[TITLE]

New susceptibility loci for severe COVID-19 by detailed GWAS analysis in European populations

[AUTHORLIST]

Frauke Degenhardt[1,*], David Ellinghaus[1,2,*], Simonas Juzenas[1,3,*], Jon Lerga-Jaso[4,*], Mareike Wendorff[1,*], Douglas Maya-Miles[5,6,7,*], Florian Uellendahl-Werth[1], Hesham ElAbd[1], Jatin Arora[8,9,10,11,12], Onur Özer[13,14], Ole Bernt Lenning[15,16], Ronny Myhre[17], May Sissel Vadla[16,18], Lars Wienbrandt[1], Aaron Blandino Ortiz[19], Adolfo de Salazar[20,21], Adolfo Garrido Chercoles[22,22], Adriana Palom[23,24], Agustín Ruiz[25,26], Alberto Mantovani[27,28], Alberto Zanella[29,30], Aleksander Rygh Holten[31,32], Alena Mayer[33], Alessandra Bandera[29,30], Alessandro Cherubini[30], Alessandro Protti[27,28], Alessio Aghemo[27,28], Alessio Gerussi[34,34,35,35], Alexander Popov[36], Alfredo Ramirez[37,38,38,39,40,41], Alice Braun[33], Almut Nebel[1], Ana Barreira[24], Ana Lleo[27,28], Ana Teles[13,14], Anders Benjamin Kildal[42], Andrea Biondi[43], Andrea Ganna[44], Andrea Gori[30,45], Andreas Glück[46], Andreas Lind[47], Anke Hinney[48], Anna Carreras Nolla[49], Anna Ludovica Fracanzani[29,30], Annalisa Cavallero[50], Anne Ma Dyrhol-Riise[32,51], Antonella Ruello[52], Antonio Julià[23], Antonio Muscatello[30], Antonio Pesenti[29,30], Antonio Voza[27,28], Ariadna Rando-Segura[53,54], Aurora Solier[55], Beatriz Cortes[49], Beatriz Mateos[6,56], Beatriz Nafria-Jimenez[22], Benedikt Schaefer[57,58], Björn Jensen[59], Carla Bellinghausen[60], Carlo Maj[61], Carlos Ferrando[62,63], Carmen de la Horra[5,7,64,65,66], Carmen Quereda[67], Carsten Skurk[33], Charlotte Thibeault[33], Chiara Scollo[68], Christian Herr[69], Christoph D. Spinner[70], Christoph Lange[71,72,73], Cinzia Hu[30], Clara Lehmann[74,75,76], Claudio Cappadona[28], Clinton Azuure[13,14], COVICAT study group, Covid-[19] Aachen Study (COVAS), Cristiana Bianco[30], Cristina Sancho[77,77], Dag Arne Lihaug Hoff[78,79], Daniela Galimberti[29,30], Daniele Prati[30], David Haschka[80], David Jiménez[55], David Pestaña[81], David Toapanta[63], Eike M. Wacker[1], Elena Azzolini[27,28], Elio Scarpini[29,30], Elisa T. Helbig[33], Eloisa Urrechaga[82,83], Elvezia Maria Paraboschi[27,28], Emanuele Pontali[84], Enric Reverter[63], Enrique J. Calderón[5,7,64,65,66], Enrique Navas[67], Erik Solligård[85,86], Ernesto Contro[87], Eunate Arana[88], Federico Garcia[20,21], Félix García Sánchez[89], Ferruccio Ceriotti[30], Filippo Martinelli-Boneschi[90,91], Flora Peyvandi[92,93], Florian Kurth[33,94], Francesco Blasi[95,96], Francesco Malvestiti[29], Francisco J. Medrano[7,64,65,66,97], Francisco Mesonero[6,56], Francisco Rodriguez-Frias[6,23,54,98], Frank Hanses[99,100], Fredrik Müller[32,47], Giacomo Bellani[101,102], Giacomo Grasselli[29,30], Gianni Pezzoli[103], Giorgio Costantino[29,30], Giovanni Albano[52], Giuseppe Bellelli[104,105], Giuseppe Citerio[104,106], Giuseppe Foti[101,107], Giuseppe Lamorte[30], Holger Neb[108], Ilaria My[27], Ingo Kurth[109], Isabel Hernández[25,26], Isabell Pink[110], Itziar de Rojas[25,26], Iván Galván-Femenia[49], Jan C. Holter[32,47], Jan Egil Afset[111,112], Jan Heyckendorf[71,72,73], Jan Kristian Damås[113,114], Jan Rybniker[74,75,115,116], Janine Altmüller[117], Javier Ampuero[5,7,64,118], Jesus M. Banales[6,119,120,121], Joan Ramon Badia[122], Joaquin Dopazo[123], Jochen Schneider[70], Johannes R. Hov[32,124,125,126], Jonas Bergan[127], Jordi Barretina[128], Jörn Walter[129], Jose Hernández Quero[20,130], Josune Goikoetxea[131,131], Juan Delgado[5,7,64,65,66], Juan M. Guerrero[5,7,64], Julia Fazaal[132], Julia Kraft[33], Julia Schröder[132], Kari Risnes[114,133], Karina Banasik[2], Karl Erik Müller[134], Karoline I. Gaede[135,135,136,137], Koldo Garcia-Etxebarria[6,121,121], Kristian Tonby[32,138], Lars Heggelund[134,139], Laura Izquierdo-Sanchez[6,121,140], Laura Rachele Bettini[43], Lauro Sumoy[128], Leif Erik Sander[33], Lena J. Lippert[33], Leonardo Terranova[30], Lindokuhle Nkambule[141,142], Lisa Knopp[59], Lise Tuset Gustad[85,143], Lucia Garbarino[144], Luigi Santoro[30], Luis Téllez[6,56], Luisa Roade[6,24,54], Mahnoosh Ostadreza[30], Maider Intxausti[77,77], Malte C. Rühlemann[1,145], Manolis Kogevinas[65,146,147,148], Mar Riveiro-Barciela[6,24,54], Marc M. Berger[149], Mari E.K. Niemi[44], María A. Gutiérrez-Stampa[150,150], Maria Grazia Valsecchi[151], María Hernandez-Tejero[63], Maria J.G.T. Vehreschild[152], Maria Manunta[30], Mariella D'Angiò[43], Marina Cazzaniga[153], Marit M Grimsrud[125,126], Marit M. Grimsrud[32], Markus Cornberg[154], Markus M. Nöthen[132], Marta Marquié[25,26], Massimo Castoldi[52], Mattia Cordioli[44],

Maurizio Cecconi[27,28], Mauro D'Amato[120,155], Max Augustin[74,75,76], Melissa Tomasi[30], Mercè Boada[25,26], Michael Dreher[156], Michael J. Seilmaier[157], Michael Joannidis[158], Michael Wittig[1], Michela Mazzocco[144], Miguel Rodríguez-Gandía[6,56], Natale Imaz Ayo[88,88], Natalia Blay[49], Natalia Chueca[21], Nicola Montano[29,30], Nicole Ludwig[159], Nikolaus Marx[160], Nilda Martínez[161], Norwegian SARS-CoV-[2] Study group, Oliver A. Cornely[115,116,162,163], Oliver Witzke[164], Orazio Palmieri[165], Pa COVID-[19] Study Group, Paola Faverio[107,166], Paolo Bonfanti[167,168], Paolo Tentorio[27], Pedro Castro[63], Pedro M. Rodrigues[6,121,140], Pedro Pablo España[82,82], Per Hoffmann[132], Philip Rosenstiel[1], Philipp Schommers[74,75,76], Phillip Suwalski[33], Raúl de Pablo[19], Ricard Ferrer[169], Robert Bals[69], Roberta Gualtierotti[29,30], Rocío Gallego-Durán[5,7,118], Rosa Nieto[55], Rossana Carpani[30], Rubén Morilla[5,7,64,65,66], Salvatore Badalamenti[27], Sammra Haider[170], Sandra Ciesek[171,172], Sandra May[1], Sara Bombace[27,173], Sara Marsal[23], Sara Pigazzini[44], Sebastian Klein[158], Selina Rolker[132], Serena Pelusi[29,30], Sibylle Wilfling[100,174,175], Silvano Bosari[29,30], Søren Brunak[2], Soumya Raychaudhuri[8,9,10,11,12,176], Stefan Schreiber[1,177], Stefanie Heilmann-Heimbach[132], Stefano Aliberti[178,179], Stephan Ripke[33], Susanne Dudman[32,47], The Humanitas COVID-[19] Task Force, The Humanitas Gavazzeni COVID-[19] Task Force, Thomas Bahmer[46], Thomas Eggermann[109], Thomas Illig[36], Thorsten Brenner[149], Tom H. Karlsen[32,125,126,180], Torsten Feldt[59], Trine Folseraas[32,124,125,126], Trinidad Gonzalez Cejudo[181], Ulf Landmesser[182], Ulrike Protzer[183,184], Ute Hehr[174], Valeria Rimoldi[28], Vegard Skogen[185,186], Verena Keitel[59], Verena Kopfnagel[36], Vicente Friaza[5,7,64,65,66], Victor Andrade[37,39], Victor Moreno[65,187,188,189], Wolfgang Poller[33], Xavier Farre[49], Xiaomin Wang[33], Yascha Khodamoradi[152], Zehra Karadeniz[33], Anna Latiano[165], Siegfried Goerg[190], Petra Bacher[191,192], Philipp Koehler[75,115,162], Florian Tran[1,177], Heinz Zoller[57,58], Eva C. Schulte[183,193,194], Bettina Heidecker[33], Kerstin U. Ludwig[132], Javier Fernández[63,195], Manuel Romero-Gómez[6,196], Agustín Albillos[6,56], Pietro Invernizzi[34,35], Maria Buti[6,24,54], Stefano Duga[27,28], Luis Bujanda[6,119,121], Tobias L Lenz[13,14,]+, Rosanna Asselta[27,28,]+, Rafael de Cid[49,]+, Luca Valenti[29,30,]+, Mario Cáceres[4,197,]+, Andre Franke[1,198,]+

[1]Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany.

[2]Novo Nordisk Foundation Center for Protein Research, Disease Systems Biology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.

[3]Institute of Biotechnology, Life Science Centre, Vilnius University, Lithuania.

[4]Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain.

[5]Hospital Universitario Virgen del Rocío de Sevilla, Sevilla, Spain.

[6]Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas (CIBEREHD), Instituto de Salud Carlos III (ISCIII), Madrid, Spain.

[7]Instituto de Biomedicina de Sevilla (IBIS), Sevilla, Spain.

[8]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA.

[9]Division of Rheumatology, Inflammation and Immunity, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.

[10]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA.

[11]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA.

[12]Center for Data Sciences, Brigham and Women's Hospital, Boston, MA, USA.

[13]Research Group for Evolutionary Immunogenomics, Max Planck Institute for Evolutionary Biology, Plön, Germany.

[14]Research Unit for Evolutionary Immunogenomics, Department of Biology, University of Hamburg, Hamburg, Germany.

[15]Research Department, Stavanger University Hospital.

[16]Randaberg Municipality, Norway.

[17]Norwegian Institute of Public Health, Division of Health Data and Digitalization, Department of Genetics and Bioinformatics (HDGB) Oslo, Norway.

[18]University of Stavanger, Faculty of Health Sciences, Department of Quality and Health Technology, Stavanger, Norway.

[19]Department of Intensive Care, Hospital Universitario Ramón y Cajal, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), University of Alcalá, Madrid, Spain.

[20]Ibs.Granada Instituto de Investigación Biosanitaria, Granada, Spain.

[21]Microbiology Unit. Hospital Univeristario Clinico San Cecilio, Granada, Spain.

[22]Osakidetza Basque Health Service, Donostialdea Integrated Health Organisation, Clinical Biochemistry Department, San Sebastian, Spain.

[23]Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Barcelona, Spain.

[24]Liver Unit, Department of Internal Medicine, Hospital Universitari Vall d'Hebron, Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain.

[25]Research Center and Memory Clinic. Ace Alzheimer Center Barcelona – Universitat Internacional de Catalunya, Spain.

[26]CIBERNED, Network Center for Biomedical Research in Neurodegenerative Diseases, National Institute of Health Carlos III, Madrid, Spain.

[27]IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy.

[28]Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy.

[29]University of Milan, Milan, Italy.

[30]Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy.

[31]Department of Acute Medicine, Oslo University Hospital, Oslo, Norway.

[32]Institute of Clinical Medicine, University of Oslo, Oslo, Norway.

[33]Charite Universitätsmedizin Berlin, Berlin, Germany.

[34]European Reference Network on Hepatological Diseases (ERN RARE-LIVER), San Gerardo Hospital, Monza, Italy.

[35]Division of Gastroenterology, Center for Autoimmune Liver Diseases, School of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy.

[36]Hannover Unified Biobank, Hannover Medical School, Hannover, Germany.

[37]Department of Neurodegenerative Diseases and Geriatric Psychiatry, University Hospital Bonn, Medical Faculty, Bonn, Germany.

[38]Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany.

[39]Division of Neurogenetics and Molecular Psychiatry, Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany.

[40]German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany.

[41]Department of Psychiatry and Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, San Antonio, TX, USA.

[42]Department of Anesthesiology and Intensive Care, University Hospital of North Norway, Tromsø, Norway.

