[TITLE]

Detailed stratified GWAS analysis for severe COVID-19 in four 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], Malte C Rühlemann[1,8], Jatin Arora[9,10,11,12,13], Onur Özer[14,15], Ole Bernt Lenning[16,17], Ronny Myhre[18], May Sissel Vadla[17,19], Eike M Wacker[1], Lars Wienbrandt[1], Aaron Blandino Ortiz[20], Adolfo de Salazar[21,22], Adolfo Garrido Chercoles[23], Adriana Palom[24,25], Agustín Ruiz[26,27], Alba-Estela Garcia-Fernandez[28], Albert Blanco-Grau[28], Alberto Mantovani[29,30], Alberto Zanella[31,32], Aleksander Rygh Holten[33,34], Alena Mayer[35], Alessandra Bandera[31,32], Alessandro Cherubini[32], Alessandro Protti[29,30], Alessio Aghemo[29,30], Alessio Gerussi[36,37], Alfredo Ramirez[38,39,40,41,42], Alice Braun[35], Almut Nebel[1], Ana Barreira[25], Ana Lleo[29,30], Ana Teles[14,15], Anders Benjamin Kildal[43], Andrea Biondi[44], Andrea Caballero-Garralda[45], Andrea Ganna[46], Andrea Gori[32,47], Andreas Glück[48], Andreas Lind[49], Anja Tanck[1], Anke Hinney[50], Anna Carreras Nolla[51], Anna Ludovica Fracanzani[31,32], Anna Peschuck[1], Annalisa Cavallero[52], Anne Ma Dyrhol-Riise[34,53], Antonella Ruello[54], Antonio Julià[24], Antonio Muscatello[32], Antonio Pesenti[31,32], Antonio Voza[29,30], Ariadna Rando-Segura[55,56], Aurora Solier[57], Axel Schmidt[58], Beatriz Cortes[51], Beatriz Mateos[6,59], Beatriz Nafria-Jimenez[23], Benedikt Schaefer[60,61], Björn Jensen[62], Carla Bellinghausen[63], Carlo Maj[64], Carlos Ferrando[65,66], Carmen de la Horra[5,7,67,68,69], Carmen Quereda[70], Carsten Skurk[35], Charlotte Thibeault[35], Chiara Scollo[71], Christian Herr[72], Christoph D Spinner[73], Christoph Gassner[1,74], Christoph Lange[75,76,77], Cinzia Hu[78], Cinzia Paccapelo[32], Clara Lehmann[79,80,81], Claudio Angelini[82], Claudio Cappadona[30], Clinton Azuure[14,15], COVICAT study group, Covid-19 Aachen Study (COVAS), Cristiana Bianco[32], Cristina Cea[28], Cristina Sancho[83], Dag Arne Lihaug Hoff[84,85], Daniela Galimberti[31,32], Daniele Prati[32], David Haschka[86], David Jiménez[57], David Pestaña[87], David Toapanta[66], Eduardo Muñiz-Diaz[88], Elena Azzolini[29,30], Elena Sandoval[66], Eleonora Binatti[36,37], Elio Scarpini[31,32], Elisa T Helbig[35], Elisabetta Casalone[89], Eloisa Urrechaga[90,91], Elvezia Maria Paraboschi[29,30], Emanuele Pontali[92], Enric Reverter[66], Enrique J Calderón[5,7,67,68,69], Enrique Navas[70],

Erik Solligård[93,94], Ernesto Contro[95], Eunate Arana-Arri[96], Fátima Aziz[66], Federico Garcia[21,22], Félix García Sánchez[97], Ferruccio Ceriotti[32], Filippo Martinelli-Boneschi[98,99], Flora Peyvandi[100,101], Florian Kurth[35,102], Francesco Blasi[103,104], Francesco Malvestiti[31], Francisco J Medrano[7,67,68,69,105], Francisco Mesonero[6,59], Francisco Rodriguez-Frias[6,24,56,106], Frank Hanses[107,108], Fredrik Müller[34,49], Georg Hemmrich-Stanisak[1], Giacomo Bellani[109,110], Giacomo Grasselli[31,32], Gianni Pezzoli[111], Giorgio Costantino[31,32], Giovanni Albano[54], Giulia Cardamone[30], Giuseppe Bellelli[110,112], Giuseppe Citerio[110,113], Giuseppe Foti[109,110], Giuseppe Lamorte[32], Giuseppe Matullo[89], Guido Baselli[32], Hayato Kurihara[82], Holger Neb[114], Ilaria My[29], Ingo Kurth[115], Isabel Hernández[26,27], Isabell Pink[116], Itziar de Rojas[26,27], Iván Galván-Femenia[51], Jan Cato Holter[34,49], Jan Egil Afset[84,117], Jan Heyckendorf[75,76,77], Jan Kässens[1], Jan Kristian Damås[118,119], Jan Rybniker[79,80,120], Janine Altmüller[121], Javier Ampuero[5,6,7,67,122], Javier Martín[123], Jeanette Erdmann[124,125,126], Jesus M Banales[6,127,128], Joan Ramon Badia[129], Joaquin Dopazo[7,130], Jochen Schneider[73], Jonas Bergan[131], Jordi Barretina[132], Jörn Walter[133], Jose Hernández Quero[21,134], Josune Goikoetxea[135], Juan Delgado[5,7,67,68,69], Juan M Guerrero[5,7,67], Julia Fazaal[58], Julia Kraft[35], Julia Schröder[58], Kari Risnes[119,136], Karina Banasik[2], Karl Erik Müller[137], Karoline I Gaede[138,139,140], Koldo Garcia-Etxebarria[6,127], Kristian Tonby[34,53], Lars Heggelund[137,141], Laura Izquierdo-Sanchez[6,127,142], Laura Rachele Bettini[44], Lauro Sumoy[132], Leif Erik Sander[35], Lena J Lippert[35], Leonardo Terranova[32], Lindokuhle Nkambule[143,144], Lisa Knopp[62], Lise Tuset Gustad[93,145], Lucia Garbarino[146], Luigi Santoro[32], Luis Téllez[6,59], Luisa Roade[6,25,56], Mahnoosh Ostadreza[32], Maider Intxausti[83], Manolis Kogevinas[68,147,148,149], Mar Riveiro-Barciela[6,25,56], Marc M Berger[150], Marco Schaefer[151], Mari EK Niemi[46], María A Gutiérrez-Stampa[152], Maria Carrabba[78], Maria E. Figuera Basso[1], Maria Grazia Valsecchi[153], María Hernandez-Tejero[66], Maria JGT Vehreschild[154], Maria Manunta[32], Marialbert Acosta-Herrera[123], Mariella D'Angiò[44], Marina Baldini[32], Marina Cazzaniga[155], Marit M Grimsrud[34,156,157], Markus Cornberg[158], Markus M Nöthen[58], Marta Marquié[26,27], Massimo Castoldi[54], Mattia Cordioli[46], Maurizio Cecconi[29,30], Mauro D'Amato[128,159], Max Augustin[79,80,81], Melissa Tomasi[32], Mercè Boada[26,27], Michael Dreher[160], Michael J Seilmaier[161], Michael Joannidis[162], Michael Wittig[1], Michela Mazzocco[146], Michele Ciccarelli[82], Miguel Rodríguez-Gandía[6,59], Monica Bocciolone[82], Monica Miozzo[31,78], Natale Imaz Ayo[96], Natalia Blay[51], Natalia Chueca[22], Nicola Montano[31,32], Nicole Braun[1,163], Nicole Ludwig[164], Nikolaus Marx[165],

Nilda Martínez[166], Norwegian SARS-CoV-2 Study group, Oliver A Cornely[40,79,81,167], Oliver Witzke[168], Orazio Palmieri[169], Pa COVID-19 Study Group, Paola Faverio[110,170], Paoletta Preatoni[82], Paolo Bonfanti[110,171], Paolo Omodei[82], Paolo Tentorio[29], Pedro Castro[66], Pedro M Rodrigues[6,127,128,142], Pedro Pablo España[90], Per Hoffmann[58], Philip Rosenstiel[1], Philipp Schommers[79,80,81], Phillip Suwalski[35], Raúl de Pablo[20], Ricard Ferrer[172], Robert Bals[72], Roberta Gualtierotti[31,32], Rocío Gallego-Durán[5,6,7], Rosa Nieto[57], Rossana Carpani[32], Rubén Morilla[5,7,67,68,69], Salvatore Badalamenti[29], Sammra Haider[173], Sandra Ciesek[174,175], Sandra May[1], Sara Bombace[29,30], Sara Marsal[24], Sara Pigazzini[46], Sebastian Klein[162], Serena Pelusi[31,32], Sibylle Wilfling[108,176,177], Silvano Bosari[31,32], Sonja Volland[178], Søren Brunak[2], Soumya Raychaudhuri[9,10,11,12,13,179], Stefan Schreiber[1,48], Stefanie Heilmann-Heimbach[58], Stefano Aliberti[29,30], Stephan Ripke[35], Susanne Dudman[34,49], Tanja Wesse[1], Tenghao Zheng[180], The Humanitas COVID-19 Task Force, The Humanitas Gavazzeni COVID-19 Task Force, Thomas Bahmer[48], Thomas Eggermann[115], Thomas Illig[178], Thorsten Brenner[150], Tomas Pumarola[181,182], Torsten Feldt[62], Trine Folseraas[34,156,157,183], Trinidad Gonzalez Cejudo[184], Ulf Landmesser[185], Ulrike Protzer[186,187], Ute Hehr[176], Valeria Rimoldi[30], Valter Monzani[32], Vegard Skogen[188,189], Verena Keitel[62], Verena Kopfnagel[178], Vicente Friaza[5,7,67,68,69], Victor Andrade[38,39], Victor Moreno[68,190,191,192], Wolfgang Albrecht[1], Wolfgang Peter[151,193], Wolfgang Poller[35], Xavier Farre[51], Xiaoli Yi[1], Xiaomin Wang[35], Yascha Khodamoradi[154], Zehra Karadeniz[35], Anna Latiano[169], Siegfried Goerg[194], Petra Bacher[1,195], Philipp Koehler[40,79,80], Florian Tran[1,48], Heinz Zoller[60,61], Eva C Schulte[186,196,197], Bettina Heidecker[35], Kerstin U Ludwig[58], Javier Fernández[66,198], Manuel Romero-Gómez[5,6,7,67,122], Agustín Albillos[6,59], Pietro Invernizzi[36,37], Maria Buti[6,25,56], Stefano Duga[29,30], Luis Bujanda[6,127] , Johannes R Hov[34,156,157,183,+], Tobias L Lenz[14,15,+], Rosanna Asselta[29,30,+], Rafael de Cid[51,+], Luca Valenti[31,32,+], Tom H Karlsen[34,156,157,183,+], Mario Cáceres[4,199,+], Andre Franke[1,200,+]

