Hum. Mol. Genet.*
876 **31**, 3789–3806 (2022).
- 877 17. Degenhardt, F. *et al.* Detailed stratified GWAS analysis for severe COVID-19 in four
878 European populations. *Hum. Mol. Genet.* **31**, 3945–3966 (2022).
- 879 18. Whole-genome sequencing reveals host factors underlying critical COVID-19 | Nature.
880 <https://www.nature.com/articles/s41586-022-04576-6>.
- 881 19. Kanai, M. *et al.* A second update on mapping the human genetic architecture of
882 COVID-19. *Nature* **621**, E7–E26 (2023).

- 883 20. Evolution-based screening enables genome-wide prioritization and discovery of DNA
884 repair genes | PNAS. <https://www.pnas.org/doi/full/10.1073/pnas.1906559116>.
- 885 21. Gioia, U. *et al.* SARS-CoV-2 infection induces DNA damage, through CHK1 degradation
886 and impaired 53BP1 recruitment, and cellular senescence. *Nat. Cell Biol.* **25**, 550–564 (2023).
- 887 22. Im, J.-Y. *et al.* DDIAS promotes STAT3 activation by preventing STAT3 recruitment to
888 PTPRM in lung cancer cells. *Oncogenesis* **9**, 1–11 (2020).
- 889 23. Im, J.-Y., Kang, M.-J., Kim, B.-K. & Won, M. DDIAS, DNA damage-induced apoptosis
890 suppressor, is a potential therapeutic target in cancer. *Exp. Mol. Med.* 1–7 (2023)
891 doi:10.1038/s12276-023-00974-6.
- 892 24. Human Noxin is an anti-apoptotic protein in response to DNA damage of A549
893 non-small cell lung carcinoma - Won - 2014 - International Journal of Cancer - Wiley Online
894 Library. <https://onlinelibrary.wiley.com/doi/10.1002/ijc.28600>.
- 895 25. Angeli, F. *et al.* The spike effect of acute respiratory syndrome coronavirus 2 and
896 coronavirus disease 2019 vaccines on blood pressure. *Eur. J. Intern. Med.* **109**, 12–21 (2023).
- 897 26. Silva-Aguiar, R. P. *et al.* Role of the renin-angiotensin system in the development of
898 severe COVID-19 in hypertensive patients. *Am. J. Physiol.-Lung Cell. Mol. Physiol.* **319**, L596–
899 L602 (2020).
- 900 27. Li, Y. *et al.* The emerging role of ISWI chromatin remodeling complexes in cancer. *J.*
901 *Exp. Clin. Cancer Res.* **40**, 346 (2021).
- 902 28. Xia, L. *et al.* Inc-BAZ2B promotes M2 macrophage activation and inflammation in
903 children with asthma through stabilizing BAZ2B pre-mRNA. *J. Allergy Clin. Immunol.* **147**, 921-
904 932.e9 (2021).
- 905 29. The Dendritic Cell Receptor for Endocytosis, Dec-205, Can Recycle and Enhance
906 Antigen Presentation via Major Histocompatibility Complex Class II-Positive Lysosomal
907 Compartments | Journal of Cell Biology | Rockefeller University Press.
908 [https://rupress.org/jcb/article/151/3/673/21295/The-Dendritic-Cell-Receptor-for-Endocytosis-](https://rupress.org/jcb/article/151/3/673/21295/The-Dendritic-Cell-Receptor-for-Endocytosis-Dec)
909 Dec.
- 910 30. Sims, A. C. *et al.* Release of Severe Acute Respiratory Syndrome Coronavirus Nuclear
911 Import Block Enhances Host Transcription in Human Lung Cells. *J. Virol.* **87**, 3885–3902 (2013).
- 912 31. A Network Integration Approach to Predict Conserved Regulators Related to
913 Pathogenicity of Influenza and SARS-CoV Respiratory Viruses | PLOS ONE.
914 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0069374>.
- 915 32. Gómez-Carballa, A. *et al.* A multi-tissue study of immune gene expression profiling
916 highlights the key role of the nasal epithelium in COVID-19 severity. *Environ. Res.* **210**, 112890
917 (2022).
- 918 33. Policard, M., Jain, S., Rego, S. & Dakshanamurthy, S. Immune characterization and
919 profiles of SARS-CoV-2 infected patients reveals potential host therapeutic targets and SARS-
920 CoV-2 oncogenesis mechanism. *Virus Res.* **301**, 198464 (2021).
- 921 34. Wei, J. *et al.* Genome-wide CRISPR Screens Reveal Host Factors Critical for SARS-CoV-2
922 Infection. *Cell* **184**, 76-91.e13 (2021).