[43]Pediatric Departement and Centro Tettamanti- European Reference Network (ERN) PaedCan, EuroBloodNet, MetabERN-University of Milano-Bicocca-Fondazione MBBM/Ospedale San Gerardo, Italy.

[44]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland.

[45]Centre for Multidisciplinary Research in Health Science (MACH), University of Milan, Milan, Italy.

[46]Klinik für Innere Medizin I, Universitätsklinikum Schleswig-Holstein, Campus Kiel.

[47]Department of Microbiology, Oslo University Hospital, Oslo, Norway.

[48]Department of Child and Adolescent Psychiatry, University Hospital Essen, University of Duisburg-Essen, Essen, Germany.

[49]Genomes for Life-GCAT lab. Germans Trias i Pujol Research Institute (IGTP), Badalona, Spain.

[50]Laboratory of Microbiology, San Gerardo Hospital, Monza, Italy.

[51]Department of Infectious diseases, Oslo University Hospital, Norway.

[52]Humanitas Gavazzeni-Castelli, Bergamo, Italy.

[53]Microbiology Department, Hospital Universitari Vall d'Hebron, Barcelona, Spain.

[54]Universitat Autònoma de Barcelona, Bellatera, Spain.

[55]Department of Respiratory Diseases, Hospital Universitario Ramón y Cajal, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), University of Alcalá, Centro de Investigación Biomédica en Red en Enfermedades Respiratorias (CIBERES), Madrid, Spain.

[56]Department of Gastroenterology, Hospital Universitario Ramón y Cajal, University of Alcalá, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), Madrid, Spain.

[57]Medical University of Innsbruck, Department of Medicine I, Gastroenterology, Hepatology and Endocrinology, Innsbruck, Austria.

[58]Christian Doppler Laboratory of Iron and Phosphate Biology at the Department of Medicine I, Medical University of Innsbruck, Innsbruck, Austria.

[59]Department of Gastroenterology, Hepatology and Infectious Diseases, University Hospital Duesseldorf, Medical Faculty Heinrich Heine University, Duesseldorf, Germany.

[60]Department of Respiratory Medicine and Allergology, University Hospital, Goethe University, Frankfurt am Main, Germany.

[61]Institute of Genomic Statistics and Bioinformatics, University Hospital Bonn, Medical Faculty University of Bonn, Venusberg-Campus 1, Bonn Germany.

[62]Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Madrid, Spain.

[63]Hospital Clinic, University of Barcelona, and IDIBAPS, Barcelona, Spain.

[64]University of Sevilla, Sevilla, Spain.

[65]Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain.

[66]Consejo Superior de Investigaciones científicas, Sevilla, Spain.

[67]Department of Infectious Diseases, Hospital Universitario Ramón y Cajal, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), University of Alcalá, Madrid, Spain.

[68]Department of Transfusion Medicine and Haematology Laboratory, San Gerardo Hospital, Monza, Italy.

[69]Department of Internal Medicine V - Pneumology, Allergology and Intensive Care Medicine, University Hospital Saarland, Homburg/Saar, Germany..

[70]Technical University of Munich, School of Medicine, University Hospital rechts der Isar, Department of Internal Medicine II, Munich, Germany.

[71]Respiratory Medicine & International Health, University of Lübeck, Lübeck, Germany.

[72]Division of Clinical Infectious Diseases, Research Center Borstel, Borstel, Germany.

[73]German Center for Infection Research (DZIF) Clinical Tuberculosis Unit, Borstel, Germany.

[74]Department I of Internal Medicine, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany.

[75]Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany.

[76]German Center for Infection Research (DZIF), Partner Site Bonn-Cologne, Cologne, Germany.

[77]Osakidetza Basque Health Service, Basurto University Hospital, Respiratory Service, Bilbao, Spain.

[78]Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Science, Norwegian University of Science and Technology, Trondheim, Norway.

[79]Department of Medicine, Møre & Romsdal Hospital Trust, Ålesund, Norway.

[80]Department of Internal Medicine II, Medical University of Innsbruck, Innsbruck, Austria.

[81]Department of Anesthesiology and Critical Care, Hospital Universitario Ramón y Cajal, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), University of Alcalá, Madrid, Spain.

[82]Osakidetza Basque Health Service, Galdakao Hospital, Respiratory Service, Galdakao, Spain.

[83]Biocruces Bizkaia Health Research Institute.

[84]Department of Infectious Diseases - E.O. Ospedali Galliera, Genova, Italy.

[85]Geminicenter for Sepsis Research, Institute of Circulation and Medical Imaging (ISB), NTNU, Trondheim, Norway.

[86]Clinic of Anesthesia and Intensive Care, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.

[87]Accident & Emergency and Emergency Medicine Unit, San Gerardo Hospital, Monza, Italy.

[88]Biocruces Bizkaia Health Research Institute, Barakaldo, Spain.

[89]Histocompatibilidad y Biologia Molecular, Centro de Transfusion de Madrid, Madrid, Spain.

[90]Dino Ferrari Center, Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy.

[91]IRCCS Fondazione Ca' Granda Ospedale Maggiore Policlinico, Neurology Unit, Milan, Italy.

[92]University of Milan, Department of Pathophysiology and Transplantation, Milan, Italy.

[93]Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Angelo Bianchi Bonomi Hemophilia and Thrombosis Center, Milan, Italy..

[94]Department of Tropical Medicine, Bernhard Nocht Institute for Tropical Medicine, and Department of Medicine I, University Medical Centre Hamburg-Eppendorf, 20359 Hamburg, Germany.

[95]Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Respiratory Unit, Milan, Italy.

[96]Department of Pathophysiology and Transplantation, Università degli Studi di Milano, Italy.

[97]Internal Medicine Department, Virgen del Rocio University Hospital, Sevilla, Spain.

[98]Biochemistry Department, University Hospital Vall d'Hebron, Barcelona, Spain.

[99]Emergency Department, University Hospital Regensburg, Regensburg, Germany.

[100]Department for Infectious Diseases and Infection Control, University Hospital Regensburg, Regensburg, Germany.

[101]Department Emergency, Anesthesia and Intensive Care, San Gerardo Hospital, Monza, Italy.

[102]School of Medicine and Surgery, University Milano-Bicocca, Milan, Italy.

[103]Fondazione Grigioni per il Morbo di Parkinson and Parkinson Institute, ASST Gaetano Pini-CTO, Milan, Italy.

[104]School of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy.

[105]Acute Geriatric Unit, San Gerardo Hospital, Monza, Italy.

[106]Neurointensive Care Unit, San Gerardo Hospital, Monza, Italy.

[107]School of Medicine and Surgery University Milano-Bicocca, Milan, Italy.

[108]Department of Anesthesiology, Intensive Care Medicine and Pain Therapy, University Hospital Frankfurt, Frankfurt am Main, Germany.

[109]Institute of Human Genetics, Medical Faculty, RWTH Aachen University, Aachen, Germany.

[110]Department of Pneumology, Hannover Medical School, Hannover, Germany.

[111]Department of Medical Microbiology, Clinic of Laboratory Medicine, St. Olavs hospital, Trondheim, Norway.

[112]Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway.

[113]Department of Infectious Diseases, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.

[114]Department of Clinical and Molecular Medicine, NTNU, Trondheim, Norway.

[115]University of Cologne, Medical Faculty and University Hospital Cologne, Department I of Internal Medicine, Cologne, Germany.

[116]University of Cologne, Medical Faculty and University Hospital Cologne, German Center for Infection Research (DZIF), Partner Site Bonn-Cologne, Cologne, Germany.

[117]Cologne Center for Genomics (CCG), University of Cologne, Cologne, Germany.

[118]Centro de Investigación Biomédica en Red Enfermedades Hepáticas y Digestivas (CIBEREHD), Sevilla, Spain.

[119]Department of Liver and Gastrointestinal Diseases, Biodonostia Health Research Institute – Donostia University Hospital, University of the Basque Country (UPV/EHU), CIBERehd, Ikerbasque, San Sebastian, Spain.

[120]Ikerbasque, Basque Foundation for Science, Bilbao, Spain.

[121]Department of Liver and Gastrointestinal Diseases, Biodonostia Health Research Institute – Donostia University Hospital, University of the Basque Country (UPV/EHU), San Sebastian, Spain.

[122]Respiratory ICU, Institut Clínic Respiratory, Hospital Clinic, University of Barcelona, and IDIBAPS, Barcelona, Spain.

[123]Bioinformatics Area, Fundación Progreso y Salud, and Instrituto of Biomedicine of Sevilla (IBIS), Sevilla, Spain.

[124]Section for Gastroenterology, Department of Transplantation Medicine, Division for Cancer medicine, Surgery and Transplantation, Oslo University Hospital Rikshospitalet, Oslo, Norway.

[125]Research Institute for Internal Medicine, Division of Surgery, Inflammatory Diseases and Transplantation, Oslo University Hospital Rikshospitalet and University of Oslo, Oslo, Norway.

[126]Norwegian PSC Research Center, Department of Transplantation Medicine, Division of Surgery, Inflammatory Diseases and Transplantation, Oslo University Hospital Rikshospitalet, Oslo, Norway.

[127]Department of Research, Ostfold Hospital Trust, Gralum, Norway.

[128]Germans Trias i Pujol Research Institute (IGTP), Badalona, Spain.

[129]Department of Genetics & Epigenetics, Saarland University, Saarbrücken, Germany.

[130]Department of Infectious Diseases, Hospital Univeristario Clinico San Cecilio, Granada, Spain.

[131]Infectious Diseases Service, Osakidetza, Biocruces Bizkaia Health Research Institute, Barakaldo, Spain.

[132]Institute of Human Genetics, University of Bonn School of Medicine & University Hospital Bonn, Bonn, Germany..

[133]Department of Research, St Olav Hospital, Trondheim University Hospital, Trondheim, Norway.

[134]Medical Department, Drammen Hospital, Vestre Viken Hospital Trust, Norway.

[135]Research Center Borstel, BioMaterialBank Nord, Germany.

[136]German Center for Lung Research (DZL), Airway Research Center North (ARCN), Germany.

[137]Popgen 2.0 network (P2N), Kiel, Germany.

[138]Department of Infectious diseases, Oslo University Hospital, Oslo, Norway.

[139]Department of Clinical Science, University of Bergen, Bergen, Norway.

[140]Biodonostia Health Research Institute, Donostia University Hospital, San Sebastian, Spain.

[141]Analytic & Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA.

[142]Stanley Center for Psychiatric Research & Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA.

[143]Clinic of Medicine and Rehabilitation, Levanger Hospital, Nord-Trondelag Hospital Trust, Levanger, Norway.

[144]HLA Laboratory – E.O. Ospedali Galliera, Genova, Italy.

[145]Institute for Medical Microbiology and Hospital Epidemiology, Hannover Medical School, Hannover, Germany.

[146]ISGlobal, Barcelona, Spain.

[147]Universitat Pompeu Fabra (UPF), Barcelona, Spain.

[148]IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain.

[149]Department of Anesthesiology and Intensive Care Medicine, University Hospital Essen, University Duisburg-Essen, Essen, Germany.

[150]Osakidetza, OSI Donostialdea, Altza Primary Care, Biodonostia Health Research Institute, San Sebastián, Spain.

[151]Center of Bioinformatics, Biostatistics and Bioimaging, School of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy.

[152]Department of Internal Medicine, Infectious Diseases, University Hospital Frankfurt & Goethe University Frankfurt, Frankfurt am Main, Germany.

[153]Phase 1 Research Centre, ASST Monza, School of Medicine and Surgery, University of Milano-Bicocca, Italy.

[154]Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany.

[155]Gastrointestinal Genetics Lab, CIC bioGUNE - BRTA, Derio, Spain.

[156]Department of Pneumology and Intensive Care Medicine, University Hospital Aachen, Germany.

[157]Munich Clinic Schwabing, Academic Teaching Hospital, Ludwig-Maximilians-University (LMU), Munich, Germany.

[158]Division of Intensive Care and Emergency Medicine, Department of Internal Medicine, Medical University Innsbruck, Innsbruck, Austria.

[159]Center of Human and Molecular Biology, Department of Human Genetics, University Hospital Saarland, Homburg/Saar, Germany.

[160]Department of Internal Medicine I, RWTH Aachen University Hospital, Aachen, Germany.

[161]Department of Anesthesiology, Hospital Universitario Ramón y Cajal, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), Madrid, Spain.

[162]University of Cologne, Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), Cologne, Germany.

[163]Clinical Trials Centre Cologne, ZKS Köln, Cologne, Germany.

[164]Department of Infectious Diseases, University Hospital Essen, University Duisburg-Essen, Essen, Germany.

[165]Gastroenterology Unit, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy.

[166]Pulmonary Unit, San Gerardo Hospital, Monza, Italy.

[167]School of Medicine and Surgery, University of Milano-Bicocca, Italy.

[168]Infectious Diseases Unit, San Gerardo Hospital, Monza, Italy.

[169]Intensive Care Department, Vall d'Hebron University Hospital, SODIR-VHIR research group, Barcelona, Spain.

[170]Department of Medicine, Møre & Romsdal Hospital Trust, Molde, Norway.

[171]Institute of Medical Virology, University Hospital Frankfurt, Goethe University, Frankfurt am Main, Germany.

[172]German Centre for Infection Research (DZIF), External Partner Site Frankfurt, Frankfurt am Main, Germany.

