



OPEN

A genome-wide association study for survival from a multi-centre European study identified variants associated with COVID-19 risk of death

Francesca Minnai^{1,2,114}, Filippo Biscarini^{3,114}, Martina Esposito¹, Tommaso A. Dragani⁴, Luis Bujanda⁵, Souad Rahmouni⁶, Marta E. Alarcón-Riquelme^{7,8}, David Bernardo^{9,10}, Elena Carnero-Montoro^{7,11}, Maria Buti¹², Hugo Zeberg¹³, Rosanna Asselta^{14,15}, Manuel Romero-Gómez¹⁶, GEN-COVID Multicenter Study^{*}, Israel Fernandez-Cadenas¹⁷, Chiara Fallerini^{18,19}, Kristina Zguro^{18,19}, Susanna Croci^{18,19}, Margherita Baldassarri^{18,19}, Mirella Bruttini^{18,19,20}, Simone Furini²¹, Alessandra Renieri^{18,19,20} & Francesca Colombo¹✉

The clinical manifestations of SARS-CoV-2 infection vary widely among patients, from asymptomatic to life-threatening. Host genetics is one of the factors that contributes to this variability as previously reported by the COVID-19 Host Genetics Initiative (HGI), which identified sixteen loci associated with COVID-19 severity. Herein, we investigated the genetic determinants of COVID-19 mortality, by performing a case-only genome-wide survival analysis, 60 days after infection, of 3904 COVID-19 patients from the GEN-COVID and other European series (EGAS00001005304 study of the COVID-19 HGI). Using imputed genotype data, we carried out a survival analysis using the Cox model adjusted for age, age², sex, series, time of infection, and the first ten principal components. We observed a genome-wide significant (P -value $< 5.0 \times 10^{-8}$) association of the rs117011822 variant, on chromosome 11, of rs7208524 on chromosome 17, approaching the genome-wide threshold (P -value = 5.19×10^{-8}). A total of 113 variants were associated with survival at P -value $< 1.0 \times 10^{-5}$ and most of them regulated the expression of genes involved in immune response (e.g., CD300 and KLR genes), or in lung repair and function (e.g., FGF19 and CDH13). Overall, our results suggest that germline variants may

¹Institute of Biomedical Technologies, National Research Council, Via F.lli Cervi, 93, 20054 Segrate, MI, Italy. ²Department of Medical Biotechnology and Translational Medicine (BioMeTra), Università degli Studi di Milano, Milan, Italy. ³Institute of Agricultural Biology and Biotechnology, National Research Council, Milan, Italy. ⁴Aspidia S.R.L., Milan, Italy. ⁵Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Biodonostia Health Research Institute, Universidad del País Vasco (UPV/EHU), San Sebastián, Spain. ⁶GIGA-Medical Genomics Unit, Uliege, Liege, Belgium. ⁷GENYO, University of Granada, Andalusian Regional Government, Granada, Spain. ⁸Institute for Environmental Medicine, Karolinska Institute, Solna, Sweden. ⁹Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas (CIBEREHD), Instituto de Salud Carlos III (ISCIII), Madrid, Spain. ¹⁰Mucosal Immunology Lab, Unit of Excellence, Institute of Biomedicine and Molecular Genetics (IBGM), University of Valladolid-CSIC, Valladolid, Spain. ¹¹University of Granada, Granada, Spain. ¹²Vall D'Hebron Institut de Recerca, Barcelona, Spain. ¹³Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden. ¹⁴Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, MI, Italy. ¹⁵IRCCS Humanitas Research Hospital, Rozzano, MI, Italy. ¹⁶Digestive Diseases Unit and CiberehdVirgen del Rocío University HospitalInstitute of Biomedicine of Seville (HUVR/CSIC/US), University of Seville, Seville, Spain. ¹⁷Stroke Pharmacogenomics and Genetics Group, Sant Pau Hospital Research Institute, Barcelona, Spain. ¹⁸Medical Genetics, University of Siena, 53100 Siena, Italy. ¹⁹Department of Medical Biotechnologies, Med Biotech Hub and Competence Center, University of Siena, 53100 Siena, Italy. ²⁰Genetica Medica, Azienda Ospedaliero-Universitaria Senese, 53100 Siena, Italy. ²¹Dipartimento di Ingegneria dell'Energia Elettrica e dell'Informazione "Guglielmo Marconi", Alma Mater Studiorum - Università di Bologna, Bologna, Italy. ¹¹⁴These authors contributed equally: Francesca Minnai and Filippo Biscarini. *A list of authors and their affiliations appears at the end of the paper. ✉email: francesca.colombo@itb.cnr.it

modulate COVID-19 risk of death, possibly through the regulation of gene expression in immune response and lung function pathways.

The clinical manifestations of COVID-19, the disease caused by the SARS-CoV-2 virus, vary widely from mild respiratory symptoms to severe organ failure and death^{1–3}. The mortality rate of COVID-19 also shows remarkable temporal and spatial heterogeneity across the world^{4,5}. Several risk factors have been associated with increased mortality, such as older age, male sex, and presence of comorbidities^{6,7}.

The role of genetics in modulating the severity and outcome of COVID-19 has been a subject of intense research and growing evidence supports the existence of individual genetic factors predisposing to a severe outcome⁸. For example, the COVID-19 Host Genetics Initiative (HGI) consortium performed large-scale meta-analyses of genome-wide data from over nine thousand critically ill cases (defined as patients who required respiratory support or died from COVID-19) and over 25 thousand hospitalized cases with moderate or severe disease, compared with up to five million controls⁹. These studies identified several genetic loci associated with either critical illness or hospitalization due to COVID-19. However, the consortium did not address the survival probability of SARS-CoV-2 infected patients, which is a relevant time-to-event phenotype that has received limited attention so far. Indeed, most studies have focused on COVID-19 severity (reviewed in¹⁰), some on mortality^{11–13} and very few on survival¹⁴, mainly investigating candidate gene polymorphisms rather than performing genome-wide analyses.

In this study, we conducted a genome-wide survival analysis to identify variants affecting the risk of death from acute SARS-CoV-2 infection. We used genotyping and clinical follow-up data (at 60 days post-infection) from about four thousand COVID-19 patients from five European cohorts. We adjusted the analyses for known non-genetic independent prognostic factors, like age, sex, and pandemic wave, which were available for all patients.

Materials and methods

Case series

The case series investigated in this study comprised 3904 COVID-19 patients molecularly tested for SARS-CoV-2 infection and enrolled for host genetics studies at several recruiting centres, in the context of the international COVID-19 HGI. Patients from GEN-COVID Multicenter Study and from the series included in the European Genome-Phenome Archive (EGA) study number EGAS00001005304 (except for BRACOVID and INMUNGEN-CoV series), with 60-days follow-up information and full data about sex, age, and infection date were analysed. BRACOVID patients were not included in the analysis since their data were not shared with us. The INMUNGEN-CoV series was not included in our study, since 90% of patients did not have survival data and, in addition, they were genotyped with a different SNP-array. Patients provided written informed consent to the use of their biological samples and data for research purposes. Personal data treatment was GDPR compliant. The research was approved by the Committees for Ethics of the recruiting centres.

Survival analysis

A multivariable Cox proportional hazard model with demographic-clinical features (i.e., age, sex, series, and time of infection) was used for survival analysis¹⁵ (age was considered as both a linear and non-linear term, the latter defined as age squared and hereafter named age²). The R *survival* package¹⁶ was used to draw Kaplan–Meier (KM) curves and run the log-rank test in R (v. 3.6.0) environment. The hazard proportionality assumptions were verified through the function “*cox.zph()*” of the *survival* package. The variables that were found to impact on survival from the log-rank test (with *P*-value < 0.05) were analysed both in multivariable Cox regression and in a weighted multivariable analysis to account for non-proportional hazards¹⁷, using the *survival* R package and the *coxphw* R package¹⁸ (applying the Average Hazards Ratio method, by setting the parameter template = “AHR”), respectively. Cox and log-rank test *P* values < 0.05 (two-sided) indicated sufficient statistical significance.

Genotyping data quality check, principal component analysis, and imputation

Genome-wide genotyping data were available at EGA (study number EGAS00001005304) and University of Siena. The LiftOver tool (<https://liftover.broadinstitute.org/>) was used to convert genomic coordinates and bring all the datasets to the same genomic build (GRCh38). PLINK v.2 software¹⁹ was used to carry out genotype quality control (QC) steps (Supplementary Figure S1). In detail, for each patient series, per-sample and per-variant QC steps were performed, excluding samples with call rate < 99% and excess of heterozygosity ($F > \pm 0.2$), removing insertions/deletions, duplicated and non-informative variants, and filtering out single nucleotide polymorphisms (SNPs) with genotyping call rate < 99% and Hardy–Weinberg equilibrium test *P*-value < 1.0×10^{-10} . Then, all the datasets were merged, and an additional round of QC was carried out to remove duplicated and related individuals and patients for whom no survival data were available. Additionally, SNPs with minor allele frequency (MAF) < 1% were filtered out, together with SNPs mapping in regions of extended linkage disequilibrium (LD)²⁰. PLINK v.2 software was also used to carry out principal components analysis (PCA): we plotted PC1 versus PC2 visualizing samples according to our five patient series (i.e., BelCovid, GENCOVID, Hostage, SPGRX, and SweCovid; Supplementary Figure S2A). Additionally, to better visualize the ancestry of our patients, we projected the first four principal components of our patients together with those of 2504 individuals from five different populations, selected from 1000 Genomes Project²¹: Africans, Americans, South-East Asians, East Asians, and Europeans (Supplementary Figure S2B and S2C). We defined as Europeans those patients that clustered together with 1000 Genomes Project European individuals. Genotype imputation to whole-genome sequence was carried out on the TopMed imputation server²² using Eagle v.2.4 for phasing²³, minimac4 algorithm²⁴, and TopMed r2

as reference panel²⁵. Finally, SNPs with a low-quality imputation ($R^2 \leq 0.3$)²⁶ and with a MAF < 0.02 were filtered out. This second MAF filter was applied after imputation to remove variants with very low frequency alleles, thus reducing the risk of the Cox model not-converging and of obtaining spurious association results when, in patients carrying the minor allele, few or no events (i.e., death) were observed.