- 923 35. Wei, J. *et al.* Pharmacological disruption of mSWI/SNF complex activity restricts SARS-
924 CoV-2 infection. *Nat. Genet.* **55**, 471–483 (2023).
- 925 36. Pereira, A. C. *et al.* Genetic risk factors and COVID-19 severity in Brazil: results from
926 BRACOVID study. *Hum. Mol. Genet.* **31**, 3021–3031 (2022).
- 927 37. Zhu, X. *et al.* ZBTB7A promotes virus-host homeostasis during human coronavirus 229E
928 infection. *Cell Rep.* **41**, 111540 (2022).
- 929 38. Gupta, S. *et al.* Emerging role of ZBTB7A as an oncogenic driver and transcriptional
930 repressor. *Cancer Lett.* **483**, 22–34 (2020).
- 931 39. Yoneyama, M. *et al.* Direct triggering of the type I interferon system by virus infection:
932 activation of a transcription factor complex containing IRF-3 and CBP/p300. *EMBO J.* **17**, 1087–
933 1095 (1998).
- 934 40. Yang, Q. *et al.* SARS-CoV-2 infection activates CREB/CBP in cellular cyclic AMP-
935 dependent pathways. *J. Med. Virol.* **95**, e28383 (2023).
- 936 41. Horowitz, J. E. *et al.* Genome-wide analysis provides genetic evidence that ACE2
937 influences COVID-19 risk and yields risk scores associated with severe disease. *Nat. Genet.* **54**,
938 382–392 (2022).
- 939 42. Harris, P. A. *et al.* Research electronic data capture (REDCap)—A metadata-driven
940 methodology and workflow process for providing translational research informatics support. *J.*
941 *Biomed. Inform.* **42**, 377–381 (2009).
- 942 43. Harris, P. A. *et al.* The REDCap consortium: Building an international community of
943 software platform partners. *J. Biomed. Inform.* **95**, 103208 (2019).
- 944 44. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-
945 Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- 946 45. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in
947 unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
- 948 46. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74
949 (2015).
- 950 47. Mao, X. *et al.* A Genomewide Admixture Mapping Panel for Hispanic/Latino
951 Populations. *Am. J. Hum. Genet.* **80**, 1171–1178 (2007).
- 952 48. Wojcik, G. L. *et al.* Genetic analyses of diverse populations improves discovery for
953 complex traits. *Nature* **570**, 514–518 (2019).
- 954 49. Zheng, X. & Davis, J. W. SAIGEgds—an efficient statistical tool for large-scale PheWAS
955 with mixed models. *Bioinformatics* **37**, 728–730 (2021).
- 956 50. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample
957 relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
- 958 51. METAL: fast and efficient meta-analysis of genomewide association scans |
959 Bioinformatics | Oxford Academic.
960 <https://academic.oup.com/bioinformatics/article/26/17/2190/198154>.