[173]Department of Biomedical Sciences, Humanitas University, Italy.

[174]Zentrum für Humangenetik Regensburg, Regensburg, Germany.

[175]Department of Neurology, Bezirksklinikum Regensburg, University of Regensburg, Regensburg, Germany.

[176]Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, Manchester Academic Health Science Centre, The University of Manchester, Manchester, UK.

[177]Klinik für Innere Medizin I, Universitätsklinikum Schleswig-Holstein, Campus Kiel, Germany.

[178]Department of Biomedical Sciences, Humanitas University, Milan, Italy.

[179]IRCCS Humanitas Research Hospital, Respiratory Unit, Rozzano, Milan, Italy.

[180]Section for Gastroenterology, Department of Transplantation Medicine, Division for Cancer Medicine, Surgery and Transplantation, Oslo University Hospital Rikshospitalet, Oslo, Norway.

[181]Biochemistry Unit. Hospital Univeristario Clinico San Cecilio, Granada, Spain.

[182]Charite Universitätsmedizin Berlin, Berlin Institute of Health, Berlin Germany.

[183]Institute of Virology, Technical University Munich/Helmholtz Zentrum München, Munich, Germany.

[184]German Center for Infection Research (DZIF), Munich partner site, Munich, Germany.

[185]Department of Infectious Diseases, University Hospital of North Norway, Tromsø, Norway.

[186]Faculty of Health Sciences, UIT The Arctic University of Norway, Norway.

[187]Catalan Institute of Oncology (ICO), Barcelona, Spain.

[188]Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain.

[189]Universitat de Barcelona (UB), Barcelona, Spain.

[190]Institute of Transfusionsmedicine, University Hospital Schleswig-Holstein (UKSH), Germany.

[191]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany.

[192]Institute of Immunology, Christian-Albrechts-University of Kiel & UKSH Schleswig-Holstein, Kiel, Germany.

[193]Institute of Psychiatric Phenomics and Genomics, University Medical Center, University of Munich, Munich, Germany.

[194]Department of Psychiatry, University Medical Center, University of Munich, Munich, Germany.

[195]European Foundation for the Study of Chronic Liver Failure (EF-CLIF), Barcelona, Spain.

[196]Digestive Diseases Unit, Virgen del Rocio University Hospital, Institute of Biomedicine of Seville, University of Seville, Seville, Spain.

[197]ICREA, Barcelona, Spain.

[198]University Hospital Schleswig-Holstein (UKSH), Campus Kiel, Germany.

[CORRESPONDING AUTHORS]

Dr. Frauke Degenhardt

Address: Institute of Clinical Molecular Biology and University Hospital of Schleswig-Holstein, Christian-Albrechts-University, Rosalind-Franklin-Str. 12, D-24105 Kiel, Germany

Mail: f.degenhardt@ikmb.uni-kiel.de

Prof. Dr. Andre Franke

Address: Institute of Clinical Molecular Biology and University Hospital of Schleswig-Holstein, Christian-Albrechts-University, Rosalind-Franklin-Str. 12, D-24105 Kiel, Germany

Mail: a.franke@mucosa.de

[ABSTRACT]

Due to the highly variable clinical phenotype of Coronavirus disease 2019 (COVID-19), deepening the host genetic contribution to severe COVID-19 may further improve our understanding about underlying disease mechanisms. Here, we describe an extended GWAS meta-analysis of 3,260 COVID-19 patients with respiratory failure and 12,483 population controls from Italy, Spain, Norway and Germany/Austria, as well as hypothesis-driven targeted analysis of the human leukocyte antigen (HLA) region and chromosome Y haplotypes. We include detailed stratified analyses based on age, sex and disease severity. In addition to already established risk loci, our data identify and replicate two genome-wide significant loci at 17q21.31 and 19q13.33 associated with severe COVID-19 with respiratory failure. These associations implicate a highly pleiotropic ~0.9-Mb 17q21.31 inversion polymorphism, which affects lung function and immune and blood cell counts, and the *NAPSA* gene, involved in lung surfactant protein production, in COVID-19 pathogenesis.

[Introduction]

In the past year, Coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has evolved into a global pandemic with more than 182 million confirmed cases and 3.9 million COVID-19 related deaths worldwide (frequencies reported by the World Health Organization, July 2nd, 2021). The clinical manifestations of COVID-19 are variable and range from complete absence of symptoms to severe respiratory failure and death. Severe COVID-19 requires intensive medical care with respiratory support and can result in long-term damages detrimental to the individual. The pathogenesis of severe COVID-19 is, however, still poorly understood. This condition has been associated with clinical risk factors such as old age, male sex and comorbidities such as diabetes, active cancer, hypertension and coronary artery disease, and solid organ transplant or other conditions that promote an immunosuppressive condition.[1–4] Studies by this group and others have shown that genetic predisposition plays a role in COVID-19 susceptibility and severity.[5–7] Previously, we reported significant associations between genetic variants at loci 3p21.31 (around *LZTFL*) and 9q34.2 (*ABO* blood group locus) to severe respiratory failure and SARS-CoV-2 infection[5], which have been replicated in subsequent studies.[7,8] While 3p21.31 was associated with disease severity, 9q34.2 was more associated with disease susceptibility[5]. Analysis of blood types of the ABO blood group additionally, showed that individuals carrying blood type A have a higher risk of COVID-19 infection. Since then, 11 additional genome-wide significant loci, associated with SARS-CoV-2 infection or COVID-19 manifestations have been reported by various studies, including the Genetics Of Mortality In

Critical Care (GenOMICC) Initiative and most recently the COVID-19 Host Genetics initiative (HGI).[6,7] Six of these loci have been linked to critical illness by COVID-19, and include loci that were previously associated with pulmonary or autoimmune and inflammatory diseases.[6] Regarding the association with blood type Mankelow *et al*.[9] report a higher disease susceptibility for blood type A secretors, determined by genetic variation at the Fucosyltransferase 2 (*FUT2)* gene. Here, we report an extended genome-wide association study (GWAS) meta-analysis of severe COVID-19 with full and rigorously quality-controlled information on age, sex and disease severity, including 3,260 COVID-19 patients with respiratory failure. The latter was defined as respiratory support with supplemental oxygen [class 1] or non-invasive and invasive ventilation [classes 2 and 3, respectively], or by extracorporeal membrane oxygenation (ECMO) [class 4])[5] and 12,483 population controls of unknown COVID-19 status from Italy, Spain, Norway and Germany/Austria. The availability of information on age, sex, and comorbidities such as hypertension, diabetes and coronary artery disease within each study cohort additionally allowed for in-depth and stratified analysis of especially age- and sex-specific risks for these loci.

The discovery study (first and second analysis) was followed by an *in-silico* replication analysis in up to 12,888 hospitalized cases (including 5,582 critically ill cases) and 1,295,966 population controls from the COVID-19 HGI, which allowed us to replicate previous findings and characterize in detail new candidate loci for disease severity. Genetic analysis was followed by a thorough *in-silico* functional characterization of new loci.

In response to clear expectations for a potential role for the human leukocyte antigens (HLA) in the disease course of COVID-19 and preliminary evidence from some smaller pilot studies[10–12], our genetic analysis includes a detailed investigation of the genetic variations in the HLA region. With male sex identified as a risk factor for severe COVID-19 and COVID-19 related death[3], we explore possible connections between genetic variants on the Y chromosome and the risk of developing severe COVID-19 in males[3]. Variations on the Y chromosome describe so-called Y-chromosome haplogroups with letters A-Z (defined by the Y Chromosome Consortium)[13] and follow a pattern of ancestral population migrations in Europe and on a global scale. HLA analysis includes classical fine-mapping of the HLA region based on local imputation of SNP, amino acid and classical allele information, as well as a broad range of other approaches, including a peptidome-wide association study (PepWAS)[14] computational prediction of SARS-CoV-2 peptide presentation, HLA class I supertype association analysis, and tests for heterozygote advantage, divergent allele advantage and molecular mimicry. Y-chromosome analysis includes analysis of known Y-chromosome haplogroups with a focus on the haplogroup R, the predominant haplogroup in Europe.

[Results]

GWAS meta-analyses for severe COVID-19 with respiratory failure

Genotyping was performed using Illumina's Global Screening Array (GSA), followed by genotype quality control (QC) analysis and TOPMed genotype imputation (**Online Methods, Supplementary Figure 1,** patient numbers before and after QC are shown in **Supplementary Tables 1**). Analogously to the COVID-19 HGI[6], we then conducted two GWAS discovery meta-analyses for two main categories of COVID-19 disease state: First, "hospitalization with respiratory support" (respiratory support classes 1-4 with a total of 3,260 patients; first analysis) and second, a more stringent definition of severe COVID-19 "hospitalization with mechanical ventilation" (classes 2-4 with 1,911 critically ill cases; second analysis). Details of per cohort patient numbers are shown in **Supplementary Table 1e**. The characteristics of patients and controls included in these analyses are shown in **Table 1** and **Supplementary Table 2**. After imputation, we carried out a GWAS of 9,223,806 high-quality genetic variants (imputation $R^2 \geq 0.6$ and minor allele frequency (MAF) ≥1%) stratified by ancestry (Italy, Spain, Norway and Germany/Austria) using logistic mixed model analysis as implemented in SAIGE[15], followed by fixed-effect inverse variance meta-analysis using METAL[16] (low genomic inflation of 1.017; **Supplementary Figures 2-3**). Genome-wide comparison of the case-control frequencies (conservatively adjusted for age, sex, age*age, sex*age and top 10 principal components from PCA as employed by the COVID-19 HGI; **Online Methods**) revealed two known loci (*LZTFL1* at 3p21.31; *ABO* at 9q34.2; *TYK2* at 19p13.2) with genome-wide significance ($P_{discovery}<5\times10^{-8}$) and an additional four (*PCDH7* at 4p15.1; *FREM1* at 9p22.3; *MAPT* at 17q21.31; *DPP9* at 19p13.3) with suggestive significance ($P_{discovery}<10^{-6}$) in our first analysis, and another five loci (*OLMF4* at 13q21.1; *TTC7B* at 14q32.11; *CPD* at 17q11.2; *PTPRM* at 18p11.23; *IFNAR2* at 21q22.11) with suggestive significance from the second analysis (**Figure 1**; **Supplementary Table 3,** regional association plots shown in **Supplementary Figures 4-5**). Across the results from the single study cohorts, the strongest association signal in the discovery analysis (besides the known association around *LZTFL1* at locus 3p21.31) was found in the Spanish cohort at locus 17q21.3 (rs8065800 at *MAPT*; $P=1.16\times10^{-8}$; OR=1.67 for minor allele G; 95%CI=1.49-1.84; **Supplementary Table 3**). These findings replicate four associations previously reported by the COVID-19 HGI.[6] Analysis of the ABO secretor status (**Online Methods**) does not statistically replicate the recently published finding that A secretors have a higher risk of COVID-19 susceptibility, though a clear trend can be observed for increased risk of COVID-19 infection for A secretors (A secretors: OR=1.32; 95%CI=1.20-1.46; A non-secretors: OR=1.08; 95%CI=0.91-1.28) (**Supplementary Table 4**). To calculate posterior probabilities of replicability (PPRs) of our genome-wide and suggestive

variants across individual discovery and replication GWAS studies, we performed hierarchical mixture model analysis with MAMBA[17] (**Online Methods**). The analysis showed a high probability (PPR>0.8), of consistent effect sizes across all analyzed cohorts for all but the variants at 14q32.11 (*TTC7B*, PPR=0.005) and at 17q11.2 (**Supplementary Table 5**) with the 17q21.3 (rs8065800) variant showing higher replicability in our first analysis compared to the second analysis (PPR=0.93/PPR=0.2).

Next, we recalculated statistics for our first and second analyses, excluding individuals that were used to calculate association statistics, submitted to the COVID-19 HGI B2 (12,888 hospitalized cases) and A2 (5,582 critically ill cases) analyses by this group and the BoSCO study group, to be used as an external replication. This left 11,906 (1,579 cases/10,327 controls) individuals for the first and 11,009 (944 cases/10,065 controls) individuals for the second analysis. Of the 13 loci reported as associated with COVID-19 by the COVID-19 HGI 8 replicated at least at nominal P-value of 0.05. 12 of 13 variants replicated in this studies' complete cohort at the same significance threshold (**Supplementary Table 3**).

Subsequently, we performed a fixed-effect inverse variance meta-analysis using METAL[16] across the first analysis and COVID-19 HGI B2 statistics as well as the second analysis and COVID-19 A2 statistics, resulting in a total of 725,601 and 1,318,140 analyzed individuals in the respective analyses as well as 9,163,457 and variants with MAF ≥ 1% (**Supplementary Figures 6-7**). This analysis prioritized rs1819040 at the 17q21.31 locus within the *KANSL1* gene as the most strongly associated variant in this region ($P_{\text{discoveryreplication}}=3.27\times10^{-11}$ for rs1819040; OR=0.88 for minor allele T; 95%CI=0.84-0.92) over rs8065800 from the first analysis (**Supplementary Table 3**). Neither of the novel associations at *PCDH7* at 4p15.1; *FREM1* at 9p22.3; *OLMF4* at 13q21.1; *TTC7B* at 14q32.11 and *CPD* at 17q11.2 had suggestive evidence after combining data with the HGI statistics. Since replicability with MAMBA, however showed high replication of the variants in this studies' cohort, subsequent analyses need to show whether this is attributable to an artefact in the cohorts of this study or maybe to high heterogeneity of cohorts and data sets included in the COVID-19 HGI analyses.

