Research paper

Increased risk of severe clinical course of COVID-19 in carriers of HLA-C*04:01

January Weiner 3rd[a], Phillip Suwalski[b,c], Manuel Holtgrewe[d], Alexander Rakitko[e], Charlotte Thibeault[f], Melina Müller[b], Dimitri Patriki[g], Claudia Quedenau[h], Ulrike Krüger[d], Valery Ilinsky[e], Iaroslav Popov[e], Joseph Balnis[i], Ariel Jaitovich[i], Elisa T Helbig[f], Lena J Lippert[f], Paula Stubbemann[f], Luis M Real[j], Juan Macías[j], Juan A Pineda[j], Marta Fernandez-Fuertes[j], Xiaomin Wang[b], Zehra Karadeniz[b], Jacopo Saccomanno[f], Jan-Moritz Doehn[f], Ralf-Harto Hübner[f], Bernd Hinzmann[k], Mauricio Salvo[k], Anja Blueher[k], Sandra Siemann[k], Stjepan Jurisic[g], Juerg H. Beer[g], Jonas Rutishauser[g], Benedikt Wiggli[g], Hansruedi Schmid[g], Kathrin Danninger[l], Ronald Binder[l], Victor M Corman[m], Barbara Mühlemann[m], Rao Arjun Arkal[n,o,x], Gabriela K. Fragiadakis[n,o,q], Eran Mick[r,s,t], Consortium COMET[u], Carolyn S. Calfee[r], David J. Erle[n,o,p,r,u,v,w], Carolyn M. Hendrickson[r], Kirsten N. Kangelaris[w], Matthew F. Krummel[n,x], Prescott G. Woodruff[m,r,x,y], Charles R. Langelier[s,t], Urmila Venkataramani[n,o], Federico García[z], Joanna Zyla[aa], Christian Drosten[m], Braun Alice[ab], Terry C Jones[m,ac,ad], Norbert Suttorp[f], Martin Witzenrath[f], Stefan Hippenstiel[f], Tomasz Zemojtel[d], Carsten Skurk[b], Wolfgang Poller[b], Tatiana Borodina[h], Study Group Pa-COVID[ae], Stephan Ripke[ab,af,ag], Leif E Sander[f], Dieter Beule[a], Ulf Landmesser[b,ah], Toumy Guettouche[k], Florian Kurth[f], Bettina Heidecker[b,*]

[a] Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Core Unit Bioinformatics Berlin, DE 10178, Germany
[b] Department of Cardiology, Charite Universitaetsmedizin Berlin, DE 12203, Germany
[c] Berliner Simulations- und Trainingszentrum, Charite, Berlin, DE 10117, Germany
[d] Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Core Unit Genomics Berlin, DE 10178, Germany
[e] Genotek Ltd., Nastavnicheskii pereulok 17/1, R 105120 Moscow, Russian Federation
[f] Charite Universitaetsmedizin Berlin, Department of Infectious Diseases and Respiratory Medicine Berlin, DE 10117, Germany
[g] Kantonsspital Baden AG, Department of Medicine, Baden, CH 5404, Switzerland
[h] Max Delbrueck Center for Molecular Medicine Berlin, DE 13125, Germany
[i] Department of Molecular and Cellular Physiology, Albany Medical College, NY, USA
[j] Unidad Clínica de Enfermedades Infecciosas y Microbiología. Hospital Universitario de Valme, Sevilla, ES 41014, Spain
[k] Roche Sequencing Solutions Pleasanton, USA 94588
[l] Department of Cardiology and Intensive Care, Klinikum Wels-Grieskirchen, Wels, Austria
[m] Charite Universitaetsmedizin Berlin, Institute of Virology Chariteplatz, 1 D-10117, Berlin, DE, 10117, Germany
[n] ImmunoX Initiative, University of California San Francisco, San Francisco, CA, USA
[o] CoLabs, University of California San Francisco, San Francisco, CA, USA
[p] Institute for Human Genetics, University of California San Francisco, San Francisco, CA, USA
[q] Division of Rheumatology, Department of Medicine, University of California San Francisco, San Francisco, CA, USA
[r] Division of Pulmonary, Critical Care, Allergy, and Sleep Medicine, University of California, San Francisco, CA, USA
[s] Division of Infectious Diseases, University of California, San Francisco, CA, USA
[t] Chan Zuckerberg Biohub, San Francisco, CA, USA
[u] COMET (COVID-19 Multiphenotyping for Effective Therapies) Consortium members are listed in the Supplementary Appendix 1
[v] Lung Biology Center, University of California, San Francisco, CA, USA
[w] Department of Medicine, University of California, San Francisco, CA, USA
[x] Department of Pathology, University of California, San Francisco, CA, USA
[y] Sandler Asthma Basic Research Center, University of California, San Francisco, CA, USA
[z] Hospital Universitario Clínico San Cecilio, Instituto de Investigación Ibs. Granada, Spain
[aa] Department of Data Science and Engineering, Silesian University of Technology, Gliwice, Poland
[ab] Charite Universitaetsmedizin Berlin, Dept. of Psychiatry and Psychotherapy Chariteplatz 1 D-10117 Berlin, DE 10117, Germany
[ac] German Center for Infection Research (DZIF), Associated Partner Site, 10117 Berlin, Germany
[ad] Centre for Pathogen Evolution, Department of Zoology, University of Cambridge, Downing St., Cambridge, CB2 3EJ, U.K

* Corresponding author.
  E-mail address: bettina.heidecker@charite.de (B. Heidecker).

[ae] *Pa-COVID Study Group, Members are listed in the Supplementary Appendix 2*
[af] *Massachusetts General Hospital, Analytic and Translational Genetics, Boston, MA 02114, USA*
[ag] *Stanley Center for Psychiatry Research, Broad Institute of MIT and Harvard Cambridge MA 02142, USA*
[ah] *Berlin Institute of Health at Charité, Berlin, Germany*

ARTICLE INFO

ABSTRACT

Background: Since the beginning of the coronavirus disease 2019 (COVID-19) pandemic, there has been increasing urgency to identify pathophysiological characteristics leading to severe clinical course in patients infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Human leukocyte antigen alleles (HLA) have been suggested as potential genetic host factors that affect individual immune response to SARS-CoV-2. We sought to evaluate this hypothesis by conducting a multicenter study using HLA sequencing.
Methods: We analyzed the association between COVID-19 severity and HLAs in 435 individuals from Germany ($n = 135$), Spain ($n = 133$), Switzerland ($n = 20$) and the United States ($n = 147$), who had been enrolled from March 2020 to August 2020. This study included patients older than 18 years, diagnosed with COVID-19 and representing the full spectrum of the disease. Finally, we tested our results by meta-analysing data from prior genome-wide association studies (GWAS).
Findings: We describe a potential association of HLA-C*04:01 with severe clinical course of COVID-19. Carriers of HLA-C*04:01 had twice the risk of intubation when infected with SARS-CoV-2 (risk ratio 1.5 [95% CI 1.1−2.1], odds ratio 3.5 [95% CI 1.9−6.6], adjusted *p*-value = 0.0074). These findings are based on data from four countries and corroborated by independent results from GWAS. Our findings are biologically plausible, as HLA-C*04:01 has fewer predicted bindings sites for relevant SARS-CoV-2 peptides compared to other HLA alleles.
Interpretation: HLA-C*04:01 carrier state is associated with severe clinical course in SARS-CoV-2. Our findings suggest that HLA class I alleles have a relevant role in immune defense against SARS-CoV-2.
Funding: Funded by Roche Sequencing Solutions, Inc.

Research in context

Evidence before this study

We have searched Pubmed for publications containing keywords "HLA", "COVID", "association" and "severity" published before June, 17th 2021 (exact search phrase: (covid) AND (severity) AND (hla) AND (association) AND (("2019/12/31"[Date - Completion]: "2021/06/17"[Date - Completion]). Among the 69 query results, there were several studies, which involved Human leukocyte antigen alleles typing and COVID-19. None of the analyses performed included a multi-center, multi-cohort study and none included additional covariates, which can potentially distort the observed association, such as sex, age or genetic background. Furthermore, none of these analyses attempted to compare the HLA typing results with the results from genome-wide association studies. All previously published analyses were based on small sample sizes.

Added value of this study

We have performed a systematic analysis of the association of type I and type II HLA alleles in over 400 patients from several countries. Furthermore, we have replicated the HLA association analysis with an analysis of GWAS data (7796 cases and 875,694 controls) and combined our analysis with results of in silico epitope modeling. Our models included covariates aimed at excluding potential confounding effects (such as alleles which are associated with age or sex rather than severity).

Implication of all the available evidence

COVID-19 severity is associated with at least one HLA allele. The association is not limited to a single population and is not the result of a confounding effect of other covariates.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first reported in Wuhan, China, at the end of 2019 [1,2]. It rapidly spread to Europe, the United States (US), and the rest of the world, manifesting as coronavirus disease 2019 (COVID-19) [3,4].

Since the beginning of the pandemic, a striking clinical spectrum has been described among patients with COVID-19. Some patients remain asymptomatic, others may become "super spreaders" or "super emitters" [5], while yet others have a severe clinical course leading to respiratory or multiorgan failure with potentially lethal outcome [6]. Various clinical risk factors for severe clinical course have been described such as age, diabetes, hypertension, and obesity [4,6]. At the genomic level, blood groups [7,8], as well as genes involved in antiviral defense mechanisms and inflammatory organ damage, may affect clinical outcomes [9].

The HLA system plays a crucial role in immune response [10,11]. Genetic polymorphisms of HLA have been reported to affect the clinical course of patients infected with various RNA viruses (e.g. influenza virus H1N1 [12], Hantaan virus [13], and SARS-CoV-1 [14]). HLA-B* 46:01 has been shown to be associated with a severe clinical course in patients infected with SARS-CoV-1 [14]. Furthermore, a recent in silico analysis of HLA data from an international database suggests that HLA-B* 46:01 predisposes patients to a more severe clinical course of SARS-CoV-2 [15], similar to that found in patients with SARS-CoV-1 [14]. While HLA-B* 46:01 frequency is 17% in Asia, where these data were obtained, the allele frequency is 0.05% in Germany and <0.01% in Spain, where the largest part of our study was conducted (http://www.allelefrequencies.net) [16]. Therefore, we suspected that if an HLA risk allele was identified in our data, it would differ from the one discovered during the SARS-CoV-1 pandemic in Asia.

Our central hypothesis was that specific HLA alleles predispose patients to a severe clinical course in COVID-19. Severe clinical course was defined as intubation, intensive care requirement, or death. We

also sought to evaluate if specific HLA alleles predispose patients to acute cardiac injury, which is another surrogate of severe illness.