- 961 52. McGuire, D. *et al.* Model-based assessment of replicability for genome-wide
962 association meta-analysis. *Nat. Commun.* **12**, 1964 (2021).
- 963 53. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and
964 annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
- 965 54. MAGMA: Generalized Gene-Set Analysis of GWAS Data | PLOS Computational Biology.
966 <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004219>.
- 967 55. Ghossaini, M. *et al.* Open Targets Genetics: systematic identification of trait-
968 associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* **49**,
969 D1311–D1320 (2021).
- 970 56. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
- 971 57. Barbeira, A. N. *et al.* Exploiting the GTEx resources to decipher the mechanisms at
972 GWAS loci. *Genome Biol.* **22**, 49 (2021).
- 973 58. Barbeira, A. N. *et al.* GWAS and GTEx QTL integration. Zenodo
974 <https://doi.org/10.5281/zenodo.3518299> (2019).
- 975 59. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene
976 expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825 (2018).
- 977 60. Barbeira, A. N. *et al.* Integrating predicted transcriptome from multiple tissues
978 improves association detection. *PLOS Genet.* **15**, e1007889 (2019).
- 979 61. Kachuri, L. *et al.* Gene expression in African Americans, Puerto Ricans and Mexican
980 Americans reveals ancestry-specific patterns of genetic architecture. *Nat. Genet.* **55**, 952–963
981 (2023).
- 982 62. Genome-wide patterns of population structure and admixture among Hispanic/Latino
983 populations | PNAS. [https://www.pnas.org/doi/10.1073/pnas.0914618107?url_ver=Z39.88-](https://www.pnas.org/doi/10.1073/pnas.0914618107?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%20pubmed)
984 [2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%20pubmed](https://www.pnas.org/doi/10.1073/pnas.0914618107?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%20pubmed).
- 985
- 986
- 987
- 988
- 989
- 990
- 991
- 992

993

994

995

996

997

998 **Table 1. Demographic characteristics of the SCOURGE Latin American cohort.**

Variable	Non Hospitalized N = 1,887	Hospitalized N = 1,625
Age – mean years ± SD	39.1 ± 11.9	54.1 ± 14.5
Sex - N (%)		
Female (%)	1253 (66.4)	668 (41.1)
GIA* – % mean ±SD		
European	54.4 ± 16.2	39.4 ± 20.7
African	15.3 ± 12.7	9.1 ± 11.6
Native American	30.3 ± 19.8	51.3 ± 26.5
Comorbidities - N (%)		
Vascular/endocrinological	488 (25.9)	888 (64.5)
Cardiac	60 (3.2)	151 (9.3)
Nervous	15 (0.8)	61 (3.8)
Digestive	14 (0.7)	33 (2.0)
Onco-hematological	21 (1.1)	48 (3.00)
Respiratory	76 (4.0)	118 (7.3)

999 *Global genetic inferred ancestry.

1000 **Table 2. Lead independent variants in the admixed AMR GWAS meta-analysis.**

SNP rsID	chr:pos	EA	NEA	OR (95% CI)	P-value	EAF	EAF	Nearest gene	Mamba
						cases	controls		PIP
rs13003835	2:159407982	T	C	1.20 (1.12-1.27)	3.66E-08	0.563	0.429	BAZ2B	0.30
rs35731912	3:45848457	T	C	1.65 (1.47-1.85)	6.30E-17	0.087	0.056	LZTFL1	0.95

<i>rs2477820</i>	6:41535254	A	T	0.84 (0.79-0.89)	1.89E-08	0.453	0.517	<i>FOXP4-ASI</i>	0.18
<i>rs77599934</i>	11:82906875	G	A	2.27 (1.7-3.04)	2.26E-08	0.016	0.011	<i>DDIAS</i>	0.95

