ORIGINAL INVESTIGATION

Common, low-frequency, rare, and ultra-rare coding variants contribute to COVID-19 severity

Chiara Fallerini[1,2] · Nicola Picchiotti[3,4] · Margherita Baldassarri[1,2] · Kristina Zguro[1] · Sergio Daga[1,2] · Francesca Fava[1,2,5] · Elisa Benetti[1] · Sara Amitrano[5] · Mirella Bruttini[1,2,5] · Maria Palmieri[1,2] · Susanna Croci[1,2] · Mirjam Lista[1,2] · Giada Beligni[1,2] · Floriana Valentino[1,2] · Ilaria Meloni[1,2] · Marco Tanfoni[3] · Francesca Minnai[6] · Francesca Colombo[6] · Enrico Cabri[7] · Maddalena Fratelli[7] · Chiara Gabbi[8] · Stefania Mantovani[9] · Elisa Frullanti[1,2] · Marco Gori[3,10] · Francis P. Crawley[11] · Guillaume Butler-Laporte[12,13] · Brent Richards[12,14,15] · Hugo Zeberg[16] · Miklos Lipcsey[17,18] · Michael Hultström[17,19] · Kerstin U. Ludwig[20] · Eva C. Schulte[21,22,23] · Erola Pairo-Castineira[24,25] · John Kenneth Baillie[24,25,26] · Axel Schmidt[20] · Robert Frithiof[17] · WES/WGS Working Group Within the HGI · GenOMICC Consortium · GEN-COVID Multicenter Study · Francesca Mari[1,2,5] · Alessandra Renieri[1,2,5,27] · Simone Furini[1]

Abstract

The combined impact of common and rare exonic variants in COVID-19 host genetics is currently insufficiently understood. Here, common and rare variants from whole-exome sequencing data of about 4000 SARS-CoV-2-positive individuals were used to define an interpretable machine-learning model for predicting COVID-19 severity. First, variants were converted into separate sets of Boolean features, depending on the absence or the presence of variants in each gene. An ensemble of LASSO logistic regression models was used to identify the most informative Boolean features with respect to the genetic bases of severity. The Boolean features selected by these logistic models were combined into an Integrated PolyGenic Score that offers a synthetic and interpretable index for describing the contribution of host genetics in COVID-19 severity, as demonstrated through testing in several independent cohorts. Selected features belong to ultra-rare, rare, low-frequency, and common variants, including those in linkage disequilibrium with known GWAS loci. Noteworthily, around one quarter of the selected genes are sex-specific. Pathway analysis of the selected genes associated with COVID-19 severity reflected the multi-organ nature of the disease. The proposed model might provide useful information for developing diagnostics and therapeutics, while also being able to guide bedside disease management.

Co-first authors: Chiara Fallerini, Nicola Picchiotti, Margherita Baldassarri.

Co-last authors: Francesca Mari, Alessandra Renieri, Simone Furini.

The members of WES/WGS Working Group Within the HGI, GenOMICC Consortium, GEN-COVID Multicenter Study are listed in Acknowledgements.

✉ Alessandra Renieri
alessandra.renieri@unisi.it

Extended author information available on the last page of the article

Introduction

For almost 2 years, COVID-19 has demonstrated itself to be a disease having a broad spectrum of clinical presentations: from asymptomatic patients to those with severe symptoms leading to death or persistent disease ("long COVID") (Livingston and Bucher 2020; Chen et al. 2019; Zhang et al. 2020a). While developing vaccination programmes and other preventive measures to significantly dampen infection transmission and reduce disease expression, a much deeper and more precise understanding of the interplay between SARS-CoV-2 and host genetics is required to support the development of treatments for new virus variants as they arise. Furthermore, advances in modelling the interplay between SARS-CoV-2 and host genetics hold significant promise for addressing other complex diseases. In this study,

we demonstrate the value of genetic modelling with its direct translatability into drug development and clinical care in the context of a severe public health crisis.