Genome-wide survival analysis

The associations between SNPs (additive model) and patient overall survival were assessed using multivariable Cox proportional hazard model, with the first 10 PCs, sex, age, age², patient series, and time of infection (before or after the first pandemic wave, that we considered finished by the end of June 2020), as covariates, using the GenABEL package in R environment²⁷. Correction for multiple testing was performed with the Benjamini–Hochberg method of false discovery rate (FDR)²⁸. Top significant polymorphisms were also tested under a dominant model using the same covariates as above. This was done for the following two reasons: first, we hypothesized that the minor allele was increasing the risk of death by acting as a dominant allele or at least codominant; second, in this way we reduced the risk of artifacts in the Kaplan–Meier curves by comparing the probability of survival of patients homozygous for the major allele with that of patients with at least one minor allele in their genotype (due to the low number of patients homozygous for the minor allele). COVID-19 severity top-significantly associated variants (<https://app.covid19hg.org/variants>; analysis B2: Hospitalized COVID19 + vs. population controls)⁹ were also investigated in our study, by comparing *P*-values of association between these variants and survival with those reported by COVID-19 HGI.

A logistic regression between a binary status phenotype (live vs. dead during the 60-days follow-up) and genome-wide imputed SNPs was performed using PLINK v.2, with the first 10 PCs, sex, age, age², time of infection, and patient series as covariates.

Functional analyses

To investigate the functional role of the identified variants associated with survival to COVID-19, we used multiple databases to obtain more reliable and robust results, by pooling together partial results from each platform.

First, we looked for the variants associated with COVID-19 survival at P -value $< 1.0 \times 10^{-5}$, in the GTEx (Analysis V8 release, [GTEx_Analysis_v8_eQTL_EUR.tar](#)) and eQTLGen²⁹ (<https://www.eqtlgen.org/cis-eqtl.html>) databases (accessed on 08/05/2023), to test whether they have already been reported as *cis* expression quantitative trait loci (eQTLs). For some variants, we also looked for eQTL SNPs in LD with them, using the LDexpress tool of LDlink³⁰. We searched all tissue eQTLs reported in the European population and in LD with query variants at $D' > 0.8$. An over-representation analysis of the gene list including all the target genes of the found eQTLs was done, using the WEB-based Gene Set Analysis Toolkit³¹, to search for enriched Reactome and KEGG pathways.

Additionally, genes that included or are close to the variants associated with COVID-19 survival (P -value $< 1.0 \times 10^{-5}$) were retrieved. The ENSEMBL gene database was used with the R Bioconductor package “biomaRt”³² to find genes that lie at a maximum distance of ± 50 kb from detected SNP variants. The resulting list of genes was then used as input for functional analysis tools. We used the Database for Annotation, Visualization and Integrated Discovery (DAVID)³³ with default functional categories and applying the high classification stringency parameters; Benjamini–Hochberg adjusted *P*-values (FDR) < 0.05 was considered as significance threshold of enrichment.

The GENE2FUNC functionality of the Functional Mapping and Annotation (FUMA) for GWAS platform³⁴ was selected to retrieve gene ontologies (biological processes, molecular functions, cell compartments), differentially expressed genes associated with COVID-19, target tissues and metabolic pathways, with the following parameters: all background genes; Ensembl version v102; GTEx v8 (54 tissue types) and GTEx v8 (30 general tissue types) gene expression datasets; Benjamini–Hochberg (FDR) correction when testing for gene-set enrichment (threshold: FDR < 0.05); minimum overlapping genes with gene set: ≥ 2 .

Ethical approval and informed consent

The research was performed in accordance with the Declaration of Helsinki and was approved by the committees for Ethics of the recruiting centres, namely, the University Hospital (Azienda ospedaliero-universitaria Senese) ethical review board, Siena, Italy (Prot n. 16917, dated March 16th, 2020); Euskadi Ethics Committee, Donostia-San Sebastian, Spain, on April 6, 2020 (approval number PI2020064); Vall d’Hebron Ethical Committee, Barcelona, Spain; ethics committee of the Junta de Andalucía, Spain (ethics id: 0886-N-20 and 1954-N-20); the ethics committee of Humanitas Clinical and Research Center, Rozzano (MI), Italy (reference number, 316/20); Valladolid Ethics Committee (PI-201716) and the Granada Ethics Committee, Spain, on March 24, 2020, and April 13, 2020, respectively; Erasme Ethics committee, Bruxelles, Belgium (protocol P2020_209); the National Ethical Review Agency, Sweden (EPM; 2020-01623). Patients provided written informed consent to the use of their biological samples and data for research purposes. Personal data treatment was GDPR compliant.

Results

Age, sex, and period of infection are associated with COVID-19 patient survival

In this study, we included a total of 3904 European COVID-19 patients recruited in Italy (GEN-COVID and, in part, Hostage series), Spain (Hostage and SPGRX), Sweden (SWECOVID) and Belgium (BelCovid) (Table 1). The median age at infection was 63 years and male patients were slightly prevalent (58%). More than two thirds of patients included in this study were infected during the first COVID-19 wave (before June 30th, 2020). Most of enrolled patients were hospitalised (86%), but 78% of the whole series did not need to be admitted to the intensive care unit (ICU). Among the 3175 patients for whom we have information about the respiratory support, just a

small fraction of patients included in the study (15%) did not need any oxygen support, whereas most of them (40%) received oxygen by mask or cannula, 11% received non-invasive ventilation, and 14% needed intubation. Information about comorbidities was available for 2887 patients and about a quarter of patients had a comorbidity (hypertension, diabetes, cancer, asthma, and heart failure, in order from the most to the least frequent) and 12% had at least two comorbidities. Considering a follow-up time of 60 days after the infection, approximately 11% of COVID-19 patients of this study have died.

To investigate the factors affecting mortality after SARS-CoV-2 infection, we carried out a survival analysis in a period of 60 days post-infection. Since the hazard proportionality assumption was not verified in our series (global Schoenfeld residuals test P -value = 6.5×10^{-6}) we carried out both a weighted multivariable Cox analysis, to deal with non-proportional hazards, and a multivariable proportional hazard Cox regression. In both models we used the following variables, for which we had full data availability: age, age², sex, and date of infection (Table 2A). The results obtained were similar. Indeed, we observed that the major mortality risk factor for COVID-19 patients was the age at infection (as expected, in both models it was a risk factor for poor prognosis, with HR > 1, with increasing age). We also explored a non-linear effect of age (i.e., age²) on survival and, although statistically significant, it was quite irrelevant (HR = 1 in both models). Female sex was associated with a slightly higher probability of survival (HR = 0.73 in both models), compared to males. Additionally, we observed that patients infected in the first pandemic wave (i.e., in the first half of 2020) had lower probabilities of survival than patients infected later (HR = 1.6 and 1.5 in Cox and weighted Cox models, respectively).

We also tested Cox and weighted Cox models with an additional covariate, i.e., the series in which patients were recruited. Indeed, we were aware that some differences in the recruitment might have confounding effects on patient survival. For instance, we knew that SweCovid patients were all admitted in intensive care unit, they probably were affected by a severe form of COVID-19 and, therefore, their probability of survival was lower than other patient series. As expected, SweCOVID patients had the highest risk of death among all patients included in the present study (Table 2B). These results prompted us to include the explored variables as covariates in the genome-wide survival analysis, in order to identify genetic variants that were independent prognostic factors.

We also drew KM curves testing the effects of age, sex, pandemic wave, and patient series, on the risk of death 60 days after infection. Regarding the age, it was discretized in five age groups, starting from patients < 55 years old, then by decades till the age of 84, and the last group comprising patients ≥ 85 years old. Highly significant associations were observed for age, patient series, and pandemic wave (Supplementary Figure S3), whereas the association with sex was weaker, but still significant (log-rank test, P -value = 0.03). As already observed with the Cox models, the probability of survival decreased with increasing age and was lower for males than females, for patients infected at the beginning of 2020, and as expected, for SweCOVID patients.

Germline variants are associated with COVID-19 overall survival 60 days after infection

Genotype data of patients from GEN-COVID Multicenter Study and from the series included in the European Genome-Phenome Archive (EGA) study number EGAS00001005304 (except for BRACOVIC and INMUNGEN-CoV, that were unavailable and with missing data, respectively, as explained above) were used for the genome-wide survival analysis. After quality controls and imputation, the dataset comprised 7,151,809 variants and 3904 patients, 91% of which were of European ancestry (Supplementary Figure S2B and C).