[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] Institute for Medical Microbiology and Hospital Epidemiology, Hannover Medical School, Hannover, Germany.
[9] Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA.
[10] Division of Rheumatology, Inflammation and Immunity, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.
[11] Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA.
[12] Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA.
[13] Center for Data Sciences, Brigham and Women's Hospital, Boston, MA, USA.
[14] Research Group for Evolutionary Immunogenomics, Max Planck Institute for Evolutionary Biology, Plön, Germany.
[15] Research Unit for Evolutionary Immunogenomics, Department of Biology, University of Hamburg, Hamburg, Germany.
[16] Research Department, Stavanger University Hospital, Stavanger, Norway.
[17] Randaberg Municipality, Norway.
[18] Norwegian Institute of Public Health, Division of Health Data and Digitalization, Department of Genetics and Bioinformatics (HDGB) Oslo, Norway.
[19] University of Stavanger, Faculty of Health Sciences, Department of Quality and Health Technology, Stavanger, Norway.
[20] Department of Intensive Care, Hospital Universitario Ramón y Cajal, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), University of Alcalá, Madrid, Spain.
[21] Ibs.Granada Instituto de Investigación Biosanitaria, Granada, Spain.
[22] Microbiology Unit. Hospital Universitario Clinico San Cecilio, Granada, Spain.
[23] Osakidetza Basque Health Service, Donostialdea Integrated Health Organisation, Clinical Biochemistry Department, San Sebastian, Spain.

[24] Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Barcelona, Spain.
[25] Liver Unit, Department of Internal Medicine, Hospital Universitari Vall d'Hebron, Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain.
[26] Research Center and Memory Clinic. Ace Alzheimer Center Barcelona – Universitat Internacional de Catalunya, Spain.
[27] CIBERNED, Network Center for Biomedical Research in Neurodegenerative Diseases, National Institute of Health Carlos III, Madrid, Spain.
[28] Department of Biochemistry, University Hospital Vall d'Hebron, Barcelona, Spain.
[29] IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy.
[30] Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy.
[31] University of Milan, Milan, Italy.
[32] Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy.
[33] Department of Acute Medicine, Oslo University Hospital, Oslo, Norway.
[34] Institute of Clinical Medicine, University of Oslo, Oslo, Norway.
[35] Charite Universitätsmedizin Berlin, Berlin, Germany.
[36] European Reference Network on Hepatological Diseases (ERN RARE-LIVER), San Gerardo Hospital, Monza, Italy.
[37] Division of Gastroenterology, Center for Autoimmune Liver Diseases, School of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy.
[38] Division of Neurogenetics and Molecular Psychiatry, Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany.
[39] Department of Neurodegenerative Diseases and Geriatric Psychiatry, University Hospital Bonn, Medical Faculty, Bonn, Germany.
[40] University of Cologne, Faculty of Medicine and University Hospital Cologne, Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), Cologne, Germany.
[41] German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany.
[42] Department of Psychiatry and Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, San Antonio, TX, USA.
[43] Department of Anesthesiology and Intensive Care, University Hospital of North Norway, Tromsø, Norway.

[44] Pediatric Departement and Centro Tettamanti- European Reference Network (ERN) PaedCan, EuroBloodNet, MetabERN-University of Milano-Bicocca-Fondazione MBBM/Ospedale San Gerardo, Italy.
[45] Biochemistry Department, Echevarne Laboratory, Sant Cugat del Vallés, Barcelona, Spain.
[46] Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland.
[47] Centre for Multidisciplinary Research in Health Science (MACH), University of Milan, Milan, Italy.
[48] Klinik für Innere Medizin I, Universitätsklinikum Schleswig-Holstein, Campus Kiel, Germany.
[49] Department of Microbiology, Oslo University Hospital, Oslo, Norway.
[50] Department of Child and Adolescent Psychiatry, University Hospital Essen, University of Duisburg-Essen, Essen, Germany.
[51] Genomes for Life-GCAT lab. Germans Trias i Pujol Research Institute (IGTP), Badalona, Spain.
[52] Laboratory of Microbiology, San Gerardo Hospital, Monza, Italy.
[53] Department of Infectious diseases, Oslo University Hospital, Oslo, Norway.
[54] Humanitas Gavazzeni-Castelli, Bergamo, Italy.
[55] Microbiology Department, Hospital Universitari Vall d'Hebron, Barcelona, Spain.
[56] Universitat Autònoma de Barcelona, Bellatera, Spain.
[57] 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.
[58] Institute of Human Genetics, University of Bonn School of Medicine & University Hospital Bonn, Bonn, Germany.
[59] Department of Gastroenterology, Hospital Universitario Ramón y Cajal, University of Alcalá, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), Madrid, Spain.
[60] Medical University of Innsbruck, Department of Medicine I, Gastroenterology, Hepatology and Endocrinology, Innsbruck, Austria.
[61] Christian Doppler Laboratory of Iron and Phosphate Biology at the Department of Medicine I, Medical University of Innsbruck, Innsbruck, Austria.

[62] Department of Gastroenterology, Hepatology and Infectious Diseases, University Hospital Duesseldorf, Medical Faculty Heinrich Heine University, Duesseldorf, Germany.
[63] Department of Respiratory Medicine and Allergology, University Hospital, Goethe University, Frankfurt am Main, Germany.
[64] Institute of Genomic Statistics and Bioinformatics, University Hospital Bonn, Medical Faculty University of Bonn, Venusberg-Campus [1], Bonn, Germany.
[65] Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Madrid, Spain.
[66] Hospital Clinic, University of Barcelona, and IDIBAPS, Barcelona, Spain.
[67] University of Sevilla, Sevilla, Spain.
[68] Ko (CIBERESP), Madrid, Spain.
[69] Consejo Superior de Investigaciones científicas, Sevilla, Spain.
[70] Department of Infectious Diseases, Hospital Universitario Ramón y Cajal, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), University of Alcalá, Madrid, Spain.
[71] Department of Transfusion Medicine and Haematology Laboratory, San Gerardo Hospital, Monza, Italy.
[72] Department of Internal Medicine V - Pneumology, Allergology and Intensive Care Medicine, University Hospital Saarland, Homburg/Saar, Germany.
[73] Technical University of Munich, School of Medicine, University Hospital rechts der Isar, Department of Internal Medicine II, Munich, Germany.
[74] Private University in the Principality of Liechtenstein.
[75] Respiratory Medicine & International Health, University of Lübeck, Lübeck, Germany.
[76] Division of Clinical Infectious Diseases, Research Center Borstel, Borstel, Germany.
[77] German Center for Infection Research (DZIF) Clinical Tuberculosis Unit, Borstel, Germany.
[78] Internal Medicine Department,Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy.
[79] Department I of Internal Medicine, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany.

[80] Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany.
[81] German Center for Infection Research (DZIF), Partner Site Bonn-Cologne, Cologne, Germany.
[82] Humanitas Clinical and Research Center, IRCCS, Milan, Italy.
[83] Osakidetza Basque Health Service, Basurto University Hospital, Respiratory Service, Bilbao, Spain.
[84] Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway.
[85] Department of Medicine, Møre & Romsdal Hospital Trust, Ålesund, Norway.
[86] Department of Internal Medicine II, Medical University of Innsbruck, Innsbruck, Austria.
[87] Department of Anesthesiology and Critical Care, Hospital Universitario Ramón y Cajal, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), University of Alcalá, Madrid, Spain.
[88] Immunohematology Department, Banc de Sang i Teixits, Autonomous University of Barcelona, Barcelona, Spain.
[89] Department of Medical Sciences, Università degli Studi di Torino, Turin, Italy.
[90] Osakidetza Basque Health Service, Galdakao Hospital, Respiratory Service, Galdakao, Spain.
[91] Biocruces Bizkaia Health Research Institute.
[92] Department of Infectious Diseases - E.O. Ospedali Galliera, Genova, Italy.
[93] Geminicenter for Sepsis Research, Institute of Circulation and Medical Imaging (ISB), NTNU, Trondheim, Norway.
[94] Clinic of Anesthesia and Intensive Care, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.
[95] Accident & Emergency and Emergency Medicine Unit, San Gerardo Hospital, Monza, Italy.
[96] Biocruces Bizkaia Health Research Institute, Barakaldo, Spain.
[97] Histocompatibilidad y Biologia Molecular, Centro de Transfusion de Madrid, Madrid, Spain.
[98] Dino Ferrari Center, Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy.

[99] IRCCS Fondazione Ca' Granda Ospedale Maggiore Policlinico, Neurology Unit, Milan, Italy.
[100] University of Milan, Department of Pathophysiology and Transplantation, Milan, Italy.
[101] Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Angelo Bianchi Bonomi Hemophilia and Thrombosis Center, Milan, Italy.
[102] Department of Tropical Medicine, Bernhard Nocht Institute for Tropical Medicine, and Department of Medicine I, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany.
[103] Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Respiratory Unit, Milan, Italy.
[104] Department of Pathophysiology and Transplantation, Università degli Studi di Milano, Italy.
[105] Internal Medicine Department, Virgen del Rocio University Hospital, Sevilla, Spain.
[106] Biochemistry Department, University Hospital Vall d'Hebron, Barcelona, Spain.
[107] Emergency Department, University Hospital Regensburg, Regensburg, Germany.
[108] Department for Infectious Diseases and Infection Control, University Hospital Regensburg, Regensburg, Germany.
[109] Department Emergency, Anesthesia and Intensive Care, San Gerardo Hospital, Monza, Italy.
[110] School of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy.
[111] Fondazione Grigioni per il Morbo di Parkinson and Parkinson Institute, ASST Gaetano Pini-CTO, Milan, Italy.
[112] Acute Geriatric Unit, San Gerardo Hospital, Monza, Italy.
[113] Neurointensive Care Unit, San Gerardo Hospital, Monza, Italy.
[114] Department of Anesthesiology, Intensive Care Medicine and Pain Therapy, University Hospital Frankfurt, Frankfurt am Main, Germany.
[115] Institute of Human Genetics, Medical Faculty, RWTH Aachen University, Aachen, Germany.
[116] Department of Pneumology, Hannover Medical School, Hannover, Germany.
[117] Department of Medical Microbiology, Clinic of Laboratory Medicine, St. Olavs hospital, Trondheim, Norway.