The identification of host genetic factors modifying disease susceptibility and/or disease severity has the potential to reveal the biological basis of disease susceptibility and outcome as well as to subsequently contribute to treatment amelioration (Elhabyan et al. 2020). From a scientific point of view, COVID-19 represents a particularly interesting and accessible complex disorder for modeling host genetic data because the environmental factor (SARS-CoV-2) can be readily identified by a PCR-based swab test. The still moderate viral genome variability has thus far been shown to have relatively low impact on disease severity (Islam et al. 2020) where currently age, sex, and comorbidities are the major factors predicting disease susceptibility and outcome (Li et al. 2021). While these factors certainly have significant value for prediction, they provide limited insights into disease pathophysiology and are of limited relevance for drug development.

Common variants in the human genome affecting the susceptibility to SARS-CoV-2 infections and COVID-19 severity have been successfully identified by Genome-Wide Association Studies (GWASs) (Severe Covid-19 GWAS Group (2020); Pairo-Castineira et al. 2020; COVID-19 Host Genetics Initiative et al. 2021). However, these variants only explain a small fraction of trait variability and, as it is well documented, GWASs are difficult to interpret because they often associate non-coding variants with phenotype; therefore, the relevant genes need to be pinpointed by deeper follow-up analyses. In contrast, next-generation sequencing-based studies have identified variants in a few genes-related to innate immunity which can solely underlie rare severe forms of COVID-19 (Zhang et al. 2020b; Van der Made et al. 2020; Fallerini et al. 2021a; Solanich et al. 2021). In these rare affected families, the predictivity is high as the susceptibility for severe COVID-19 follows Mendelian inheritance patterns. However, these patients represent only a small proportion of those severely affected by COVID-19. Taken together, genetic findings can currently only explain a limited proportion of COVID-19 susceptibility and severity, in spite of the relatively high predicted heritability of COVID-19 and COVID-19 symptoms (Williams et al. 2020). A better and more holistic understanding of host genetics could support the development of more specific, or even targeted drugs and treatment interventions leading to less morbidity and mortality.

The Italian GEN-COVID Multicenter Study collected more than 2000 biospecimens and clinical data from SARS-CoV-2-positive individuals (Daga et al. 2021), and whole-exome sequencing (WES) analysis contributed to the identification of rare variants (Fallerini et al. 2021a) and common polymorphisms (Baldassarri et al. 2021a; Croci et al. 2021;

Fallerini et al. 2021b) associated with COVID-19 severity. In 2020, we started to investigate how common variants may combine with rare variants to determine COVID-19 severity in WES data using a first small cohort of hospitalized patients. This pilot analysis revealed that the combination of rare and common variants could potentially impact clinical outcome (Benetti et al. 2020a). We then proposed a new post-Mendelian model for a genetic characterization of the disorder (Picchiotti et al. 2021) based on an adapted Polygenic Risk Score (PRS) (Mars et al. 2020), called Integrated PolyGenic Score (IPGS). This allowed us to reach a more precise disease severity prediction than that based on sex and age alone. In this article, we substantially improve this post-Mendelian model to include ultra-rare and low-frequency variants while also demonstrating that IPGS significantly contributes to predictivity in combination with—as well as alongside—age and sex, and is able to extract patient-specific genes. The IPGS predictivity was also sustained in three independent European cohorts of the WES/Whole-Genome Sequencing study working group within the COVID-19 Host Genetics Initiative (COVID-19 Host Genetics Initiative 2021).

Materials and methods

Contributing cohorts

Five different cohorts (from Germany, Italy, Quebec, Sweden, UK) contributed to this study as described in Supplementary Table 1. For multi ancestry cohorts (Quebec and UK), the subpopulation of European Ancestry was included in the study. Institutional Review Board approval was obtained for each study (Supplementary Table 1).