We run a GWAS Cox model, with the first 10 PCs, age, age², sex, period of infection and patient series as covariates and we found one variant (rs117011822) associated with survival at P -value < 5.0×10^{-8} (genome-wide significance level) and another one (rs7208524) nearly significant (P -value = 5.19×10^{-8}). The results for all the variants tested in the multivariable Cox model are reported in a Manhattan plot (Fig. 1). The minor alleles of both these variants were risk factors for poor prognosis, showing hazard ratios > 1. Both are low frequency variants, in our dataset (MAF < 5%). The top one is a 2 kb upstream variant of *FGF19* gene on chromosome 11, and the other one maps in an intron of the *GPRC5C* gene on chromosome 17.

Looking at all the variants with a P -value < 1.0×10^{-5} ($n = 113$; Supplementary Table S1), we found that 7 variants mapped in the *GPRC5C* gene locus, 16 polymorphisms were on chromosome 8, in the *PSD3* gene locus, 9 SNPs on chromosome 9, in an intergenic region of 359 kb between *PBX3* and *MVB12B* genes, and 9 variants in the *CDH13* locus, on chromosome 16. Additionally, other 27 mapped on chromosome 6, 12 of which were in an intergenic region near the *EPHA7* gene, in proximity of an enhancer region (ENSR00000798782), and 13 mapped in the locus of the *PERP* gene, mostly in its 5' regulatory region. A zoomed plot for each of these loci is reported in Fig. 2.

Beside the additive model, we tested the two top-significant variants in a dominant model (using the same covariates), where the survival of patients heterozygous and homozygous for the alternative allele was compared with that of patients homozygous for the common allele. Also this analysis indicated that having at least one alternative allele of these variants conferred a higher risk of poor COVID-19 prognosis than having a wild-type genotype (rs117011822: HR = 2.47, P -value = 3.43×10^{-8} ; rs7208524: HR = 2.70, P -value = 1.10×10^{-7}). We plotted the KM curves to visualize the probability of survival according to the genotypes of both variants (Fig. 3).

We also looked for the top-significant variants previously identified by the meta-analysis, reported by the COVID-19 HGI, as associated with COVID-19 severity⁹, in our results. Of note, the phenotypes under investigation in that study were quite different, since the COVID-19 HGI defined the severity phenotype as a binary variable representing COVID-19 patients' risk of hospitalisation. As shown in Supplementary Table S2, none of the COVID-19 HGI top variants were significant, neither at nominal P -value < 0.01, in our survival model.

Finally, we tested the risk of mortality in the 60 days after infection, using a logistic regression model, with the same covariates used in the Cox analysis. In this GWAS, no SNPs reached the genome-wide significance threshold. However, 30% of the 113 top significant SNPs associated with survival probability were confirmed

also in this GWAS, as associated with mortality risk, although at a higher P -value ($< 1.0 \times 10^{-5}$). These results are reported in Supplementary Table S1, for comparison with Cox analysis results.

Variants associated with COVID-19 patient survival are involved in immune or lung functions

We queried two different eQTL databases, namely GTEx and eQTLGen, to test whether our most significant variants (P -value $< 1.0 \times 10^{-5}$) were previously reported as eQTLs. In GTEx, 43 out of 113 SNPs were eQTLs of 33 target genes, in several different tissues (e.g., brain, lung, muscle, heart, spleen, whole blood) and in eQTLGen 54 out of 113 SNPs acted as *cis* eQTLs for 30 genes, in the whole blood (Supplementary Table S3 and S4). Eight target genes (i.e., *TSPYL1*, *EPHA7*, *MOSPD3*, *UFSP1*, *GIGYF1*, *SLC12A9*, *MVB12B*, and *KLRC1*) were found in both databases. The list of the 55 unique eQTL target genes was enriched for the Reactome pathway R-HSA-198933: Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell (enrichment ratio = 16.7; FDR = 0.017), including the CD300 genes (A, C, and LB) and two other genes (*KLRC1* and *KLRF1*).

In addition, we explored the possibility that the quite numerous SNPs identified on chromosome 6 and 8, next to *EPHA7* and *PSD3* genes, respectively, might be in LD with some near variants reported as eQTLs of these two genes. Indeed, none of these SNPs were reported as eQTLs in the two searched databases. Looking at data available in the European population (using LDlink), we found several variants in strong LD ($D' > 0.8$) with our SNPs and acting as *EPHA7* and *PSD3* eQTLs, in many tissues (Supplementary Table S5).

Considering the list of 38 genes (Supplementary Table S6) mapping within 50 kbps of the 113 top significant variants, functional annotation analysis with DAVID identified 19 significant terms (FDR < 0.05 , Supplementary Table S7). Among these, the top significant ones are two Gene Ontology (GO) biological processes, namely, “positive regulation of natural killer (NK) cell mediated cytotoxicity” and “stimulatory C-type lectin receptor signalling pathway”, both involving the same five genes (*KLRK1*, *KLRC3*, *KLRC4*, *KLRC4-KLRK1*, and *KLRD1*). Three of these genes (*KLRC3*, *KLRC4*, and *KLRD1*) were also annotated in the Biocarta “h_nkcellsPathway: Ras-Independent pathway in NK cell-mediated cytotoxicity”, the only pathway that reached the statistical significance threshold.

Partially overlapping results were obtained in the overrepresentation analysis carried out with the FUMA platform (Supplementary Table S8), that identified 36 significantly enriched functional gene sets. Among them, we observed GO biological processes, KEGG and Biocarta pathways related to NK cell regulation and other immune functions, in which approximately the same genes as above are involved. In addition, looking at tissue specificity analysis by FUMA, we observed that the 38-gene list was enriched of genes over-expressed in the lung and in the brain's putamen basal ganglia (Supplementary Figure S4).

Discussion

In this study we investigated the effects of host germline variants on the overall survival of COVID-19 patients, 60 days after the infection. With a case-only approach and using a multivariable Cox model to look for variants associated with the probability of survival after infection, we aimed to dissect the genetics bases to develop a severe COVID-19 from a different point of view, as compared to several other genetic studies on COVID-19 severity (reviewed in¹⁰). The analysis considered, as covariate of the multivariable model, the most widely known prognostic factors for COVID-19 survival, i.e., patient age at infection, sex, and period of infection (at the very beginning of the pandemic or after).

We identified a genome-wide level significant association between survival and the SNP rs117011822, on chromosome 11. We observed that individuals with an increasing number of minor alleles of this variant in their genotype (both under additive and dominant model) had a worse prognosis than patients homozygous for the major allele, 60 days after SARS-CoV-2 infection. This variant maps in a regulatory region (ENSR00000958007), specifically a CTCF binding site, upstream the *FGF19* gene. However, we did not find evidence for a regulation of *FGF19* expression by this variant in GTEx or eQTLGen, but it might be interesting to investigate this aspect further. Indeed, it was reported that serum levels of the FGF19 protein were lower in asymptomatic than symptomatic COVID-19 patients³⁵. In that study, the authors discussed a possible role of FGF19 (together with other proteins) in lung tissue repair, with differences between asymptomatic and symptomatic patients. Therefore, it might be interesting to further investigate the role of the minor allele of rs117011822 in the regulation on *FGF19* levels with the aim to understand the functional mechanism underlying the statistical association observed in our study.

Additional loci on chromosomes 17, 8, 6, 9, and 16 were suggestively associated with COVID-19 patient overall survival. On chromosome 17 we identified a locus of seven variants that were reported to act as regulators of the expression of CD300 genes (*CD300A*, *CD300C*, and *CD300LB*). Of note, CD300 is a family of leukocyte surface proteins involved in immune response signalling pathways and it has been observed that shifts in the expression pattern of CD300 molecules in T-cells of COVID-19 patients correlated with COVID-19 severity³⁶.

The 16 variants on chromosome 8 mapped in intronic regions of the *PSD3* gene, also known as *EFA6R*, a member of the family of guanine nucleotide exchange factors, that activate ADP-ribosylation factor 6 (ARF6)³⁷. This protein is involved in endocytosis, and it has been reported to play a role in SARS-CoV-2 cell entry^{38,39}. So far, it has not yet been reported any role of *PSD3* variants in the regulation of ARF6 activation. Our variants showed a strong LD with eQTL SNPs of *PSD3*.

The six variants we found associated with COVID-19 patient survival, mapping in an intergenic region of chromosome 9, were reported to be eQTLs of the near the *MVB12B* gene. No evidence for a role of this gene in COVID-19 is available, but it is interesting to underline that it codes for a subunit of the ESCRT-I complex that mediates HIV virus budding⁴⁰.