To test our hypothesis, we collected blood samples from patients infected with SARS-CoV-2 for HLA sequencing during the first peak of the European COVID-19 pandemic. Enrollment centers included hospitals in Germany, Switzerland, and Spain. Furthermore, we leveraged existing RNA-Seq data obtained in the US to validate our findings. Then, we tested our findings through meta-analysis of data from several, large, recent GWAS studies to evaluate the reproducibility of the identified HLA risk allele in patients with severe clinical course in COVID-19 and to investigate possible evidence of increased susceptibility to COVID-19 in carriers of this risk allele.

2. Methods

2.1. Study participants and recruitment

We analyzed the associations between all HLA class I and important HLA class II loci and disease severity in 435 patients with mild to severe symptoms of COVID-19 for HLA sequencing analysis.

All patients of the Charité Universitätsmedizin Berlin are included in data set 1 (DS1) and patients from Switzerland and Spain in data set 2 (DS2) according to the sequence of their enrollment into the study. In addition to the 288 patients recruited for this study (DS1 and DS2), we used data from a public data set containing RNA sequencing data for a total of 99 patients with COVID-19 containing phenotypic information on requirement for intensive care or intubation (https://www.ncbi.nlm.nih.gov/geo; GSE157103 [17], GSE174818 [18]) and for 48 patients with COVID-19 enrolled in the UCSF COVID-19 Multiphenotyping for Effective Therapies (COMET) Cohort with available RNA sequencing data (data set 3, DS3; Supplemental Table 1). Through inclusion of multiple international data sets we tested for broad applicability of our results.

Participants in our cohort were representative of the full spectrum of clinical presentation of COVID-19 from asymptomatic to multi organ failure and death. Diagnosis was made based on detection of SARS-CoV-2 viral RNA using a reverse transcription-polymerase chain reaction (RT-PCR) test from nasopharyngeal swabs and/or detection of SARS-CoV-2 specific antibodies in the blood through enzyme-linked immunosorbent assay (ELISA). Sample collection took place during the first peak of the pandemic at six European Hospitals: Germany: Charite Universitaetsmedizin Berlin (3 sites, total 135 patients, DS1) [19,20]; Spain: Hospital Universitario de Valme, Sevilla (total 115 patients, first part of DS2), Hospital Universitario San Cecilio, Granada (total 18 patients, second part of DS2) and Switzerland: Kantonsspital Baden AG, Baden (total 20 patients, third part of DS2). All patients with COVID-19, who provided informed consent and were 18 years or older were included. Whole-blood samples, buffy coats, or peripheral blood mononuclear cells (PBMCs) were collected from all patients for deoxyribonucleic acid (DNA) extraction during routine diagnostic venipuncture. In addition to clinical and laboratory data, we had access to first-measured viral load data for the cohort from Germany (based on RT-PCR cycle threshold) [21].

We used requirement for intensive care unit (ICU) admission and intubation as categories for HLA analysis and standard reference range for troponin T hs (<14 ng/l) to define troponin elevation. Acute cardiac injury was defined as elevation of troponin T hs (>14 ng/l) without obvious signs for acute coronary syndrome. We directly used the troponin T hs levels to test for association. Troponin T hs levels were available for 66 patients in whom the value was determined as part of clinical routine, when indicated. Finally, we tested if the identified risk allele was associated with viral load. Dr. habil. January Weiner, PD Dr. med. Bettina Heidecker and Mr. Phillip Suwalski had full access to the data.

Our protocol included blood samples obtained during routine, clinically indicated venipuncture, which were then frozen at −80 °C

for subsequent DNA extraction and sequencing. A minimal set of deidentified clinical data was collected to protect data privacy. At Charite Universitaetsmedizin Berlin procedures were performed within the framework of the Pa-COVID-19 protocol [19].

Given the exceptional situation of the pandemic and the international character of this study, there were differences amongst local ethics committees in consent procedures [8]. At each center, written informed consent was obtained from patients or representatives when possible, sometimes retrospectively if the patient was not able to give consent at the moment of sample collection.

This study was approved by the ethics committee of Charite Universitaetsmedizin Berlin, Germany (EA2/066/20), the Ethics Committee Nordwest- und Zentralschweiz, Basel, Switzerland (2020−00,952), the Ethics Committee of the Hospital Universitario de Valme, Sevilla Spain (1346-N-16), the Comité Ético de Investigación (CEIM/CEI) Provincial de Granada (2577-N-20), and the Ethics Committee at the University of California, San Francisco, IRB protocol number 20−30,497.

Data sets differ in the depth of clinical annotation and sites used different criteria, e.g. for admitting patients to the ICU. The body-mass index (BMI) was not available in the cohort from Sevilla.

2.2. DNA extraction, HLA sequencing, and quality control

2.2.1. DNA extraction

DNA was extracted either manually using the Quick-DNA™ Miniprep Plus Kit (Zymogen, number D4068) or automatically on a Roche LC2 robot according to the manufacturer's instructions. On the robot The MagNA Pure LC DNA Isolation Kit - Large Volume (Roche, number 03,310,515,001) was used. DNA concentration was measured using Nanodrop (Thermo Fisher) for the Zymogen extracted samples and Qubit™ dsDNA HS Assay Kit (Thermo Fisher, number Q32851) for robot-extracted samples.

2.3. HLA typing by Pacbio SMRT sequencing, PCR amplification of full-length HLA genes

Aliquots of patient genomic DNA samples were diluted to the same concentrations in 96-well plates. To meet the challenges of a high-throughput project, further pipetting steps were performed with multichannel pipettes or on the epMotion 5075 pipetting robot (Eppendorf).

For each patient sample, eleven HLA genes were amplified using NGSgo®-AmpX v2 HLAGeneSuite kit (#7371,662, GenDx). According to the design of this protocol, amplicons cover all class I HLA genes (HLA-A, -B, -C) and the most relevant parts of the class II HLA genes (HLA-DRB* 1, -DRB* 3/4/5, -DQB* 1, -DPA* 1, -DPB* 1, -DQA* 1), which have very long introns.

Amplification was performed according to the manufacturer's instructions (https://www.gendx.com/product_line/ngsgo-ampx-v2). To improve PCR efficiency, the starting amount of template DNA was increased from 45 ng to 80 ng per reaction for loci HLA-DRB* 1, -DQB* 1, -DPA* 1, -DPB* 1 and −DQA* 1. In addition, elongation time for HLA-DQA* 1 was extended to 5 min. Nevertheless, assay performance differed between the loci. Amplification was typically unproblematic for HLA-A, -B, -C, and DQA1 genes and less consistent in terms of output for other loci.

Amplicons were purified with Agencourt AMPure XP beads (#A63881, Beckman Coulter). Quality and quantity of purified PCR amplicons were checked using D5000 Screen Tape assay (#5067−5593, Agilent) on TapeStation (Agilent) and Qubit dsDNA assay kit (#33,120, Thermofisher) on FLUOStar Omega microplate reader (BMG Labtech), respectively. Confirmed amplicons were equimolar combined into two− for HLA class I and HLA class II − pools per patient.

2.4. HLA SMRTbell library preparation and sequencing

Single molecule real-time (SMRT) sequencing libraries were prepared according to the manufacturer's protocol (Procedure & Checklist -Preparing SMRTbell[TM] Libraries using PacBio® Barcoded Adapters for Multiplex SMRT® Sequencing PacBio®). Barcoded Adapters were used to label patient amplicon pools. Up to 72 amplicon pools were combined per HLA class I and HLA class II library. Libraries were validated using 12,000 DNA Assay (#50,697−1508, Agilent) on Agilent Bioanalyzer and quantified using Qubit dsDNA assay on Qubit Fluorometer.

Libraries were loaded on SMRT cells and sequenced on a Sequel instrument (Pacific Biosciences) using 10 h / SMRT Cell 1 M sequencing runtime. Loading setup and run design was planned with SMRTLink v9.0 (https://www.pacb.com/wp-content/uploads/Quick-Reference-Card-Loading-and-Pre-Extension-Recommendations-for-the-Sequel-System.pdf). In total, HLA genes amplicons for 229 patients were sequenced on 12 SMRT cells.

Circular consensus sequence (CCS) data analysis and demultiplexing was performed within SMRT Link 9.0.0.92188 GUI.

2.5. HLA typing of RNA-Seq data (data set 3, DS 3)

Publicly available data were downloaded from the Gene Expression Omnibus (GEO identifier GSE157103). All RNA-Seq data were analysed using the arcasHLA package, version 0.2.0 [22].

Finally, correlation analysis was performed to compare HLA allele frequencies between all three data sets (DS1, DS2, DS3).

2.5.1. Exome sequencing

Genomic DNAs were quantified by Quant-iT[TM] dsDNA Assay Kit, broad range (ThermoFisher Scientific) with a FLUOStar Omega microplate reader (BMG Labtech). Illumina-compatible libraries were prepared on a Bravo NGS Workstation Option B according to the manual KAPA HyperPrep/HyperPlus Automated Workflow for Agilent Bravo B NGS Instructions for Use, v3.0 (Roche Sequencing) using the KAPA Hyper Plus workflow with the reagents included in the KAPA Hyper Plus kit (Roche Sequencing). In short, 100 ng of genomic DNA was fragmented at 37 °C in an Eppendorf Mastercycler Plus S for 25 min with subsequent end repair. Following end repair, fragments were ligated to the KAPA Universal Adapter (Roche Sequencing) and purified by KAPA HyperPure Beads (Roche Sequencing). Purified libraries were amplified by 6 cycles of PCR utilizing the KAPA UDI Primer Mix* (Roche Sequencing) and again purified using KAPA HyperPure Beads*. Amplified sample library concentration and size distribution was determined by Qubit dsDNA HS assay kit (ThermoFisher Scientific) with a Qubit[TM] 3 Fluorometer (ThermoFisher Scientific) and D5000 ScreenTape assay (Agilent Technologies) with a 4200 TapeStation system (Agilent Technologies). Eight amplified sample libraries were pooled prior to enrichment. For this purpose, 187,5 ng of each library was pooled into one well and the final volume was adjusted to 45 μl with nuclease-free water as designated by the manual KAPA HyperCap Workflow v3.0 (https://pim-eservices.roche.com/eLD/web/pi/en/home) and all further steps were performed following the recommendations of that manual.