Phenotype definitions

The training of the model proposed for predicting the severity of COVID-19 requires as inputs the exome variants, age, sex, and COVID-19 severity assessed using a modified version of the WHO COVID-19 Outcome Scale (COVID-19 Therapeutic Trial Synopsis 2020) as coded into the following six classifications: (1) death; (2) hospitalized receiving invasive mechanical ventilation; (3) hospitalized, receiving continuous positive airway pressure (CPAP) or bilevel positive airway pressure (BiPAP) ventilation; (4) hospitalized, receiving low-flow supplemental oxygen; (5) hospitalized, not receiving supplemental oxygen; and (6) not hospitalized. The aim of the model is to predict a binary classification of patients into mild and severe cases, where a patient is considered severe if hospitalized and receiving any form of respiratory support (WHO severity-grading equal to 4 or higher in 8 points classification). The next section describes

how the annotation of exome variants and the selection of patients were performed in the GEN-COVID cohort. Following this, the training and testing of the model are described.

Massive parallel sequencing

GEN-COVID cohort

Whole-exome sequencing with at least 97% coverage at 20x was performed using the Illumina NovaSeq6000 System (Illumina, San Diego, CA, USA). Library preparation was performed using the Illumina Exome Panel (Illumina) according to the manufacturer's protocol. Library enrichment was tested by qPCR, and the size distribution and concentration were determined using Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). The Novaseq6000 System (Illumina) was used for DNA sequencing through 150 bp paired-end reads. Variant calling was performed according to the GATK4 (O'Connor and Auwera 2020) best practice guidelines, using BWA (Li and Durbin 2010) for mapping and ANNOVAR (Wang et al. 2010) for annotating.

Swedish cohort

Whole-exome sequencing was performed using the Twist Bioscience exome capture probe and was sequenced on the Illumina NovaSeq6000 platform. Data were then analyzed using the McGill Genome Center bioinformatics pipeline (https://doi.org/10.1093/gigascience/giz037) in accordance with GATK best practices.

DeCOI Germany

800–1000 ng of genomic DNA of each individual was fragmented to an average length of 350 bp. Library preparation was performed using the TruSeq DNA PCR-free kit (Illumina, San Diego, CA, USA) according to the manufacturer's protocol. Whole-genome sequences were obtained as 150 bp paired-end reads on S4 flow cells using the NovaSeq6000 system (Illumina). The intended average sequencing depth was 30X. The DRAGEN pipeline (Illumina, version 3.6.3 or 3.5.7) was used for alignment and joint variant calling was performed with the Glnexus software (version 1.3.2). Individuals with a 20-fold coverage in less than 96% of the protein-coding sequence were removed as well as related individuals to retain only from related pairs. Variant QC was performed using hail (version 0.2.58). European individuals were selected by performing PCA analysis along with the 1000 genomes data. Finally, annotation was performed using Variant Effect Predictor (VEP, version 101). In the present study, only individuals from the BoSCO and CoMRI substudies were included.

BQC-19

Whole-genome sequencing at mean coverage of 30x was performed on the Illumina NovaSeq6000 platform, then analyzed using the McGill Genome Center bioinformatics pipeline (https://doi.org/10.1093/gigascience/giz037), in accordance with GATK best practice guidelines.

GenOMICC/ISARIC4C

Whole-genome sequencing at mean coverage of 20x was performed on the Illumina NovaSeq6000 platform and then analyzed using the DRAGEN pipeline (software v01.011.269.3.2.22 , hardware v01.011.269). Variants were genotyped with the GATK GenotypeGVCFs tool v4.1.8.1.