The 12 intergenic variants on chromosome 6 (at position 93 Mb) were near the *EPHA7* gene, which was suggested to be a downstream mediator of cytokine production, induced by the N-terminal domain of the

Characteristic	BelCovid (n = 381)	GENCOVID (n = 1727)	Hostage (n = 1303)	SPGRX (n = 364)	SweCovid (n = 129)	Total (n = 3904)
Age at diagnosis, years, median (range)	59 (23–104)	60 (18–99)	65 (1–96)	71 (23–103)	61 (24–86)	63 (1–104)
Age at diagnosis, years, n (%)						
< 55	151 (39.6)	615 (35.6)	318 (24.4)	66 (18.1)	42 (32.6)	1192 (30.5)
55–64	89 (23.4)	416 (24.1)	321 (24.6)	59 (16.2)	31 (24.0)	916 (23.5)
65–74	83 (21.8)	345 (20.0)	363 (27.9)	88 (24.2)	30 (23.3)	909 (23.3)
75–84	43 (11.3)	227 (13.1)	202 (15.5)	88 (24.2)	24 (18.6)	584 (15.0)
> 85	15 (3.94)	124 (7.18)	99 (7.60)	63 (17.3)	2 (1.55)	303 (7.76)
Sex, n (%)						
Male	199 (52.2)	1032 (59.8)	773 (59.3)	188 (51.6)	91 (70.5)	2283 (58.5)
Female	182 (47.8)	695 (40.2)	530 (40.7)	176 (48.4)	38 (29.5)	1621 (41.5)
Infection before June 30th, 2020, n (%)						
No	41 (10.8)	1075 (62.2)	118 (9.06)	0	17 (13.2)	1251 (32.0)
Yes	340 (89.2)	652 (37.8)	1185 (90.9)	364 (100)	112 (86.8)	2653 (68.0)
Hospitalization, n (%)						
No	125 (32.8)	357 (20.7)	9 (0.691)	52 (14.3)	0	543 (13.9)
Yes	256 (67.2)	1370 (79.3)	1294 (99.3)	312 (85.7)	129 (100)	3361 (86.1)
ICU admission, n (%)						
No	238 (62.5)	1566 (90.7)	961 (73.8)	263 (72.2)	0	3028 (77.6)
Yes	143 (37.5)	161 (9.32)	342 (26.2)	101 (27.7)	129 (100)	876 (22.4)
Highest respiratory support, n (%)						
None	0	605 (35.0)	0	0	0	605 (15.5)
Oxygen (mask or cannula)	124 (32.5)	592 (34.3)	854 (65.5)	0	0	1570 (40.2)
Non-invasive ventilation *	6 (1.57)	369 (21.4)	63 (4.83)	0	0	438 (11.2)
Intubation	103 (27.0)	161 (9.32)	298 (22.9)	0	0	562 (14.4)
Not available	148 (38.8)	0	88 (6.75)	364 (100)	129 (100)	729 (18.7)
Number of comorbidities, n (%)						
None	64 (16.8)	865 (50.1)	323 (24.8)	0	36 (27.9)	1322 (33.9)
1	97 (25.5)	522 (30.2)	305 (23.4)	0	48 (37.2)	964 (24.7)
2	72 (18.9)	205 (11.9)	174 (13.4)	0	36 (27.9)	469 (12.0)
3 or more	22 (5.77)	55 (3.18)	54 (4.14)	0	9 (6.98)	132 (3.38)
Not available	126 (33.1)	80 (4.63)	447 (34.3)	364 (100)	0	1013 (26.1)
Comorbidity, n (%)						
Asthma	34 (8.92)	148 (8.57)	91 (6.98)		22 (17.1)	295 (7.56)
Cancer	51 (13.4)	126 (7.30)	154 (11.8)		15 (11.6)	346 (8.86)
Diabetes	89 (23.4)	207 (12.0)	266 (20.4)		35 (27.1)	597 (15.3)
Heart failure	39 (10.2)	160 (9.26)	87 (6.68)		6 (4.65)	292 (7.48)
Hypertension	156 (40.9)	484 (28.0)	554 (42.5)		70 (54.3)	1264 (32.4)
Alive status (at 60 days of follow-up), n (%)						
Alive	335 (87.9)	1622 (93.9)	1147 (88.0)	286 (78.6)	95 (73.6)	3485 (89.3)
Deceased	46 (12.1)	105 (6.08)	156 (12.0)	78 (21.4)	34 (26.4)	419 (10.7)
Series, n (%)						
BelCovid						381 (9.76)
GEN-COVID						1727 (44.2)
Hostage						1303 (33.4)
SPRGX						364 (9.32)
SweCovid						129 (3.33)

Table 1. Clinical characteristics of patients included in the survival genome-wide association study. *CPAP, BiPAP, or high-flow cannula.

SARS-CoV-2 spike protein⁴¹. These variants mapped in an enhancer region and might affect the expression of the *EPHA7* gene, although we did not find them among the eQTLs of this gene. However, these variants are in strong LD with eQTL SNPs of *EPHA7*. The other 11 variants on chromosome 6, instead, mapped in the *PERP* gene locus and some of them were already annotated as eQTLs of this gene. Although no functions related to

	Multivariable Cox's test		Multivariable weighted Cox's test	
	HR (95% CI)	P-value	HR (95% CI)	P-value
A				
Age at diagnosis (linear)	1.20 (1.10–1.30)	2.59×10^{-5}	1.20 (1.09–1.32)	1.54×10^{-4}
Age at diagnosis (quadratic)	1.00 (1.00–1.00)	0.0135	1.00 (1.00–1.00)	0.0235
Sex				
Male	1		1	
Female	0.730 (0.596–0.894)	2.38×10^{-3}	0.733 (0.597–0.902)	3.23×10^{-3}
Infection before June 30th, 2020				
No	1		1	
Yes	1.55 (1.21–1.99)	6.13×10^{-4}	1.53 (1.19–1.96)	9.35×10^{-4}
B				
Age at diagnosis (linear)	1.20 (1.10–1.30)	2.68×10^{-5}	1.20 (1.10–1.31)	1.32×10^{-4}
Age at diagnosis (quadratic)	1.00 (1.00–1.00)	0.0159	1.00 (1.00–1.00)	0.025
Sex				
Male	1		1	
Female	0.730 (0.595–0.895)	2.43×10^{-3}	0.733 (0.595–0.903)	3.56×10^{-3}
Series				
BelCovid	1		1	
GEN-COVID	0.736 (0.508–1.07)	0.104	0.739 (0.503–1.09)	0.124
Hostage	0.760 (0.547–1.06)	0.103	0.745 (0.529–1.05)	0.0931
SPRGX	1.05 (0.727–1.52)	0.795	1.03 (0.702–1.51)	0.884
SweCovid	2.57 (1.64–4.01)	3.52×10^{-5}	2.54 (1.62–3.99)	4.79×10^{-5}
Infection before June 30th, 2020				
No	1		1	
Yes	1.36 (1.01–1.83)	0.0405	1.36 (1.01–1.82)	0.0424

Table 2. Clinical and personal prognostic factors for COVID-19 patients, as resulted from Cox's and weighted Cox's multivariable tests. HR, hazard ratio; CI, confidential interval.

COVID-19 have been reported for the *PERP* gene, so far, it encodes a protein that is a p53 apoptosis effector, and, recently, Wang and colleagues reviewed the possible roles of p53 in mediating host-virus interactions in infections caused by Coronaviruses⁴².

Finally, the nine SNPs on chromosome 16, intronic to the *CDH13* gene, were eQTLs of this gene, which encodes the T-cadherin protein, a regulator of vascular permeability, but also receptor of adiponectin and LDL. It is involved in lung function, and reportedly associated with several metabolic disorders as atherosclerosis, dyslipidemia, obesity, and also diabetes (as reviewed in⁴³).

Interestingly, the list of genes, whose expression was reported to be regulated by our top-significant variants, was enriched for genes with immunoregulatory functions. These included the already mentioned CD300 genes, but also *KLRC1* and *KLRF1* genes. *KLRC1* (alias *NKG2A*) expression, in SARS-CoV-2 infected patients was suggested to correlate with functional exhaustion of cytotoxic lymphocytes and with a severe COVID-19 outcome⁴⁴. In addition, functional annotation of the genes near or where top-significant variants mapped resulted in an enrichment of immune-related terms and pathways, in particular those involved in regulation of NK cell-mediated cytotoxicity. This finding is interesting in the light of previous results by Maucorant et al.⁴⁵ showing that distinct NK-immunotypes were related to COVID-19 severity. Among the genes in this pathway, there is *KLRK1*, also known as *NKG2D*, coding for a NK cell activating receptor. It has been previously reported that SARS-CoV-2 non-structural protein 1 can downregulate ligands of the NKG2D receptor, thus escaping NK cells cytotoxicity⁴⁶.

Regarding the findings from FUMA tissue specificity analysis, the observed enrichment of our gene list in genes whose expression is altered in lung is not unexpected: as the major clinical manifestation of SARS-CoV-2 infection and severity is at respiratory level, we expected that variants associated with COVID-19 survival were in genes expressed in the lungs. On the other hand, it is more difficult to speculate on the finding of an enrichment of genes upregulated in brain putamen basal ganglia. In a recent paper⁴⁷, Balsak et al. reported that basal ganglia can be damaged after COVID-19, due to microstructural alterations caused by hypoxia. However, further studies are needed to understand if the variants/genes identified in our study might be involved in the hypoxia induced brain alterations after SARS-CoV-2 infection and, also, if this kind of damage might affect COVID-19 patient survival.