1001 EA: effect allele; NEA: noneffect allele; EAF: effect allele frequency in the SCOURGE study.

1002

1003

1004

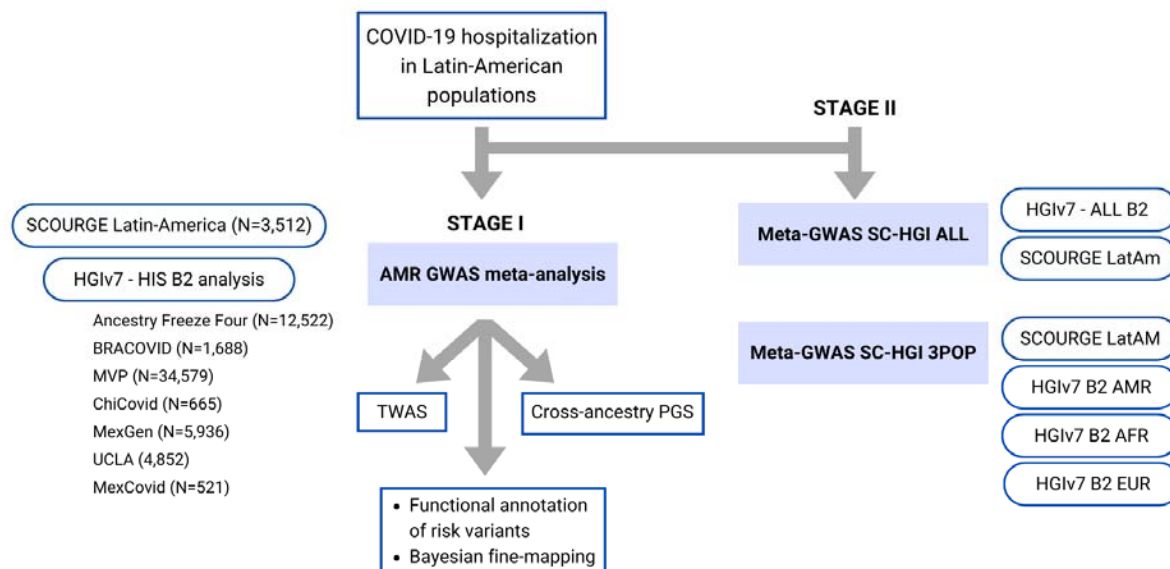
1005 **Table 3. Novel variants in the SC-HGI_{ALL} and SC-HGI_{3POP} meta-analyses (with**

1006 **respect to HGIv7). Independent signals after LD clumping.**

<i>SNP rsID</i>	<i>chr:pos</i>	<i>EA</i>	<i>NEA</i>	<i>OR (95% CI)</i>	<i>P-value</i>	<i>Nearest gene</i>	<i>Analysis</i>
<i>rs76564172</i>	16:3892266	T	G	1.31 (1.19-1.44)	9.64E-09	<i>CREBBP</i>	SC-HGI _{3POP}
<i>rs66833742</i>	19:4063488	T	C	0.94 (0.92-0.96)	1.89E-08	<i>ZBTB7A</i>	SC-HGI _{3POP}
<i>rs66833742</i>	19:4063488	T	C	0.94 (0.92-0.96)	2.50E-08	<i>ZBTB7A</i>	SC-HGI _{ALL}
<i>rs2876034</i>	20:6492834	A	T	0.95 (0.93-0.97)	2.83E-08	<i>CASC20</i>	SC-HGI _{ALL}

1007 EA: effect allele; NEA: non-effect allele.

1008 **Figure 1. Flow chart of this study.**



1009

1010

1011 **Figure 2. A) Manhattan plot for the admixed AMR GWAS meta-analysis.**

1012 Probability thresholds at $p=5 \times 10^{-8}$ and $p=5 \times 10^{-5}$ are indicated by the horizontal lines.

1013 Genome-wide significant associations with COVID-19 hospitalizations were found on

1014 chromosome 2 (within *BAZ2B*), chromosome 3 (within *LZTFL1*), chromosome 6

1015 (within *FOXP4*), and chromosome 11 (within *DDIAS*). A Quantile-Quantile plot is