PC analysis

The genetic ancestry of the patients was estimated using a random forest classifier. SNPs of autosomes with MAF above 10% and in linkage disequilibrium were extracted from the 1000 genome project using BCFTOOLS (Danecek et al. 2021), and intersected with variants from the GEN-COVID cohort. The resulting set of variants was used to compute 6 Principal Components by PLINK (Purcell et al. 2007) using samples from the 1000 genome projects. GEN-COVID samples were projected along the same Principal Components. The Random Forest classifier, as implemented in the R library randomForest (Liaw and Wiener 2002), was trained using samples from 1000 genomes with known ancestry, and then used to predict ancestry for the GEN-COVID cohort. To avoid bias in the analysis due to the different ethnicity, only patients of genetic European ancestry were retained for further analyses.

Definition of the Boolean features

Variants were converted into 12 sets of Boolean features to better represent the variability at the gene-level. First, any variant not impacting on the protein sequence was discarded. Then, the remaining variants were classified according to their minor allele frequency (MAF) as reported in gnomAD for the reference population as: ultra-rare, MAF < 0.1%; rare, $0.1\% \leq MAF < 1\%$; low-frequency, $1\% \leq MAF < 5\%$; and common, $MAF \geq 5\%$. Non-Finnish European (NFE) was used as a reference population. SNPs with MAF not available in gnomAD were treated as ultra-rare. INDELs with frequency not available in gnomAD were treated as ultra-rare when present only once in the cohort and otherwise discarded as possible artefacts of sequencing. For the ultra-rare variants, 3 alternative Boolean representations were defined, which are designed to capture the autosomal dominant (AD), autosomal recessive (AR), and X-linked

(XL) model of inheritance, respectively. The AD and AR representations included a feature for all the genes on autosomes. These features were equal to 1 when the corresponding gene presented at least 1 for the AD model, or 2 for the AR model, variants in the ultra-rare frequency range and 0 otherwise. The XL representation included only genes belonging to the X chromosome. These features were equal to 1 when the corresponding gene presented at least 1 variant in the ultra-rare frequency range and 0 otherwise. The same approach was used to define AD, AR, and XL Boolean features for the rare and low-frequency variants. Common variants were represented using a different approach that is designed to better capture the presence of alternative haplotypes. For each gene, all the possible combinations of common variants were computed. For instance, in the case of a gene belonging to an autosome with 2 common variants (named A and B), 3 combinations are possible (A, B, and AB), and (consequently) 3 Boolean features were defined both for the AD and AR model. In the AR model each of these 3 features was equal to 1 if all the variants in that particular combination were present in the homozygous state and 0 otherwise. The same rule was used for the AD model, but setting the feature to 1 even if the variants in that particular combination are in the heterozygous state. In both models, AD and AR, a further feature was defined for each gene to represent the absence of any of the previously defined combinations. In the AD model, this feature was equal to 1 if no common variant is present and 0 otherwise; in the AR model, it is equal to 1 if no common variant is present in the homozygous state and 0 otherwise. The same approach was used to define the set of Boolean features for common variants in genes belonging to the X chromosome. The full list of Boolean representations is reported in Supplementary Table 2.

Connectome analysis

To measure the relatedness of the genes in our list with known genes related to viral susceptibility and disease severity, we used the human gene connectome (HGC—http://hgc.rockefeller.edu/). To define the lists of core genes we used three different steps. To probe the initial processes of virus entry and intracellular replication, we used genes coding for the virus–human proteins interactome found by Gordon et al. (2020) and we added a small set of genes known to be important for virus entry (https://viralzone.expasy.org/9077). To extend the study to the whole host response to the viral infection, we used the PanelApp https://panelapp.genomicsengland.co.uk). Finally, we used the union of GWAS loci (Severe Covid-19 GWAS Group (2020); Pairo-Castineira et al. 2020; COVID-19 Host Genetics Initiative et al. (2021)) and the genes in eQTL with them (by GTEX analysis), the genes already demonstrated to be associated

with severity in functional studies and the list deduced from integrative genomic analyses by Pathak et al. (2021).