Despite the above cited evidence found in the literature, without appropriate functional studies, we cannot assert that the variants we identified as associated with survival play a role in predisposing patients to a worse COVID-19 outcome. However, we believe that our findings are worthy of further investigation and are of interest, since we explored the genetics of COVID-19 severity with an unconventional approach that might have led to new results. Indeed, not surprisingly, none of the previously reported variants associated with severity in⁹ were

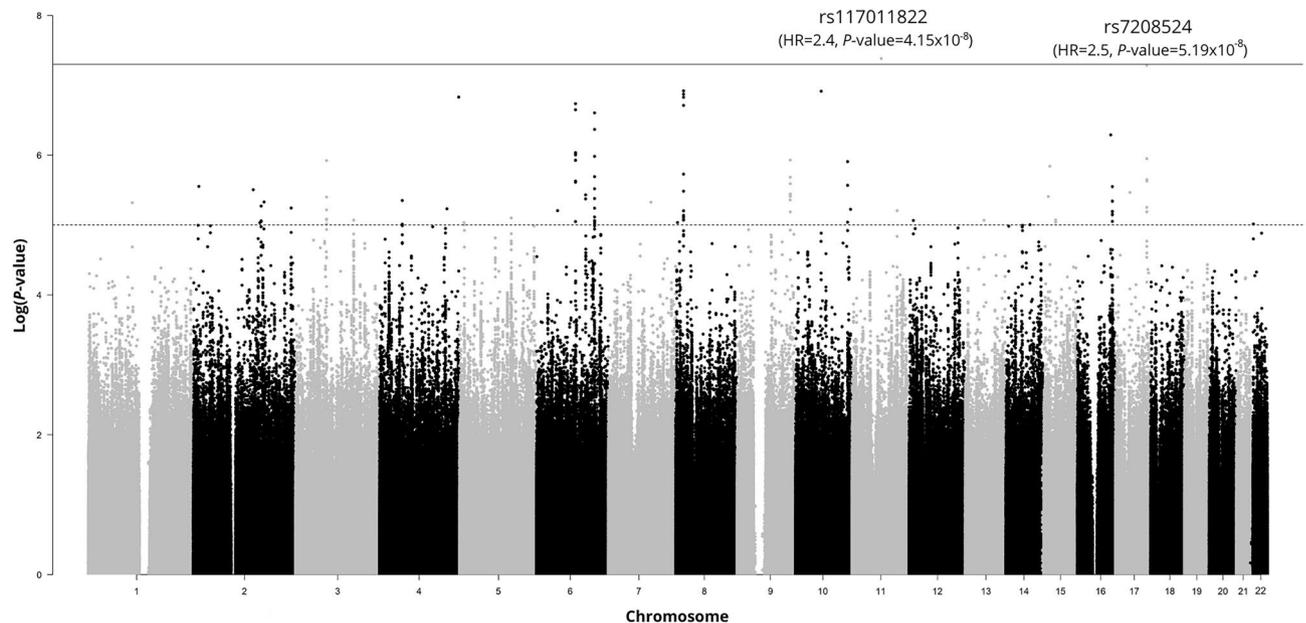


Figure 1. Manhattan plot of the results of the GWAS for survival of COVID-19 patients. SNPs are plotted on the x-axis according to their genomic position (GChr 38, hg38 release), and P -values ($-\log_{10}(P)$) for their association with survival probability are plotted on the y-axis. The horizontal solid line represents the threshold of significance (P -value $< 5.0 \times 10^{-8}$), whereas the dashed one represents a suggestive threshold (P -value $< 1.0 \times 10^{-5}$). Names, hazard ratios and P -values of the two most significant SNPs are shown on top.

significant in our study. Indeed, both the phenotype (severity vs. overall survival 60 days after the infection) and the analysis model (regression vs. Cox) were completely different.

We are aware of some limitations in our study. First, validation in an independent, wider, but possibly genetically homogeneous patient series should be needed. Indeed, we were able to detect only one variant associated with survival at a genome-wide significance level. In addition, our series, is mainly composed of patients of European ancestry and our results, although controlled for population stratification, would not be directly generalizable to patients from different ethnicities. Additionally, since we did not adjust for patients' comorbidities (as no full data were available), we cannot exclude that some of the identified associations might result from such potential confounding factors. For instance, the variants in the *CDH13* gene, involved in metabolic disorders, might be this case. Although we were aware of this limitation, we preferred not to further reduce the already relatively small sample size of our analysis, by excluding patients with unavailable comorbidity data, to avoid losing statistical power.

Overall, our results shed new light on the genetics of COVID-19 severity, having identified some loci associated with patient survival, at 60 days after infection. Although our findings suggest that genetics plays a limited role in affecting mortality probability after SARS-CoV-2 infection, the identified variants are worthy to be further investigated as possible prognostic factors for COVID-19.

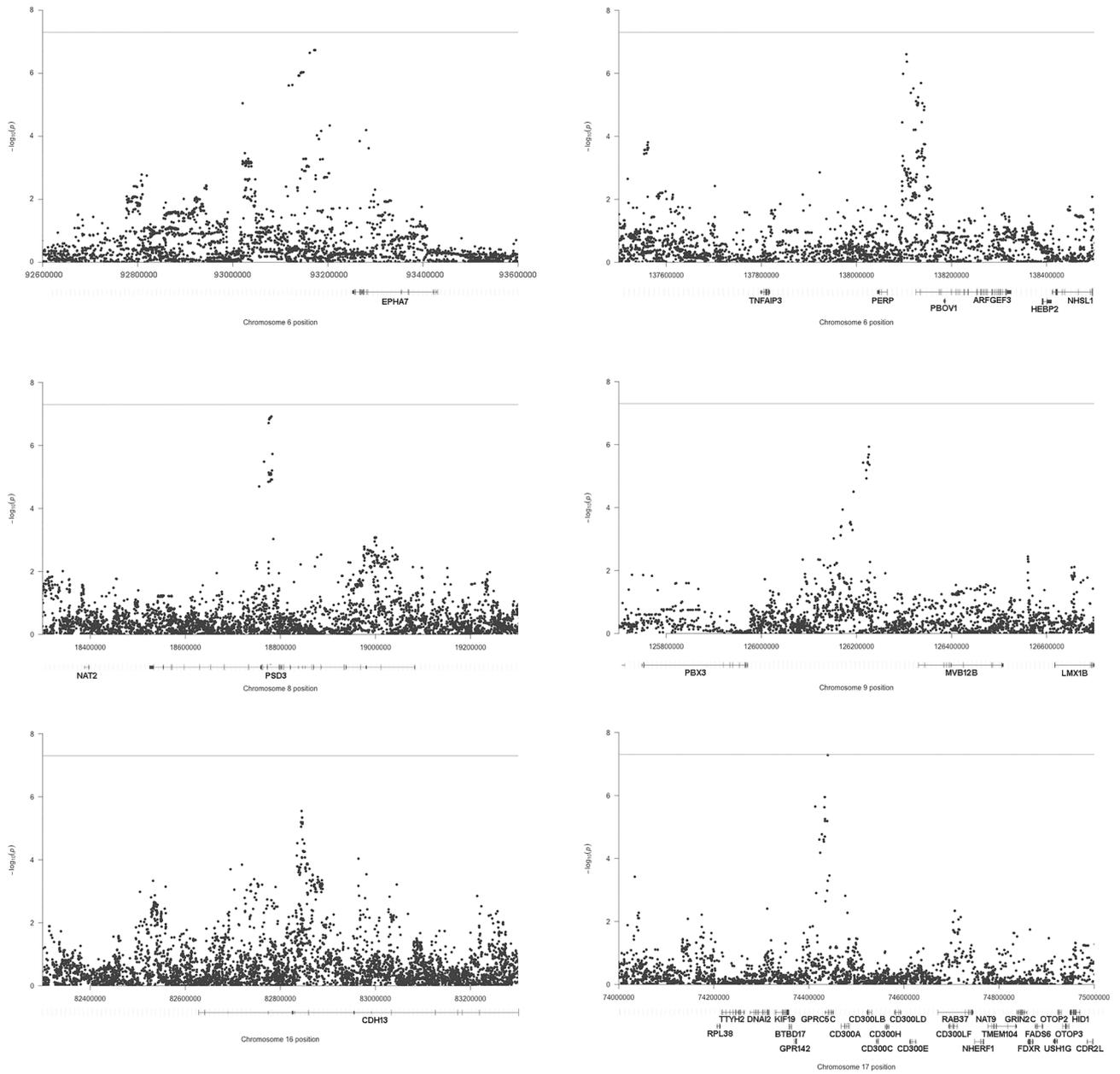


Figure 2. Zoomed plots of six loci associated with patient survival. SNPs are plotted on the x-axis according to their chromosome position (GChr 38, hg38 release), and P -values ($-\log_{10}(P)$) for their association with survival probability are plotted on the y-axis. The horizontal line represents the threshold of significance (P -value $< 5.0 \times 10^{-8}$). Below the x axis, the mapped genes are plotted (according to University of California Santa Cruz Genome Browser notation).

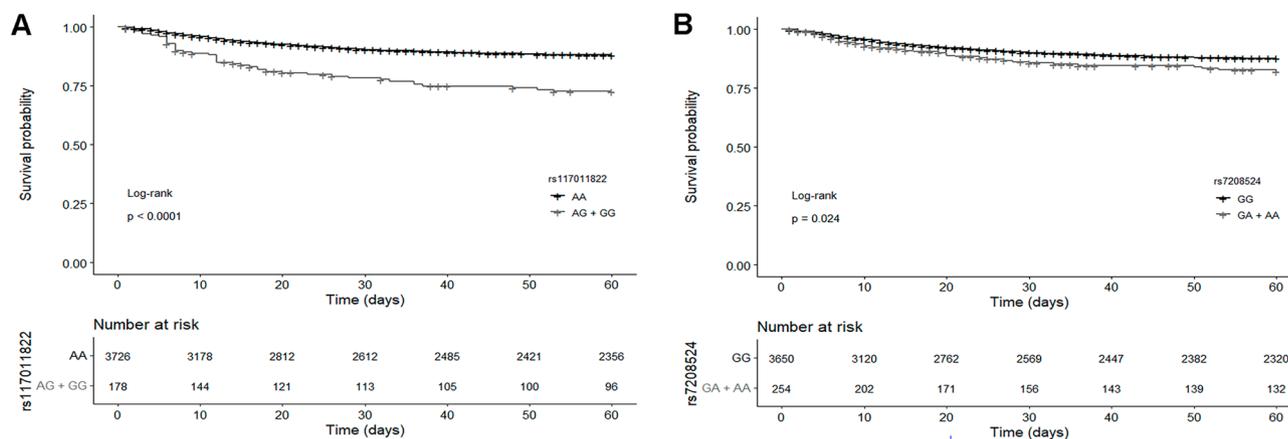


Figure 3. Kaplan–Meier survival curves for COVID-19 patients according to the genotype of the top significant variants (A) rs117011822 and (B) rs7208524. Black line represents patients homozygous for the major allele and grey line represents patients with at least one copy of the minor allele (according to the dominant model, patients with heterozygous genotype and homozygous for the minor allele were grouped together). Crosses denote censored samples. Numbers of patients at risk are shown below the plot. Log–rank test *P*-value is shown.