The meta-analysis of our second discovery cohort (critically ill only) with the COVID-19 HGI summary statistics revealed an additional genome-wide significant locus not previously associated with severe COVID-19 (*NR1H2* at 19q13.33; $P_{\text{discoveryreplication}}=3.25\times10^{-8}$ for rs1405655; OR=1.09 for minor allele C; 95%CI=1.06-1.1) (regional association plot is shown in **Supplementary Figure 8, Supplementary Table 3**).

Stratified analysis on genome-wide and suggestive loci

To estimate possible hidden genetic effects from age, sex and disease severity, we performed an in-depth stratified analysis of the 3 genome-wide significant and 7 replicable suggestive loci from our first and second analyses as well as the 19q13.33 variant from the meta-analysis with COVID-19 HGI summary statistics. We additionally investigated association of these variants to known comorbidities such as hypertension, coronary artery disease and diabetes **Supplementary Table 6, Supplementary Figures 9/10**). We confirmed our previously described[5] association with age (first analysis; ages: 41-60; $P=2.15x10^{-23}$, OR=2.20, 95%CI=1.88-2.57; ages: 61-80*;* $P=6.41x10^{-6}$; OR=1.49, 95%CI=1.25-1.77) and disease severity for 3p21.31 ($P=9.73\times10^{-7}$, OR=1.65; 95%CI=1.35-2.03) and additionally observed a strong trend for association with age for rs12610495 at the 19p13.3 locus (first analysis: ages: 41-60; OR=1.33, 95%CI=1.19-1.48; ages: 61-80; OR=1.14, 95%CI=1.02-1.28). None of the other loci showed genome-wide or suggestive association to any of the analyzed categories.

Additional stratified analysis on candidate SNPs from the COVID-19 HGI analyses[6], showed trends for association with age for rs1819040 at 17q21.31 (see also below) and of rs74956615 with gender and age (first analysis; ages: 41-60; OR=1.68, 95%CI=1.33-2.13; ages: 61-80*;* OR=1.04, 95%CI=0.80-1.35; male: OR=1.53, 95%CI=1.23-1.89; female: OR=1.03, 95%CI=0.75-1.40) which was not observed for rs11085725 (*TYK2* at 19p13.2) at the same locus (between variant LD is r^2=0.09 or D'=0.94) (**Supplementary Table 6, Supplementary Figure 11**).

Fine-mapping of signals at 17q21.31 and 19q13.33

We next performed an in-depth characterization of the previously unknown genome-wide significant locus 19q13.33 (*NR1H2)* and the 17q21.31 locus, which contains a known ~0.9-Mb inversion polymorphism spanning across several different genes, whose implication in COVID-19 has not been accurately investigated.

Bayesian fine-mapping analysis with FINEMAP[18] (**Online Methods**) identified a total of 1,531 (log_{10}(Bayes factor)=10.46) and 15 (log_{10}(Bayes factor)=5.95) variants that belong to the 95% credible sets of variants most likely to be causal at 17q21.31 or 19q13.33, respectively (**Figure 1, Supplementary Table 7**). For 17q21.31, the 95% credible set included rs1819040 as the best SNP candidate with only 1.1% certainty, followed by 1,530 variants in high linkage disequilibrium (LD) (mean(LD)=0.997 and min(LD)=0.954) with certainty <0.3%, indicating that the individual SNP associations are only proxy variants for the actual causal variant (see below). For 19q13.33, the 95% credible set included rs1405655 as the best SNP candidate

with 59.1% certainty, followed by rs1274514 and rs1274510 (5.3% certainty), and 12 variants with certainty <4%, so we assume that rs1405655 represents the candidate causal variant

Imputation and statistical analysis of the 17q21.31 inversion

The lead variant rs1819040 and the most strongly associated variants of the 95% credible set at this locus point directly to the common 17q21.31 inversion polymorphism, spanning across several different genes, including and mapping to two highly divergent haplotypes, H1 and H2, which are estimated to have evolved separately >2 million years ago[19] (**Figure 1a**). The inversion haplotypes H1 and H2, were determined more accurately for COVID-19 respiratory support failure cases and controls by genotype imputation with IMPUTE2[20], employing as reference 109 individuals from the 1000 Genomes Project for which 17q21.31 inversion genotypes were obtained experimentally by FISH and droplet digital PCR (**Online Methods**). LD between the rs1819040 variant, prioritized in the meta-analysis with the COVID-19 HGI summary statistics and the inversion in our cohorts is near perfect (r^2=0.98, D'=0.99). rs8065800-G, observed as associated in our first and second analysis as a risk allele in the Spanish population, is inherited together with the major H1 haplotype, although being in low LD (in our cohorts: r^2=0.12-0.18; D'=0.97-1), which also points to the effect of the inversion. Genome-wide significant association with severe respiratory COVID-19 for the inversion was confirmed using logistic regression followed by meta-analysis across this study's discovery and replication panels from the COVID-19 HGI (meta-analysis first discovery panel and COVID-19 HGI release 5 B2: $P=7.61\times10^{-10}$, OR=0.89; 95%CI= 0.84-0.92; meta-analysis second discovery panel and COVID-19 HGI release 5 A2: $P=1.5\times10^{-4}$, OR=0.90; 95%CI= 0.85-0.95; **Figure 1b, Supplementary Table 8a**). For deduction of P-values from the HGI replication summary datasets we used the published method Fast and Accurate P-value Imputation for genome-wide association study (FAPI)[21]. Using our first and second discovery analyses only, the 17q21.31 inversion association with a protective OR was consistently observed across the Italian, Spanish, and German/Austrian study cohorts in the 40-60 years age group (Norway was excluded for this analysis, having $N_{Case} < 50$), while association was inconsistent in the equally powered 60-80 years age groups (**Figure 1b, Supplementary Table 8b**).

Functional analysis of 17q21.31 and 19q13.33 using publicly available datasets

We next performed several follow-up analyses to understand better possible functional implications of associations at the 17q21.31 and 19q13.33 loci. A phenome-wide association study (PheWAS) for 17q21.31 and 19q13.33 using a wide range of phenotypes from the

NHGRI GWAS Catalog[22] and other available GWAS data revealed no known phenotypes to be linked with the rs1405655 lead SNP at 19q13.33, while 162 GWAS associations were identified for the 17q21.31 inversion in the GWAS Catalog, illustrating its pleiotropic effects. These associations included several traits potentially related to COVID-19 pathology, such as blood and immune cell composition or lung function (**Figure 1c**). Credible sets from Bayesian fine mapping at 17q21.31 and 19q13.33, overlap with several genes including *MAPT*, *KANSL1, FMNL1* and *CRHR1* at the inversion locus 17q21.31, and *NAPSA*, *NR1H2*, *KCNC3* at locus 19q13.33 (**Supplementary Figures 4, 5 and 8)**. We performed an exploratory gene expression analysis using several publicly available datasets to: 1) Identify in which tissues or cell types our candidate genes are expressed by analyzing their RNA expression at bulk and single-cell level; 2) Examine the direct effect of both loci on gene expression by using expression and splicing quantitative trait loci (eQTL and sQTL) and; 3) Infer the possible contribution of these genes to COVID-19 pathology by looking at their expression patterns in a) monocytes exposed to different viral and non-viral immune stimulators; b) organoids infected with COVID-19 and c) single cell RNA-seq of several tissues including lung coming from patients who died after experiencing a SARS-CoV-2 severe disease (**Online Methods**).

The potential functional role of the 17q21.31 inversion is supported by the analysis of 2,902 linked variants (r^2>0.9) that are already reported as eQTL or sQTL in the GTEx Project.[23] The inversion is in LD with lead eQTLs and sQTLs (i.e. displaying the strongest association with target), for 24 and 7 genes, respectively, in at least one tissue (**Supplementary Table 9**, **Supplementary Figure 12**). Expression patterns of coding genes affected by the inversion are shown in **Supplementary Figure 13**, with the best candidates to play a role in the effects of the inversion on their known function and expression, including *MAPT*, *KANSL1, FMNL1* and *CRHR1,* being summarized in **Figure 2**. Many of these genes are highly expressed in neural and male-specific tissues (**Figure 2a**, **Supplementary Figure 13**). However, several of them are also expressed in major immune cell types in COVID-19 relevant tissues (**Figure 2b; Supplementary Figure 13a/b**), such as *KANSL1*, which is expressed in lung tissue-resident alveolar macrophages, or *FMNL1* (**Figure 2b**, **Supplementary Figure 13b**). These two genes especially show significantly higher expression in different lung cell types in COVID-19 patients versus healthy controls[24] (**Figure 2c**, **Supplementary Figure 13c**). Moreover, RNA-seq data of monocytes under different bacterial and viral stimuli[25] show expression changes in several genes affected by the inversion (**Supplementary Figure 14, Supplementary Table 10**). For example, *KANSL1,* which could have anti-inflammatory effects, shows higher expression in H2 carrier monocytes stimulated with Influenza A virus infection-like conditions, which appears to be related to a significant increase of coding and non-coding isoforms (**Supplementary Figure 15**). Similarly, in SARS-CoV-2 infected brain

organoids, the expression of *MAPT* was significantly downregulated in premature and mature neuronal cells (**Supplementary Figure 16, Supplementary Table 11**).

In the case of the 19q13.33 locus, expression of candidate genes shows high tissue specificity, with the *NAPSA* mRNA being specific to lung and lung parenchyma and *KCNC3* being highly expressed in brain and thyroid tissue, while *NR1H2* is more broadly expressed among human tissues, including many immune cell types (**Figure 2b**, **Supplementary Figure 13**). Of those candidates, expression of *KCNC3* and especially *NAPSA* appears to be clearly affected by the rs1405655 lead SNP (**Figure 2a**). Single SNP Mendelian randomization analysis using rs1405655 and eQTL data identified higher expression of *NAPSA* as a potentially causal protective factor for severe COVID-19 (Beta=-0.39; $P=1.26\times10^{-7}$) **(Supplementary Methods, Supplementary Table 12)**. Notably, *NAPSA* shows significantly increased expression in type 1 alveolar cells of COVID-19 patients as compared to healthy controls[24] (**Figure 2**, **Supplementary Figure 13**). The *NR1H2* gene is significantly down-regulated in some parenchymal (basal, ciliated, club) and endothelial (pericyte) lung cells and is up-regulated in monocytes of COVID-19 patients as compared to healthy controls (**Figure 2c**, **Supplementary Figure 13**).[24]

Association analysis of specific candidate genomic regions: HLA locus & Y-chromosome haplogroups

The HLA fine-mapping approach yielded no association at the genome-wide or nominal ($P<10^{-5}$) significance threshold, neither in the overall meta-analysis across the four cohorts, nor within the separate cohorts (**Supplementary Table 13, Supplementary Figure 17**). Furthermore, we found no significant association for any HLA-presented viral peptide in a so-called PepWAS approach (**Supplementary Table 14**, **Supplementary Figures 18-19**), where associations between HLA-presented peptides and disease is unravelled by integrating similarities and differences in peptide binding among HLA alleles across patients nor robust statistical associations with any of the other tested HLA parameters (**Supplementary Table 15**).

Results of the Y-chromosome haplogroup analysis are shown in **Supplementary Table 16**. We observed a significant risk association of the R-haplogroup R (M207) in the Italian population for individuals aged > 80 years with respiratory failure COVID-19 (P=0.0014, OR=4.29, 95%CI=1.76-10.47; N_{Cases} or $N_{Controls}$ < 50 in all other cohorts for this age group) in the first analysis. At different haplogroup levels this signal was driven by haplogroup R1b (M343) (P=0.0026, OR=3.92; 95%CI=1.61-9.54) and its subgroup R1b1a2a2 (P312) ($P=8.11\times10^{-5}$, OR=16.18, 95%CI=4.05-64.58). This association was however not seen in any

other age group or cohort. COVID-19 related mortality was significantly associated with haplogroup R1b1a2a1 (U106) (P=0.01, OR=2.8, 95%CI=1.27-6.12). This association remained after adjusting for the comorbidities hypertension, coronary artery disease and diabetes. COVID-19 related death remains, however, a challenging endpoint influenced by many factors but did not surpass correction for multiple testing.

[DISCUSSION]

We here present a large collaborative COVID-19 genetics study of different centers from Italy, Spain, Norway and Germany/Austria. With our clearly defined phenotype of severe respiratory COVID-19 and a centralized genotyping and rigorous quality control, we have generated a valuable resource for further COVID-19 related genetic analyses. We identified and analyzed in detail two new loci of interest associated with severe COVID-19, the 17q21.31 inversion and the 19q13.33 locus. Furthermore, we examined for the first time effects of age and sex on various variants identified as genome-wide significant or suggestive in this study.