Model training

The dataset was divided into a training set and a testing set (90/10), and the entire procedure described in this section was performed using only samples in the training set. A bootstrap approach with 100 iterations was adopted to train the model. At each bootstrap iteration, 90% of the samples were selected (without replication), and the following two steps were performed: (step 1) selection of the most relevant features for each Boolean representations; and (step 2) definition of the weights of the Integrated Polygenic Score (IPGS). After the 100 bootstrap iterations, the information extracted on relevant features and weighting factors are merged to define the final IPGS (step 3). The IPGS is then used, together with age and sex, for training a model that predicts the COVID-19 severity (step 4). These four steps of the training procedure are described in detail in the next subsections. Since the model is based on a combination of rare and common variants, the training procedure should be performed using a dataset with homogeneous ancestry.

Step 1: Features' selection

The subsets of the most relevant features were identified using logistic regression models with least absolute shrinkage and selection operator (LASSO) regularization. Separate logistic models were trained for each of the 12 sets of Boolean features. The predicted outcome variable for each of these models was a re-classified phenotype adjusted by age and sex. To compute these re-classified phenotypes as adjusted by age and sex, the patients were first divided into males and females. Then, for each sex, an ordinal logistic regression model was fitted using the age to predict the WHO phenotype classification into six grades. The ordinal logistic regression model was chosen as: it imposes a simple monotone relation between input feature and target variable; and it provides easily interpretable thresholds between the predicted classes. The patients with a predicted grading equal to the actual grading were excluded. The remaining patients were divided into two classes depending on whether their actual phenotype was milder or more severe than the one expected for a patient of that age and sex. This procedure has the benefit of isolating patients whose genetic factors are most important for predicting COVID-19 severity. This binary trait, i.e. phenotype more/less severe than expected, was used as the outcome variable for the 12 LASSO logistic models based on the 12 separate Boolean representations. For each LASSO model, the regularization strength was optimized by tenfold cross-validation with 50

equally spaced values in the logarithmic scale in the range $[10^{-2}, 10^1]$. The optimal regularization strength was selected as the one with the best trade-off between the simplicity of the model and the cross-validation score, i.e. as the highest regularization strength providing an average score closer to the highest average score than 0.5 standard deviations. Once the regularization strength was defined, the LASSO model was re-trained using all the samples in that particular bootstrap iteration. The features with non-null coefficients are the ones selected for the next step. In summary, for each bootstrap iteration, this procedure returns 12 lists of features (one for each Boolean representation) that are expected to be the most important features for predicting the phenotype adjusted by age and sex (in that particular bootstrap iteration).

Step 2: Weights of the Integrated Polygenic Score (IPGS)

In the previous step, the Boolean representations are considered isolated from each other. The aim of the IPGS is to combine information from different representations (Eq. 1). To reach this goal, it is necessary to compute the relative weights of the different contributions. For each bootstrap iteration, the list of relevant features extracted as described in the previous section are used to compute the number of features that are associated with mildness or severity for the different frequency ranges. For instance, n corresponds to the number of features associated with r the mild phenotype coming from Boolean features computed for variants in the frequency range [0.1%, 1%]. A feature is considered associated with the mild phenotype when its coefficient in the LASSO model estimated in step 1 is negative, i.e. it contributes to the prediction of the phenotype adjusted by age and sex in the direction of a phenotype less severe than what expected at that particular age and sex. The same rule, applied to the corresponding Boolean representation, is used to define the other feature-counts appearing in Eq. (1). The weighting factors in Eq. (1) were estimated as the ones that maximize the Silhouette coefficient of the separation between the clusters of patients more/less severe than expected. The minimization was performed with weighting factors restricted to the following ranges: $F_{LF} \in [1, 4]$, $F_R \in [2, 8]$, and $F_{UR} \in [5, 100]$.

This procedure returns three optimal values for the weighing factors associated with each bootstrap iteration.