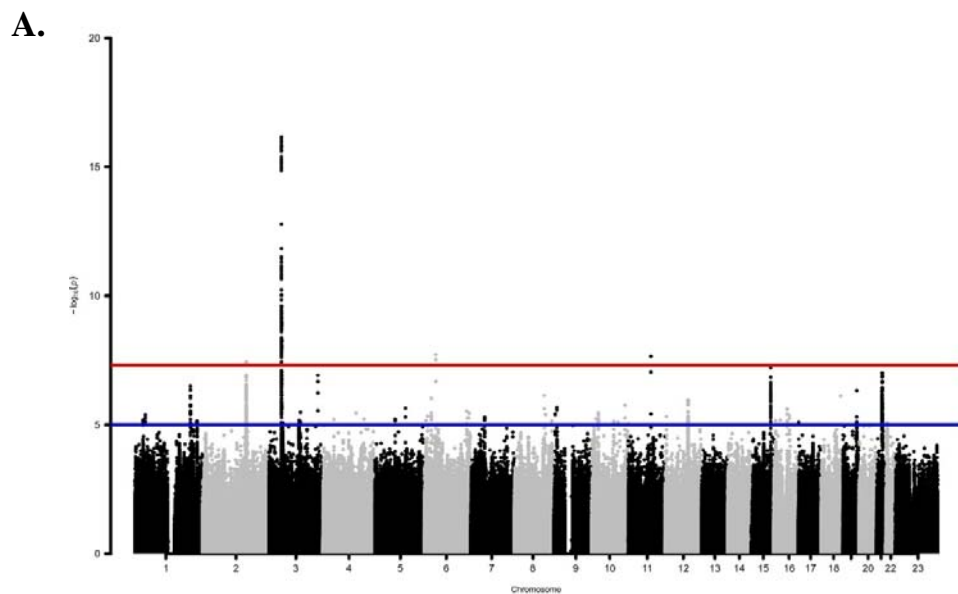
1016 shown in supplementary Figure 2. B) **Regional association plots for rs1003835 at**

1017 **chromosome 2 and rs77599934 at chromosome 11; C) Allele frequency distribution**

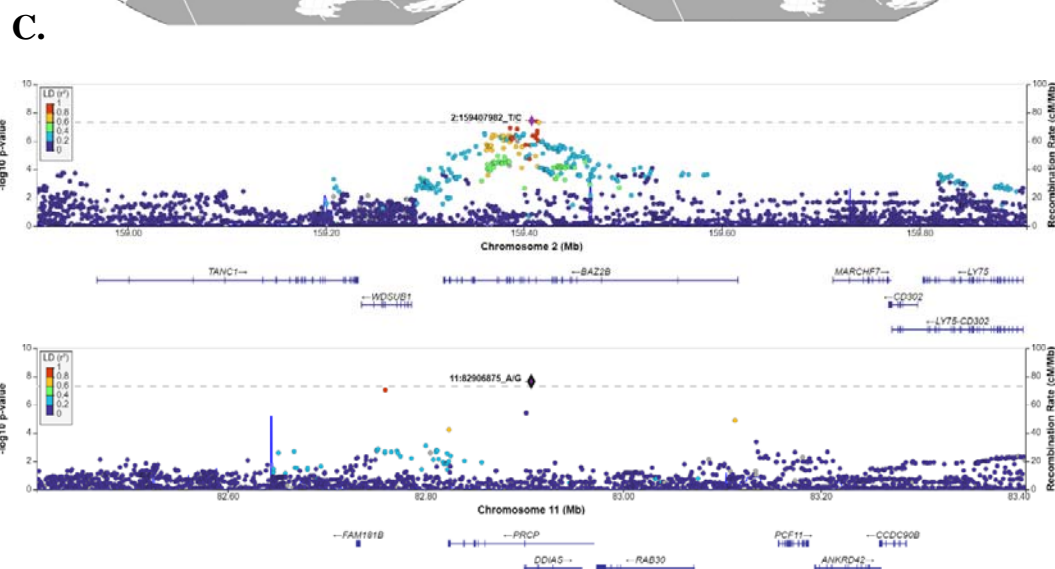
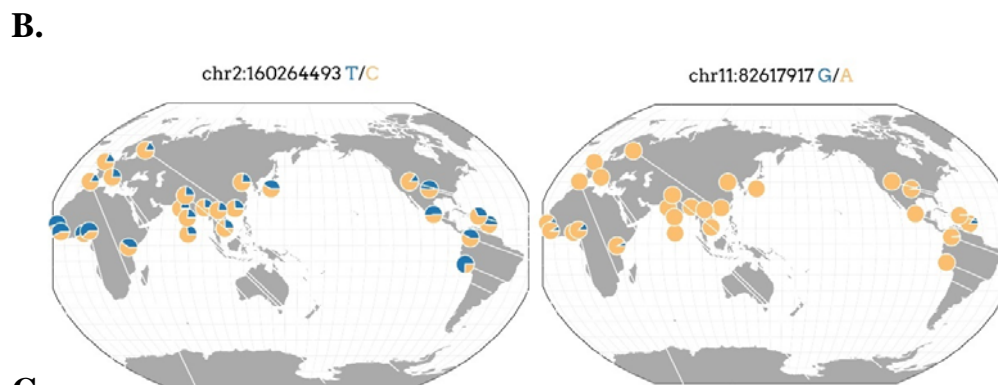
1018 **across the 1000 Genomes Project populations for the lead variants rs1003835 and**