These included hybridization of the amplified sample library pool to KAPA HyperExome Probes (Roche Sequencing) washing and recovery of the captured multiplex sample library, as well as amplification and purification of the enriched multiplex sample library. Final library pools were quantified by Qubit dsDNA HS assay kit with a Qubit[TM] 3 Fluorometer and correct size distribution was validated by High Sensitivity D5000 ScreenTape assay (Agilent Technologies) with a 4200 TapeStation system. A total of 12 multiplex enriched sample libraries were pooled equimolar utilizing the NextGen HyperCap Pooling Guide and Calculator template (Roche Sequencing) and diluted to 1.3 nM. Each final pool was sequenced on two lanes of an S4 FlowCell (Illumina, Inc.) for 2 × 151 cycles on a NovaSeq 6000 system (Illumina, Inc.)

*For Research Use Only. Not for use in diagnostic procedures.

2.5.2. Exome analysis

Exome data was aligned using BWA-MEM to the reference GRCh37 (hs357d.fa) [53]. Samples were identified as KIR2DS4f-positive [54,55] by screening the GATK HaplotypeCaller v3.7 [23] variant calls for presence of an insertion of CCCGGAGCTCCTATGACATGTA in exon 4 of KIR2DS4. For a more detailed description of the analysis please see methods in the data supplement.

Principal components analysis based on SNP genotyping from the obtained exome data was performed with the R package SNPrelate v. 1.26.0 [24] as described below.

2.6. Statistical analysis

2.6.1. Composite disease score

DS1 consisted of patients of the Charité cohort. For patients in DS1, we defined a composite score, including troponin T hs levels, death, ICU and intubation status (binned in five categories). Then, we used multiple correspondence analysis (MCA; R package FactoMineR [25], v. 2.3) to obtain a score derived from the selected variables, defined as the first dimension of the individual coordinates.

2.6.2. HLA association testing

The dominant model of association (h/h and h/- vs -/-) with the selected response variables was tested using a generalized linear model as implemented in the HIBAG R package [26]. For DS1 (German cohort), sex, age, and reported region of origin were used as covariates; for DS2, sex and age, and for DS3, sex, age, and reported ethnicity. In an additional analysis, BMI was used as a covariate. For categorical variables such as ICU status, logistic regression was used. In the analysis, we filtered alleles by minor allele frequency (MAF), i.e. considered alleles with frequency in all data sets above 5% and present in at least five individuals. Altogether, 35 alleles from 10 loci were tested for association (detailed information about tested alleles in Supplemental Table 2). The response variables differed between the three data sets analysed. DS 1 (Germany) included the following continuous variables: Troponin T hs (logarithmitized), maximum WHO ordinal scale for clinical improvement score, as well as categorical: Troponin T hs elevation above 14 ng/l, ICU, and intubation status. For DS 2 and 3 (DS2: Switzerland/Spain and DS3: US, respectively) we tested only for association with ICU and intubation status. For intubation and ICU status, we tested the association with alleles in all three data sets separately and then integrated the results using meta-analysis (see below). Troponin T (as continuous or categorical variable) and WHO score were only tested in DS1 and the results were corrected for multiple testing using the Bonferroni correction.

In addition, as a follow-up analysis, we tested for an association between viral load and alleles in DS1.

2.7. Meta-analysis of association results

The results of association tests in DS1-−3 were next subjected to a meta-analysis. Given that only intubation status and ICU status were present in all three data sets, our primary analysis was focused on these two response variables. The log-odds ratios (beta-coefficients from the logistic regression models) were combined, and p-values for each allele were obtained using a weighted Z-test [27] (weight equal to the reciprocal of standard error). Family-wise error rates (p-values corrected for multiple comparisons) were derived using the Bonferroni method over all meta-analysis results.

2.8. Genetic structure of DS1- DS3

To test presence of genetic subpopulations in DS 1 and 3, we leveraged exome derived gene variants. We used the R package SNPrelate, version 1.26.0, to process the VCF file and perform principal components (PC) analyses. Next, we tested whether any of the PCs

were associated with the presence / absence of the identified risk allele using a *t*-test. Since exome typing was not available for roughly half of the samples in DS1 and for no samples in DS3, we have used self-reported region of origin for DS1 and ethnicity for DS3 as covariates in the linear regression models. In addition, we tested whether removing potential outliers based on exome data-derived PC analysis and including the PCs as covariates in the models of DS1 and DS2 has a major impact on the results of the meta-analysis. To this end, we identified outliers using interquartile range method as implemented in the R function boxplot.stats and removed them (as well as all individuals with no genetic background information) from DS1 and DS2. Next, we recalculated the PC analysis using the selected samples only and repeated the association tests, but now using PC1 and PC2 in lieu of self-reported region of origin. Finally, we repeated again the meta-analysis and compared its results with the results of the original analysis based on all data.

2.9. Association of HLA-C*04:01 with susceptibility to COVID-19 in external cohort

Recent reports demonstrated that some genetic risk alleles for COVID-19 susceptibility are also associated with severity [28]. Studies that investigate susceptibility to COVID-19 by comparing participants who are COVID-19 negative vs. positive are generally much larger in sample size than studies that investigate disease severity in COVID-19. Therefore, we investigated data derived from studies of susceptibility and severity.

We hypothesized that HLA-C*04:01 might also be associated with susceptibility to COVID-19. To test this hypothesis, we used an external cohort of 12,139 clients of the genetic testing company Genotek Ltd (Moscow, Russia). Genotek Ltd is part of the COVID-19 Host Genetics Initiative (https://www.covid19hg.org/results). All participants were genotyped using Illumina Infinium Global Screening Array. The HLA alleles were imputed using HIBAG software [26]. Cases were defined as individuals with self-reported positive COVID-19 RT-PCR or antibody test. Controls were defined as individuals with self-reported COVID-19 negative status.

Phenotyping was performed during November 2020 − April 2021. Samples from participants age > 18 years, and BMI >15 and <60 kg/m^2 were included in the final data set. The cohort consisted of 2113 cases and 10,026 controls (51.6% were females; average age 36.4 ± 13.5 years). To test the association of HLA-C*04:01 with COVID-19 susceptibility a logistic regression was fitted with age, sex, and BMI as covariates.

2.10. Meta-analysis of previous GWAS studies

We analyzed data from HLA-related studies and GWAS analyses that investigated COVID-19 susceptibility and severity, including studies in preprint stage. The list of studies included in the meta-analysis is provided as a table in supplemental file 2.

In most GWAS analyses, authors did not impute HLA alleles from microarray data and report association results for the individual SNPs. To show that a certain SNP is a tag-SNP for HLA-C*04:01, we calculated D', the corresponding normalized coefficient of linkage disequilibrium and the squared correlation coefficient between pairs of loci (r^2). These linkage disequilibrium metrics were computed using R package genetics and genotypes from the 1000 Genomes (1KG) Project.

Consortium [29]. HLA-alleles for the 1KG data set were derived from a study by Abi-Rached and colleagues [30]. To detect the possible tag-SNPs we calculated linkage disequilibrium-metrics (D',r^2) for all SNPs from 1KG EUR which are within 500 kb of the HLA-C gene (provided as a table in supplemental file 3). We find that rs5010528 was presented in the summary statistics and could be used as tag-SNP for HLA-C*04:01 allele ($r^2 = 0.966$, D' = 1 in 1KG EUR cohort). To

estimate the overall effect, we performed the meta-analysis using R package meta, assuming the fixed-effect model.

2.11. Role of funding source

3. Results

3.1. Patients

Descriptive data on age, sex, intensive care management, maximum respiratory support during hospitalization and relevant comorbidities (hypertension, coronary artery disease, diabetes) of patients included in all three data sets are provided in table 1 and supplemental Table 1.

The median age was similar in all three data sets, ranging from 57 to 60 years. The Spanish cohort consisted of more women (59%) than the cohorts from Germany (33%), Switzerland (20%) and the US (35%). In the Spanish cohort the prevalence of hypertension and diabetes was lower, and patients were less likely to have been intubated. In DS1, the majority of the patients were of European ancestry (81% of patients who reported ancestry). However, this information was not available for 43% of the patients.

In DS3, data on ethnicity was available for the majority of patients. Patients were classified as being Asian, African American, Hispanic or Caucasian.

We found that HLA allele frequencies amongst the three data sets (DS1, DS2, DS3, supplemental Figure 1 and supplemental Figure 2) were correlated and no major differences in the frequencies of the alleles were apparent.

3.2. Association of HLA alleles with disease severity

The allele HLA-C*04:01 was associated with intubation (adjusted p-value = 0.0074, Table 2, supplemental file 1) in all 3 data sets using age, sex and ethnicity as covariates. The calculated odds ratio for intubation was 3.5 [95% CI 1.9−6.6] while the risk ratio was 1.5 [95% CI 1.1−2.1]. The OR for the association between HLA-C*04:01 with intubation when no other covariates were included in the model was 2.9 [1.6−5.2], adjusted p-value = 0.02.

In the individual data sets, HLA-C*04:01 was associated with a higher percentage of intubated patients in DS1 (69% vs 35%) and in DS3 (63% vs 37%), but not in DS2 (7% vs 6%, Fig. 1 and Fig. 2). In total, out of 127 intubated patients in this study, 38 (30%) carried the allele HLA-C*04:01, while among the 308 remaining patients, 56 (18%) were carriers (for 7 patients, the HLA-C locus could not be determined).

When BMI was used as an additional covariate, the adjusted p-value for the association of HLA-C*04:01 with intubation remained significant (adjusted p-value = 0.025; supplemental Table 3). After removing the cohort from Sevilla from the main analysis, the association of HLA-C*04:01 with intubation remained significant with an adjusted p-value = 0.0035 (supplemental Table 4).

3.3. Association of HLA-C*04:01 with COVID-19 severity and genetic background

One of the potential error sources of HLA analysis may be that an allele occurs with a higher frequency in individuals with certain genetic background and, incidentally, a different profile of the disease

Table 1
Clinical parameters of patients in data set 1, 2, and 3.