Data availability

Genotype data that support the findings of this study are available at The European Genome-phenome Archive (EGA) (study number: EGAS00001005304) and upon request to Prof. Renieri at University of Siena.

Code availability

No custom software or algorithm were used. Packages and tools used for the analyses described in the manuscript are all publicly available. Anyway, used scripts are available from the corresponding author upon request.

Received: 20 November 2023; Accepted: 30 January 2024

Published online: 06 February 2024

References

- Long, Q. X. *et al.* Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat. Med.* **26**, 1200–1204 (2020).
- Guan, W. *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *N. Engl. J. Med.* **382**, 1708–1720 (2020).
- White-Dzuro, G. *et al.* Multisystem effects of COVID-19: A concise review for practitioners. *Postgrad. Med.* **133**, 1 (2020).
- Michelozzi, P. *et al.* Temporal dynamics in total excess mortality and COVID-19 deaths in Italian cities. *BMC Public Health* **20**, 1–8. <https://doi.org/10.1186/S12889-020-09335-8> (2020).
- Rostami, A. *et al.* SARS-CoV-2 seroprevalence worldwide: A systematic review and meta-analysis. *Clin. Microbiol. Infect.* **27**, 331 (2021).
- Elliott, J. *et al.* COVID-19 mortality in the UK Biobank cohort: Revisiting and evaluating risk factors. *Eur. J. Epidemiol.* **36**, 299–309 (2021).
- Minnai, F., De Bellis, G., Dragani, T. A. & Colombo, F. COVID-19 mortality in Italy varies by patient age, sex and pandemic wave. *Sci. Rep.* **12**, 4604. <https://doi.org/10.1038/s41598-022-08573-7> (2022).
- Onoja, A. *et al.* An explainable model of host genetic interactions linked to COVID-19 severity. *Commun. Biol.* **5**, 1133. <https://doi.org/10.1038/s42003-022-04073-6> (2022).
- Pathak, G. A. *et al.* A first update on mapping the human genetic architecture of COVID-19. *Nature* **608**, E1–E10 (2022).
- Cappadona, C., Rimoldi, V., Paraboschi, E. M. & Asselta, R. Genetic susceptibility to severe COVID-19. *Infect. Genet. Evol.* **110**, 105426 (2023).
- Lehrer, S. & Rheinstein, P. H. ABO blood groups, COVID-19 infection and mortality. *Blood Cells Mol. Dis.* **89**, 102571 (2021).
- Fricke-Galindo, I. *et al.* IFNAR2 relevance in the clinical outcome of individuals with severe COVID-19. *Front. Immunol.* **13**, 949413. <https://doi.org/10.3389/fimmu.2022.949413> (2022).
- Hu, J., Li, C., Wang, S., Li, T. & Zhang, H. Genetic variants are identified to increase risk of COVID-19 related mortality from UK Biobank data. *Hum. Genom.* **15**, 10 (2021).
- de Andrade, C. C. *et al.* A polymorphism in the TMPRSS2 gene increases the risk of death in older patients hospitalized with COVID-19. *Viruses* **14**, 2557 (2022).
- Clark, T. G., Bradburn, M. J., Love, S. B. & Altman, D. G. Survival analysis part I: Basic concepts and first analyses. *Br. J. Cancer* **89**, 232–238 (2003).
- Therneau, T. M. & Grambsch, P. M. *Modeling Survival Data: Extending the Cox Model* (Springer-Verlag, 2000).
- Schemper, M. Cox analysis of survival data with non-proportional hazard functions. *The Statistician* **41**, 455 (1992).
- Dunkler, D., Ploner, M., Schemper, M. & Heinze, G. Weighted cox regression using the R package coxphw. *J. Stat. Softw.* **84**, 1–26 (2018).
- Chang, C. C. *et al.* Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
- Price, A. L. *et al.* Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83**, 132–135 (2008).
- Delaneau, O. *et al.* Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat. Commun.* **5**, 3934 (2014).
- Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
- Loh, P. R. *et al.* Reference-based phasing using the haplotype reference consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
- Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minimac2: Faster genotype imputation. *Bioinformatics* **31**, 782–784 (2015).

25. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed program. *Nature* **590**, 290–299 (2021).
26. Verlouw, J. A. M. *et al.* A comparison of genotyping arrays. *Eur. J. Hum. Genet.* **29**, 1611–1624 (2021).
27. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: An R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–1296 (2007).
28. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **57**, 289–300 (1995).
29. Vösa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300–1310 (2021).
30. Machiela, M. J. & Chanock, S. J. LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555–3557 (2015).
31. Wang, J., Vasaiakar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: A more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* **45**, W130–W137 (2017).
32. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
33. Huang, D. W. *et al.* DAVID bioinformatics resources: Expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.* **35**, W169–W175 (2007).
34. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
35. Soares-Schanoski, A. *et al.* Asymptomatic SARS-CoV-2 infection is associated with higher levels of serum IL-17C, matrix metalloproteinase 10 and fibroblast growth factors than mild symptomatic COVID-19. *Front. Immunol.* **13**, 821730. <https://doi.org/10.3389/fimmu.2022.821730> (2022).
36. Zenarruzabeitia, O. *et al.* T cell activation, highly armed cytotoxic cells and a shift in monocytes CD300 receptors expression is characteristic of patients with severe COVID-19. *Front. Immunol.* **12**, 655934. <https://doi.org/10.3389/fimmu.2021.655934> (2021).
37. Kanamarlapudi, V. Exchange factor EFA6R requires C-terminal targeting to the plasma membrane to promote cytoskeletal rearrangement through the activation of ADP-ribosylation factor 6 (ARF6). *J. Biol. Chem.* **289**, 33378–33390 (2014).
38. Zhou, Y.-Q. *et al.* SARS-CoV-2 pseudovirus enters the host cells through spike protein-CD147 in an Arf6-dependent manner. *Emerg. Microbes Infect.* **11**, 1135–1144 (2022).
39. Mirabelli, C. *et al.* ARF6 is a host factor for SARS-CoV-2 infection in vitro. *J. Gener. Virol.* <https://doi.org/10.1099/jgv.0.001868> (2023).
40. Morita, E. *et al.* Identification of human MVB12 proteins as ESCRT-I subunits that function in HIV budding. *Cell Host. Microbe* **2**, 41–53 (2007).
41. Chan, M. *et al.* Machine learning identifies molecular regulators and therapeutics for targeting SARS-CoV2-induced cytokine release. *Mol. Syst. Biol.* **17**, e10426. <https://doi.org/10.15252/msb.202110426> (2021).
42. Wang, X., Liu, Y., Li, K. & Hao, Z. Roles of p53-mediated host-virus interaction in coronavirus infection. *Int. J. Mol. Sci.* **24**, 6371 (2023).
43. Rubina, K. A. *et al.* Revisiting the multiple roles of T-cadherin in health and disease. *Eur. J. Cell Biol.* **100**, 151183 (2021).
44. Zheng, M. *et al.* Functional exhaustion of antiviral lymphocytes in COVID-19 patients. *Cell Mol. Immunol.* **17**, 533–535 (2020).
45. Maucourant, C. *et al.* Natural killer cell immunotypes related to COVID-19 disease severity. *Sci. Immunol.* **5**, eabd6832. <https://doi.org/10.1126/sciimmunol.abd6832> (2020).
46. Lee, M. J. *et al.* SARS-CoV-2 escapes direct NK cell killing through Nsp1-mediated downregulation of ligands for NKG2D. *Cell Rep.* **41**, 111892 (2022).
47. Balsak, S. *et al.* Microstructural alterations in hypoxia-related BRAIN centers after COVID-19 by using DTI: A preliminary study. *J. Clin. Ultrasound* **51**, 1276–1283 (2023).

Acknowledgements

The authors acknowledge Dr. Andrea Ganna who granted access to the EGA datasets. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 08/05/2023. Specimens collected within the GEN-COVID consortium were provided by the COVID-19 Biobank of Siena, which is part of the Genetic Biobank of Siena, member of BBMRI-IT, Telethon Network of Genetic Biobanks (project no. GTB18001), Euro-BioBank, and RD-Connect.