Step 3: IPGS definition

In this step, the data extracted at each bootstrap iteration in steps 1 and 2 are combined to define the IPGS. First, for each of the Boolean features, of all the 12 representations, the number of times this feature was selected in the 100 bootstrap iterations is computed. Then, the entire bootstrap procedure is repeated using random input phenotypes, and the 5th percentile of the number of times that a feature is associated with a random phenotype is estimated. This threshold, computed separately for each Boolean representation, was used to select which Boolean features are included in the final model. As no significant association is expected among the Boolean features and the random phenotype, the threshold of the 5th percentile is expected to exclude with a 95% level of confidence the possible false positive associations. For the GEN-COVID cohort, the features selected correspond to ~ 4.4% of the initial number of Boolean features. The weighing factors in Eq. (1) were computed as the median values of the estimates obtained in the 100 bootstrap iterations.

Step 4: Training of the predictive model based on age, sex, and IPGS

The procedure described in the previous sections completely defines how to calculate the IPGS. The predictive model of the binary COVID-19 severity (hospitalized patients with any form of respiratory support versus all other patients) was defined as a logistic model that uses as input features IPGS, age, and sex. It should be noted that in steps 1-3, only patients that deviates from their expected severity based on age and sex were used. The procedure was designed to isolate the genetic basis of COVID-19 severity. Instead, in this final step, IPGS, age and sex are combined to predict the actual COVID-19 severity, i.e. hospitalized patients with any form of respiratory support. To prevent overfitting, the model was fitted using 466 samples different from the training set adopted in steps 1–3. During the fitting procedure, the class unbalancing is tackled by penalizing the misclassification of the minority class with a multiplicative factor inversely proportional to the class frequencies. The percentile normalization of the IPGS scores is performed within each cohort. An alternative logistic model that used as input features only age and sex was also fitted on the same training set. The comparison between the two models is intended to evaluate if the genetic information summarized in the IPGS improves the prediction of severity compared to a model based on age and sex alone. A further logistic regression model is fitted by only considering the IPGS variable.

Model testing

The training procedure returned 2 logistic models to be compared: one using as input features only age and sex, and the other one using as input features age, sex, and the IPGS. These models were tested, without any further adjustment, using other cohorts of European ancestry. The performances of the two models, with and without IPGS, were evaluated and compared in terms of accuracy, precision, sensitivity,

and specificity. The increases of the performances are evaluated with respect to the performances of a model where the values of the IPGS feature have been shuffled, by computing the p value on the empirical null distribution. In addition, the empirical probability density function of IPGS has been estimated for the severe and non-severe patients of the cohort including both train and test sets and a *t* test is carried out to evaluate whether the means of the two distributions were significantly different. As a further evaluation of the importance of the IPGS score on the severity prediction, univariate logistic regression models using as independent variables age (continuous represented in decades), sex, and IPGS were fitted to the dataset that combines both the training set and the testing sets for a total of 2240 patients. These models were used to estimate the odds ratios and the *p* values of the association with the severe phenotype. Furthermore, a multivariable logistic regression was fitted using IPGS, age, and sex together. Finally, a multivariable logistic regression was performed using as predictor variables: IPGS, age, sex and comorbidities (congestive/ischemic heart failure; asthma/COPD/OSAS; diabetes; hypertension; cancer). This latter model has been fitted in the training set, where the information on comorbidities was available.

Pathway analysis

Pathway analysis was made using a ranked GSEA approach (Subramanian et al. 2005; Mootha et al. 2003), modified according to the specificity of our data. The metrics for gene ranking was calculated on the basis of the results of feature selection models, weighting in each Boolean feature both beta values and the number of bootstrap iterations where it was found significantly associated with severity/mildness (Supplementary Tables 3–6). All the Boolean features that were found significant in at least one of the models were included in the list. As the sign of beta depends on which allele is taken as reference (which is relative for common variants), we decided to use absolute beta values for all the features. To also weight the importance according to variant frequency, we used the *F* values from the IPGS score for the four categories (ultra-rare 5, rare 4, low frequency 2, common 1). Finally, we summed all the weights of each Boolean feature by gene. Briefly,

$$W_{\text{geneA}} = \sum_{\text{Features Gene A}} \text{ABS}(\text{mean}\beta) \times \text{count} \times F.$$