[118] Department of Infectious Diseases, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.
[119] Department of Clinical and Molecular Medicine, NTNU, Trondheim, Norway.
[120] German Centre for Infection Research (DZIF), Partner Site Bonn-Cologne, Cologne, Germany.
[121] Cologne Center for Genomics (CCG), University of Cologne, Cologne, Germany.
[122] Digestive Diseases Unit, Virgen del Rocio University Hospital, Institute of Biomedicine of Seville, University of Seville, Seville, Spain.
[123] Institute of Parasitology and Biomedicine Lopez-Neyra, CSIC, Granada, Spain.
[124] Institute for Cardiogenetics, University of Lübeck, Lübeck, Germany.
[125] German Research Center for Cardiovascular Research, partner site Hamburg–Lübeck–Kiel, Lübeck, Germany.
[126] University Heart Center Lübeck, Lübeck, Germany.
[127] 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.
[128] Ikerbasque, Basque Foundation for Science, Bilbao, Spain.
[129] Respiratory ICU, Institut Clínic Respiratory, Hospital Clinic, University of Barcelona, and IDIBAPS, Barcelona, Spain.
[130] Bioinformatics Area, Fundación Progreso y Salud, and Institute of Biomedicine of Sevilla (IBIS), Sevilla, Spain.
[131] Department of Research, Ostfold Hospital Trust, Gralum, Norway.
[132] Germans Trias i Pujol Research Institute (IGTP), Badalona, Spain.
[133] Department of Genetics & Epigenetics, Saarland University, Saarbrücken, Germany.
[134] Department of Infectious Diseases, Hospital Universitario Clinico San Cecilio, Granada, Spain.
[135] Infectious Diseases Service, Osakidetza, Biocruces Bizkaia Health Research Institute, Barakaldo, Spain.
[136] Department of Research, St Olav Hospital, Trondheim University Hospital, Trondheim, Norway.
[137] Medical Department, Drammen Hospital, Vestre Viken Hospital Trust, Norway.
[138] Research Center Borstel, BioMaterialBank Nord, Germany.

[139] German Center for Lung Research (DZL), Airway Research Center North (ARCN), Germany.
[140] Popgen 2.0 network (P2N), Kiel, Germany.
[141] Department of Clinical Science, University of Bergen, Bergen, Norway.
[142] Biodonostia Health Research Institute, Donostia University Hospital, San Sebastian, Spain.
[143] Analytic & Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA.
[144] Stanley Center for Psychiatric Research & Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA.
[145] Clinic of Medicine and Rehabilitation, Levanger Hospital, Nord-Trondelag Hospital Trust, Levanger, Norway.
[146] HLA Laboratory – E.O. Ospedali Galliera, Genova, Italy.
[147] ISGlobal, Barcelona, Spain.
[148] Universitat Pompeu Fabra (UPF), Barcelona, Spain.
[149] IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain.
[150] Department of Anesthesiology and Intensive Care Medicine, University Hospital Essen, University Duisburg-Essen, Essen, Germany.
[151] Stefan-Morsch-Stiftung, Birkenfeld, Germany.
[152] Osakidetza, OSI Donostialdea, Altza Primary Care, Biodonostia Health Research Institute, San Sebastián, Spain.
[153] Center of Bioinformatics, Biostatistics and Bioimaging, School of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy.
[154] Department of Internal Medicine, Infectious Diseases, University Hospital Frankfurt & Goethe University Frankfurt, Frankfurt am Main, Germany.
[155] Phase [1] Research Centre, ASST Monza, School of Medicine and Surgery, University of Milano-Bicocca, Italy.
[156] Research Institute for Internal Medicine, Division of Surgery, Inflammatory Diseases and Transplantation, Oslo University Hospital Rikshospitalet and University of Oslo, Oslo, Norway.
[157] Norwegian PSC Research Center, Department of Transplantation Medicine, Division of Surgery, Inflammatory Diseases and Transplantation, Oslo University Hospital Rikshospitalet, Oslo, Norway.

[158] Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany.
[159] Gastrointestinal Genetics Lab, CIC bioGUNE - BRTA, Derio, Spain.
[160] Department of Pneumology and Intensive Care Medicine, University Hospital Aachen, Germany.
[161] Munich Clinic Schwabing, Academic Teaching Hospital, Ludwig-Maximilians-University (LMU), Munich, Germany.
[162] Division of Intensive Care and Emergency Medicine, Department of Internal Medicine, Medical University Innsbruck, Innsbruck, Austria.
[163] University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany.
[164] Center of Human and Molecular Biology, Department of Human Genetics, University Hospital Saarland, Homburg/Saar, Germany.
[165] Department of Internal Medicine I, RWTH Aachen University Hospital, Aachen, Germany.
[166] Department of Anesthesiology, Hospital Universitario Ramón y Cajal, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), Madrid, Spain.
[167] University of Cologne, Faculty of Medicine and University Hospital Cologne, Clinical Trials Centre Cologne (ZKS Köln), Cologne, Germany.
[168] Department of Infectious Diseases, University Hospital Essen, University Duisburg-Essen, Essen, Germany.
[169] Gastroenterology Unit, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy.
[170] Pulmonary Unit, San Gerardo Hospital, Monza, Italy.
[171] Infectious Diseases Unit, San Gerardo Hospital, Monza, Italy.
[172] Intensive Care Department, Vall d'Hebron University Hospital, SODIR-VHIR research group, Barcelona, Spain.
[173] Department of Medicine, Møre & Romsdal Hospital Trust, Molde, Norway.
[174] Institute of Medical Virology, University Hospital Frankfurt, Goethe University, Frankfurt am Main, Germany.
[175] German Centre for Infection Research (DZIF), External Partner Site Frankfurt, Frankfurt am Main, Germany.
[176] Zentrum für Humangenetik Regensburg, Regensburg, Germany.
[177] Department of Neurology, Bezirksklinikum Regensburg, University of Regensburg, Regensburg, Germany.

[178] Hannover Unified Biobank, Hannover Medical School, Hannover, Germany.
[179] Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, Manchester Academic Health Science Centre, The University of Manchester, Manchester, UK.
[180] School of Biological Sciences, Monash University, Clayton, VIC, Australia.
[181] Department of Microbiology, University Hospital Vall d'Hebron, Barcelona, Spain.
[182] Autonoma University of Barcelona, Barcelona, Spain.
[183] Section for Gastroenterology, Department of Transplantation Medicine, Division for Cancer Medicine, Surgery and Transplantation, Oslo University Hospital Rikshospitalet, Oslo, Norway.
[184] Biochemistry Unit. Hospital Universitario Clinico San Cecilio, Granada, Spain.
[185] Charite Universitätsmedizin Berlin, Berlin Institute of Health, Berlin Germany.
[186] Institute of Virology, Technical University Munich/Helmholtz Zentrum München, Munich, Germany.
[187] German Center for Infection Research (DZIF), Munich partner site, Munich, Germany.
[188] Department of Infectious Diseases, University Hospital of North Norway, Tromsø, Norway.
[189] Faculty of Health Sciences, UIT The Arctic University of Norway, Norway.
[190] Catalan Institute of Oncology (ICO), Barcelona, Spain.
[191] Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain.
[192] Universitat de Barcelona (UB), Barcelona, Spain.
[193] University of Cologne, Faculty of Medicine and University Hospital Cologne, Institute for Transfusion Medicine, Cologne, Germany.
[194] Institute of Transfusionsmedicine, University Hospital Schleswig-Holstein (UKSH), Germany.
[195] Institute of Immunology, Christian-Albrechts-University of Kiel & UKSH Schleswig-Holstein, Kiel, Germany.
[196] Institute of Psychiatric Phenomics and Genomics, University Medical Center, University of Munich, Munich, Germany.
[197] Department of Psychiatry, University Medical Center, University of Munich, Munich, Germany.
[198] European Foundation for the Study of Chronic Liver Failure (EF-CLIF), Barcelona, Spain.

[199] ICREA, Barcelona, Spain.

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

* The authors wish it to be known that, in their opinion, the first 6 authors should be regarded as joint First Authors
+ The authors wish it to be known that, in their opinion, the last 7 authors should be regarded as joint Last Authors.

[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]

Given the highly variable clinical phenotype of Coronavirus disease 2019 (COVID-19), a deeper analysis of the host genetic contribution to severe COVID-19 is important to improve our understanding of underlying disease mechanisms. Here, we describe an extended GWAS meta-analysis of a well-characterized cohort of 3,260 COVID-19 patients with respiratory failure and 12,483 population controls from Italy, Spain, Norway and Germany/Austria, including stratified analyses based on age, sex and disease severity, as well as targeted analyses of chromosome Y haplotypes, the human leukocyte antigen (HLA) region and the SARS-CoV-2 peptidome. By inversion imputation, we traced a reported association at 17q21.31 to a highly pleiotropic ~0.9-Mb inversion polymorphism and characterized the potential effects of the inversion in detail. Our data, together with the 5th release of summary statistics from the COVID-19 Host Genetics Initiative, also identified a new locus at 19q13.33, including *NAPSA,* a gene which is expressed primarily in alveolar cells responsible for gas exchange in the lung.