Author contributions

Conceptualization, F.C., F.M., F.B. and T.A.D.; formal analysis, F.M., F.B., M.E., and F.C.; resources, A.R., M.Bruttini, S.C., L.B. S.R., M.E.A.R., D.B., E.C.M., M.Buti, H.Z., R.A., M.R.G, GEN-COVID Multicenter Study, I.F.C.; data curation, C.F., K.Z, M.Baldassarri, S.F, F.M, F.B., and F.C; writing—original draft preparation, F.M., F.B., and F.C.; writing—review and editing F.M., F.B., T.A.D., R.A., and F.C.; funding acquisition, F.C., R.A., and A.R.. All authors have read and agreed to the published version of the manuscript.

Funding

Istituto Buddhista Italiano Soka Gakkai funded A.R.'s and F.C.'s project “PAT-COVID: Host genetics and pathogenic mechanisms of COVID-19” with the 8 × 1000 funds (ID n. 2020-226 2016_RIC_3), that, in part, included the present study. We also thank the generous contribution of Banca Intesa San Paolo to R.A. The funding organisations had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-53310-x>.

Correspondence and requests for materials should be addressed to F.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

GEN-COVID Multicenter Study

Francesca Mari^{18,19,20}, Sergio Daga^{18,19}, Iaria Meloni^{18,19}, Giulia Brunelli^{18,19}, Mirjam Lista^{18,19}, Debora Maffeo^{18,19}, Elena Pasquinelli^{18,19}, Enrica Antolini^{18,19}, Simona Letizia Basso^{18,19}, Samantha Minetto^{18,19}, Giulia Rollo^{18,19}, Angela Rina^{18,19}, Martina Rozza^{18,19}, Rossella Tita²⁰, Maria Antonietta Mencarelli²⁰, Caterina Lo Rizzo²⁰, Anna Maria Pinto²⁰, Francesca Ariani^{18,19,20}, Francesca Montagnani^{19,22}, Mario Tumbarello^{19,22}, Iaria Rancan^{19,22}, Massimiliano Fabbiani²², Paolo Cameli²³, David Bennett²³, Federico Anedda²⁴, Simona Marcantonio²⁴, Sabino Scolletta²⁴, Federico Franchi²⁴, Maria Antonietta Mazzei²⁵, Susanna Guerrini²⁵, Edoardo Conticini²⁶, Luca Cantarini²⁶, Bruno Frediani²⁶, Danilo Tacconi²⁷, Chiara Spertilli Raffaelli²⁷, Arianna Emiliozzi²⁷, Marco Feri²⁸, Alice Donati²⁸, Raffaele Scala²⁹, Luca Guidelli²⁹, Genni Spargi³⁰, Marta Corridi³⁰, Cesira Nencioni³¹, Leonardo Croci³¹, Gian Piero Caldarelli³², Davide Romani³³, Paolo Piacentini³³, Maria Bandini³³, Elena Desanctis³³, Silvia Cappelli³³, Anna Canaccini³⁴, Agnese Verzuri³⁴, Valentina Anemoli³⁴, Agostino Ognibene³⁵, Maria Lorubbio³⁵, Alessandro Pancrazzi³⁵, Massimo Vaghi³⁶, Antonella D.'Arminio Monforte³⁷, Federica Gaia Miraglia³⁷, Mario U. Mondelli^{38,39}, Stefania Mantovani³⁸, Raffaele Bruno^{38,40}, Marco Vecchia³⁸, Marcello Maffezzoni⁴⁰, Enrico Martinelli⁴¹, Massimo Girardis⁴², Stefano Busani⁴², Sophie Venturelli⁴², Andrea Cossarizza⁴³, Andrea Antinori⁴⁴, Alessandra Vergori⁴⁴, Stefano Rusconi^{45,46}, Matteo Siano⁴⁶, Arianna Gabrielli⁴⁶, Agostino Riva^{45,46}, Daniela Francisci⁴⁷, Elisabetta Schiaroli⁴⁷, Carlo Pallotto⁴⁷, Saverio Giuseppe Parisi⁴⁸, Monica Basso⁴⁸, Sandro Panese⁴⁹, Stefano Baratti⁴⁹, Pier Giorgio Scotton⁵⁰, Francesca Andretta⁵⁰, Mario Giobbia⁵⁰, Renzo Scaggiante⁵¹, Francesca Gatti⁵¹, Francesco Castelli⁵², Eugenia Quiros-Roldan⁵², Melania Degli Antoni⁵², Isabella Zanella^{53,54}, Matteo della Monica⁵⁵, Carmelo Piscopo⁵⁵, Mario Capasso^{56,57}, Roberta Russo^{56,57}, Immacolata Andolfo^{56,57}, Achille Iolascon^{56,57}, Giuseppe Fiorentino⁵⁸, Massimo Carella⁵⁹, Marco Castori⁵⁹, Giuseppe Merla^{56,60}, Gabriella Maria Squeo⁶⁰, Filippo Aucella⁶¹, Pamela Raggi⁶², Rita Perna⁶², Matteo Bassetti^{63,64}, Antonio Di Biagio^{63,64}, Maurizio Sanguinetti^{65,66}, Luca Masucci^{65,66}, Alessandra Guarnaccia⁶⁵, Serafina Valente⁶⁷, Alex Di Florio⁶⁷, Marco Mandalà⁶⁸, Alessia Giorli⁶⁸, Lorenzo Salerno⁶⁸, Patrizia Zucchi⁶⁹, Pierpaolo Parravicini⁶⁹, Elisabetta Menatti⁷⁰, Tullio Trotta⁷¹, Ferdinando Giannattasio⁷¹, Gabriella Coiro⁷¹, Gianluca Lacerenza⁷², Cristina Mussini⁷³, Luisa Tavecchia⁷⁴, Lia Crotti^{75,76,77,78,79}, Gianfranco Parati^{75,76}, Roberto Menè^{75,76}, Maurizio Sanarico⁸⁰, Marco Gori^{81,82}, Francesco Raimondi⁸³, Alessandra Stella⁸³, Filippo Biscarini⁸⁴, Tiziana Bachetti⁸⁵, Maria Teresa La Rovere⁸⁶, Maurizio Bussotti⁸⁷, Serena Ludovisi⁸⁸, Katia Capitani⁸⁹, Simona Dei⁹⁰, Sabrina Ravaglia⁹¹, Annarita Giliberti⁹², Giulia Gori⁹², Rosangela Artuso⁹², Elena Andreucci⁹², Antonio Perrella⁹³, Francesco Bianchi^{19,93}, Paola Bergomi⁹⁴, Emanuele Catena⁹⁴, Riccardo Colombo⁹⁴, Sauro Luchi⁹⁵, Giovanna Morelli⁹⁵, Paola Petrocelli⁹⁵, Sarah Iacopini⁹⁵, Sara Modica⁹⁵, Silvia Baroni⁹⁶, Giulia Micheli⁹⁷, Marco Falcone⁹⁸, Donato Urso⁹⁸, Giusy Tiseo⁹⁸, Tommaso Matucci⁹⁸, Alice Pulcinelli⁹⁸, Davide Grassi⁹⁹, Claudio Ferri⁹⁹, Franco Marinangeli¹⁰⁰, Francesco Brancati^{101,102}, Antonella Vincenti¹⁰³, Valentina Borgo¹⁰³, Stefania Lombardi¹⁰³, Mirco Lenzi¹⁰³, Massimo Antonio Di Pietro¹⁰⁴, Letizia Attala¹⁰⁴, Cecilia Costa¹⁰⁴, Andrea Gabbuti¹⁰⁴,

**Alessio Bellucci¹⁰⁴, Marta Colaneri⁴⁰, Patrizia Casprini¹⁰⁵, Cristoforo Pomara¹⁰⁶,
Massimiliano Esposito¹⁰⁶, Roberto Leoncini¹⁰⁷, Michele Cirianni¹⁰⁷, Lucrezia Galasso¹⁰⁷,
Marco Antonio Bellini¹⁰⁸, Chiara Gabbi¹⁰⁹, Nicola Picchiotti⁸¹, Simone Furini^{19,110},
Elisabetta Pelo¹¹¹, Barbara Minuti¹¹¹, Francesca Gerundino¹¹¹, Chiara Lazzeri¹¹¹,
Arianna Vecchi¹¹¹, Leila Bianchi¹¹², Elisabetta Venturini¹¹², Carlotta Montagnani¹¹²,
Elena Chiappini^{112,113}, Cristina Beltrami¹¹² & Luisa Galli^{112,113}**