As for the novel association at the 19q13.33 locus, additional analyses provided first hints for a functional involvement in COVID-19 through its regulation of the *NAPSA* gene, a gene encoding a protease highly expressed in Type 1 (AT1) and Type 2 (AT2) alveolar cells, two cell types required for the gas exchange at the lung surface and the secretion of surfactant proteins as well as immunomodulatory factors (AT2).[26] Complementary findings were observed in another recent study dissecting the lung transcriptome of COVID-19 infected patients in which *NAPSA* expression is increased also in AT1 cells. This study also linked *NAPSA* to the marker gene expression signature of “damage-associated transient progenitors” (DAMPs), an intermediate cell state between AT1 and AT2 cells promoted by inflammation, characterized by a failure of AT2 cells to differentiate to AT1 cells.[27] Thus, given that the NAPSA protein is involved in lung surfactant production, which is dysregulated in COVID-19[28], the fact that the COVID-19 risk allele is associated with decreased *NAPSA* expression while increased expression of *NAPSA* is a protective factor for severe COVID-19, suggests a potential role of *NAPSA* in susceptibility for severe COVID-19.

Moreover, we detected a clear association of the well-known and pleiotropic 17q21.31 inversion polymorphism, which is linked to many traits potentially relevant to COVID-19 outcome. For instance, the inverted haplotype H2 was previously associated with higher number of red blood cells and hemoglobin levels, whereas each haplotype correlates with different proportions of lymphocytes and granulocytes, which could potentially modulate the immune response during SARS-CoV-2 infection[29]. In addition, the H2 protective allele is also associated with decreased lung function and increased risk of chronic obstructive pulmonary

disease but protects against development of pulmonary fibrosis[30], and is associated with higher ventilatory response to corticosteroids in individuals with asthma[31], showing potential trade-offs and shared pathways that may be important in lung health. This variant had has been proposed to be under positive selection in Europeans through its effect on fertility[19]. Our results point to a role of this polymorphism in immunity and virus infection defense. Interestingly, inversion effects were found to be stronger in the younger age group in both severity classes, which could explain the weaker association in the HGI more severe A2 phenotype due likely to a larger proportion of older individuals[32]. The inversion probably affects COVID-19 disease course through its large effects on gene expression shown by us and others[6]. Although the function of many of the affected genes are not well known, the inversion acts as an eQTL and sQTL of several interesting candidate genes for severe COVID-19. In particular, there are several genes potentially associated with immune function and immune response. For example, *KANSL1*, involved in histone acetylation, is broadly expressed in many types of immune cells in upper airways and lung tissue (**Figure 2**) and has been proposed to play a role in the macrophage transition to an anti-inflammatory phenotype in mice[33,34]. Here, we have found that its expression decrease in infection-like stimulated monocytes is partially compensated in homozygotes for the H2 inversion haplotype. *CHRH1*, associated to higher expression in the H2 haplotype in several tissues (**Figure 2**), encodes a receptor that binds to corticotropin-releasing hormones, which are major regulators of the hypothalamic-pituitary-adrenal axis, and regulates immune and inflammatory responses[35]. Finally, *FMNL1*, which is also located in the inversion locus, shows high expression levels in macrophages, dendritic cells and B and T lymphocytes in different COVID-19 related tissues (**Figure 2**) and it is involved in cell motility and T cell trafficking[36]. In addition, many of the genes in the 17q21.31 inversion regions show a predominant expression in brain tissues, which could also play an important role in COVID-19. The clearest example is *MAPT*, which is downregulated in SARS-CoV-2 infected neuronal cells (**Supplementary Figure 16**) and lung-related cells and tissues[37]. The H2 haplotype is linked to increased expression of *MAPT* in the lung and its *MAPT* exon 3 in brain tissues[38], which could compensate for the downregulation during viral infection and have a protective effect against COVID-19. However, despite the potential implication of these and other genes, it is not possible to single out just one as the most likely candidate. In fact, inversions are well-known for keeping together a combination of alleles from different genes that generate complex phenotypic traits in different organisms[39].

Our hypothesis-driven analysis of associations in the HLA, as well as COVID-19 specific PepWAS analyses yielded no significant results, indicating no major role for HLA variability in mediating the severity of COVID-19 in our cohorts. These results are in line with a recent, and

the so-far largest, HLA analysis from Shachar *et al*.[40] Interestingly, we observed statistical association of the Y-haplogroup R with COVID-19 disease and mortality, however none of the results remain significant after correction for multiple testing. To gain more knowledge regarding the potential role of the Y-chromosome haplogroups in the COVID-19 pandemic, larger study samples are necessary as well as studies following the pandemic over time investigating whether the associations weaken or strengthen for different haplogroups.

In summary, our findings add to the number of genome-wide significant hits for COVID-19 – totaling now around 16 independent loci – and provide new insights to the molecular basis of COVID-19 severity that could potentially trigger subsequent and more targeted experiments to develop therapies for severe COVID-19.

[ONLINE METHODS]

Study Participants and Recruitment

We recruited 5,228 patients with mild to severe COVID-19, which was defined as hospitalization only (mild) or with respiratory failure (severe) with a confirmed SARS-CoV-2 viral RNA polymerase-chain-reaction (PCR) test from nasopharyngeal swabs or other relevant biologic fluids, cross sectionally, from intensive care units and general wards at different hospitals from Italy (4 centers, N=1,857), Spain (6 centers, N=2,795), Norway (7 centers, N=127) and Germany/Austria (8 German, 1 Austrian center, N=449). For comparison, we included 13,705 control participants from Italy (4 centers, N=5,247), Spain (3 centers, N= 4,552), Norway (1 center, N = 288) and Germany (1 center, N= 3,582). Details on the centers and origin of the control panels are shown in **Supplementary Table 1a/b**. Though all patient samples that were sent to our study center were processed, only the severe COVID-19 individuals were analyzed in this study (**Supplementary Table 1e)**. Respiratory failure was defined in the simplest possible manner to ensure feasibility: the use of oxygen supplementation or mechanical ventilation, with severity graded according to the maximum respiratory support received at any point during hospitalization (1: supplemental oxygen therapy only, 2: noninvasive ventilatory support, 3: invasive ventilatory support, or 4: extracorporeal membrane oxygenation).[5]

Recruiting Centers and Ethics Committee Approval IDs

The project protocol involved the rapid recruitment of patient-participants and no additional project-related procedures (we primarily used material from clinically indicated venipunctures) and afforded anonymity, owing to the minimal dataset collected. Differences in recruitment and consent procedures among the centers arose because some centers integrated the project into larger COVID-19 biobanking efforts, whereas other centers did not, and because there were differences in how local ethics committees provided guidance on the handling of anonymization or deidentification of data as well as consent procedures. Written informed consent was obtained, sometimes in a delayed fashion, from the study patients at each center when possible. In some instances, informed consent was provided verbally or by the next of kin, depending on local ethics committee regulations and special policies issued for COVID-19 research. For some severely ill patients, an exemption from informed consent was obtained from a local ethics committee or according to local regulations to allow the use of completely anonymized surplus material from diagnostic venipuncture.

Centers from which samples were obtained are listed together with their ethics approval reference numbers from each ethic committee in the **Supplementary Table 1b.**

Sample Processing, Genotyping, Quality control, and Imputation

Detailed description on sample processing, genotyping, genotype quality control and genotype imputation can be found in the **Supplementary Methods**. In brief: The majority of samples were processed at the DNA laboratory of the Institute of Clinical Molecular Biology (Christian-Albrechts-University of Kiel, Germany), with a subset of the German samples (BoSCO study) processed and typed at the Genomics Department of Life&Brain Center, Bonn, and another subset (COMRI study) prepped at the Technical University Munich, Munich, Germany and genotyped at the Genotyping laboratory of the Institute for Molecular Medicine Finland FIMM Technology Centre, University of Helsinki, Finland. The German control samples were prepped at the Institute of Clinical Molecular Biology and genotyped at the Regeneron Genetics Center, U.S.A. In brief, prepped DNA extracts or non-prepped whole blood, buffy coat samples (and for an exceedingly small subset also saliva) genotyped on Illumina's Global Screening Array (GSA), version 1.0 (German controls), 2.0 or 3.0 (BoSCO study, COMRI study and 378 cases and 1,180 controls from the Italian cohort) with a SNP coverage of 700,078 to 730,059 variants. We performed a uniform sample and single-nucleotide polymorphism (SNP) quality control across the German, Italian, Norwegian and Spanish populations respectively. After the exclusion of samples during quality control (the majority due to population outliers) (**Supplementary Table 1**), the final case–control datasets comprised 1,536, patients and 4,759 control participants from Italy, 1,421 patients and 4,377 control participants from Spain, 62 patients and 262 controls from Norway and 241 patients and 3,110 control participants from Germany. The number of SNPs pre-imputation were 664,969 for the Italian cohort, 669,359 for the Spanish cohort 663,411 for the Norwegian cohort and 568,542 for the German cohort. To maximize genetic coverage, we performed SNP imputation on genome build GRCh38 using the Michigan Imputation Server and 194,512 haplotypes generated by the Trans-Omics for Precision Medicine (TOPMed) program (freeze 5)[41,42] (**Supplementary Methods**). The number of SNPs with a post-imputation score (R^2) of > 0.1 were 77,767,912 for the Italian cohort, 72,504,622 for the Spanish cohort 19,466,514, for the Norwegian cohort and 51,399,774 for the German cohort pre-imputation.

Statistical Analysis

Genome-wide association analysis

To take imputation uncertainty into account, we tested for phenotypic associations with allele dosage data separately for the Italian, Spanish, German, and Norwegian case-control data. We carried out a logistic regression analysis corrected for potential population stratification, age and sex bias using the SAIGE software[43] (**Supplementary Methods**). An inverse-variance weighted fixed-effects meta-analysis was conducted with the meta-analysis tool METAL[15] on the first discovery cohort including 3,260 cases and 12,483 population controls of unknown COVID-19 status from Italy, Spain, Norway and Germany and the second discovery cohort including 1,911 critically ill cases and 12,483 population controls (**Supplementary Table 1).** This was followed by an *in-silico* replication analysis in up to 12,888 hospitalized cases (including 5,582 critically ill cases) and 1,295,966 population controls of the COVID-19 HGI. Only variants that were common to at least 2 datasets with respective post-imputation $R^2 \geq 0.6$ and that had an overall minor allele frequency (MAF) of ≥ 1% were considered in the analysis. For each variant, we computed across-cohort heterogeneity P-values and I^2 values using METAL[16]. We used a significance threshold of $P=10^{-6}$ for joint P-values to determine statistical significance.

Candidate SNPs for each respective main meta-analysis were fine mapped using Bayesian fine mapping with the tool FINEMAP[16] and analyzed using Meta-Analysis Model-based Assessment of replicability[18] (MAMBA) (**Supplementary Methods**) to assess replicability at each locus.

Association analysis of candidate SNPs

We additionally performed age- and sex-stratified as well as severity analyses on candidate SNPs from our first and second discovery cohorts, as well as candidate SNPs from the COVID-19 HGI analysis[6]. We carried out a logistic regression analysis corrected for potential population stratification, age and sex bias using the software R version 3.6.2 (**Supplementary Methods**). The inverse-variance weighted fixed-effects meta-analysis was conducted using the R-package metafor[44] including only statistics from cohorts with N_{Case} and $N_{Control} > 50$. Sub-analyses in age groups of 20-40, 41-60, 61-80 and > 80 years were carried out, with the highest sample numbers and statistical power in the age groups of 41-60 and 61-80 years, such that only these are reported (**Supplementary Table 1e**).

Association analysis of the 17q21.31 inversion

17q21.31 inversion genotypes were imputed using IMPUTE v2.3.2[45,46] from quality controlled SNP data (first and second analysis) based on experimentally-validated inversion genotypes from 109 individuals from the 1000 Genomes Project or from P-values (COVID-19 HGI A2/B2 release 5 data) using Fast and accurate P-value Imputation for genome-wide association study (FAPI) (**Supplementary Methods**).[21] The inversion was coded as 0 for the major allele H1 and 1 for the minor allele H2. All association analyses were carried out as described above on the minor allele H2.

Functional analysis

Phenotype associations with the different variants were obtained from the NHGRI-EBI GWAS Catalog.[47] Similarly, tissue-specific expression or splicing effects were obtained by searching for SNPs in high LD (r^2>0.9) that have been already identified as expression quantitative trait loci (eQTLs) or splicing quantitative trait loci (sQTLs) in cis by the GTEx Project (GTEx Analysis Release v8).[22] Candidate coding genes were selected based on their inclusion in the GWAS credible sets and/or if any of the variants had been identified as lead eQTL or sQTL. Exploratory gene expression analysis of selected candidates was performed on publicly available pre-processed RNA-seq datasets generated from organ tissues (GTEx Analysis Release v8 immune cell types (BLUEPRINT[23]), as well as respiratory tract[48] and brain cells[49] (COVID-19 Cell Atlas)[50]. Differential expression of candidate genes in COVID-19 infected lung cells were obtained from pseudo-bulk differential expression analysis performed by Delorey *et al*.[51] Since several candidate genes (including *MAPT*, *CRHR1* and *KANSL1*) were highly expressed in the neural system, differential gene expression was also analyzed on single-cell RNA-seq dataset of COVID-19 infected brain organoid cells from Song *et al.*[24] (obtained upon request). The analysis was carried out using hurdle modeling, implemented in the R package MAST.[52] Finally, to check the effect of the 17q21.31 inversion on monocytes stimulated by infection-like conditions we also performed a differential expression analysis in the RNA-seq data from Quach *et al*.[53] by imputing the inversion genotypes with IMPUTE v2.3.2[25] and quantifying gene and transcript expression differences with Kallisto v0.46.0[20] and QTLtools v1.1[54]. For detailed description, refer to **Supplementary Methods**.