[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 312 million confirmed cases and 5.5 million COVID-19 related deaths worldwide (frequencies reported by the World Health Organization, January 12th, 2022). 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 including diabetes, active cancer, hypertension and coronary artery disease as well as solid organ transplant or other conditions that promote an immunosuppressive state.(1–4)

Different studies, including the first genetic analysis of our Severe COVID-19 GWAS study group, have shown that genetic predisposition plays a role in COVID-19 susceptibility and severity.(5–7) In particular, we reported significant associations

between genetic variants at loci 3p21.31 (around *LZTF1*) and 9q34.2 (*ABO* blood group locus) to severe respiratory failure and SARS-CoV-2 infection in 1,980 severe COVID-19 patients and 2,160 population controls of European ancestry. Since then, these loci have been replicated in subsequent studies and extended also to non-European cohorts.(8–10) 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, more recently, the largest genetic consortium for COVID-19, the COVID-19 Host Genetics initiative (HGI), which contains 3,815 individuals from the Severe COVID-19 GWAS study group.(6,7) Six of these loci have been linked to critical illness by COVID-19, and include genes that were previously associated with pulmonary or autoimmune and inflammatory diseases.(7)

Since our primary publication, the Severe COVID-19 GWAS study group dataset has been extended to 3,260 severe COVID-19 patients with respiratory failure(5) and 12,483 controls with unknown COVID-19 status from 30 study centers across Italy, Spain, Norway and Germany/Austria (**Supplementary Table 1a**). Further, detailed information on age, sex and comorbidities as hypertension, diabetes, and coronary artery disease is available in this dataset. The control cohort included predominantly population controls recruited at the respective centers as blood or bone marrow donors with unknown or negative COVID-19 status (**Supplementary Table 1c**). With this unique resource at hand, we performed an extended genome-wide association study (GWAS) and meta-analysis of severe COVID-19 as well as a meta-analysis with the COVID-19 HGI release 5 GWAS summary statistics, followed by stratified analyses of age, sex and disease and a detailed analysis of ABO secretor status. Using Y-chromosomal genotype calls from the Global Screening Array (GSA), we additionally performed an imputation and disease association of chromosome Y haplotypes. Compared to our first study, we also performed an even more detailed analysis of the HLA which 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)(11) 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. Finally, by inversion imputation, we

present a functional analysis of the established 17q21.31 locus(7), which has been described by COVID-19 HGI but not characterized in depth.

[Results]

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

Genotyping was carried out using Illumina's 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 Table 1**). As before, respiratory failure 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] Analogously to the COVID-19 HGI(7), we 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 and 12,483 controls; first analysis) and second, a more stringent definition of severe COVID-19 "hospitalization with mechanical ventilation" (classes 2-4 with 1,911 critically ill individuals and 12,226 controls; second analysis). Compared to our previous study, we increased the number of patients and controls by approximate factors of 2 and 5.5 respectively (from 1,610 and 2,205). 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(12), followed by fixed-effect inverse variance meta-analysis using METAL(13) (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 principal component analysis (PCA) as employed by the COVID-19 HGI; **Online Methods**) revealed genome-wide significant associations ($P<5\times10^{-8}$; **Supplementary Table 3**) at four known loci: chr3:45848457:C:T (rs35731912) within *LZTFL1* at the 3p21.31 locus, chr9:133261703:A:G (rs687289*)* within *ABO* at 9q34.2, chr19:10351837:C:T (rs687289) within *TYK2* at 19p13.2 and chr19:4717660:A:G (rs12610495) within *DPP9* at 19p13.3 in the first or second analysis. Variants at four loci (*PCDH7* at 4p15.1; *FREM1* at 9p22.3; *MAPT* at 17q21.31; *DPP9* at 19p13.3) had suggestive

significance ($P<10^{-6}$) in our first analysis, and an additional five loci (*OLMF4* at 13q21.1; *TTC7B* at 14q32.11; *CPD* at 17q11.2; *PTPRM* at 18p11.23; *IFNAR2* at 21q22.11) had suggestive significance in the second analysis (**Figure 1**; **Supplementary Table 3**, regional association plots shown in **Supplementary Figures 4-5**). 3 loci not included in the list above had evidence for excessive heterogeneity ($P <10^{-5}$). Of note, associations at 17q21.31 (*MAPT*), 19p13.3 (*DPP9*) and 21q22.11 (*IFNAR2*) have been previously reported by the COVID-19 HGI.(7) The respective lead risk variants (this study vs. COVID-19 HGI) at reported loci are linked with r^2 or $D' >0.9$ across our study cohorts (**Supplementary Table 4**). Hierarchical mixture model analysis with MAMBA(14) (**Online Methods**) showed a high posterior probability of replication (PPR>0.8) of consistent effect sizes across all analyzed cohorts for all but chr14:90679345:C:T at 14q32.11 (*TTC7B*, PPR=0.005) and chr17:30409086:A:T at 17q11.2 (CPD, PPR=0.038) (**Supplementary Table 3**). The chr17:45933112:G:A *MAPT* variant (rs8065800), showed higher replicability in our first analysis compared to the second analysis (*MAPT*, PPR=0.93/PPR=0.2). None of the novel and replicable associations at *PCDH7* at 4p15.1; *FREM1* at 9p22.3; *OLMF4* at 13q21.1 and *PTPRM* at 18p11.23 had suggestive associations in the COVID-19 HGI 5 release dataset, which is why we did not consider them further. Future analyses will explore whether these are attributable to artefacts in the cohorts of this study or maybe a too high heterogeneity of cohorts and data sets included in the COVID-19 HGI analyses.

Next, we performed a replication analysis of the 13 variants reported to be genome-wide significantly associated with COVID-19 by the COVID-19 HGI using only individuals from the Severe COVID-19 GWAS group not included in the COVID-19 HGI release 5 datasets (**Supplementary Methods**). We replicated 8 of the 13 variants reported as genome-wide significant by the COVID-19 HGI(7) at least at nominal P-value of 0.05 (located near *LZTFL1*, *ABO*, *TYK2*, *DDP9*, *IFNAR2*, MAPT/*KANSL1*, *OAS1*) **Supplementary Table 5**).

Subsequently, we performed a fixed-effect inverse variance meta-analysis using METAL(13) 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 (6,526 cases and 719,075 controls) and 1,320,760 (14,467 and cases and 1,306,293 controls) analyzed individuals in the respective analyses as well as 9,163,456 and

9,309,373 high quality variants (imputation $R^2 \geq 0.6$ in our cohorts and overall minor allele frequency (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 ($P=3.27\times10^{-11}$ for rs1819040; OR=0.88 for minor allele T; 95%CI=0.84-0.92) over rs8065800 from the first analysis (**Supplementary Tables 3,4,6**). 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, chr19:50379362:T:C (rs1405655, $P=3.25\times10^{-8}$; OR=1.09; 95%CI=1.06-1.1) located in the *NAPSA* gene at 19q13.33; (regional association plot is shown in **Supplementary Figure 8, Supplementary Table 6**). Bayesian fine-mapping analysis of this locus with FINEMAP(15) (**Online Methods**) identified a total of 15 ($\log_{10}$(Bayes factor)=5.95) variants that belong to the 95% most likely to be causal (**Figure 1, Supplementary Table 7**).

In-depth stratified analysis of lead variants

To estimate differences in genetic effects from age, sex and disease severity groups, we performed an in-depth stratified analysis for our discovery populations of the four replicable genome-wide significant hits from our first and second analyses, all of the 13 (including different variants at the aforementioned four loci) variants from the COVID-19 HGI release 5 analysis(7) and the novel association, chr19:50379362:T:C, at the 19q13.33 locus (**Supplementary Methods**). We additionally investigated association of these variants to known comorbidities such as hypertension, coronary artery disease and diabetes in cases only (**Supplementary Table 8**) and performed sex-stratified analysis within age and severity groups. We compared sex-specific and age-specific 95%CIs of ORs to determine statistical significance when CIs did not overlap. We confirmed our previously described association with age and disease severity at 3p21.31 for chr3:45848457:C:T (first analysis; ages 41-60; $P=2.15\times10^{-23}$, OR=2.20, 95%CI=1.88-2.57; ages 61-80; $P=6.41\times10^{-6}$; OR=1.49, 95%CI=1.25-1.77; severity: $P=9.73\times10^{-7}$, OR=1.65; 95%CI=1.35-2.03), which is also discussed in detail in a study by Nakanishi *et al*.(16) In addition, the chr17:46142465:T:A the (rs1819040) variant from the COVID-19 HGI study was associated with a protective OR that was consistently observed across the Italian, Spanish, and German/Austrian study cohorts in the 41-60 years age group (first analysis: $P=4.59\times10^{-7}$, OR=0.74, 95%CI=0.66-

0.83), while no association was observed in the equally powered 61-80 years age group (P=0.491, OR=0.96, 95%CI=0.86-1.07, **Supplementary Table 8, Supplementary Figures 9-10**). This was especially pronounced in the male individuals. Non-significant (here comparing the overlap of the 95%CIs of the ORs), strong trends for association with age were observed for chr19:4717660:A:G (rs12610495; *DPP9*) at the 19p13.3 locus (first analysis: ages 41-60; P=3.8x10^{-7}; OR=1.33, 95%CI=1.19-1.48; ages 61-80; P=0.0137; OR=1.15, 95%CI=1.03-1.28) and chr21:33242905:T:C (rs13050728; *IFNAR2*) which was significantly associated with severe COVID-19 only in younger individuals (first analysis; age <=60; P=5.01x10^{-7}; OR=1.27, 95%CI=1.16-1.4; age > 60: P=0.109 OR=1.08; 95%CI=0.98-1.18). We did not observe any statistically significant associations with age, sex, severity or comorbidities for the remaining analyzed lead variants.

The genome-wide significant association near *LTZF1*, *TYK2*, *DDP9* and *ABO* have been followed up in detail before. Therefore, we next focused on the novel association at the 19q13.33 (*NAPSA*) locus from the meta-analysis with COVID-19 HGI release 5 summary statistics and an in-depth analysis of the 17q21.31 (*MAPT, KANSL1*) locus associated with age.

The 17q21.31 locus has been previously reported to be linked to a common inversion polymorphism.(17) Indeed, Bayesian fine-mapping(15) at this locus shows that the 95% credible set includes 1,530 variants in high linkage disequilibrium (LD; r^2) (mean(r^2)=0.997 and min(r^2)=0.954) (among these chr17:45933112:G:A (rs8065800) from our initial discovery analysis and chr17:46142465:T:A (rs1819040) from the COVID-19 HGI analysis with certainty <0.3%, indicating that the individual SNP associations are only proxy variants for the likely causal variant, the inversion polymorphism (**Figure 1a**, **Supplementary Table 9**). We therefore imputed the inversion haplotypes H1 and H2 for our cohorts by genotype imputation with IMPUTE2(18), 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 chr17:46142465:T:A (rs1819040) variant, which was prioritized over chr17:45933112:G:A (rs8065800) 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). 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 9**). Using stratified analysis, we also confirmed the age effect for the inversion (**Supplementary Table 7 and Figure 1b**). We did not observe any association of the inversion with disease severity.