²²Infectious and Tropical Diseases Unit, Department of Medical Sciences, Azienda Ospedaliera Universitaria Senese, 53100 Siena, Italy. ²³Unit of Respiratory Diseases and Lung Transplantation, Department of Internal and Specialist Medicine, University of Siena, 53100 Siena, Italy. ²⁴Unit of Intensive Care Medicine, Department of Emergency and Urgency, Medicine, Surgery and Neurosciences, Siena University Hospital, 53100 Siena, Italy. ²⁵Unit of Diagnostic Imaging, Department of Medical, Surgical and Neuro Sciences and Radiological Sciences, University of Siena, 53100 Siena, Italy. ²⁶Rheumatology Unit, Department of Medicine, Surgery and Neurosciences, University of Siena, Policlinico Le Scotte, 53100 Siena, Italy. ²⁷Infectious Diseases Unit, Department of Specialized and Internal Medicine, San Donato Hospital, 52100 Arezzo, Italy. ²⁸Anesthesia Unit, Department of Emergency, San Donato Hospital, Arezzo, Italy. ²⁹Pneumology Unit and UTIP, Department of Specialized and Internal Medicine, San Donato Hospital, 52100 Arezzo, Italy. ³⁰Anesthesia Unit, Department of Emergency, Misericordia Hospital, 58100 Grosseto, Italy. ³¹Infectious Diseases Unit, Department of Specialized and Internal Medicine, Misericordia Hospital, 58100 Grosseto, Italy. ³²Clinical Chemical Analysis Laboratory, Misericordia Hospital, 58100 Grosseto, Italy. ³³Dipartimento di Prevenzione, Azienda USL Toscana Sud Est, 53100 Arezzo, Italy. ³⁴Dipartimento Tecnico-Scientifico Territoriale, Azienda USL Toscana Sud Est, 53100 Arezzo, Italy. ³⁵UOC Laboratorio Analisi Chimico Cliniche, 52100 Arezzo, Italy. ³⁶Chirurgia Vascolare, Ospedale Maggiore di Crema, 26013 Crema, Italy. ³⁷Department of Health Sciences, Clinic of Infectious Diseases, ASST Santi Paolo e Carlo, University of Milan, 20142 Milan, Italy. ³⁸Division of Clinical Immunology - Infectious Diseases, Department of Medicine, Fondazione IRCCS Policlinico San Matteo, 27100 Pavia, Italy. ³⁹Department of Internal Medicine and Therapeutics, University of Pavia, 27100 Pavia, Italy. ⁴⁰University of Pavia, 27100 Pavia, Italy. ⁴¹Department of Respiratory Diseases, Azienda Ospedaliera di Cremona, 26100 Cremona, Italy. ⁴²Department of Anesthesia and Intensive Care, University of Modena and Reggio Emilia, 41121 Modena, Italy. ⁴³Department of Medical and Surgical Sciences for Children and Adults, University of Modena and Reggio Emilia, 41121 Modena, Italy. ⁴⁴HIV/AIDS Department, National Institute for Infectious Diseases, IRCCS, Lazzaro Spallanzani, 00161 Rome, Italy. ⁴⁵III Infectious Diseases Unit, ASST-FBF-Sacco, 20146 Milan, Italy. ⁴⁶Department of Biomedical and Clinical Sciences Luigi Sacco, University of Milan, 20146 Milan, Italy. ⁴⁷Infectious Diseases Clinic, "Santa Maria Della Misericordia" Hospital, University of Perugia, 06100 Perugia, Italy. ⁴⁸Department of Molecular Medicine, University of Padova, Padua, Italy. ⁴⁹Clinical Infectious Diseases, Mestre Hospital, Venezia, Italy. ⁵⁰Department of Infectious Diseases, Treviso Hospital, Local Health Unit 2 Marca Trevigiana, Treviso, Italy. ⁵¹Infectious Diseases Clinic, ULSS1, Belluno, Italy. ⁵²Department of Infectious and Tropical Diseases, University of Brescia and ASST Spedali Civili Hospital, Brescia, Italy. ⁵³Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy. ⁵⁴Clinical Chemistry Laboratory, Cytogenetics and Molecular Genetics Section, Diagnostic Department, ASST Spedali Civili di Brescia, Brescia, Italy. ⁵⁵Medical Genetics and Laboratory of Medical Genetics Unit, A.O.R.N. "Antonio Cardarelli", Naples, Italy. ⁵⁶Department of Molecular Medicine and Medical Biotechnology, University of Naples Federico II, Naples, Italy. ⁵⁷CEINGE Biotecnologie Avanzate, Naples, Italy. ⁵⁸Unit of Respiratory Physiopathology, AORN dei Colli, Monaldi Hospital, Naples, Italy. ⁵⁹Division of Medical Genetics, Fondazione IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Italy. ⁶⁰Laboratory of Regulatory and Functional Genomics, Fondazione IRCCS Casa Sollievo della Sofferenza, Foggia, Italy. ⁶¹Department of Medical Sciences, Fondazione IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Italy. ⁶²Clinical Trial Office, Fondazione IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Italy. ⁶³Department of Health Sciences, University of Genova, Genoa, Italy. ⁶⁴Infectious Diseases Clinic, Policlinico San Martino Hospital, IRCCS for Cancer Research Genova, Genoa, Italy. ⁶⁵Microbiology, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Catholic University of Medicine, Rome, Italy. ⁶⁶Department of Laboratory Sciences and Infectious Diseases, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy. ⁶⁷Department of Cardiovascular Diseases, University of Siena, Siena, Italy. ⁶⁸Otolaryngology Unit, University of Siena, Siena, Italy. ⁶⁹Department of Internal Medicine, ASST Valtellina e Alto Lario, Sondrio, Italy. ⁷⁰Study Coordinator Oncologia Medica e Ufficio Flussi Sondrio, Sondrio, Italy. ⁷¹First Aid Department, Luigi Curto Hospital, Polla, Salerno, Italy. ⁷²Department of Pharmaceutical Medicine, Misericordia Hospital, Grosseto, Italy. ⁷³Infectious Diseases Clinics, University of Modena and Reggio Emilia, Modena, Italy. ⁷⁴U.O.C. Medicina, ASST Nord Milano, Ospedale Bassini, Cinisello Balsamo, MI, Italy. ⁷⁵Department of Cardiovascular, Neural and Metabolic Sciences, San Luca Hospital, Istituto Auxologico Italiano, IRCCS, Milan, Italy. ⁷⁶Department of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy. ⁷⁷Center for Cardiac Arrhythmias of Genetic Origin, Istituto Auxologico Italiano, IRCCS, Milan, Italy. ⁷⁸Laboratory of Cardiovascular Genetics, Istituto Auxologico Italiano, IRCCS, Milan, Italy. ⁷⁹Member of the European Reference Network for Rare, Low Prevalence and Complex Diseases of the Heart-ERN GUARD-Heart, Amsterdam, The Netherlands. ⁸⁰Milan, Italy. ⁸¹DIISM-SAILAB, University of Siena, Siena, Italy. ⁸²Maasai, I3S CNRS, Université Côte d'Azur, Nice, France. ⁸³Laboratorio di Biologia Bio@SNS, Scuola Normale Superiore, Pisa, Italy. ⁸⁴CNR-Consiglio Nazionale delle Ricerche, Istituto di Biologia e Biotecnologia Agraria (IBBA), Milan, Italy. ⁸⁵Direzione Scientifica, Istituti Clinici Scientifici Maugeri IRCCS, Pavia, Italy. ⁸⁶Department of Cardiology, Institute of Montescano, Istituti Clinici Scientifici Maugeri IRCCS, Pavia, Italy. ⁸⁷Department of Cardiology, Institute of Milan, Istituti Clinici Scientifici Maugeri IRCCS, Pavia, Italy. ⁸⁸Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. ⁸⁹Core Research Laboratory, ISPRO, Florence, Italy. ⁹⁰Health Management, Azienda USL Toscana Sud Est, Tuscany, Italy. ⁹¹IRCCS C. Mondino Foundation, Pavia, Italy. ⁹²Medical Genetics Unit, Meyer Children's University Hospital,

Florence, Italy. ⁹³Pneumology Unit, Department of Medicine, Misericordia Hospital, Grosseto, Italy. ⁹⁴Department of Anesthesia and Intensive Care Unit, ASST Fatebenefratelli Sacco, Luigi Sacco Hospital, Polo Universitario, University of Milan, Milan, Italy. ⁹⁵Infectious Disease Unit, Hospital of Lucca, Lucca, Italy. ⁹⁶Department of Diagnostic and Laboratory Medicine, Institute of Biochemistry and Clinical Biochemistry, Fondazione Policlinico Universitario A. Gemelli IRCCS, Catholic University of the Sacred Heart, Rome, Italy. ⁹⁷Clinic of Infectious Diseases, Catholic University of the Sacred Heart, Rome, Italy. ⁹⁸Infectious Diseases Unit, Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy. ⁹⁹Department of Clinical Medicine, Public Health, Life and Environment Sciences, University of L'Aquila, L'Aquila, Italy. ¹⁰⁰Anesthesiology and Intensive Care, University of L'Aquila, L'Aquila, Italy. ¹⁰¹Department of Life, Health and Environmental Sciences, University of L'Aquila, 67100 L'Aquila, Italy. ¹⁰²Human Functional Genomics Laboratory, IRCCS San Raffaele Roma, 00167 Rome, Italy. ¹⁰³Infectious Disease Unit, Hospital of Massa, Massa, Italy. ¹⁰⁴Infectious Diseases Unit, USL Centro, Santa Maria Annunziata Hospital, Florence, Italy. ¹⁰⁵Laboratory of Clinical Pathology and Immunoallergy, Florence-Prato, Italy. ¹⁰⁶Department of Medical, Surgical and Advanced Technologies "G.F. Ingrassia", University of Catania, Catania, Italy. ¹⁰⁷Laboratorio Patologia Clinica, Azienda Ospedaliero-Universitaria Senese, Siena, Italy. ¹⁰⁸Ambulatory Chronic Polipathology of Siena, Department of Medicine, Surgery and Neurosciences, University of Siena, Siena, Italy. ¹⁰⁹Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden. ¹¹⁰Bioinformatics, University of Bologna, Bologna, Italy. ¹¹¹SOD Diagnostica Genetica, Azienda Ospedaliero Universitaria Careggi, Florence, Italy. ¹¹²Infectious Diseases Unit, Department of Pediatrics, Meyer Children's Hospital, Florence, Italy. ¹¹³Department of Health Sciences, University of Florence, Florence, Italy.