Analysis and fine mapping of the HLA

Detailed descriptions of this analysis can be found in the **Supplementary Methods**. In brief, quality-controlled genotypes at the HLA region (chr6:29-34Mb) were extracted. HLA allele, amino acid, and SNP imputation was performed using the random-forest based HLA genotype

imputation with attribute bagging (HIBAG) and applying specially tailored as well as publicly available reference panels.[5,55,56] The resulting data were used as a basis for several subsequent analyses, including: 1) basic association analysis (fine mapping) as described in the section **Statistical Analysis** and the **Supplementary Methods**, 2) a peptidome-wide association study[14] (pepWAS), to screen for disease-relevant peptides from SARS-CoV-2, that may present a possible functional link between severe COVID-19 and variation at classical HLA loci, 3) quantitative HLA analyses directed at the number of peptides bound by an HLA allele, as well as 4) an analysis of HLA-presentation of shared peptides ('molecular mimicry').

Analysis of the Y-chromosome haplotypes

First, we produced high quality Y-chromosome genotypes by manually calling and visually inspecting Y-chromosome SNPs in the male fraction of the cohorts only. Next, we used 22 Y-chromosome SNPs to distinguish known Y-chromosome haplogroups as described in the **Supplementary Methods** at different haplogroup resolutions. We here focused on haplogroup R, the most prevalent Y-chromosome haplogroup across Europe. Association analysis was carried out as described in the Section **Statistical Analysis – Association analysis of candidate SNPs** on Y-chromosome haplogroups coded as absent (0) or present (1). We additionally analyzed the end-point mortality (**Supplementary Methods**).

Analysis of the ABO secretor status

ABO blood group typing was performed as described by Ellinghaus *et al*.[5]. Briefly, genotypes of the SNPs rs8176747, rs41302905 and rs8176719 were extracted from the imputed data (R^2=1 for all SNPs and cohorts) and used to infer the A, B and O blood types. The ABO-"secretor" status was inferred from the genotypes of the rs601338 SNP (G>A) at the *FUT2* gene, located at 19q13.33, extracted from the imputed data (R^2=0.98-0.99 for all cohorts). Individuals carrying genotypes GA or GG were assigned secretor status and individuals carrying genotype AA were assigned non-secretor status based on the genotype dosages, ranging from 0 to 2, retrieved from the imputed data. Individuals with allelic dosages 1.3-1.7 were called as "no call", individuals with dosages $\leq$1.3 were called "secretors" and individuals with dosages $\geq$1.7 were called "non-secretors". Association analysis was carried out as described in the Section **Statistical Analysis – Association analysis of candidate SNPs** on blood type or blood type secretor status coded as absent (0) or present (1)**.** Blood types A and AB were also analyzed combined as "A and AB".

Availability of Summary Statistics

Genome-wide summary statistics of our analyses will be made available upon reasonable request to the corresponding author.

Tables

Table 1. Overview of patients included in the genome-wide discovery analysis. Overview of patients included in our first analysis (3,260 patients) and second analysis (1,911 patients). Individuals of the Italian, Spanish, Norwegian and German cohorts were recruited at five, seven, eight and ten different hospitals/centers, respectively. Shown are respiratory support status groups 1-4, age and median age across all individuals as well as within each respiratory support group, percentage of females within each cohort, as well as percentage of individuals affected by known comorbidities of COVID-19, factors related to lung health and mortality. Commonly reported comorbidities in COVID-19 are shown, hypertension, coronary artery disease and diabetes. Characteristics of control individuals are shown in **Supplementary Table 1**.

Characteristic				
	N=1536	**N=1421**	**N=62**	**N=241**
Respiratory support — (%)	**Italy**	**Spain**	**Norway**	**Germany**
Supplemental oxygen only (1)	17.97	63.35	80.65	45.23
Noninvasive ventilation (2)	60.74	8.62	3.23	8.71
Ventilator (3)	20.83	26.91	16.13	35.68
ECMO (4)	0.46	1.13	0	10.37
Median age (IQR) — yr	67 (22)	68 (18)	59.5 (21)	63 (19)
Median age (IQR) — yr (1)	75 (21)	69 (21)	59 (20)	63 (23)
Median age (IQR) — yr (2)	66 (23)	72 (17)	68.5 (10)	68 (20)
Median age (IQR) — yr (3)	63 (14)	65 (15)	66 (19)	64 (18)
Median age (IQR) — yr (4)	49 (12)	57 (10)	/	56 (12)
Female sex — (%)	32.81	35.88	33.87	26.97
First analysis				
Hypertension — (% affected) (% missing)	40.55 (21.16)	49.6 (1.91)	41.3 (25.81)	53.09 (32.78)
Coronary artery disease — (% affected) (% missing)	17.09 (21.16)	10.16 (1.98)	25.81 (0)	15.23 (37.34)
Diabetes — (% affected) (% missing)	14.86 (21.16)	24.55 (1.91)	14.52 (0)	19.63 (32.37)

Smoking — (% affected) (% missing)	11.45 (22.07)	21.26 (1.69)	3.33 (3.23)	43.59 (83.82)
Lung disease — (% affected) (% missing)	13.3 (43.23)	22.86 (32.34)	/	16.07 (30.29)
Mortality — (% affected) (% missing)	15.73 (37.11)	15.4 (18.36)	9.68 (0)	12.63 (21.16)
Second analysis				
Female sex — (%)	29.68	27.55	/	22.73
Hypertension — (% affected) (% missing)	38.26 (23.25)	52.19 (3.28)	/	62.96 (38.64)
Coronary artery disease — (% affected) (% missing)	16.03 (23.25)	12.55 (3.28)	/	18.75 (39.39)
Diabetes — (% affected) (% missing)	14.58 (23.25)	25.3 (3.28)	/	19.75 (38.64)
Smoking — (% affected) (% missing)	11.52 (24.21)	20.44 (2.89)	/	42.11 (85.61)
Lung disease — (% affected) (% missing)	12.88 (48.25)	27.08 (44.51)	/	18.29 (37.88)
Mortality — (% affected) (% missing)	15.68 (40.79)	25.07 (26.2)	/	26.14 (33.33)

Figure 1. Association of the 17q21.31 locus with severe COVID-19 with respiratory failure. (a) Regional association plot showing the variant most strongly associated with severe COVID-19 (rs1819040, purple diamond) a ~0.9-Mb inversion polymorphism at 17q21.21[33] (white line with blue rectangles representing the variable segmental duplication (SD) blocks at the breakpoints), and the large credible set obtained by statistical fine-mapping including 2,178 SNPs in high LD (median(LD)=0.97) with the inversion (**Supplementary Table 7**), Pairwise LD values (r^2) with lead variant rs1819040 were calculated from merged Italian, Spanish, German and Norwegian GWAS discovery datasets. Below, organization of the 17q21.31 inversion genomic region, with the extended haplotypes associated with each orientation (H1 and H2) shown as red and blue arrows, respectively, and breakpoint SDs as dark rectangles. Protein-coding genes for which the inversion is a lead eQTL in at least one GTEx tissue are shown as pointed rectangles indicating the direction of transcription. **(b)** Forest plot for an extended meta-analysis of our first discovery analysis and COVID-19 HGI release 5 analysis B2 datasets (**Online Methods**) of the association between severe COVID-19 and the 17q21.31. Shown are OR and 95% CIs of the main (all), age-stratified (41-60 and 61-80 years (yrs)) and sex-stratified analysis across all analyzed cohorts. **(c)** Phenome-wide association study (PheWAS) results for the 17q21.31 inversion showing only potentially COVID-19 related phenotypes from GWAS Catalog ($P=10^{-7}$) grouped by disease categories using different colors. The effect direction of known SNP-trait associations is shown using triangles pointing upward (increase) and downward (decrease), whereas dots represent unknown effect direction. The dotted line indicates the genome-wide significance threshold ($P=5\times10^{-8}$).

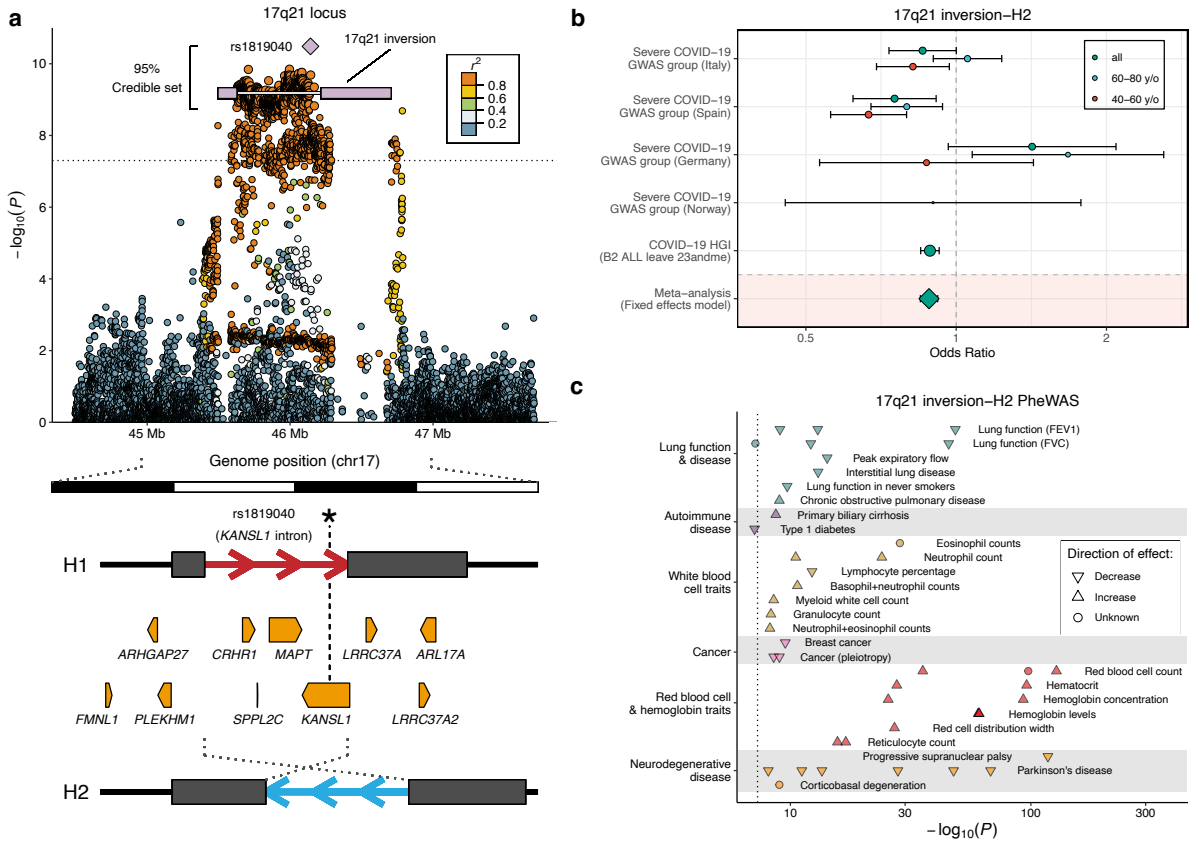
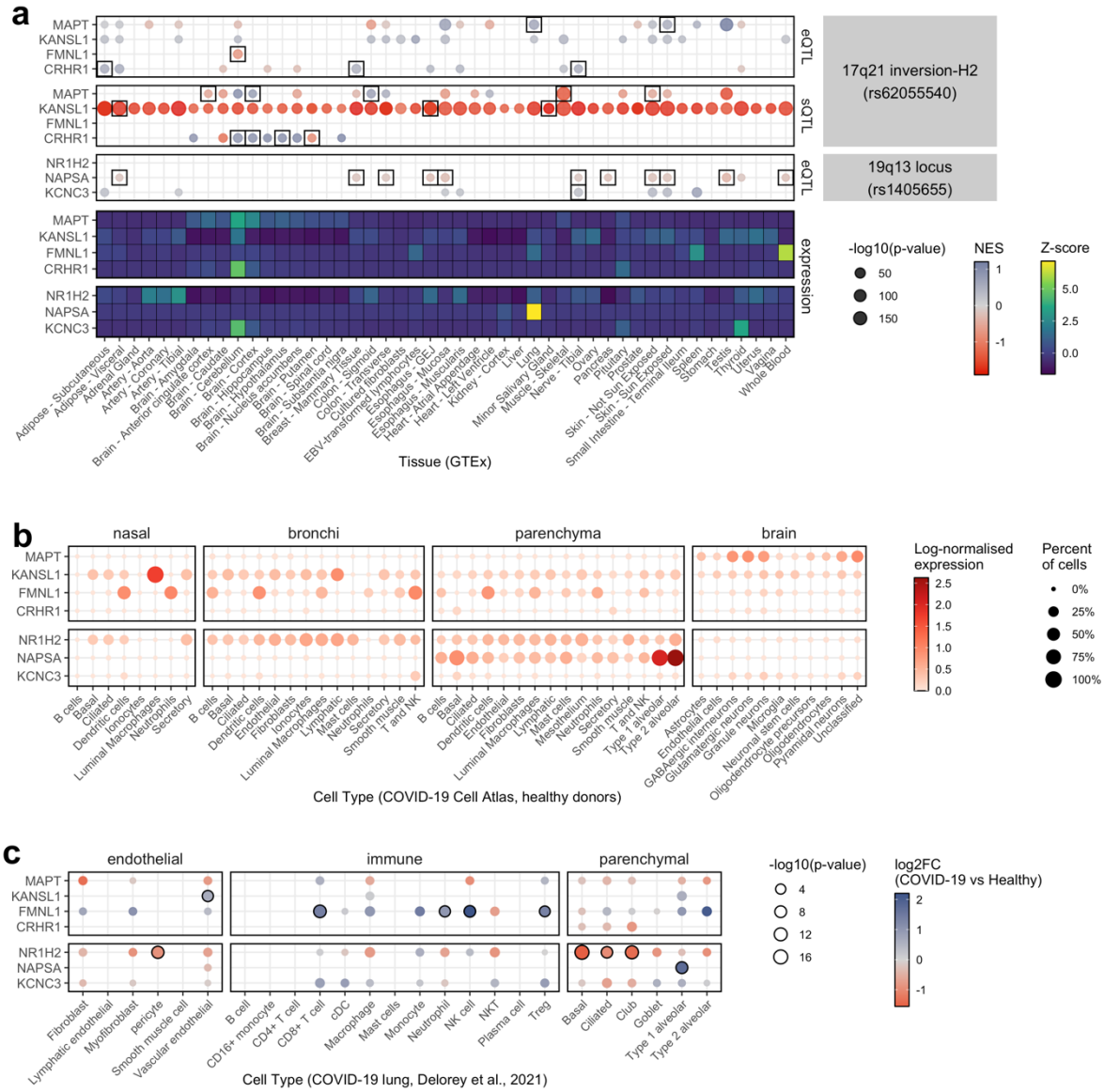