1019 **rs77599934.** Retrieved from *The Geography of Genetic Variants Web* or GGv.

1020



1021



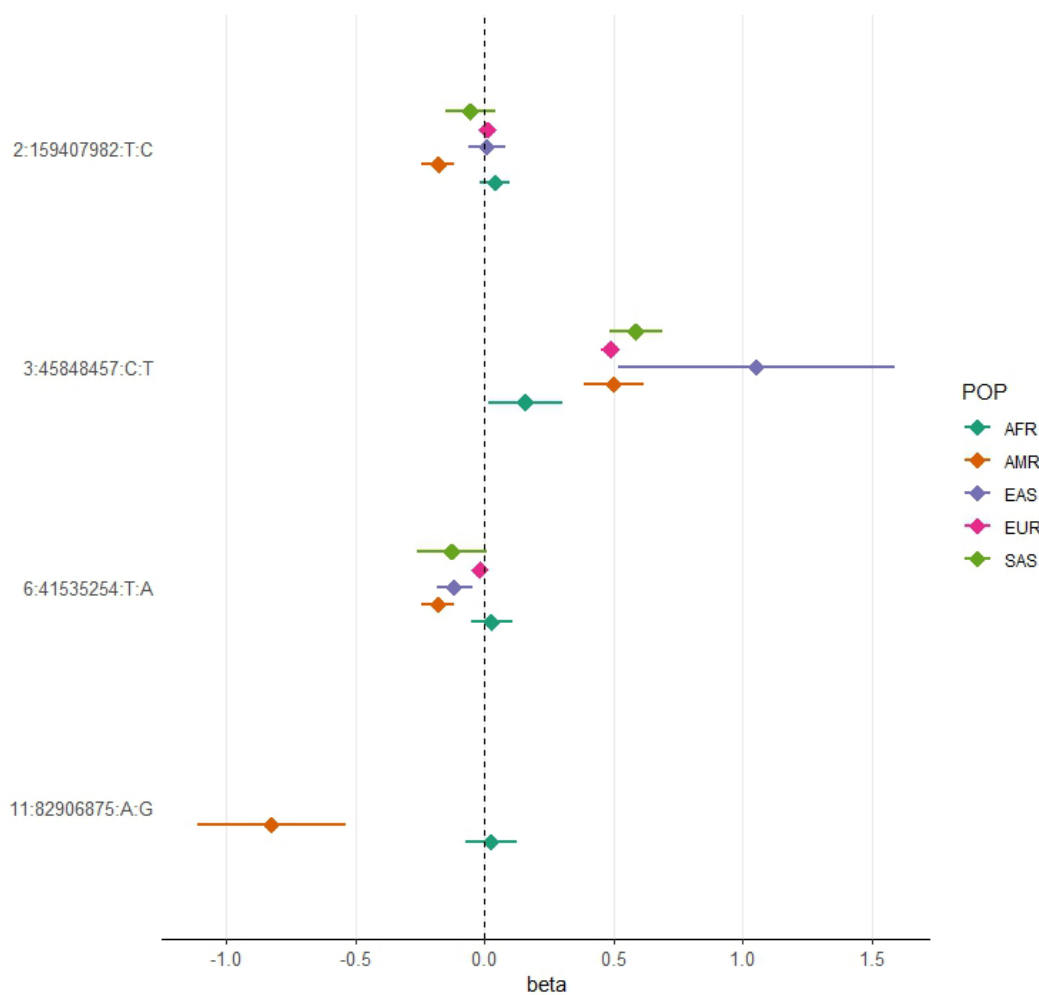
1024

1025

1026

1027 **Figure 3. Forest plot showing effect sizes and the corresponding confidence**
1028 **intervals for the sentinel variants identified in the AMR meta-analysis across**
1029 **populations.** All beta values with their corresponding CIs were retrieved from the B2
1030 population-specific meta-analysis from the HGI v7 release, except for AMR, for which
1031 the beta value and IC from the HGI_{AMR}-SCOURGE meta-analysis are represented.