	Data set 1	Data set 2	Data set 3	Total
Characteristics	Germany (n = 135)	Spain, Switzerland (n = 153)	US (n = 147)	All data (n = 435)
Female sex − no. (%)	45 (33)	82 (54)	52 (35)	179 (41)
Median age (IQR) − years	60 (48−71)	57 (44−67)	60 (48−72)	58 (46−71)
Intensive Care Unit − no. (%)	77 (57)	13 (8)	85 (58)	175 (40)
Intubation − no. (%)	56 (41)	10 (7)	61 (41)	127 (29)
Diabetes − no. (%)	36 (27)	13 (10)	54 (37)	127 (29)
Hypertension − no. (%)	72 (54)	23 (17)	N.a.	95 (35)
Coronary artery disease − no. (%)	17 (13)	11 (7)	6 (12)	34 (10)
Median BMI (IQR)	27 (23−31)	28 (26−31)	29 (26−34)	29 (24−32)
Death − no. (%)	17 (13)	3 (2)	5 (3)	25 (6)

due to genetic or socioeconomic factors. For DS1 and DS2, partial genetic data was available from exome sequencing for 153 out of 302 patients (51%). Rather than excluding all individuals with missing genetic background information, we sought other means of testing if genetic background has impact on the association. First, we did a PC analysis based on SNP variants in the available data (Supplemental Figure 3). Next, we tested whether there is any significant difference in the first four components between individuals who are or are not treated at the ICU, intubated, as well as difference between carriers of the HLA C*04:01 allele and patients lacking that allele (Supplemental Table 5). No significant associations of genetic background were found. Then, we identified outliers in the PC analysis plot, removed them and repeated the whole procedure again to detect any further outliers (Supplemental Figure 4). Finally, we repeated the full meta-analysis using only those samples from DS1 and DS2 for which genetic data was available and which were not outliers on the PC analysis (as determined by the IQR method of detecting outliers). Then, we calculated the associations between alleles and ICU / intubation status using PC1 and PC2 as covariates and compared the results of the meta-analysis with the results from the total data set (Supplemental Figure 5). The calculated effect sizes were highly similar (Pearson r^2 = 0.90 for intubation status and 0.87 for ICU), indicating that the effect did not depend on the genetic background in DS1 and DS2. Moreover, the HLA-C*04:01 allele remained the allele with the highest odds ratio (OR) in both meta-analyses.

While the majority of patients in DS1 were of European ancestry (Supplemental Figure 6), the data set from the United States (DS3) included patients from Asian, African American, Hispanic, and Caucasian ethnicities (Supplemental Figure 7). While such designations can be used at best as proxies of a genetic background [31], it is known that among these ethnicities, severity and incidence of COVID-19 as well as frequencies of alleles vary. Therefore, there was a possibility that the association between HLA-C*04:01 (known to differ between various populations) and COVID-19 severity is a statistical artifact in DS3. The association of HLA-C*04:01 remained significant despite including ethnicity designation as a covariate for the association test

in the DS3. On the level of individual ethnicities, we found the association of HLA-C*04:01 with COVID-19 severity to be present in all groups. Due to low sample size resulting from this subgroup analysis, this association did not reach a level of significance in any ethnicity except for Caucasian. However, all patients of African American, Hispanic, and Caucasian ethnicity, who were carriers of HLA- C*04:01 were hospitalized in the ICU, and all African American and Hispanic, as well as 66% of Caucasian carriers of HLA-C*04:01 were intubated (Supplemental Figure 7).

3.4. Association of alleles with viral load

For DS1, first-measured viral load data were available. Notably, measured viral load was not associated with severity of symptoms such as ICU/intubation status or troponin T HS levels (Supplementary Table 6). There were five significant associations of initial viral load with HLA alleles at $p < 0.05$, none of which were significant after correction for false discovery rate, but with at least medium effect size. The HLA-C*04:01 allele was not significantly associated with viral load. Given that viral load was not associated with disease severity in DS1 and that viral load data were not available for DS2 or DS3, we did not attempt to confirm whether the lack of association between HLA and viral load in DS1 was mirrored in the two other data sets.

3.5. Meta-analysis of HLA-C*04:01 allele and COVID-19 severity / susceptibility

After having identified HLA-C*04:01 as a risk allele for severe clinical course in COVID-19, we sought to test the hypothesis that it is also associated with COVID-19 susceptibility. To test this hypothesis, we analyzed data from the Genotek cohort of 2113 self-reported COVID-19 positive individuals and 10,026 controls. In the Genotek cohort, the frequency of HLA-C*04:01 was 13%. An association analysis using logistic regression with age, sex, and BMI as covariates showed that HLA-C*04:01 significantly increased the risk of infection with COVID-19 (OR = 1.16, p-value = 0.005).

Table 2
Top five results for the association of human leukocyte antigen (HLA) alleles with admission to the intensive care unit (ICU) and intubation. There was a significant association (adjusted p-value, p Adjusted) of HLA-C*04:01 with intubation. OR 1−3, odds ratios (with 95% CI) in data sets 1−3; OR Meta-analysis, odds ratio from meta-analysis; RR, risk ratio with 95% CI; p-value, p-value from meta-analysis; p Adjusted, p-value corrected for multiple testing using Bonferroni correction.

| trait	Allele	OR DS1	OR DS2	OR DS3	OR Meta-analysis	RR	p-value	p Adjusted
icu	HLA_C 04:01	3.4 [1.2−9.8]	0.9 [0.2−3.6]	3.2 [1.1−8.8]	2.5 [1.3−4.8]	1.2 [0.9−1.5]	0.0038	0.26
„	DQA1 01:02	0.8 [0.3−1.7]	5.2 [1.5−18.4]	0.6 [0.3−1.3]	0.9 [0.6−1.6]	1.1 [0.8−1.4]	0.021	1
„	HLA_C 06:02	0.5 [0.2−1.3]	4.9 [1.1−21.5]	1.5 [0.6−2.2]	1.2 [0.6−2.2]	1.0 [0.8−1.4]	0.022	1
„	DRB1 15:01	0.6 [0.2−1.6]	18.7 [2.6−135.9]	0.7 [0.2−2.1]	1.0 [0.5−1.9]	1.1 [0.8−1.5]	0.032	1
„	HLA_C 07:02	1.0 [0.4−2.2]	2.3 [0.5−10.1]	0.4 [0.2−0.9]	0.7 [0.4−1.3]	1.1 [0.8−1.4]	0.063	1
intub	HLA_C 04:01	5.4 [1.9−15.1]	1.5 [0.3−6.9]	3.3 [1.3−8.5]	3.5 [1.9−6.6]	1.5 [1.1−2.1]	0.00011	0.0074
„	DQA1 01:02	0.7 [0.3−1.5]	3.8 [1.0−15.0]	0.5 [0.2−1.2]	0.8 [0.5−1.3]	1.0 [0.7−1.4]	0.018	1
„	DQA1 01:01	2.7 [1.0−7.1]	0.0 [0.0−Inf]	1.6 [0.5−4.7]	2.1 [1.0−4.4]	1.1 [0.8−1.6]	0.039	1
„	DQA1 03:01	2.0 [0.7−6.2]	0.7 [0.1−6.1]	2.1 [0.9−5.0]	1.9 [1.0 - 3.6]	1.5 [1.1−2.1]	0.039	1
„	HLA_A 03:01	0.7 [0.3−1.8]	1.6 [0.4−6.9]	4.3 [1.2−14.9]	1.4 [0.7−2.7]	0.9 [0.6−1.4]	0.046	1
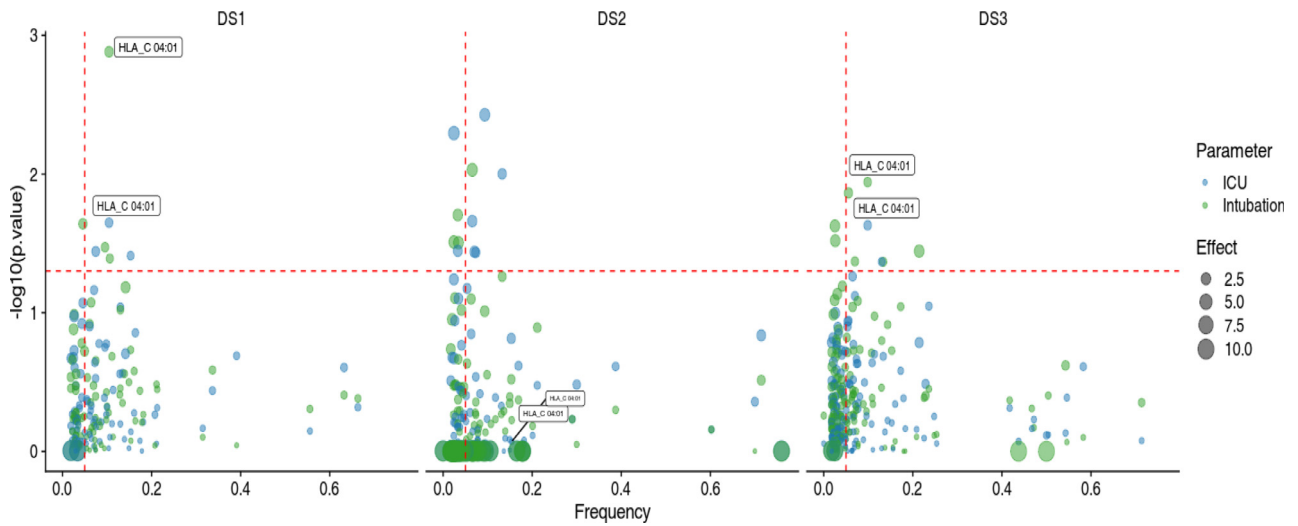
Fig. 1. Association between alleles and categorical response parameters in data sets, DS 1−3. Each dot represents one allele / response parameter association. Colors correspond to response parameters in the meta-analysis (intensive care unit, ICU status and intubation status). The Y-axis represents the negative logarithm of the *p*-value obtained from the logistic regression test. Sizes of dots correspond to the calculated effect size of the associations (log-odds ratios). Dashed vertical line corresponds to the minor allele frequency selection threshold (0.05). Dashed horizontal line corresponds to *p*-value of 0.05 (not corrected for multiple testing).

Next, we studied if HLA-C*04:01 was associated with COVID-19 susceptibility and severity in previous GWAS and HLA analyses reported in the literature. We conducted two meta-analyses on HLA-C*04:01 and COVID-19 severity/susceptibility (Fig. 3). For the severity meta-analysis we included association results from four cohorts with 7796 cases and 875,694 controls in total. The overall effect for HLA-C*04:01 on COVID-19 severity was OR = 1.1 (95%-CI: 1.04−1.17, *p*-value = 5.8e-04). The leading variant, rs143334143 (CCHCR1), was significantly associated with COVID-19 severity (OR = 1.9, *p*-value = $8.8 \times 10{-}18$). This variant is in linkage disequilibrium with HLA-C*04:01 (r^2 = 0.77, D' = 0.97) in the 1KG European cohort.

For the susceptibility meta-analysis, we included association results from four cohorts with 39,044 cases and 1,668,157 controls in total. The overall effect for HLA-C*04:01 on COVID-19 susceptibility was OR = 1.06 (95%-CI: 1.04−1.09, *p*-value = 1.4e-05). However, in this meta-analysis significant heterogeneity between studies was observed (I^2 = 78%, τ^2 = 0.0115, *p*-value = 3.2e-03).