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 the associations at the 17q21.31 inversion and at the novel 19q13.33 locus, including a phenome-wide association study (PheWas) and exploratory gene expression analysis.

We queried variants in high LD (r^2>0.9) with chr19:50379362:T:C (rs1405655) or the 17q21.31 inversion, using a wide range of phenotypes from the NHGRI GWAS Catalog (**Online Methods**).(19) While no known phenotypes were found to be linked to chr19:50379362:T:C (rs140565) or its proxy variants, 161 GWAS associations were identified for variants in high LD (r^2>0.9) with the 17q21.31 inversion, illustrating its pleiotropic effects (**Supplementary Table 10**). These associations included several traits potentially related to COVID-19 pathology, such as blood and immune cell composition or lung function (**Figure 1c**).

Variants within the credible set from Bayesian fine mapping at 19q13.33 overlap with several genes, including for instance *NAPSA*, *NR1H2* and *KCNC3,* while the 17q21.31 inversion spans multiple genes including *MAPT*, *KANSL1, FMNL1* and *CRHR1* (**Supplementary Figures 4, 5 and 8** and **Figure 1a**). To identify possible target genes relevant in COVID-19 disease pathology, we performed an exploratory gene expression analysis using several publicly available datasets to: 1) Examine the direct effect of both loci on gene expression by using expression and splicing quantitative trait loci (eQTL and sQTL), 2) Identify in which tissues or cell types our candidate genes are expressed by analyzing their RNA expression at bulk and single-cell levels;

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 lung and other tissues from patients who died after experiencing a SARS-CoV-2 severe disease (**Online Methods**). Results of these analyses are displayed in **Figure 2**.

The potential functional role of the 17q21.31 inversion is supported by the identification of 2,902 linked variants (r^2>0.9) that are already reported as eQTL or sQTL in the GTEx Project.(20) The inversion is in LD with lead eQTLs and sQTLs (i.e. displaying the strongest association with target), for 24 (eQTL) and 7 (sQTL) genes, respectively (**Supplementary Table 11**, **Supplementary Figure 11**). Expression patterns of coding genes affected by the inversion are shown in **Supplementary Figure 12**, with the best candidates to play a role in the effects of the inversion on COVID-19, including *MAPT*, *KANSL1, FMNL1* and *CRHR1,* being summarized in **Figure 2***.* Many of genes the coding genes affected by the inversion are highly expressed in neural and male-specific tissues (i.e. testis, **Supplementary Figure 12**). However, several of them are also expressed in major immune cell types in COVID-19 relevant tissues (**Figure 2b; Supplementary Figure 12a/b**)**,** such as *KANSL1*, which is expressed in lung tissue-resident alveolar macrophages, or *FMNL1* (**Figure 2b**, **Supplementary Figure 12b**). These two genes especially show significantly higher expression in different lung cell types in COVID-19 patients versus healthy controls(21) (**Figure 2c**, **Supplementary Figure 12c**). Moreover, RNA-seq data of monocytes under different bacterial and viral stimuli(22) show expression changes in several genes affected by the inversion (**Supplementary Figure 13, Supplementary Table 12**). 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 14**). Similarly, in SARS-CoV-2 infected brain organoids, the expression of *MAPT* was significantly downregulated in premature and mature neuronal cells (**Supplementary Figure 15, Supplementary Table 13**).

In the case of the 19q13.33 locus, expression of candidate genes shows high tissue specificity, with *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 12**). Of those candidates, expression of *KCNC3* and especially *NAPSA* appears to be clearly affected by the rs1405655 lead SNP (or variants on high LD, r^2>0.9) (**Figure 2a**). Single SNP Mendelian randomization analysis using cis-eQTL data and assessment of effect heterogeneity identified higher expression of *NAPSA* as a potentially causal protective factor for severe COVID-19 (Beta=-0.39; P=1.26×10^{-7}) (**Supplementary Methods, Supplementary Table 14**). Notably, *NAPSA* shows significantly increased expression in type 1 alveolar cells of COVID-19 patients as compared to healthy controls(21) (**Figure 2**, **Supplementary Figure 12**). On the other hand, in COVID-19 patients, the *NR1H2* gene is significantly down-regulated in some parenchymal (basal, ciliated, club) and endothelial (pericyte) lung cells and is up-regulated in monocytes compared to healthy controls (**Figure 2c**, **Supplementary Figure 12**).(21)

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

The ABO locus is one of the most significant associations with COVID-19 infection, as highlighted in several publications.(5,7) In our initial publication, we showed that the genetic association at the ABO locus could be explained by differences in the actual ABO blood type. ABO secretor status, i.e. secretion of ABO antigens into body fluids, is determined by genetic variation in the *FUT2* gene. Recent studies suggest that A secretors have a higher risk of COVID-19 susceptibility.(23) Here we sought to replicate this finding (**Online Methods**) and compared 95%CIs for secretors and non-secretors of the A and B subtypes. 81% of all individuals were secretors irrespective of blood group on average across all control cohorts. In our study, though not statistically significant (95%CIs overlap), a clear trend was 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 15**). Neither blood type B nor B secretion were significantly associated with COVID-19 susceptibility. When performing the stratified analysis as described above on the ABO locus, we saw a strong tendency for a more severe disease course in female A non-secretors in the Italian and Spanish cohorts, and a more severe disease course for

males with blood type O, which did however not replicate in the comparably smaller German cohort. This should be followed-up in future studies.

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 16, Supplementary Figure 16**). Furthermore, we found no significant association for any HLA-presented viral peptide in a so-called PepWAS approach (**Supplementary Table 17**, **Supplementary Figures 17-18**), 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 18**).

With male sex identified as a risk factor for severe COVID-19 and COVID-19-related death(3), we explored possible associations between genetic variants on the Y chromosome and the risk of developing severe COVID-19 and COVID-19 mortality in males.(3) Variations on the Y chromosome describe so-called Y-chromosome haplogroups with letters A-Z (defined by the Y Chromosome Consortium)(24) and follow a pattern of ancestral population migrations in Europe and on a global scale.
Results of the Y-chromosome haplogroup analysis are shown in **Supplementary Table 19**. We did not observe any statistically significant or consistent results for Y chromosomal haplotype association with COVID-19 across the Spanish, Italian and German populations in our first or second analysis after correction of P-values for multiple testing using Bonferroni. We found that cohort-specific suggestive ($P<0.05$) associations within the Italian population for the R1b haplogroup were partially driven by single genotyping batches with diverging haplogroup frequencies (i.e. high coefficient of variation between control batches). COVID-19 related mortality showed a strong trend for association with haplogroup R1b1a2a1 (U106) ($P=9.49\times10^{-3}$, OR=2.8, 95%CI=1.29-6.2) in both the Spanish and Italian populations. 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.

[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, a centralized genotyping and rigorous quality control, we have generated a valuable resource for further COVID-19 related genetic studies. By conservatively including only severely affected COVID-19 patients in our analysis we aimed to balance out a potential bias from selection of controls from the general population, which is however general practice in GWA analyses. Our study replicated known associations with COVID-19 from our previous and other studies and suggest associations at four novel loci, *PCDH7* at 4p15.1; *FREM1* at 9p22.3; *OLMF4* at 13q21.1 and *PTPRM* at 18p11.23, which showed a high level of replicability across our cohorts but could not be replicated in the COVID-19 HGI release 5 statistics. Future studies in independent cohorts need to show whether these signals are true.
Genome-wide meta-analysis of COVID19 HGI release 5 summary statistics with our data revealed a so-far unreported association the 19q13.33 locus, near the gene *NAPSA*. Our in-depth stratified analysis showed age-associations or trends for an association for *LZTFL1* at the 3p21.31 locus, *DPP9* at 19p13, the inversion at 17q21.31 and *IFNAR2* at 21q22.11. Here younger age groups (<= 60 years old; or aged 41-60) carrying the minor allele of variants associated at these loci, were more susceptible to COVID-19 than older age groups (>60 years old; or aged 61-80) carrying this allele. Effects of age in genetic variation are frequently observed in genetic studies, however reasons for this are not clear. Possible models for this are discussed in detail in Jiang *et al*.(25)

Our functional analysis focused on two loci of interest associated with severe COVID-19, the previously known 17q21.31 inversion and the 19q13.33 locus, including the *NAPSA* gene.

For 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. *NAPSA* encodes for 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 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. These findings are in line with a recent report in which *NAPSA* has been identified as a candidate gene for COVID-19 severity through cross methylome omnibus test combined with S-PrediXcan analyses in blood tissue and integrative multiomics(29), as well as with another recent report in which summary based Mendelian randomization analysis showed an inverse correlation between the expression of *NAPSA* in blood and COVID-19 severity.(30)

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.(31) 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(32), and is associated with higher ventilatory response to corticosteroids in individuals with asthma(33), showing potential trade-offs and shared pathways that may be important in lung health. This variant has been proposed to be under positive selection in Europeans through its effect on fertility.(17) Our results point to a role of this polymorphism in immunity and virus infection defence as well. 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.(7) The inversion probably affects COVID-19 disease course through its large effects on gene expression shown by us and others.(7) 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.(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 15**) and lung-related cells and tissues. The H2 haplotype is linked to increased expression of *MAPT* in the lung and its *MAPT* exon 3 in brain tissues(37), 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.(38,39)

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 so-far the largest, HLA analysis from Shachar et *al*.(40) Within our analysis of Y-chromosome haplogroups, none of the results remain significant after correction for multiple testing, though single haplogroups showed trends for association (**Supplementary Text**) and larger study samples are necessary to obtain reliable conclusions. For individual haplogroups, we observed different frequencies between batches that may also arise from different versions of the same genotyping

platform. In general, the analysis of Y-chromosomal disease-relevant SNPs is mainly neglected in GWAS, hence the curation of Y-chromosomal SNPs on genotyping arrays is potentially also a confounding factor. Therefore, to gain more knowledge regarding the potential role of the Y-chromosome haplogroups in disease in general, this should be investigated in more depth.