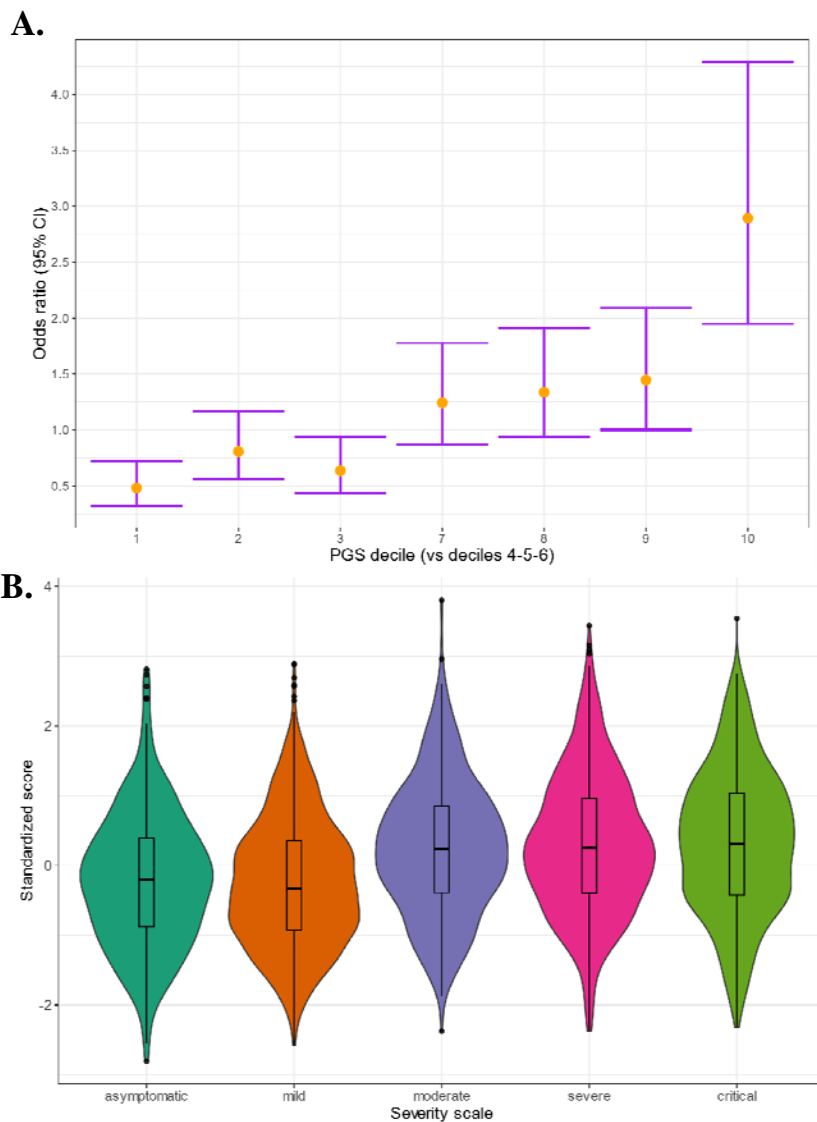
1032



1033

1034

1035 **Figure 4. (A) Polygenic risk stratified by PGS deciles comparing each risk group**
1036 **against the lowest risk group (OR-95% CI); (B) Distribution of the PGS scores in**
1037 **each of the severity scale classes . 0-Asymptomatic, 1-Mild disease, 2-Moderate**
1038 **disease, 3-Severe disease, 4-Critical disease.**