3.6. Affinity analysis

A possible explanation for the effect of an HLA allele on disease severity might be abnormal binding affinity with SARS-CoV-2 peptides. To test that hypothesis, we used a previously described data set of putative (computed in silico) binding affinities [15]. Following an approach described by Iturrieta-Zuazo and colleagues [32], we calculated the number of SARS-CoV-2 peptides, which bind either "tightly" (at < 50 mM) or "loosely" (at < 500 mM) for each allele. Using this approach, we discovered that HLA-C*04:01 is among the ten alleles with the fewest predicted binding SARS-CoV-2 peptides (supplemental Table 7). Therefore, we sought to systematically test the hypothesis that the putative affinity with SARS-CoV-2 peptides is a correlate of disease severity. First, for each of the response variables (ICU and intubation status, troponin T hs levels, composite score) we tested whether the effect size is correlated with the number of potentially binding peptides. We included all alleles (not filtered for frequency) in this comparison. While in all data sets combined there were no
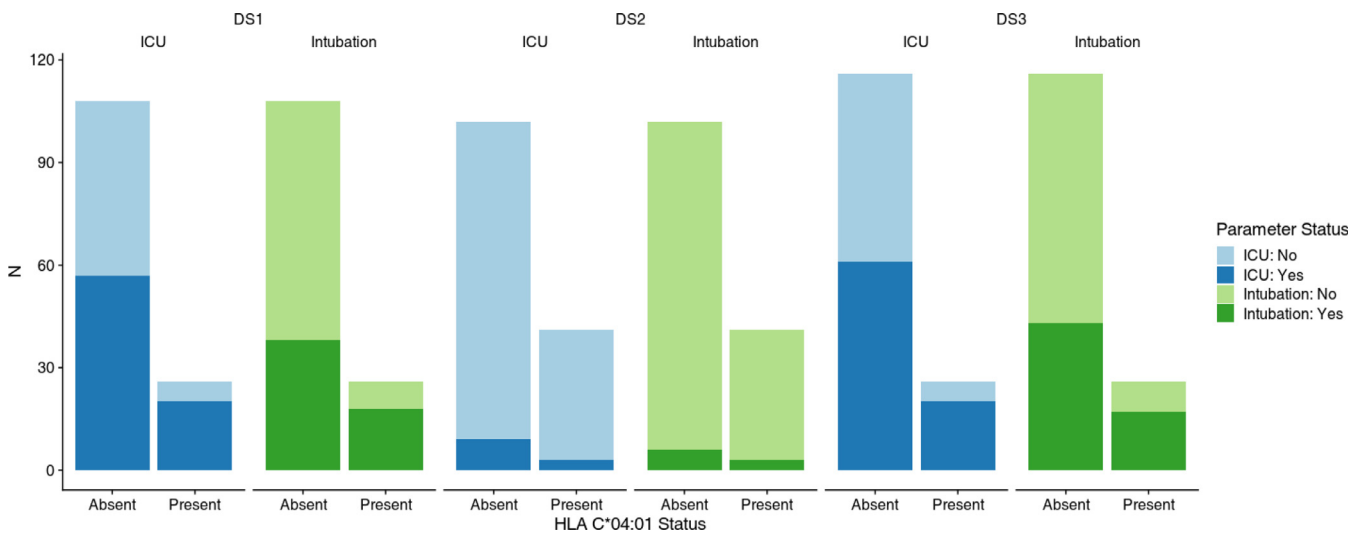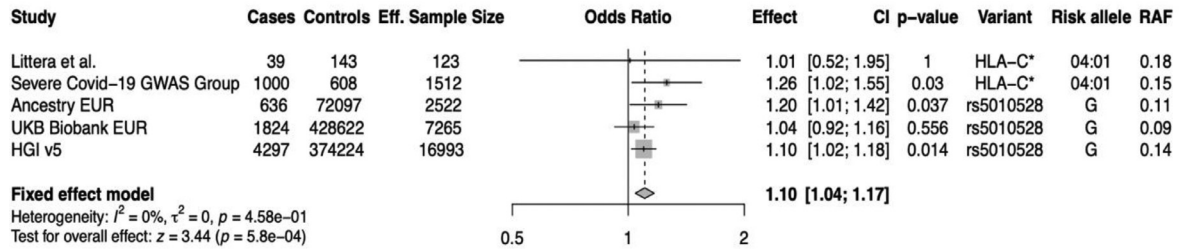

Fig. 2. Associations of HLA-C*04:01 with intensive care unit (ICU) and intubation status. Each panel shows the difference between patients who were carrying the allele (HLA C*04:01 status: Present) and patients who did not carry the allele (HLA C*04:01 status: Absent). Colors correspond to variable status. The vertical axis shows the absolute number of patients. Data sets, DS 1−3.
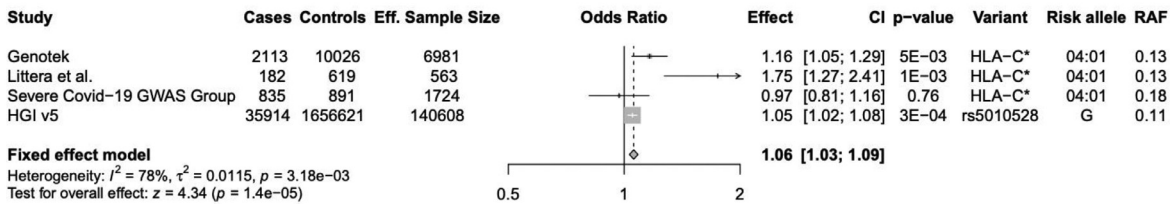
a)

Study	Cases	Controls	Eff. Sample Size	Odds Ratio	Effect	CI	p-value	Variant	Risk allele	RAF
Littera et al.	39	143	123		1.01	[0.52; 1.95]	1	HLA-C*	04:01	0.18
Severe Covid-19 GWAS Group	1000	608	1512		1.26	[1.02; 1.55]	0.03	HLA-C*	04:01	0.15
Ancestry EUR	636	72097	2522		1.20	[1.01; 1.42]	0.037	rs5010528	G	0.11
UKB Biobank EUR	1824	428622	7265		1.04	[0.92; 1.16]	0.556	rs5010528	G	0.09
HGI v5	4297	374224	16993		1.10	[1.02; 1.18]	0.014	rs5010528	G	0.14
Fixed effect model					**1.10**	**[1.04; 1.17]**				

Heterogeneity: $I^2 = 0\%$, $\tau^2 = 0$, $p = 4.58\text{e}{-01}$
Test for overall effect: $z = 3.44$ ($p = 5.8\text{e}{-04}$)

b)

Study	Cases	Controls	Eff. Sample Size	Odds Ratio	Effect	CI	p-value	Variant	Risk allele	RAF
Genotek	2113	10026	6981		1.16	[1.05; 1.29]	5E-03	HLA-C*	04:01	0.13
Littera et al.	182	619	563		1.75	[1.27; 2.41]	1E-03	HLA-C*	04:01	0.13
Severe Covid-19 GWAS Group	835	891	1724		0.97	[0.81; 1.16]	0.76	HLA-C*	04:01	0.18
HGI v5	35914	1656621	140608		1.05	[1.02; 1.08]	3E-04	rs5010528	G	0.11
Fixed effect model					**1.06**	**[1.03; 1.09]**				

Heterogeneity: $I^2 = 78\%$, $\tau^2 = 0.0115$, $p = 3.18\text{e}{-03}$
Test for overall effect: $z = 4.34$ ($p = 1.4\text{e}{-05}$)

Fig. 3. a. Meta-analysis on HLA-C*04:01 and COVID-19 severity.b. Meta-analysis on HLA-C*04:01 and COVID-19 susceptibility.Odds ratios (Effect column) for each variant and the corresponding 95% confidence intervals (CI) are plotted as horizontal bars with a square in the middle. The size of the square is proportional to the weight of the corresponding cohort in the meta-analysis. The overall effect for the fixed-effect model is plotted as a diamond and vertical dotted line. For those cohorts where HLA alleles were not studied, we consider rs5010528 as tag SNP for HLA-C*04:01. Effective sample size (Eff. Sample Size) was calculated as 4/(1/N_cases + 1/N_controls). RAF: risk allele frequency.

large effects in alleles with many potential binding peptides, negative correlation was only significant in DS3 (Spearman test, $p < 0.007$, $\rho = 0.27$).

Next, for each patient, we calculated how many SARS-CoV-2 peptides can the type I alleles in that patient potentially bind and tested whether there is an association between disease severity (using troponin T hs levels, intubation and ICU status and composite score as proxies) and that number. However, none of these tests showed a significant association in this subgroup analysis.

Furthermore, we tested if HLA-C*04:01 is associated with higher C-reactive protein (CRP) levels as a surrogate for systemic inflammation. The differences were significant, (Wilcoxon test, $p = 0.021$; Supplemental Figure 8; Supplemental Table 8) but the effect was small (Wilcoxon effect size $r = 0.2$). CRP was, however, significantly different between patients who were treated at the ICU and those who were not ($p < 10{-5}$, $r = 0.53$), as well as between those who were intubated and those who were not ($p < 10{-5}$; $r = 0.6$).

3.7. Exome sequencing data

Since polymorphisms in the KIR2DS4 gene in combination with the HLA-C*04:01 allele have been reported to increase viral load and lead to severe clinical course in patients with human immunodeficiency virus (HIV) [33], we evaluated if presence of the KIR2DS4f variant was associated with our outcome variables (e.g. viral load, intubation). However, no association was identified. A combination of a KIR2DS4f variant with HLA-C*04:01 was identified in only four patients, of which only one was homozygous at the KIR2DS4 locus. This patient had a severe clinical course, was intubated, and had highly elevated troponin T hs and maximum WHO ordinal scale for clinical improvement score.

4. Discussion

In the beginning of the outbreak of the COVID-19 pandemic in Europe in spring 2020, an international collaboration of European centers was established to address the question of whether there were potential genetic host factors associated with severe clinical course of SARS-CoV-2 infection. Samples were collected in Spain,

Switzerland, and Germany [19] and subsequently processed in Germany for full-length sequencing of HLA genes.

We identified HLA-C*04:01 as a potential risk allele, which was associated with a two-fold increased risk of intubation when present in the form of at least one allele. This is the first significant description of this HLA allele as a relevant allele for the clinical course.

Importantly these findings were reproduced in an independent public data set obtained from patients with COVID-19 at Albany Medical Center, in the United States using RNA sequencing [17,18] and a data set from the University of California, San Francisco, United States. Using this approach, we demonstrated broad applicability of our findings and confirmed reproducibility in data gathered through a distinct method – the analysis of the transcriptome [34]. This is particularly relevant in the context of regional differences in allele frequencies [35,36]. In that regard, HLA-B* 46:01 was shown to be associated with severe clinical course in SARS-CoV in Asia [14]. However, this allele is very rare in Germany, Spain, Switzerland and the United States [16]. In line with that, HLA-B* 46:01 was not represented in our European data set. In a correlation analysis we demonstrated that the frequencies of most alleles of the German cohort were similar to those in the cohorts from Switzerland, Spain, and the United States.