In summary, our findings add to the number of genome-wide significant hits for COVID-19 – totalling now 14 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.

[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 **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/Austrian, 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/Austrian 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) (**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/Austrian cohort.

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/Austrian, 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(12) (**Supplementary Methods**). An inverse-variance weighted fixed-effects meta-analysis was conducted with the meta-analysis tool METAL(13) 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/Austria and the second discovery cohort including 1,911 critically ill cases and 12,483 population controls (**Supplementary Table 1.** Only variants that were common to at least 2 datasets with respective post-imputation R^2 ≥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. We used a significance threshold of $P<10^{-6}$ for joint P-values to determine suggestive statistical significance and the common threshold of $P<5x10^{-8}$ to determine genome-wide significance. Meta-Analysis Model-based Assessment of replicability(14) (MAMBA) was used (**Supplementary Methods**) to assess replicability at each locus.

To replicate 13 genome-wide significant associations reported by the COVID-19 HGI(7), we excluded 3,837 individuals overlapping between our current study and COVID-19 HGI release 5. We subsequently carried our logistic regression and meta-analysis as described above in 1,579 cases and 10,372 population controls for the first analysis and 944 cases and 10,065 population controls for the second analysis (**Supplementary Table 1e**). A threshold of 0.05 was used to define replicability of the results. Meta-analysis with the COVID-19 HGI release 5 summary statistics included

up to 12,888 hospitalized cases (including 5,582 critically ill cases) and 1,295,966 population controls of the COVID-19 HGI. To unify case definitions by our study and the COVID-19 HGI, we performed meta-analysis of summary statistics from our first analysis with COVID-19 release 5 A2 summary statistics and meta-analysis of our second analysis and COVID-19 release 5 B2 summary statistics. We used the common threshold of $P<5x10^{-8}$ to determine genome-wide significance

Fine mapping of candidate SNPs was performed using Bayesian fine mapping with the tool FINEMAP(15) (**Supplementary Methods**).

Association analysis of candidate SNPs

We performed age- and sex-stratified and severity analyses on 4 candidate SNPs from our first and second discovery cohorts, and 13 candidate SNPs from the COVID-19 HGI analysis.(7) We additionally analysed comorbidities hypertension, coronary artery disease and diabetes in cases only. 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(42) including only statistics from cohorts with N_{Case} and $N_{Control}$ > 50. Sub-analyses in age groups of >60 , <=60, 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 and the binarized age groups of <=60 and >60 (**Supplementary Table 1e**).

Association analysis of the 17q21.31 inversion

17q21.31 inversion genotypes were imputed using IMPUTE v2.3.2(43) from quality controlled SNP data (first and second analysis) based on experimentally-validated inversion genotypes from 109 individuals from the 1000 Genomes Project(44) 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**).(45) 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(19). We queried GWAS hits in high LD ($r^2 > 0.9$) with the inversion as well as hits in high LD ($r^2 > 0.9$) with rs1405655 (19q13.33). Similarly, tissue-specific expression or splicing effects were obtained by searching for SNPs in high LD ($r^2>0.9$) with the inversion or rs1405655 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).(20) 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 (**Supplementary Table 19)**. 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(46)), as well as respiratory tract(47) and brain cells(48) (COVID-19 Cell Atlas).(49) Differential expression of candidate genes in COVID-19 infected lung cells were obtained from pseudo-bulk differential expression analysis performed by Delorey *et al*.(21) 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.*(50) (obtained upon request). The analysis was carried out using hurdle modeling, implemented in the R package MAST.(51) 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*.(22) by imputing the inversion genotypes with IMPUTE v2.3.2(18) and quantifying gene and transcript expression differences with Kallisto v0.46.0(52) and QTLtools v1.1.(53) 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.(5,54,55) 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(11) (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**). The coefficient of variation (CV) across frequencies was calculated for Y-chromosome haplogroups between individual batches (i.e. contributing center/hospital or different versions of the GSA). Here only frequencies that were observed in more than 100 individuals were considered.

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 (c.803C>T, *ABO*B*), rs41302905 (c.802G>A, *ABO*O2)* and rs8176719 (c.261delG, *ABO*O1*) 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 (c.428G>A, *FUT2*01N.02*) 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 ≤1.3 were called “secretors” and individuals with dosages ≥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)**.**

Availability of Summary Statistics

Genome-wide summary statistics of our analyses will be made available upon acceptance of this manuscript.

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 reports grants or contracts from Amplyx, Basilea, BMBF, Cidara, DZIF, EU-DG RTD (101037867), F2G, Gilead, Matinas, MedPace, MSD, Mundipharma, Octapharma, Pfizer, Scynexis; Consulting fees from Amplyx, Biocon, Biosys, Cidara, Da Volterra, Gilead, Matinas, MedPace, Menarini, Molecular Partners, MSG-ERC, Noxxon, Octapharma, PSI, Scynexis, Seres; Honoraria for lectures from Abbott, Al-Jazeera Pharmaceuticals, Astellas, Grupo Biotoscana/United Medical/Knight, Hikma, MedScape, MedUpdate, Merck/MSD, Mylan, Pfizer; Payment for expert testimony from Cidara; Participation on a Data Safety Monitoring Board or Advisory Board from Actelion, Allecra, Cidara, Entasis, IQVIA, Jannsen, MedPace, Paratek, PSI, Shionogi; A pending patent currently reviewed at the German Patent and Trade Mark Office; Other interests from DGHO, DGI, ECMM, ISHAM, MSG-ERC, Wiley. 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, BioveloIta, 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 Pleguezuelos 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). 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 research was partly funded by a MIUR grant to the Department of Medical Sciences, under the program "Dipartimenti di Eccellenza 2018–2022". This study makes use of data generated by the GCAT-Genomes for Life. Cohort study of the Genomes of Catalonia, Fundació 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 Cáceres 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). Thorsen Brenner, Marc M Berger, Oliver Witzke und Anke Hinney are supported by the Stiftung Universitätsmedizin Essen. Marialbert Acosta-Herrera was supported by Juan de la Cierva Incorporacion program, grant IJC2018-035131-I funded by MCIN/AEI/10.13039/501100011033. Eva C Schulte is supported by the Deutsche Forschungsgemeinschaft (DFG; SCHU 2419/2-1).

[REFERENCES]

1. Booth, A., Reed, A. B., Ponzo, S., et al. (2021) Population risk factors for severe disease and mortality in COVID-19: A global systematic review and meta-analysis. Population risk factors for severe disease and mortality in COVID-19: A global systematic review and meta-analysis. *PLoS One* **(2021)** , *16*.
2. Wong, C. K. H., Wong, J. Y. H., Tang, E. H. M., et al. (2020) Clinical presentations, laboratory and radiological findings, and treatments for 11,028 COVID-19 patients: a systematic review and meta-analysis. *Sci. Rep.*, **10**.
3. Peckham, H., de Gruijter, N. M., Raine, C., et al. (2020) Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission. *Nat. Commun.*, **11**.
4. Hu, Y., Sun, J., Dai, Z., et al. (2020) Prevalence and severity of corona virus disease 2019 (COVID-19): A systematic review and meta-analysis. Prevalence and severity of corona virus disease 2019 (COVID-19): A systematic review and meta-analysis. *J. Clin. Virol.* **(2020)** , *127*.
5. Severe Covid-19 GWAS Group (2020) Genomewide association study of severe Covid-19 with respiratory failure. *N. Engl. J. Med.*, **383**.
6. Pairo-Castineira, E., Clohisey, S., Klaric, L., et al. (2021) Genetic mechanisms of critical illness in COVID-19. *Nature*, **591**.
7. COVID-19 Host Genetics Initiative, C. (2021) Mapping the human genetic architecture of COVID-19. *Nature*.
8. Velavan, T. P., Pallerla, S. R., Rüter, J., et al. (2021) Host genetic factors determining COVID-19 susceptibility and severity. *EBioMedicine*, **72**, 103629.
9. Namkoong, H., Edahiro, R., Fukunaga, K., et al. (2021) Japan COVID-19 Task Force: a nation-wide consortium to elucidate host genetics of COVID-19 pandemic in Japan. *medRxiv*.
10. Li, Y., Ke, Y., Xia, X., et al. (2021) Genome-wide association study of COVID-19 severity among the Chinese population. *Cell Discov.*, **7**.
11. Arora, J., McLaren, P. J., Chaturvedi, N., et al. (2019) HIV peptidome-wide association study reveals patient-specific epitope repertoires associated with HIV control. *Proc. Natl. Acad. Sci. U. S. A.*, **116**.
12. Zhou, W., Nielsen, J. B., Fritsche, L. G., et al. (2018) Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic

association studies. *Nat. Genet.*, **50**.

13. Willer, C. J., Li, Y. and Abecasis, G. R. (2010) METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**.
14. McGuire, D., Jiang, Y., Liu, M., et al. (2021) Model-based assessment of replicability for genome-wide association meta-analysis. *Nat. Commun.*, **12**.
15. Benner, C., Spencer, C. C. A., Havulinna, A. S., et al. (2016) FINEMAP: Efficient variable selection using summary data from genome-wide association studies. *Bioinformatics*, **32**.
16. Nakanishi, T., Pigazzini, S., Degenhardt, F., et al. (2021) Age-dependent impact of the major common genetic risk factor for COVID-19 on severity and mortality. *medRxiv Prepr. Serv. Heal. Sci.*
17. Stefansson, H., Helgason, A., Thorleifsson, G., et al. (2005) A common inversion under selection in Europeans. *Nat. Genet.*, **37**.
18. Howie, B., Marchini, J. and Stephens, M. (2011) Genotype imputation with thousands of genomes. *G3*, **1**, 457–470.
19. MacArthur, J., Bowler, E., Cerezo, M., et al. (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.*, **45**.
20. Aguet, F., Barbeira, A. N., Bonazzola, R., et al. (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (80-.).*, **369**.
21. Delorey, T. M., Ziegler, C. G. K., Heimberg, G., et al. (2021) COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. *Nature*.
22. Quach, H., Rotival, M., Pothlichet, J., et al. (2016) Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell*, **167**.
23. Mankelow, T. J., Singleton, B. K., Moura, P. L., et al. (2021) Blood group type A secretors are associated with a higher risk of COVID-19 cardiovascular disease complications. *eJHaem*, **2**.
24. Hammer, M. (2002) A nomenclature system for the tree of human Y-Chromosomal binary haplogroups. *Genome Res.*, **12**.
25. Jiang, X., Holmes, C. and McVean, G. (2021) The impact of age on genetic risk for common diseases. *PLoS Genet.*, **17**.
26. WARD, H. E. and NICHOLAS, T. E. (1984) ALVEOLAR TYPE I AND TYPE II CELLS. *Aust. N. Z. J. Med.*, **14**.