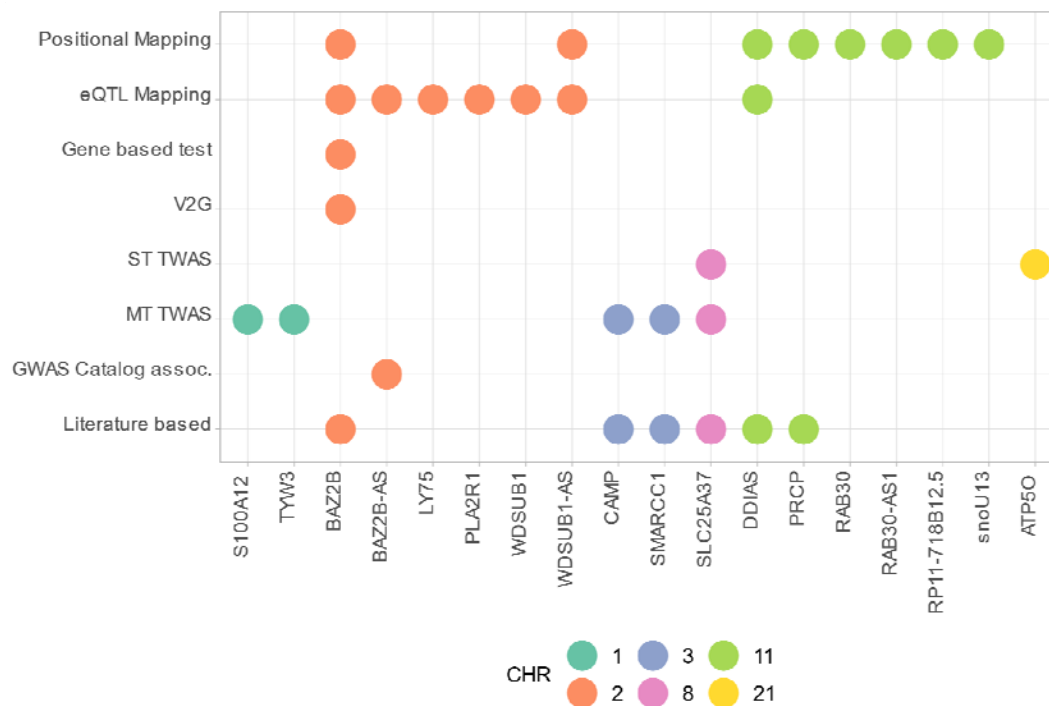
1039

1040

1041

1042

1043 **Figure 5. Summary of the results from gene prioritization strategies used for**
 1044 **genetic associations in AMR populations. GWAS catalog association for *BAZ2B-AS***
 1045 **was with FEV/FCV ratio. Literature based evidence is further explored in discussion.**



1046