However, an observation that warrants mention is that the Spanish population in DS2 consisted of a larger percentage of women (63% in Spain vs 33% in Germany). Importantly, DS2 was a representative sample of patients in Sevilla, where our recruiting center collected samples consecutively at the peak of the pandemic and by now has seen a similar proportion of women (60%) in all patients, who have since been recruited ($n = 842$). While our analysis did not reveal any association of specific HLA alleles with sex (supplemental Table 8), it has been shown that female sex is associated with better outcomes [37]. This effect of sex may have masked the effect of HLA-C*04:01 validation in DS 2. A higher percentage of women in this cohort may have contributed to the low rate of patients who required treatment in the intensive care unit. Future research is necessary to investigate this possibility.

The reported OR for HLA-C*04:01 were high with very broad CI (OR in meta-analysis 3.5 [1.9−6.6]). The reason for this was inclusion

of several covariates (age, sex, ethnicity) in relation with a relatively modest number of samples in each data set.

In addition to the robust association between HLA-C*04:01 and intubation, previous literature suggests a potential biological role of HLA-C*04:01 in defending viral infections. Specifically, it has been shown that in patients infected with the HIV, HLA-C*04:01 in combination with polymorphisms in the KIR2DS4 gene is associated with high viral load and severe clinical course [33]. Moreover, HLA-C*04:01 has been reported to occur at a higher frequency in COVID-19 patients than in the healthy population [38].

The allele frequency of HLA-C*04:01 is approximately 13% in Germany overall and 19% in German residents of Turkish origin [16], suggesting that the presented findings may impact a relevant fraction of the German population. Similar numbers are found in Switzerland, where the allele frequency is about 16% and Spain, where an allele frequency of 15% has been reported [16]. Notably, the difference in HLA-C*04:01 frequency in populations with different genetic background may suggest the presence of a potential confounder such as that the observed effect may be due to a combination of differences in allele frequencies and socioeconomic factors influencing the disease severity. However, despite testing this hypothesis with multiple approaches, we were unable to find confirmation of this effect. Association between HLA-C*04:01 and disease severity did not change when genetic background or ethnicity were used as covariates, or when the effect was calculated for a genetically homogeneous population.

Since HLA-C*04:01 is associated with viral load in HIV, we sought to evaluate if this was true for SARS-CoV-2. We did not find any association of HLA-C*04:01 with first-measured viral load of patients during their hospitalization. In addition, the first-measured viral load in intubated patients did not differ from patients with mild disease. We support the common hypothesis that by the time the patients present with severe respiratory failure, the virus may have been already partially cleared from the body in some of them and the disease may be maintained through inflammatory or autoimmune mechanisms similar to other viral respiratory illnesses [39]. The dynamics of viral load in SARS-CoV 2 have been reported comprehensively in the literature [40−42]. It may be beneficial to test for HLA-C*04:01 in clinical trials to identify from which drug class this high-risk population benefits most.

It is also possible that viral load of patients with HLA-C*04:01 may have been higher during the very first days of infection [43], but since most patients only become symptomatic after approximately a week [42,44,45], this time window may have been missed in some patients of our cohort to see a potential relationship of HLA-C*04:01 and viral loads measured during early infection. In general, there was no association of any specific HLA allele with first-measured viral load in our cohort.

Exome sequencing revealed a combination of a KIR2DS4f variant with HLA-C*04:01 in only four patients. One of those patients was homozygous at the KIR2DS4 locus and had a severe clinical course. While these findings are interesting, the small number of cases preclude drawing meaningful conclusions.

Notably, HLA affinity analysis revealed that HLA-C*04:01 binding affinity with SARS-CoV-2 peptides was amongst the ten lowest HLA alleles (supplemental Table 7). Delayed immune response due to low HLA binding affinity may be one potential biological explanation for the severe clinical course observed in patients with HLA-C*04:01.

Alternatively, HLA-C*04:01 may have led to worse outcomes by predisposing patients to a more severe inflammatory state. To test that, we evaluated for a potential association of HLA-C*04:01 with CRP serum levels. Indeed, CRP levels were higher in carriers of HLA-C*04:01 (p = 0.02, supplemental Figure 8). CRP correlated with the risk of intubation, as shown in prior studies [46,47].

The number of samples in our study was relatively small. To avoid the detrimental effects of small sample size on the statistical power,

we have limited our study only to alleles filtered by MAF, resulting in 35 alleles tested for 2 phenotypes (ICU and intubation status) in all three data sets, and three additional phenotypes in DS1 only. Furthermore, we have included several different cohorts and compared our results with publicly available data sets. Nonetheless, small sample size might result in overestimated effect sizes. This could be clearly seen when comparing models including covariates (such as sex, and ethnicity) and models without such covariates.

We found significant association of the HLA-C*04:01 allele with COVID-19 susceptibility in the independent cohort. Moreover, we conducted two meta-analyses to combine the association results of HLA-C*04:01 or linked SNPs with COVID-19 severity and susceptibility. In these analyses, HLA-C*04:01 was significantly associated with COVID-19 severity and susceptibility. It should be noted that some studies included in the meta-analyses did not detect the association between HLA-C*04:01 and COVID-19, which can be explained by several factors. First, small sample size was a major limiting factor in some studies, resulting in a design and test combination that was underpowered for detecting hypothetical effect sizes of interest. Second, there were inconsistent definitions of phenotypes between different studies. In many studies the control groups were defined as "general population" with limited details. HLA alleles have different frequencies in different populations, which could affect the power of the studies with heterogeneous cohorts (e.g. COVID-19 Host Genetics Initiative) and lead to false positive findings. Some studies did not genotype or impute HLA alleles. Therefore, we had to use the tag SNPs which are in strong linkage disequilibrium with the considered HLA allele to include these studies in the meta-analysis. However, a high variability of linkage disequilibrium between certain SNP and HLA alleles for various populations in 1KG data can be observed.

The study by Pairo-Castineira and colleagues included 2244 cases from the GenOMICC data base who had a confirmed diagnosis of COVID-19 based on clinical testing and who required continuous cardio-respiratory monitoring, and ancestry-matched (5 to 1) controls from the UK Biobank in the discovery cohort. We found that the leading variant rs143334143 (CCHCR1) was significantly associated with COVID-19 severity and that this variant is in linkage disequilibrium with HLA-C*04:01 in the 1KG European cohort. However, in the meta-analysis [48] this variant was removed from the final results due to high heterogeneity (mainly driven by the GenOMICC cohort). The possible reasons of such heterogeneity include:

1) Different definitions of the controls (population control, PCR negative controls etc.); 2) various ancestry of the cohorts; 3) small sample sizes of certain studies.

Some other SNPs which are in the same linkage disequilibrium-block with rs143334143 and HLA-C*04:01 did not demonstrate a similar effect. The different analysis method of cases and controls in this study might explain that mentioned difference and therefore we opted for not including this study into the meta-analysis.

Our study is an important addition to prior HLA research in the field of COVID-19 that provided cumulative evidence for a potential predisposing susceptibility of HLA alleles for infection with SARS-CoV-2. It has been shown that specific HLA alleles were more common in patients infected with SARS-CoV-2 as compared to healthy individuals (e.g. HLA-DRB1* 15:01, HLA-DQB1*06:02, HLA-DRB1×08, HLA-C*04:01) and it was extrapolated that these alleles may predispose to infection [38,49−51]. It is not surprising that some of these alleles differed from ours, as we identified a genetic factor specifically for severe clinical course amongst patients with COVID-19, while other groups compared healthy versus diseased. A recent GWAS study identified blood group A as a major risk factor for COVID-19, but did not detect associations of the classical HLA loci with either COVID-19 infection or disease severity [8]. That study used target short-read sequencing of HLA regions for a sub-cohort and investigated association with HLA-C:04 only on a 2-digit rather than on the 4-digit level used in our study. Both factors may contribute to a reduced power that made it impossible to detect the

association that our approach finds supported in three data sets from different studies and in different populations.

In addition to host factors, it has been shown that genetic factors of the virus may affect affinity of host HLA to viral structures. In that regard, de Sousa and colleagues have shown that a single mutation in the wild type sequence of SARS-CoV-2 could affect peptide binding of the virus to HLA Class II alleles [52].

In summary, we applied full-length HLA sequencing to patients with COVID-19 to enable identification of HLA alleles with high precision in a large international data set and validated our findings successfully in an independent cohort. We used analysis of clinical and laboratory parameters, viral load data, and whole exome sequencing to evaluate biological plausibility.

A limitation of this study was that in order to promptly respond to the COVID-19 pandemic with an international multi-center study, even patients with limited metadata were enrolled in the study. However, risk factors and clinical outcomes that we considered most relevant based on prior literature were collected and analyzed for all patients included in our cohorts.

Reproducibility of our data in large GWAS analyses supports our findings despite this limitation.

Furthermore, it warrants mention that any study recruiting patients from a given clinical center (such as our center in Berlin) will be somewhat biased based on its patient selection – the Berlin population will not be representative for all of Germany or Europe, and the DS3 populations are not likely to be representative for all of US. However, we evaluated whether allele frequencies are similar to the allele frequencies reported based on larger population studies. Furthermore, we tested our findings in multiple international data sets and were able to replicate our findings, suggesting broad applicability of our results.

Larger cohorts with consistent pheno typing are required to draw more precise estimates about effect-sizes for accurate translation into further research and clinical care.

Author contributions

BH, FK, JW, LE, PS, TG and UL conceptualized the trial.

AB, ABr, BH, BHi, BM, BW, CD, CS, CT, DB, DP, FK, GF, HB, HS, JM, JMD, JP, JR, JS, JW, KD, LE, LJ, LR, ME, MFF, MH, MM, MS, MW, NS, PCSG, PS, PSt, RAA, RB, RHH, SH, SJ, SR, SSi, TB, TCJ, TEH, TG, TZ, UL, VC, XW, ZK and CC contributed to study design and data collection.

BH, RA, BHi, CS, DB, FK, LE, MH, PS, JZ, TCJ, TG and TZ contributed to methodologies and analysis.

Contribution to methodologies, data curation and software conception for study data analysis: RA, IV, IP, JW, MS, JZ and SR.

BH, BHi, CC, CQ, PCSG, PS, UL and WP contributed to study supervision and organization.

BH, BHi, CD, CQ, JW, PS, TB, TCJ, UK, XW and ZK conducted the laboratory processing and data preparation.