27. Melms, J. C., Biermann, J., Huang, H., et al. (2021) A molecular single-cell lung atlas of lethal COVID-19. *Nature*.
28. Islam, A. B. M. M. K. and Khan, M. A. A. K. (2020) 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**.
29. Wu, L., Zhu, J., Liu, D., et al. (2021) An integrative multiomics analysis identifies putative causal genes for COVID-19 severity. *Genet. Med.*
30. Hernandez Cordero, A. I., Li, X., Milne, S., et al. (2021) Integrative Genomic Analysis Highlights Potential Genetic Risk Factors for Covid-19. .
31. Astle, W. J., Elding, H., Jiang, T., et al. (2016) The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell*, **167**.
32. Sakornsakolpat, P., Prokopenko, D., Lamontagne, M., et al. (2019) Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nat. Genet.*, **51**.
33. Tantisira, K. G., Lazarus, R., Litonjua, A. A., et al. (2008) Chromosome 17: Association of a large inversion polymorphism with corticosteroid response in asthma. *Pharmacogenet. Genomics*, **18**.
34. den Dekker, A. D., Davis, F. M., Joshi, A. D., et al. (2020) TNF-α regulates diabetic macrophage function through the histone acetyltransferase MOF. *JCI Insight*, **5**.
35. Cursano, S., Battaglia, C. R., Urrutia-Ruiz, C., et al. (2020) A CRHR1 antagonist prevents synaptic loss and memory deficits in a trauma-induced delirium-like syndrome. *Mol. Psychiatry*.
36. Thompson, S. B., Sandor, A. M., Lui, V., et al. (2020) Formin-like 1 mediates effector t cell trafficking to inflammatory sites to enable t cell-mediated autoimmunity. *Elife*, **9**.
37. Pickett, B. E., Greer, D. S., Zhang, Y., et al. (2012) Virus pathogen Database and Analysis Resource (ViPR): A comprehensive bioinformatics Database and Analysis Resource for the Coronavirus research community. *Viruses*, **4**.
38. Kirkpatrick, M. (2010) How and why chromosome inversions evolve. *PLoS Biol.*, **8**.
39. Wellenreuther, M. and Bernatchez, L. (2018) Eco-Evolutionary Genomics of Chromosomal Inversions. Eco-Evolutionary Genomics of Chromosomal Inversions. *Trends Ecol. Evol.* **(2018)** , *33*.

40. Ben Shachar, S., Barda, N., Manor, S., et al. (2021) MHC Haplotyping of SARS-CoV-2 Patients: HLA Subtypes Are Not Associated with the Presence and Severity of COVID-19 in the Israeli Population. *J. Clin. Immunol.*, **41**.
41. Taliun, D., Harris, D. N., Kessler, M. D., et al. (2021) Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*, **590**.
42. Viechtbauer, W. (2010) Conducting meta-analyses in R with the metafor. *J. Stat. Softw.*, **36**.
43. Browning, B. L. and Browning, S. R. (2016) Genotype Imputation with Millions of Reference Samples. *Am J Hum Genet*, **98**, 116–126.
44. Auton, A., Brooks, L. D., Durbin, R. M., et al. (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.
45. Kwan, J. S. H., Li, M. X., Deng, J. E., et al. (2016) FAPI: Fast and accurate P-value Imputation for genome-wide association study. *Eur. J. Hum. Genet.*, **24**.
46. Stunnenberg, H. G., Abrignani, S., Adams, D., et al. (2016) The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell* **(2016)** , *167*.
47. Vieira Braga, F. A., Kar, G., Berg, M., et al. (2019) A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat. Med.*, **25**.
48. Habib, N., Avraham-Davidi, I., Basu, A., et al. (2017) Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods*, **14**.
49. Sungnak, W., Huang, N., Bécavin, C., et al. (2020) SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med.*, **26**.
50. Song, E., Zhang, C., Israelow, B., et al. (2021) Neuroinvasion of SARS-CoV-2 in human and mouse brain. *J. Exp. Med.*, **218**.
51. Finak, G., McDavid, A., Yajima, M., et al. (2015) MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.*, **16**.
52. Bray, N. L., Pimentel, H., Melsted, P., et al. (2016) Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.*, **34**.
53. Delaneau, O., Ongen, H., Brown, A. A., et al. (2017) A complete tool set for molecular QTL discovery and analysis. *Nat. Commun.*, **8**.
54. Degenhardt, F., Wendorff, M., Wittig, M., et al. (2019) Construction and

benchmarking of a multi-ethnic reference panel for the imputation of HLA class I and II alleles. *Hum. Mol. Genet.*

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

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

[LEGENDS TO FIGURES]

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(56) (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/Austrian and Norwegian GWAS discovery datasets. The dotted line indicates the genome-wide significance threshold ($P=5\times10^{-8}$). 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 and extended meta-analysis of our first discovery analysis and the COVID-19 HGI release 5 analysis B2 dataset (**Online Methods**) of the association between severe COVID-19 and the 17q21.31 inversion haplotype H2. Shown are OR and 95% CIs of the main (all), age-stratified (41-60 and 61-80 years (y/o)) across all analyzed cohorts. **(c)** Phenome-wide association study (PheWAS) results for the 17q21.31 inversion showing only potentially COVID-19 related phenotypes from the 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. Phenotypes shown were selected according to previously reported COVID-19 links with lung damage, blood cell alterations and exacerbated immune response, as well as some potential co-morbidities. The whole list of phenotypic associations is included in **Supplementary Table 10**.

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(20) 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 H2 (here using the tag SNP rs62055540-C, LD (r^2)=1), rs1405655-T or variants in high LD (r^2>0.9) with either variant. Black rectangles indicate genes for which these variants or proxy variants in high LD (r^2) are lead QTL variants in that tissue. If no box is shown other variants (i.e. not rs62055540-C, rs1405655-T or their proxy variants) 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(47) and adult human brain cells from recently deceased, non-diseased donors.(48) 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.(49) **(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*.(21) More detailed figures are shown in **Supplementary Figures 11,12** and **15.**

[Tables]

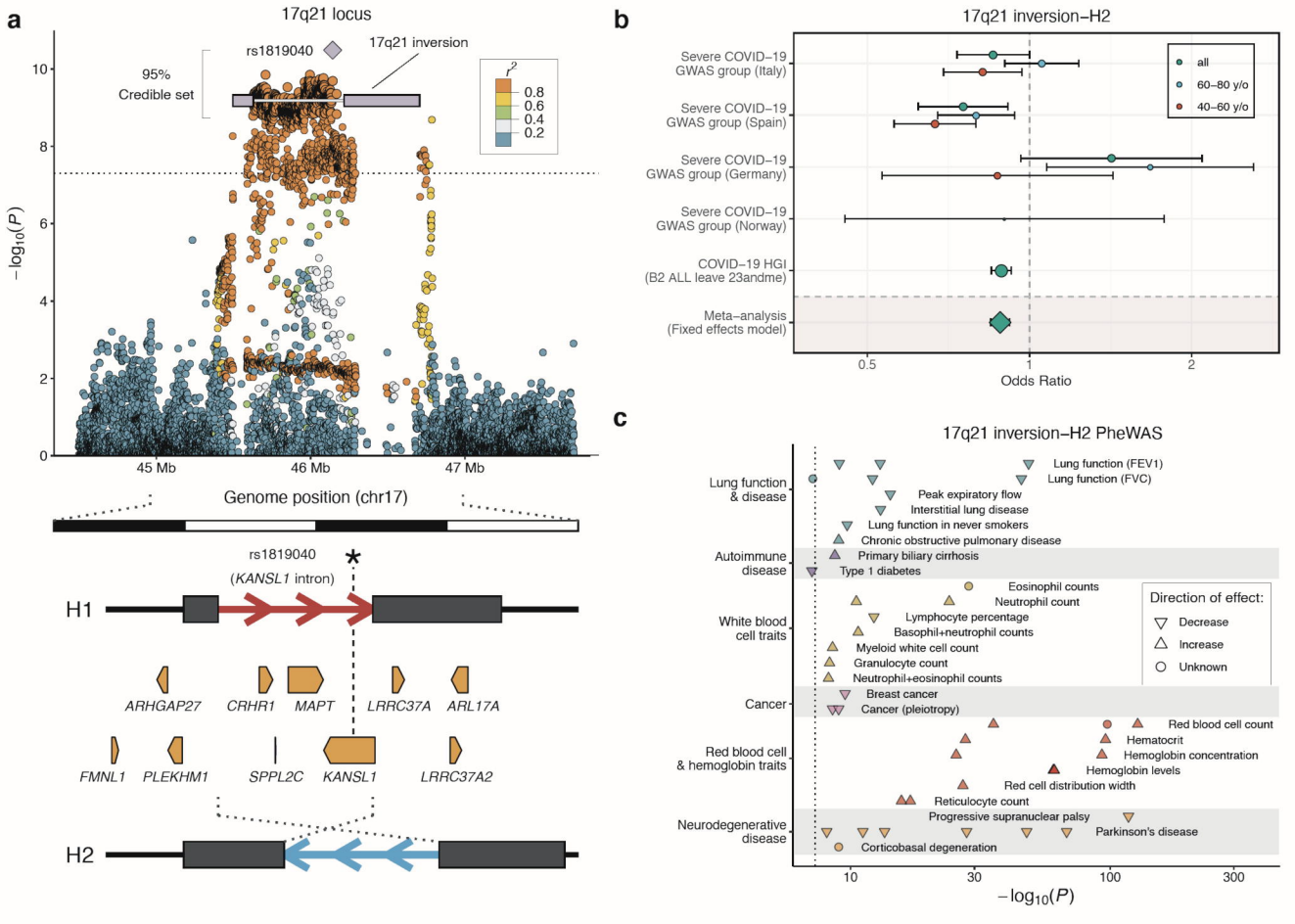
Table 1. Overview of patients included in the genome-wide discovery analysis. Overview of patients included in our first analysis (3,260 patients) and second analysis (1,911 patients). Individuals of the Italian, Spanish, Norwegian and German/Austrian cohorts were recruited at five, seven, eight and 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. Commonly reported comorbidities in COVID-19 are shown, hypertension, coronary artery disease (CAD) and diabetes. Characteristics of control individuals are shown in **Supplementary Table 2**.