Pathway enrichment analysis was made using the GSEA-preranked module (v. 7.2.4) of the Genepattern platform (Wang et al. 2010), on several pathway categories (BIOCARTA, KEGG, REACTOME, GOBP, HALLMARKS, C7 and C8), limiting the size of gene sets to the 10-300 range and performing 10,000 permutations. The networks showing similarity of significant pathways were built using the EnrichmentMap algorithm (Subramanian et al. 2005) in the Cytoscape suite (v. 3.8.2) (Merico et al. 2011; Shannon et al. 2003). Parameters used for network creations are: Jaccard Overlap Combined Index (*k* constant = 0.5), edge cutoff 0.05.

Website and data distribution

The coordination of international partners has been possible through the Host Genetics Initiative (HGI) (https://www.covid19hg.org/projects/).

Results can be shared through the Gen-Covid website (https://sites.google.com/dbm.unisi.it/gen-covid).

Code availability

Data analyses were performed using Python with the Scipy ecosystem (Virtanen et al. 2020), and the scikit-learn library (Pedregosa et al. 2011). Statistical association was done with the statsmodel Python library. The code is freely available at the github repository: https://github.com/gen-covid/pmm.

Results

The post-Mendelian paradigm for COVID-19 modelization for combining interpretability with predictivity based on ultra-rare, rare, low-frequency, and common variants

The aim of the present study was to develop an easily interpretable model that could be used to predict the severity of COVID-19 from host genetic data. Patients were considered severe when hospitalized and receiving any form of respiratory support. The focus on this target variable is motivated by the practical importance of rapidly identifying which patients are more likely to require oxygen support, in an effort to prevent further complications. Interpretability has been a guiding principle in the definition of the machine-learning model, as only a readily interpretable model can provide useful and reliable information for clinical practice while also contributing significantly to diagnostic, and therapeutic targeting. The high dimensionality of host genetic data poses a serious challenge to evident and reliable interpretability. So far, the development of a robust predictive model able to make a direct association between single variants and disease severity grading based on an accurate analysis of the vast number of host genetic variants compared to a much smaller number of individual patients has proven to be too complex and ultimately unreliable. To address the complexity with predictive reliability, an enriched gene-level representation of host genetic data was modeled in a

machine-learning framework. The complexity of COVID-19 immediately suggests that both common and rare variants are expected to contribute to the likelihood of developing a severe form of the disease. However, the contribution of common and rare variants to the severe phenotype is not expected to be the same. A single rare variant that impairs the protein function might cause a severe phenotype by itself after viral infection, while this is not so probable for a common polymorphism, which is likely to have a less marked effect on protein functionality. These observations led to the definition of a score, named IPGS, that includes data regarding the variants at different frequencies:

$$\text{IPGS} = \left(n_{\text{C}}^{\text{s}} - n_{\text{C}}^{\text{m}}\right) + F_{\text{LF}} \cdot \left(n_{\text{LF}}^{\text{s}} - n_{\text{LF}}^{\text{m}}\right) \\ + {}_{R} \cdot \left(n_{\text{R}}^{\text{s}} - n_{\text{R}}^{\text{m}}\right) + F_{\text{UR}} \cdot \left(n_{\text{UR}}^{\text{s}} - n_{\text{UR}}^{\text{m}}\right). \quad (1)$$