BH, BHi, CMH, CQ, CRL, CSC, DJE, DP, FG, IP, IV, JA, JB, JW, JZ, KNK, MFK, MH, PGW, RA, SR, UK, UV contributed to study data validation.

All authors were involved in the drafting and reviewing of the manuscript and provided approval for submission.

Declaration of Competing Interest

Acknowledgements

Data sharing statement

Patient genetic data underlying the study can be made available upon request pending necessary ethics committee and / or

confidentiality statements approvals. Summary statistics are provided in supplemental Table 1.

Additional encrypted / anonymized data are available upon request. Custom R scripts and functions are available from https://github.com/bihealth/covidhla/.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.eclinm.2021.101099.

References

[1] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020;395:497–506.

[2] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W. China Novel Coronavirus I and Research T. a novel coronavirus from patients with Pneumonia in China, 2019. N Engl J Med 2020;382:727–33.

[3] Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, Spitters C, Ericson K, Wilkerson S, Tural A, Diaz G, Cohn A, Fox L, Patel A, Gerber SI, Kim L, Tong S, Lu X, Lindstrom S, Pallansch MA, Weldon WC, Biggs HM, Uyeki TM. Pillai SK and Washington State -nCo VCIT. first case of 2019 novel coronavirus in the United States. N Engl J Med 2020;382:929–36.

[4] Wu Z, McGoogan JM. Characteristics of and important lessons from the Coronavirus Disease 2019 (COVID-19) Outbreak in China: summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention. JAMA 2020;323:1239–42.

[5] Asadi S, Wexler AS, Cappa CD, Barreda S, Bouvier NM, Ristenpart WD. Aerosol emission and supere mission during human speech increase with voice loudness. Sci Rep 2019;9:2348.

[6] Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DSC, Du B, Li LJ, Zeng G, Yuen KY, Chen RC, Tang CL, Wang T, Chen PY, Xiang J, Li SY, Wang JL, Liang ZJ, Peng YX, Wei L, Liu Y, Hu YH, Peng P, Wang JM, Liu JY, Chen Z, Li G, Zheng ZJ, Qiu SQ, Luo J, Ye CJ, Zhu SY. Zhong NS and China Medical Treatment Expert Group for C. clinical characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020;382:1708–20.

[7] Zietz M, Zucker J, Tatonetti NP. Associations between blood type and COVID-19 infection, intubation, and death. Nat Commun 2020;11:5761.

[8] Severe Covid GG, Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, Fernandez J, Prati D, Baselli G, Asselta R, Grimsrud MM, Milani C, Aziz F, Kassens J, May S, Wendorff M, Wienbrandt L, Uellendahl-Werth F, Zheng T, Yi X, de Pablo R, Chercoles AG, Palom A, Garcia-Fernandez AE, Rodriguez-Frias F, Zanella A, Bandera A, Protti A, Aghemo A, Lleo A, Biondi A, Caballero-Garralda A, Gori A, Tanck A, Carreras Nolla A, Latiano A, Fracanzani AL, Peschuck A, Julia A, Pesenti A, Voza A, Jimenez D, Mateos B, Nafria Jimenez B, Quereda C, Paccapelo C, Gassner C, Angelini C, Cea C, Solier A, Pestana D, Muniz-Diaz E, Sandoval E, Paraboschi EM, Navas E, Garcia Sanchez F, Ceriotti F, Martinelli-Boneschi F, Peyvandi F, Blasi F, Tellez L, Blanco-Grau A, Hemmrich-Stanisak G, Grasselli G, Costantino G, Cardamone G, Foti G, Aneli S, Kurihara H, ElAbd H, My I, Galvan-Femenia I, Martin J, Erdmann J, Ferrusquia-Acosta J, Garcia-Etxebarria K, Izquierdo-Sanchez L, Bettini LR, Sumoy L, Terranova L, Moreira L, Santoro L, Scudeller L, Mesonero F, Roade L, Ruhlemann MC, Schaefer M, Carrabba M, Riveiro-Barciela M, Figuera Basso ME, Valsecchi MG, Hernandez-Tejero M, Acosta-Herrera M, D'Angio M, Baldini M, Cazzaniga M, Schulzky M, Cecconi M, Wittig M, Ciccarelli M, Rodriguez-Gandia M, Bocciolone M, Miozzo M, Montano N, Braun N, Sacchi N, Martinez N, Ozer O, Palmieri O, Faverio P, Preatoni P, Bonfanti P, Omodei P, Tentorio P, Castro P, Rodrigues PM, Blandino Ortiz A, de Cid R, Ferrer R, Gualtierotti R, Nieto R, Goerg S, Badalamenti S, Marsal S, Matullo G, Pelusi S, Juzenas S, Aliberti S, Monzani V, Moreno V, Wesse T, Lenz TL, Pumarola T, Rimoldi V, Bosari S, Albrecht W, Peter W, Romero-Gomez M, D'Amato M, Duga S, Banales JM, Hov JR, Folseraas T, Valenti L, Franke A. Karlsen TH. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. N Engl J Med 2020;383:1522–34.

[9] Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, Walker S, Parkinson N, Fourman MH, Russell CD, Furniss J, Richmond A, Gountouna E, Wrobel N, Harrison D, Wang B, Wu Y, Meynert A, Griffiths F, Oosthuyzen W, Kousathanas A, Moutsianas L, Yang Z, Zhai R, Zheng C, Grimes G, Beale R, Millar J, Shih B, Keating S, Zechner M, Haley C, Porteous DJ, Hayward C, Yang J, Knight J, Summers C, Shankar-Hari M, Klenerman P, Turtle L, Ho A, Moore SC, Hinds C, Horby P, Nichol A, Maslove D, Ling L, McAuley D, Montgomery H, Walsh T, Pereira A, Renieri A, Gen OI, Investigators I, Initiative C-HG, andMe I, Investigators B, Gen CI, Shen X, Ponting CP, Fawkes A, Tenesa A, Caulfield M, Scott R, Rowan K, Murphy L, Openshaw PJM, Semple MG, Law A, Vitart V, Wilson JF, Baillie JK. Genetic mechanisms of critical illness in Covid-19. Nature 2020.

[10] Heidecker B, Kittleson MM, Kasper EK, Wittstein IS, Champion HC, Russell SD, Hruban RH, Rodriguez ER, Baughman KL, Hare JM. Transcriptomic biomarkers for the accurate diagnosis of myocarditis. Circulation 2011;123:1174–84.

[11] Cho JH, Gregersen PK. Genomics and the multifactorial nature of human autoimmune disease. N Engl J Med 2011;365:1612–23.

[12] Dutta M, Dutta P, Medhi S, Borkakoty B, Biswas D. Polymorphism of HLA class I and class II alleles in influenza A(H1N1)pdm09 virus infected population of Assam, Northeast India. J Med Virol 2018;90:854–60.

[13] Ma Y, Yuan B, Yi J, Zhuang R, Wang J, Zhang Y, Xu Z, Zhang Y, Liu B, Wei C, Zhang C, Yang A, Jin B. The genetic polymorphisms of HLA are strongly correlated with the disease severity after Hantaan virus infection in the Chinese Han population. Clin Dev Immunol 2012;2012:308237.

[14] Lin M, Tseng HK, Trejaut JA, Lee JH, Loo JH, Chu CC, Chen PJ, Su YW, Lim KH, Tsai ZU, Lin RY, Lin RS, Huang CH. Association of HLA class I with severe acute respiratory syndrome coronavirus infection. BMC Med Genet 2003;4:9.

[15] Nguyen A, David JK, Maden SK, Wood MA, Weeder BR, Nellore A, Thompson RF. Human leukocyte antigen susceptibility map for SARS-CoV-2. J Virol 2020.

[16] Gonzalez-Galarza FF, McCabe A, Melo Dos Santos EJ, Takeshita L, Ghattaoraya G, Jones AR, Middleton D. Allele Frequency Net Database. Methods Mol Biol 2018;1802:49–62.

[17] Overmyer KA, Shishkova E, Miller IJ, Balnis J, Bernstein MN, Peters-Clarke TM, Meyer JG, Quan Q, Muehlbauer LK, Trujillo EA, He Y, Chopra A, Chieng HC, Tiwari A, Judson MA, Paulson B, Brademan DR, Zhu Y, Serrano LR, Linke V, Drake LA, Adam AP, Schwartz BS, Singer HA, Swanson S, Mosher DF, Stewart R, JJ Coon, Large-Scale Jaitovich A. Multiomic analysis of COVID-19 Severity. Cell Syst 2021.

[18] Balnis J, Madrid A, Hogan KJ, Drake LA, Chieng HC, Tiwari A, Vincent CE, Chopra A, Vincent PA, Robek MD, Singer HA, RS Alisch, Jaitovich A. Blood DNA methylation and COVID-19 outcomes. Clin Epigenetics 2021;13:118.

[19] Kurth F, Roennefarth M, Thibeault C, Corman VM, Muller-Redetzky H, Mittermaier M, Ruwwe-Glosenkamp C, Heim KM, Krannich A, Zvorc S, Schmidt S, Kretzler L, Dang-Heine C, Rose M, Hummel M, Hocke A, Hubner RH, Opitz B, Mall MA, Rohmel J, Landmesser U, Pieske B, Knauss S, Endres M, Spranger J, Mockenhaupt FP, Tacke F, Treskatsch S, Angermair S, Siegmund B, Spies C, Weber-Carstens S, Eckardt KU, Schurmann D, Uhrig A, Stegemann MS, Zoller T, Drosten C, Suttorp N, Witzenrath M, Hippenstiel S, von Kalle C, Sander LE. Studying the pathophysiology of coronavirus disease 2019: a protocol for the Berlin prospective COVID-19 patient cohort (Pa-COVID-19). Infection 2020;48:619–26.

[20] Thibeault C, Muhlemann B, Helbig ET, Mittermaier M, Lingscheid T, Tober-Lau P, Meyer-Arndt LA, Meiners L, Stubbemann P, Haenel SS, Bosquillon de Jarcy L, Lippert L, Pfeiffer M, Stegemann MS, Roehle R, Wiebach J, Hippenstiel S, Zoller T, Muller-Redetzky H, Uhrig A, Balzer F, von Kalle C, Suttorp N, Jones TC, Drosten C, Witzenrath M, Sander LE, Pa CSG, Corman VM, Kurth F. Clinical and virological characteristics of hospitalised COVID-19 patients in a German tertiary care centre during the first wave of the SARS-CoV-2 pandemic: a prospective observational study. Infection 2021.

[21] Wolfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Muller MA, Niemeyer D, Jones TC, Vollmar P, Rothe C, Hoelscher M, Bleicker T, Brunink S, Schneider J, Ehmann R, Zwirglmaier K, Drosten C, Wendtner C. Virological assessment of hospitalized patients with COVID-2019. Nature 2020;581:465–9.