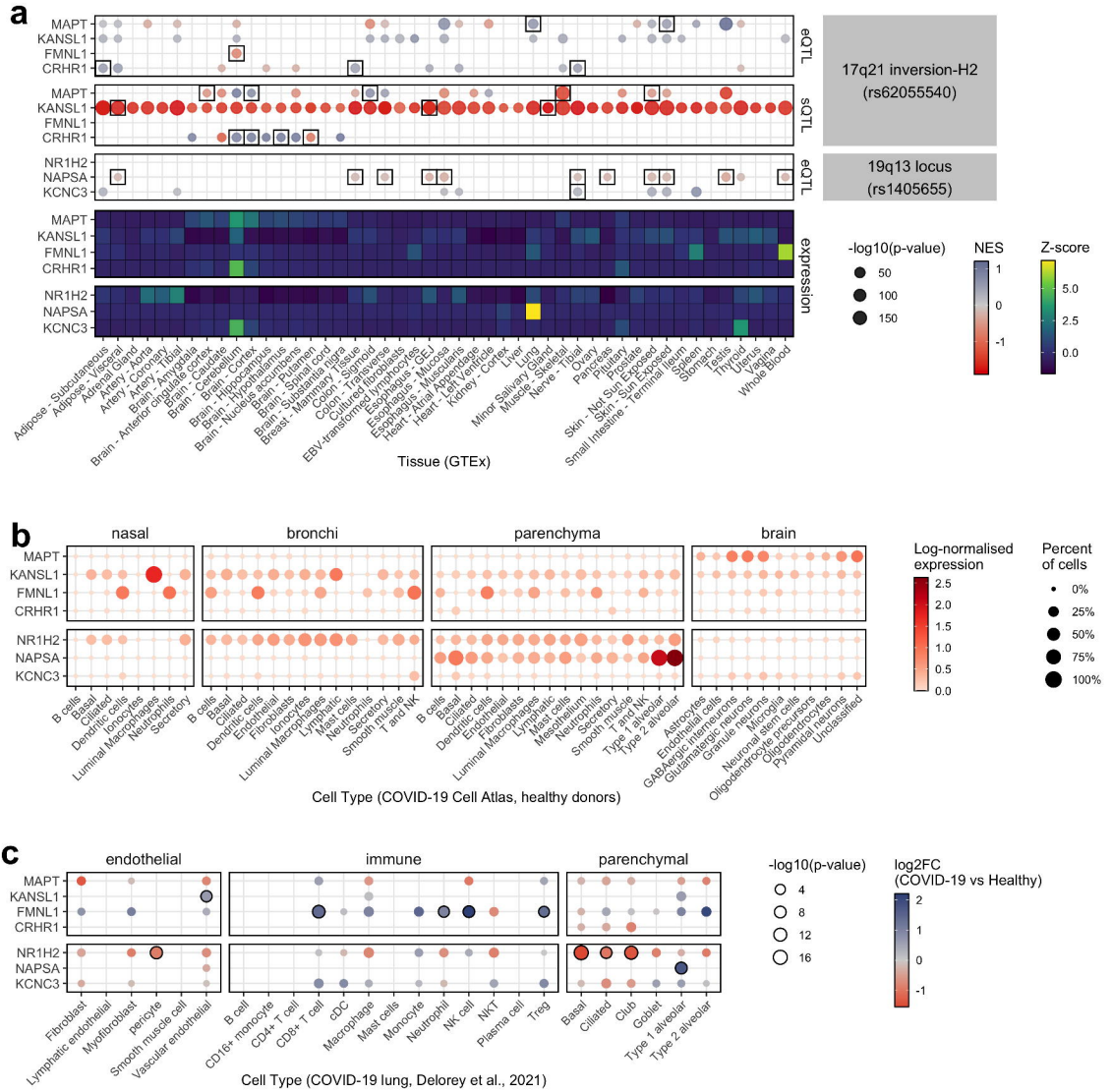
	Italy	**Spain**	**Norway**	**Germany /Austria**
Total N severe COVID-19 patients	1,536	1,421	262	241
Respiratory support categories				
N (%)				
Supplemental oxygen only (1)	276 (17.97)	897 (63.53)	80.65	109 (45.23)
Noninvasive ventilation (2)	933 (60.74)	122 (8.62)	3.23	21 (8.71)
Ventilator (3)	320 (20.83)	381 (26.91)	16.13	86 (35.68)
ECMO (4)	7 (0.46)	16 (1.13)	0	25 (10.37)
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)
First analysis				
N severe COVID-19 patients	1,536	1,421	262	241
Female sex - %	32.81	35.88	33.87	26.97
Median age (IQR) — yr	67 (22)	68 (18)	59.5 (21)	63 (19)
Hypertension -% affected (% missing)	40.55 (21.16)	49.6 (1.91)	41.3 (25.81)	53.09 (32.78)
CAD - % 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)

Second analysis				
N severe COVID-19 patients	1,260	519		132
Female sex - %	29.68	27.55	/	22.73
Hypertension % affected				
(% missing)	38.26 (23.25)	52.19 (3.28)	/	62.96 (38.64)
CAD - % 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)

[ABBREVIATIONS]

95%CI	95% confidence interval
AT1	Type 1 alveolar cells
AT2	Type 2 alveolar cells
CAD	Coronary artery disease
CHRH1	Corticotropin Releasing Hormone Receptor 1
COVID-19	Coronavirus disease 2019
COVID-19 HGI	COVID-19 Host Genetics Initiative
DAMPs	Damage-associated transient progenitors
DNA	Deoxyribonucleic acid
ECMO	Extracorporeal membrane oxygenation
eQTL	Expression quantitative trait loci
FISH	Fluorescence in situ hybridization
FMNL1	Formin Like 1
GenOMICC	Genetics Of Mortality In Critical Care Initiative
GSA	Global Screening Array
GTEx	Genotype-Tissue Expression project
GWAS	Genome-wide association analysis
HLA	Human Leukocyte Antigen
KANSL1	KAT8 Regulatory NSL Complex Subunit 1
KCNC3	Potassium Voltage-Gated Channel Subfamily C Member 3
LD	Linkage disequilibrium
MAF	Minor allele frequency
MAPT	Microtubule Associated Protein Tau
Mb	Megabase pairs
NAPSA	Napsin A Aspartic Peptidase
NES	normalized effect size
NR1H2	Nuclear Receptor Subfamily 1 Group H Member 2
OR	Odds ratio
P	P-value of association
PCA	Principal component analysis
PCR	Polymerase-chain reaction
PepWAS	Peptidome-wide association study
PheWAS	Phenome wide association study
PPR	Posterior probability of replication
QC	Quality control
RNA	Ribonucelic acid
RNA-seq	RNA sequencing
scRNA-seq	Single sell RNA-seq
SNP	Single nucleotide Polymorphism
sQTL	Quantitative splicing trait loci


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)


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
Adipose - Subcutaneous
Adipose - Visceral
Adrenal Gland
Artery - Aorta
Artery - Coronary
Artery - Tibial
Brain - Amygdala
Brain - Anterior cingulate cortex
Brain - Caudate
Brain - Cerebellum
Brain - Cortex
Brain - Hippocampus
Brain - Hypothalamus
Brain - Nucleus accumbens
Brain - Putamen
Brain - Substantia nigra
Brain - Spinal cord
Brain - Cerebellum
Brain - Cortex
Brain - Hippocampus
Brain - Hypothalamus
Brain - Nucleus accumbens
Brain - Putamen
Brain - Substantia nigra
Brain - Spinal cord
Brain - Cerebellum
Breast - Mammary Tissue
Colon - Transverse
Colon - Sigmoid
Cultured fibroblasts
EBV-transformed lymphocytes
Esophagus - GEJ
Esophagus - Mucosa
Esophagus - Muscularis
Heart - Atrial Appendage
Heart - Left Ventricle
Kidney - Cortex
Liver
Lung
Minor Salivary Gland
Muscle - Skeletal
Nerve - Tibial
Ovary
Pancreas
Pituitary
Prostate
Skin - Not Sun Exposed
Skin - Sun Exposed
Small Intestine - Terminal Ileum
Spleen
Stomach
Testis
Thyroid
Uterus
Vagina
Whole Blood
Tissue (GTEx)
b
nasal
bronchi
parenchyma
brain
MAPT
KANSL1
FMNL1
CRHR1
NR1H2
NAPSA
KCNC3
B cells
Basal
Ciliated
Dendritic cells
Endothelial
Fibroblasts
Ionocytes
Luminal Macrophages
Lymphatic
Mast cells
Neutrophils
Secretory
Smooth muscle
T and NK
Type 1 alveolar
Type 2 alveolar
Astrocytes
GABAergic interneurons
Glutamatergic neurons
Granule neurons
Microglia
Neuronal stem cells
Oligodendrocyte precursors
Oligodendrocytes
Pyramidal neurons
Unclassified
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
MAPT
KANSL1
FMNL1
CRHR1
NR1H2
NAPSA
KCNC3
Fibroblast
Lymphatic endothelial
Myofibroblast
pericyte
Smooth muscle cell
Vascular endothelial
B cell
CD16+ monocyte
CD4+ T cell
CD8+ T cell
cDC
Macrophage
Mast cells
Monocyte
Neutrophil
NK cell
NKT
Plasma cell
Treg
Basal
Ciliated
Club
Goblet
Type 1 alveolar
Type 2 alveolar
-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)