a
17q21 locus
rs1819040
17q21 inversion
95%
Credible set
r2
0.8
0.6
0.4
0.2
−log10(P)
45 Mb
46 Mb
47 Mb
Genome position (chr17)
rs1819040
(KANSL1 intron)
H1
ARHGAP27
CRHR1
MAPT
LRRC37A
ARL17A
FMNL1
PLEKHM1
SPPL2C
KANSL1
LRRC37A2
H2
b
17q21 inversion−H2
Severe COVID−19 GWAS group (Italy)
Severe COVID−19 GWAS group (Spain)
Severe COVID−19 GWAS group (Germany)
Severe COVID−19 GWAS group (Norway)
COVID−19 HGI (B2 ALL leave 23andme)
Meta−analysis (Fixed effects model)
all
60−80 y/o
40−60 y/o
Odds Ratio
c
17q21 inversion−H2 PheWAS
Lung function & disease
Lung function (FEV1)
Lung function (FVC)
Peak expiratory flow
Interstitial lung disease
Lung function in never smokers
Chronic obstructive pulmonary disease
Autoimmune disease
Primary biliary cirrhosis
Type 1 diabetes
White blood cell traits
Eosinophil counts
Neutrophil count
Lymphocyte percentage
Basophil+neutrophil counts
Myeloid white cell count
Granulocyte count
Neutrophil+eosinophil counts
Cancer
Breast cancer
Cancer (pleiotropy)
Red blood cell & hemoglobin traits
Red blood cell count
Hematocrit
Hemoglobin concentration
Hemoglobin levels
Red cell distribution width
Reticulocyte count
Neurodegenerative disease
Progressive supranuclear palsy
Parkinson's disease
Corticobasal degeneration
Direction of effect:
Decrease
Increase
Unknown
−log10(P)

Figure 2. **Expression analysis of the most plausible candidate genes associated with the 17q.21.31 and 19q.13.33 loci in organ tissues and COVID-19 relevant cell types.**

(a) GTEx tissue-specific expression QTL (eQTL, upper panel) and splicing QTL (sQTL, middle panel) effects of the 17q21.31 and 19q.13.33 loci on selected candidate genes, as well as expression of these genes in GTEx[23] tissues (lower panel). Direction of normalized eQTL and sQTL effect size (NES) is represented by color intensities, and statistical significance by dot size. Effects are calculated on the 17q21.31 inversion haplotype or 19q.13.33 allele (rs1405655-T). Black rectangles indicate genes for which these variants or proxy variants in high LD (r^2) are lead QTL variant in that tissue. If no box is shown other variants (i.e. not inversion or rs1405655-T) are lead QTLs. Heatmap displays gene-wise centered median by tissue expression values (represented by color intensities), showing in which tissues candidate genes are mostly enriched. **(b)** Expression levels of candidate genes in scRNA-seq datasets from healthy upper airways (nasal, bronchi) and lung (parenchyma) cells[48] and adult human brain cells from recently deceased, non-diseased donors[49]. Figure displays log-normalized mean expression (represented by color) and fraction of cells expressing those genes (indicated by dot size). Processed and cell-type annotated gene expression levels from studies were retrieved from COVID-19 Cell Atlas[50]. **(c)** The figure shows differential expression of candidate genes in lung cells of COVID-19 patients compared to healthy controls. Log_2 fold change (log2FC) values are presented as color gradient. Nominal p-values in $-log_{10}$ scale are shown proportionally to dot size. Black-bordered circles indicate significantly differentially expressed genes after FDR correction. Results were obtained from pseudo-bulk differential expression analysis by Delorey *et al*.[24] More in detailed figures are shown in **Supplementary Figures 13 and 15.**


a
MAPT
KANSL1
FMNL1
CRHR1
NR1H2
NAPSA
KCNC3
eQTL
sQTL
eQTL
expression
17q21 inversion-H2
(rs62055540)
19q13 locus
(rs1405655)
-log10(p-value)
50
100
150
NES
1
0
-1
Z-score
5.0
2.5
0.0
Tissue (GTEx)
b
nasal
bronchi
parenchyma
brain
Log-normalised
expression
2.5
2.0
1.5
1.0
0.5
0.0
Percent
of cells
0%
25%
50%
75%
100%
Cell Type (COVID-19 Cell Atlas, healthy donors)
c
endothelial
immune
parenchymal
-log10(p-value)
4
8
12
16
log2FC
(COVID-19 vs Healthy)
2
1
0
-1
Cell Type (COVID-19 lung, Delorey et al., 2021)

CONTRIBUTIONS

Contributions are shown in the **Supplementary Note.**

CONFLICTS OF INTEREST

Philipp Koehler has received lecture honoraria from or is advisor to Akademie für Infektionsmedizin e.V., Astellas Pharma, European Confederation of Medical Mycology, Gilead Sciences, GPR Academy Ruesselsheim, MSD Sharp & Dohme GmbH, Noxxon N.V., and University Hospital, LMU Munich outside the submitted work. Oliver A. Cornely has received research grants from, is an advisor to, or received lecture honoraria from Actelion, Allecra Therapeutics, Al-Jazeera Pharmaceuticals, Amplyx, Astellas, Basilea, Biosys, Cidara, Da Volterra, Entasis, F2G, Gilead, Grupo Biotoscana, IQVIA, Janssen, Matinas, Medicines Company, MedPace, Melinta Therapeutics, Menarini, Merck/MSD, Mylan, Nabriva, Noxxon, Octapharma, Paratek, Pfizer, PSI, Roche Diagnostics, Scynexis, and Shionogi. Stefano Duga received funding from the Banca Intesa San Paolo. Alberto Mantovani has received research funding from or reports personal fees from Dolce&Gabbana Fashion Firm, Ventana, Pierre Fabre, Verily, AbbVie, Astra Zeneca, Verseau Therapeutics, Compugen, Myeloid Therapeutics, Third Rock Venture, Imcheck Therapeutics, Ellipses, Novartis, Roche, Macrophage Pharma, BiovelocIta, Merck, Principia, Cedarlane Laboratories Ltd, HyCult Biotechnology, eBioscience, Biolegend, ABCAM Plc, Novus Biologicals, Enzo Life (ex Alexis Corp.), Affymetrix, BMS, Johnson&Johnson; in addition, he has a patent WO2019057780 "Anti-human migration stimulating factor (MSF) and uses thereof" issued, a patent WO2019081591 "NK or T cells and uses thereof" issued, a patent WO2020127471 "Use of SAP for the treatment of Euromycetes fungi infections" issued, and a patent EP20182181.6 "PTX3 as prognostic marker in Covid-19" pending. Ana Lleo has consulted for Intercept Pharma, Takeda, AlfaSigma, Abbvie, Gilead, and Merck Sharp, and Dohme.Manuel Romero Gómez has served as a speaker for AbbVie, Bristol-Myers Squibb, GENFIT, Gilead Sciences, Intercept, MSD and Roche; an advisory board member for GENFIT, Gilead Sciences, Intercept, Janssen-Cilag, Kaleido, NovoNordisk, Medimmune and Prosceinto; and has received research grants from Abbvie, Gilead Sciences and Intercept.Jan Cato Holter received a philanthropic donation from Vivaldi Invest A/S owned by Jon Stephenson von Tetzchner during the conduct of this study. Christoph Spinner reports grants, personal fees, and non-financial support from AbbVie; grants, personal fees, and non-financial support from Apeiron; grants, personal fees from BBraun, grants from Cepheid, personal fees from Formycon, grants, personal fees, and non-financial support from Gilead Sciences; grants and personal fees from Eli Lilly; grants, personal fees, and non-financial support from Janssen-Cilag; grants, personal fees, and non-financial support from GSK/ViiV Healthcare; grants, personal fees, and

non-financial support from MSD, outside the submitted work. Dr. Blasi reports grants and personal fees from Astrazeneca, grants and personal fees from Chiesi, grants and personal fees from Gsk, personal fees from Grifols, personal fees from Guidotti, personal fees from Insmed, grants and personal fees from Menarini, personal fees from Novartis, grants and personal fees from Pfizer, personal fees from Zambon, personal fees from Vertex, personal fee from Viatris outside the submitted work. Christoph Lange reports personal fees from Chiesi, Gilead, Janssen, Novartis, Oxfordimmunotec and Insmed outside the submitted work. Jan Heyckendorf reports personal fees from Chiesi, Gilead, and Janssen outside the submitted work.

ACKNOWLEDGEMENTS

We would like to thank the COVID-19 Human Genetics Initiative (HGI) for collecting and openly sharing extensive summary statistics with the community.

For the cohort from Milan Fondazione Ca' Granda we would like to acknowledge specifically Rossana Carpani. A full list of the investigators who contributed to the generation of the GCAT data is available from www.genomesforlife.com/. We thank Dr. Luis Puig and Vanessa Plegezuelos on behalf of the Blood and Tissue Bank from Catalonia (BST) who collaborated in the GCAT/COVICAT recruitment, and all the GCAT volunteers that participated in the study. We would also like to acknowledge the Task Force Humanitas and the Task Force Humanitas Gavazzeni Castelli (**Supplementary Note**). We would like to acknowledge all the participants from the "Grupo de Trabajo en Medicina Personalizada contra el COVID-19 de Andalucia". We also thank "Consejeria de Salud y Familias" and "Junta de Andalucía" for their support and funding that is currently supporting the COVID-19-GWAS and the COVID-PREMED initiatives. We are indebted to the HCB-IDIBAPS Biobank for the biological human samples and data procurement and to the Fundació Glòria Soler for its support to the COVIDBANK collection. We acknowledge the work of the Norwegian SARS-CoV-2 Study group (**Supplementary Note**). We also acknowledge Benedicte A. Lie at the Department of Immunology, Oslo University Hospital and the Norwegian Bone Marrow Donor Registry are acknowledged for providing healthy controls. We thank the support of the German COVID-19 multiomics initiative (decoi.eu). We thank the members of the COVID-19 use and access committee (Prof. Reinhold Förster, Prof. Christine Falk, Prof. Thomas Fühner, Prof. Christoph Höner zu Siederdissen, Prof. Marius Höper, Prof. Thomas Friedrich Schulz, Prof. Markus Cornberg, Prof. Thomas Illig and Prof. Michael Manns), as well as the study nurses, medical doctors and biobank staff members involved in the project. For funding we thank the Ministry of Science and Culture (MWK) of Lower-Saxony. We thank everybody involved in the COVID-19 registry at the University Hospital Regensburg (COVUR). We thank Tanja Wesse, Nicole

Braun, Anja Tanck and Xiaoli Yi for performing DNA extraction and genotyping at the Institute of Clinical Molecular Biology, Kiel, Germany. We thank Wolfgang Peter and Marco Schäfer from the Stefan Morsch Stiftung, Birkenfeld, Germany for their work on the HLA genotyping. Genotyping of the German control dataset was performed at the Regeneron Genetics Center. Genotyping of the COMPRI study was performed by the Genotyping laboratory of Institute for Molecular Medicine Finland FIMM Technology Centre, University of Helsinki. We are very grateful to Professor Akiko Iwasaki and Eric Song for generously sharing pre-processed single-cell RNA-seq data generated from SARS-CoV-2 infected brain organoids.