[22] Orenbuch R, Filip I, Comito D, Shaman J, Pe'er I, Rabadan R. arcasHLA: high-resolution HLA typing from RNAseq. Bioinformatics 2020;36:33–40.

[23] DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491–8.

[24] Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 2012;28:3326–8.

[25] Sébastien Lê JJ, Husson François. FactoMineR: an R package for multivariate analysis. J Stat Softw 2008.

[26] Zheng X, Shen J, Cox C, Wakefield JC, Ehm MG, MR Nelson, Weir BS. HIBAG—HLA genotype imputation with attribute bagging. Pharmacogenomics J 2014;14:192–200.

[27] Zaykin DV. Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. J Evol Biol 2011;24:1836–41.

[28] Horowitz JE, Kosmicki JA, Damask A, Sharma D, Roberts GHL, Justice AE, Banerjee N, Coignet MV, Yadav A, Leader JB, Marcketta A, Park DS, Lanche R, Maxwell E, Knight SC, Bai X, Guturu H, Sun D, Baltzell A, Kury FSP, Backman JD, Girshick AR, O'Dushlaine C, McCurdy SR, Partha R, Mansfield AJ, Turissini DA, Li AH, Zhang M, Mbatchou J, Watanabe K, Gurski L, McCarthy SE, Verma A, Sirugo G, Center RG, Ritchie MD, Jones M, Balasubramanian S, Salerno WJ, Shuldiner AR, Rader DJ, Mirshahi T, Locke AE, Marchini J, Overton JD, Carey DJ, Habegger L, Cantor MN, Rand KA, Hong EL, Reid JG, Ball CA, Baras A, Abecasis GR, Ferreira MA. Common genetic variants identify therapeutic targets for COVID-19 and individuals at high risk of severe disease. medRxiv. 2020 2020.12.14.20248176.

[29] Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. Nature 2015;526:68–74.

[30] Abi-Rached L, Gouret P, Yeh JH, Di Cristofaro J, Pontarotti P, Picard C, Paganini J. Immune diversity sheds light on missing variation in worldwide genetic diversity panels. PLoS ONE 2018;13:e0206512.

[31] Yudell M, Roberts D, DeSalle R, Tishkoff S. Science and Society. Taking race out of human genetics. Science 2016;351:564–5.

[32] Iturrieta-Zuazo I, Rita CG, Garcia-Soidan A, de Malet, Pintos-Fonseca A, Alonso-Alarcon N, Pariente-Rodriguez R, Tejeda-Velarde A, Serrano-Villar S, Castaner-Alabau JL, Nieto-Ganan I. Possible role of HLA class-I genotype in SARS-CoV-2 infection and progression: a pilot study in a cohort of Covid-19 Spanish patients. Clin Immunol 2020;219:108572.

[33] Olvera A, Perez-Alvarez S, Ibarrondo J, Ganoza C, Lama JR, Lucchetti A, Cate S, Hildebrand W, Bernard N, Gomez L, Sanchez J, Brander C. The HLA-C*04: 01/KIR2DS4 gene combination and human leukocyte antigen alleles with high population frequency drive rate of HIV disease progression. AIDS 2015;29:507–17.

[34] Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, Gonzalez-Porta M, Kurbatova N, Griebel T, Ferreira PG, Barann M, Wieland T, Greger L, van Iterson M, Almlof J, Ribeca P, Pulyakhina I, Esser D, Giger T, Tikhonov A, Sultan M, Bertier G, MacArthur DG, Lek M, Lizano E, Buermans HP, Padioleau I, Schwarzmayr T, Karlberg O, Ongen H, Kilpinen H, Beltran S, Gut M, Kahlem K, Amstislavskiy V, Stegle O, Pirinen M, Montgomery SB, Donnelly P, McCarthy MI, Flicek P, Strom TM, Geuvadis C, Lehrach H, Schreiber S, Sudbrak R, Carracedo A, Antonarakis SE, Hasler R, Syvanen AC, van Ommen GJ, Brazma A, Meitinger T, Rosenstiel P, Guigo R, Gut IG, Estivill X, Dermitzakis ET. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 2013;501:506–11.

[35] Meyer D, Thomson G. How selection shapes variation of the human major histocompatibility complex: a review. Ann Hum Genet 2001;65:1–26.

[36] Meyer D, VR CA, Bitarello BD, DY CB, Nunes K. A genomic perspective on HLA evolution. Immunogenetics 2018;70:5–27.

[37] Peckham H, de Gruijter NM, Raine C, Radziszewska A, Ciurtin C, Wedderburn LR, Rosser EC, Webb K, Deakin CT. Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission. Nat Commun 2020;11:6317.

[38] Littera R, Campagna M, Deidda S, Angioni G, Cipri S, Melis M, Firinu D, Santus S, Lai A, Porcella R, Lai S, Rassu S, Scioscia R, Meloni F, Schirru D, Cordeddu W, Kowa-lik MA, Serra M, Ragatzu P, Carta MG, Del Giacco S, Restivo A, Deidda S, Orrù S, Palimodde A, Perra R, Orrù G, Conti M, Balestrieri C, Serra G, Onali S, Marongiu F, Perra A, Chessa L. Human Leukocyte Antigen Complex and Other Immunogenetic and Clinical Factors Influence Susceptibility or Protection to SARS-CoV-2 Infection and Severity of the Disease Course. The Sardinian Experience. Front Immunol 2020:11.

[39] Cowling BJ, Chan KH, Fang VJ, Lau LLH, So HC, Fung ROP, Ma ESK, Kwong ASK, Chan CW, Tsui WWS, Ngai HY, Chu DWS, Lee PWY, Chiu MC, Leung GM, Peiris JSM. Comparative epidemiology of pandemic and seasonal influenza A in households. N Engl J Med 2010;362:2175–84.

[40] Arons MM, Hatfield KM, Reddy SC, Kimball A, James A, Jacobs JR, Taylor J, Spicer K, Bardossy AC, Oakley LP, Tanwar S, Dyal JW, Harney J, Chisty Z, Bell JM, Methner M, Paul P, Carlson CM, McLaughlin HP, Thornburg N, Tong S, Tamin A, Tao Y, Uehara A, Harcourt J, Clark S, Brostrom-Smith C, Page LC, Kay M, Lewis J, Montgomery P, Stone ND, Clark TA, Honein MA, Duchin JS, Jernigan JA, Public H-S, King C, Team CC-I. Presymptomatic SARS-CoV-2 Infections and transmission in a skilled nursing facility. N Engl J Med 2020;382:2081–90.

[41] Singanayagam A, Patel M, Charlett A, Lopez Bernal J, Saliba V, Ellis J, Ladhani S, Zambon M, Gopal R. Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. Euro Surveill 2020:25.

[42] Jones TC, Biele G, Muhlemann B, Veith T, Schneider J, Beheim-Schwarzbach J, Bleicker T, Tesch J, Schmidt ML, Sander LE, Kurth F, Menzel P, Schwarzer R, Zuchowski M, Hofmann J, Krumbholz A, Stein A, Edelmann A, Corman VM, Drosten C. Estimating infectiousness throughout SARS-CoV-2 infection course. Science 2021.

[43] Peiris JS, Chu CM, Cheng VC, Chan KS, Hung IF, Poon LL, Law KI, Tang BS, Hon TY, Chan CS, Chan KH, Ng JS, Zheng BJ, Ng WL, Lai RW, Guan Y, Yuen KY, Group HUSS. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. Lancet 2003;361:1767–72.

[44] Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, Yu J, Kang M, Song Y, Xia J, Guo Q, Song T, He J, Yen HL, Peiris M, Wu J. SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. N Engl J Med 2020;382:1177–9.

[45] Tindale LC, Stockdale JE, Coombe M, Garlock ES, Lau WYV, Saraswat M, Zhang L, Chen D, Wallinga J, Colijn C. Evidence for transmission of COVID-19 prior to symptom onset. Elife 2020;9.

[46] Xu JB, Xu C, Zhang RB, Wu M, Pan CK, Li XJ, Wang Q, Zeng FF, Zhu S. Associations of procalcitonin, C-reaction protein and neutrophil-to-lymphocyte ratio with mortality in hospitalized COVID-19 patients in China. Sci Rep 2020;10:15058.

[47] Mueller AA, Tamura T, Crowley CP, DeGrado JR, Haider H, Jezmir JL, Keras G, Penn EH, Massaro AF, Kim EY. Inflammatory biomarker trends predict respiratory decline in COVID-19 patients. Cell Rep Med 2020;1:100144.

[48] Ganna A. Mapping the human genetic architecture of COVID-19 by worldwide meta-analysis. medRxiv 2021 2021.03.10.21252820.

[49] Correale P, Mutti L, Pentimalli F, Baglio G, Saladino RE, Sileri P, Giordano A. HLA-B*44 and C*01 prevalence correlates with Covid19 spreading across Italy. Int J Mol Sci 2020;21.

[50] Amoroso A, Magistroni P, Vespasiano F, Bella A, Bellino S, Puoti F, Alizzi S, Vaisitti T, Boros S, Grossi PA, Trapani S, Lombardini L, Pezzotti P, Deaglio S, Brusaferro S, Cardillo M, of Italian network. Regional transplant coordinating C. HLA and AB0 Polymorphisms May Influence SARS-CoV-2 Infection and COVID-19 Severity. Transplantation 2020.

[51] Novelli A, Andreani M, Biancolella M, Liberatoscioli L, Passarelli C, Colona VL, Rogliani P, Leonardis F, Campana A, Carsetti R, Andreoni M, Bernardini S, Novelli G, Locatelli F. HLA allele frequencies and susceptibility to COVID-19 in a group of 99 Italian patients. HLA 2020;96:610–4.

[52] de Sousa E, Ligeiro D, Lerias JR, Zhang C, Agrati C, Osman M, El-Kafrawy SA, Azhar EI, Ippolito G, Wang FS, Zumla A, Maeurer M. Mortality in COVID-19 disease patients: correlating the association of major histocompatibility complex (MHC) with severe acute respiratory syndrome 2 (SARS-CoV-2) variants. Int J Infect Dis 2020;98:454–9.

[53] Heng L. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv 2013.

[54] Roe D, Kuang Rui. Accurate and Efficient KIR Gene and Haplotype Inference from Genome Sequencing Reads with Novel K-mer Signatures. bioRxiv 2020.

[55] Hsu KC, Chida S, DE Geraghty, Dupont B. The killer cell immunoglobulin-like receptor (KIR) genomic region: gene-order, haplotypes and allelic polymorphism. Immunol. Rev. 2002;190:40–52.