FUNDING

Andre Franke and David Ellinghaus were supported by a grant from the German Federal Ministry of Education and Research (01KI20197), Andre Franke, David Ellinghaus and Frauke Degenhardt were supported by the Deutsche Forschungsgemeinschaft Cluster of Excellence “Precision Medicine in Chronic Inflammation” (EXC2167). David Ellinghaus was supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the Computational Life Sciences funding concept (CompLS grant 031L0165). David Ellinghaus, Karina Banasik and Søren Brunak acknowledge the Novo Nordisk Foundation (grant NNF14CC0001 and NNF17OC0027594). Tobias L. Lenz, Ana Teles and Onur Özer were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project numbers 279645989; 433116033; 437857095. Mareike Wendorff and Hesham ElAbd are supported by the German Research Foundation (DFG) through the Research Training Group 1743, "Genes, Environment and Inflammation". This project was supported by a Covid-19 grant from the German Federal Ministry of Education and Research (BMBF; ID: 01KI20197). Luca Valenti received funding from: Ricerca Finalizzata Ministero della Salute RF-2016-02364358, Italian Ministry of Health ""CV PREVITAL – strategie di prevenzione primaria cardiovascolare primaria nella popolazione italiana; The European Union (EU) Programme Horizon 2020 (under grant agreement No. 777377) for the project LITMUS- and for the project ""REVEAL""; Fondazione IRCCS Ca' Granda ""Ricerca corrente"", Fondazione Sviluppo Ca' Granda ""Liver-BIBLE"" (PR-0391), Fondazione IRCCS Ca' Granda ""5permille"" ""COVID-19 Biobank"" (RC100017A). Andrea Biondi was supported by the grant from Fondazione Cariplo to Fondazione Tettamanti: "Bio-banking of Covid-19 patient samples to support national and international research (Covid-Bank). This study makes use of data generated by the GCAT-Genomes for Life. Cohort study of the Genomes of Catalonia, Fundacio IGTP. IGTP is part of the CERCA Program / Generalitat de Catalunya. GCAT is supported by Acción de Dinamización del ISCIII-MINECO and the Ministry of Health of the Generalitat of Catalunya (ADE 10/00026); the Agència de Gestió d’Ajuts Universitaris i de

Recerca (AGAUR) (2017-SGR 529). Marta Marquié received research funding from ant PI19/00335 Acción Estratégica en Salud, integrated in the Spanish National RDI Plan and financed by ISCIII-Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional (FEDER-Una manera de hacer Europa"). Beatriz Cortes is supported by national grants PI18/01512. Xavier Farre is supported by VEIS project (001-P-001647) (co-funded by European Regional Development Fund (ERDF), “A way to build Europe”). Additional data included in this study was obtained in part by the COVICAT Study Group (Cohort Covid de Catalunya) supported by IsGlobal and IGTP, EIT COVID-19 Rapid Response activity 73A and SR20-01024 La Caixa Foundation. Antonio Julià and Sara Marsal were supported by the Spanish Ministry of Economy and Competitiveness (grant numbers: PSE-010000-2006-6 and IPT-010000-2010-36). Antonio Julià was also supported the by national grant PI17/00019 from the Acción Estratégica en Salud (ISCIII) and the FEDER. The Basque Biobank is a hospital-related platform that also involves all Osakidetza health centres, the Basque government's Department of Health and Onkologikoa, is operated by the Basque Foundation for Health Innovation and Research-BIOEF. Mario Cacéres received Grants BFU2016-77244-R and PID2019-107836RB-I00 funded by the Agencia Estatal de Investigación (AEI, Spain) and the European Regional Development Fund (FEDER, EU). Manuel Romero Gómez, Javier Ampuero Herrojo, Rocío Gallego Durán and Douglas Maya Miles are supported by the “Spanish Ministry of Economy, Innovation and Competition, the Instituto de Salud Carlos III” (PI19/01404, PI16/01842, PI19/00589, PI17/00535 and GLD19/00100), and by the Andalussian government (Proyectos Estratégicos-Fondos Feder PE-0451-2018, COVID-Premed, COVID GWAs). The position held by Itziar de Rojas Salarich is funded by grant FI20/00215, PFIS Contratos Predoctorales de Formación en Investigación en Salud. Enrique Calderón's team is supported by CIBER of Epidemiology and Public Health (CIBERESP), "Instituto de Salud Carlos III". Jan Cato Holter reports grants from Research Council of Norway grant no 312780 during the conduct of the study. Dr. Solligård: reports grants from Research Council of Norway grant no 312769. The BioMaterialBank Nord is supported by the German Center for Lung Research (DZL), Airway Research Center North (ARCN). The BioMaterialBank Nord is member of popgen 2.0 network (P2N). Philipp Koehler has received non-financial scientific grants from Miltenyi Biotec GmbH, Bergisch Gladbach, Germany, and the Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases, University of Cologne, Cologne, Germany. He is supported by the German Federal Ministry of Education and Research (BMBF). Oliver A. Cornely is supported by the German Federal Ministry of Research and Education and is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – CECAD, EXC 2030 – 390661388. The COMRI cohort is funded by Technical University of Munich, Munich, Germany. Genotyping was performed by the Genotyping laboratory of Institute for Molecular

Medicine Finland FIMM Technology Centre, University of Helsinki. This work was supported by grants of the Rolf M. Schwiete Stiftung, the Saarland University, BMBF and The States of Saarland and Lower Saxony. Kerstin U. Ludwig is supported by the German Research Foundation (DFG, LU-1944/3-1). Genotyping for the BoSCO study is funded by the Institute of Human Genetics, University Hospital Bonn. Frank Hanses was supported by the Bavarian State Ministry for Science and Arts. Part of the genotyping was supported by a grant to Alfredo Ramirez from the German Federal Ministry of Education and Research (BMBF, grant: 01ED1619A, European Alzheimer DNA BioBank, EADB) within the context of the EU Joint Programme – Neurodegenerative Disease Research (JPND). Additional funding was derived from the German Research Foundation (DFG) grant: RA 1971/6-1 to Alfredo Ramirez. Philip Rosenstiel is supported by the DFG (CCGA Sequencing Centre and DFG ExC2167 PMI and by SH state funds for COVID19 research). Florian Tran is supported by the Clinician Scientist Program of the Deutsche Forschungsgemeinschaft Cluster of Excellence “Precision Medicine in Chronic Inflammation” (EXC2167). Christoph Lange and Jan Heyckendorf are supported by the German Center for Infection Research (DZIF).

[REFERENCES]

1. Booth, A. *et al.* Population risk factors for severe disease and mortality in COVID-19: A global systematic review and meta-analysis. *PLoS ONE* **16**, (2021).

2. Wong, C. K. H., Wong, J. Y. H., Tang, E. H. M., Au, C. H. & Wai, A. K. C. Clinical presentations, laboratory and radiological findings, and treatments for 11,028 COVID-19 patients: a systematic review and meta-analysis. *Sci. Rep.* **10**, (2020).

3. Peckham, H. *et al.* Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission. *Nat. Commun.* **11**, (2020).

4. Hu, Y. *et al.* Prevalence and severity of corona virus disease 2019 (COVID-19): A systematic review and meta-analysis. *Journal of Clinical Virology* **127**, (2020).

5. Severe Covid-19 GWAS Group. Genomewide association study of severe Covid-19 with respiratory failure. *N. Engl. J. Med.* **383**, (2020).

6. Ganna, A. Mapping the human genetic architecture of COVID-19 by worldwide meta-analysis. *medRxiv* (2021).

7. Pairo-Castineira, E. *et al.* Genetic mechanisms of critical illness in COVID-19. *Nature* **591**, (2021).

8. Shelton, J. F. *et al.* Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nat. Genet.* **53**, 801–808 (2021).

9. Mankelow, T. J. *et al.* Blood group type A secretors are associated with a higher risk of COVID-19 cardiovascular disease complications. *eJHaem* **2**, (2021).

10. Iturrieta-Zuazo, I. *et al.* Possible role of HLA class-I genotype in SARS-CoV-2 infection and progression: A pilot study in a cohort of Covid-19 Spanish patients. *Clin. Immunol.* **219**, (2020).

11. Novelli, A. *et al.* HLA allele frequencies and susceptibility to COVID-19 in a group of 99 Italian patients. *HLA* **96**, (2020).

12. Sanchez-Mazas, A. HLA studies in the context of coronavirus outbreaks. *Swiss medical weekly* **150**, (2020).

13. Hammer, M. A nomenclature system for the tree of human Y-Chromosomal binary haplogroups. *Genome Res.* **12**, (2002).

14. Arora, J. *et al.* HIV peptidome-wide association study reveals patient-specific epitope repertoires associated with HIV control. *Proc. Natl. Acad. Sci. U. S. A.* **116**, (2019).

15. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, (2018).

16. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, (2010).

17. McGuire, D. *et al.* Model-based assessment of replicability for genome-wide association meta-analysis. *Nat. Commun.* **12**, (2021).

18. Benner, C. *et al.* FINEMAP: Efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, (2016).

19. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nat. Genet.* **37**, (2005).

20. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3* **1**, 457–470 (2011).

21. Kwan, J. S. H., Li, M. X., Deng, J. E. & Sham, P. C. FAPI: Fast and accurate P-value Imputation for genome-wide association study. *Eur. J. Hum. Genet.* **24**, (2016).

22. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, (2017).

23. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (80-.).* **369**, (2020).

24. Delorey, T. M. *et al.* COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. *Nature* (2021). doi:10.1038/s41586-021-03570-8

25. Quach, H. *et al.* Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell* **167**, (2016).

26. WARD, H. E. & NICHOLAS, T. E. ALVEOLAR TYPE I AND TYPE II CELLS. *Aust. N. Z. J. Med.* **14**, (1984).

27. Melms, J. C. *et al.* A molecular single-cell lung atlas of lethal COVID-19. *Nature* (2021). doi:10.1038/s41586-021-03569-1

28. Islam, A. B. M. M. K. & Khan, M. A. A. K. Lung transcriptome of a COVID-19 patient and systems biology predictions suggest impaired surfactant production which may be druggable by surfactant therapy. *Sci. Rep.* **10**, (2020).

29. Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, (2016).

30. Sakornsakolpat, P. *et al.* Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nat. Genet.* **51**, (2019).

31. Tantisira, K. G., Lazarus, R., Litonjua, A. A., Klanderman, B. & Weiss, S. T. Chromosome 17: Association of a large inversion polymorphism with corticosteroid response in asthma. *Pharmacogenet. Genomics* **18**, (2008).

32. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).

33. de Jong, S. *et al.* Common inversion polymorphism at 17q21.31 affects expression of multiple genes in tissue-specific manner. *BMC Genomics* **13**, (2012).

34. Puig, M. *et al.* Determining the impact of uncharacterized inversions in the human genome by droplet digital PCR. *Genome Res.* **30**, (2020).

35. den Dekker, A. D. *et al.* TNF-α regulates diabetic macrophage function through the histone acetyltransferase MOF. *JCI Insight* **5**, (2020).

36. Cursano, S. *et al.* A CRHR1 antagonist prevents synaptic loss and memory deficits in a trauma-induced delirium-like syndrome. *Mol. Psychiatry* (2020). doi:10.1038/s41380-020-0659-y

37. Thompson, S. B. *et al.* Formin-like 1 mediates effector t cell trafficking to inflammatory sites to enable t cell-mediated autoimmunity. *Elife* **9**, (2020).

38. Pickett, B. E. *et al.* Virus pathogen Database and Analysis Resource (ViPR): A comprehensive bioinformatics Database and Analysis Resource for the Coronavirus research community. *Viruses* **4**, (2012).

39. Caffrey, T. M. & Wade-Martins, R. The role of MAPT sequence variation in mechanisms of disease susceptibility. in *Biochemical Society Transactions* **40**, (2012).

40. Shachar, S. Ben *et al.* MHC haplotyping of SARS-CoV-2 patients: HLA subtypes are not associated with the presence and severity of Covid-19 in the Israeli population. *medRxiv* **413**, (2020).

41. Kirkpatrick, M. How and why chromosome inversions evolve. *PLoS Biol.* **8**, (2010).

42. Wellenreuther, M. & Bernatchez, L. Eco-Evolutionary Genomics of Chromosomal Inversions. *Trends in Ecology and Evolution* **33**, (2018).

43. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

44. Viechtbauer, W. Conducting meta-analyses in R with the metafor. *J. Stat. Softw.* **36**, (2010).

45. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, (2021).

46. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, (2016).

47. So, H. C., Li, M. & Sham, P. C. Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. *Genet. Epidemiol.* **35**, (2011).

48. Vieira Braga, F. A. *et al.* A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat. Med.* **25**, (2019).

49. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, (2017).

50. Sungnak, W. *et al.* SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med.* **26**, (2020).

51. Stunnenberg, H. G. *et al.* The International Human Epigenome Consortium: A Blueprint for

Scientific Collaboration and Discovery. *Cell* **167**, (2016).

52. Song, E. *et al.* Neuroinvasion of SARS-CoV-2 in human and mouse brain. *J. Exp. Med.* **218**, (2021).

53. Finak, G. *et al.* MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, (2015).

54. Delaneau, O. *et al.* A complete tool set for molecular QTL discovery and analysis. *Nat. Commun.* **8**, (2017).

55. Degenhardt, F. *et al.* Construction and benchmarking of a multi-ethnic reference panel for the imputation of HLA class I and II alleles. *Hum. Mol. Genet.* (2019). doi:10.1093/hmg/ddy443

56. Degenhardt, F. *et al.* Trans-ethnic analysis of the human leukocyte antigen region for ulcerative colitis reveals common disease signatures. *Hum. Mol. Genet.* **ahead of p**, (2021).