In Eq. (1), n variables are used to indicate the number of input features of the predictive model that promote the severe outcome (superscript s) or that protect from a severe outcome (superscript m) and with genetic variants having Minor Allele Frequency (MAF) $\geq 5\%$ (common, subscript C), $1\% < \text{MAF} \leq 5\%$ (low-frequency, subscript LF), $0.1\% < \text{MAF} \leq 1\%$ (rare, subscript R), and $\text{MAF} < 0.1\%$ (ultra-rare, subscript UR). The features promoting or preventing severity were identified by an ensemble of logistic models, as described in the next section. The weighting factors F_{LF}, $_{\text{R}}$, and F_{UR} were included to model the different penetrant effects of low-frequency, rare, and ultra-rare variants, compared to common variants. Thus, the 4 terms of Eq. (1) can be interpreted as the contributions of common, low-frequency, rare, and ultra-rare variants to a score that represents the genetic propensity of a patient to develop a severe form of COVID-19.

Feature selection and gene discovery

The definition of the single terms of the *IPGS* formula requires 4 separate steps (Fig. 1): (1) the definition of a severity phenotype adjusted by age and sex; (2) the conversion of genetic variants into Boolean features that represent the presence of variants in different frequency ranges in each gene; (3) the selection of those features that are associated with disease severity; and (4) the optimization of the weighting factors appearing in Equation 1. These 4 steps were executed using data from a training set extracted from the GEN-COVID dataset (90% of the patients, n=1780, see Methods). The phenotype adjusted by age and sex was computed using an ordered logistic regression model, with the purpose of facilitating the extraction of features associated with the genetic basis of COVID-19 severity (Fig. 1B). The conversion of genetic variants into Boolean features led to the definition of 12 separate sets of input features (Supplementary Table 2). The set of input features

"ultra-rare_autosomal dominant" (UR_AD) is designed to represent in a binary way an autosomal dominant hereditary model associated with variants with MAF lower than 0.1%, i.e. these Boolean features are equal to 1 for genes presenting at least one variant in this frequency range. Similarly, the set of input features "ultra-rare_autosomal recessive" (UR_AR) and "ultra-rare_X-linked" (UR_X) were designed to describe the autosomal recessive and X-linked models of inheritance of ultra-rare variants. Analogous principles were used for rare and low-frequency variants. In the case of common variants, the same 3 sets of Boolean features representing the autosomal dominant, autosomal recessive, and X-linked models of inheritance were used. However, instead of simply defining the binary variables as "absence/presence of variants", the absence/presence of variant combinations was tested (Fig. 1C).

The Boolean representation of the genetic variability described in the previous paragraph significantly reduces the dimensionality of the problem. However, the number of input features is still orders of magnitude higher than the number of patients that can be reasonably collected for training the model. To reduce the number of input features, a feature selection strategy based on logistic models with the validated Least Absolute Shrinkage and Selection Operator (LASSO) regularization was employed (Fig. 1D). The aim of LASSO regularization is to minimize (shrink) the number of coefficients of the model, consequently minimizing the number of input features used for predicting outcomes. Separate logistic models with LASSO regularization were trained for the 12 sets of Boolean features for predicting COVID-19 severity, allowing us to identify the relevant features for each set. About 4% of the cumulative tested features were found to contribute to COVID-19 variability in severity (Fig. 1D). Twenty-six percent identity match between extracted genes and known viral susceptibility genes was found (Supplementary Table 12). We further investigate extracted genes using the Human Gene Connectome (HGC). Interestingly, of the 4943 genes of our model that are mapped in HGC, 4401 (89%) are biologically significantly connected ($p < 0.05$) and 2847 (57.6%) with a degree of 0 (overlap) or 1 (direct connection) with one of the genes of the three Core Lists (Supplementary Table 13).

Biological interpretability of extracted genetic features

Selected genes contributed by ultra-rare, rare, low-frequency variants, or/and common variants (Fig. 2A–D and Supplementary Tables 3a–g). Specifically, 54% contributed by only one, 29% by two, 11% by three, and 6% by four types of variants. Around 25% of the genes were sex-specific. The latter group includes either genes located on the X chromosome, such as *TLR7* and *TLR8* in males, or genes regulated
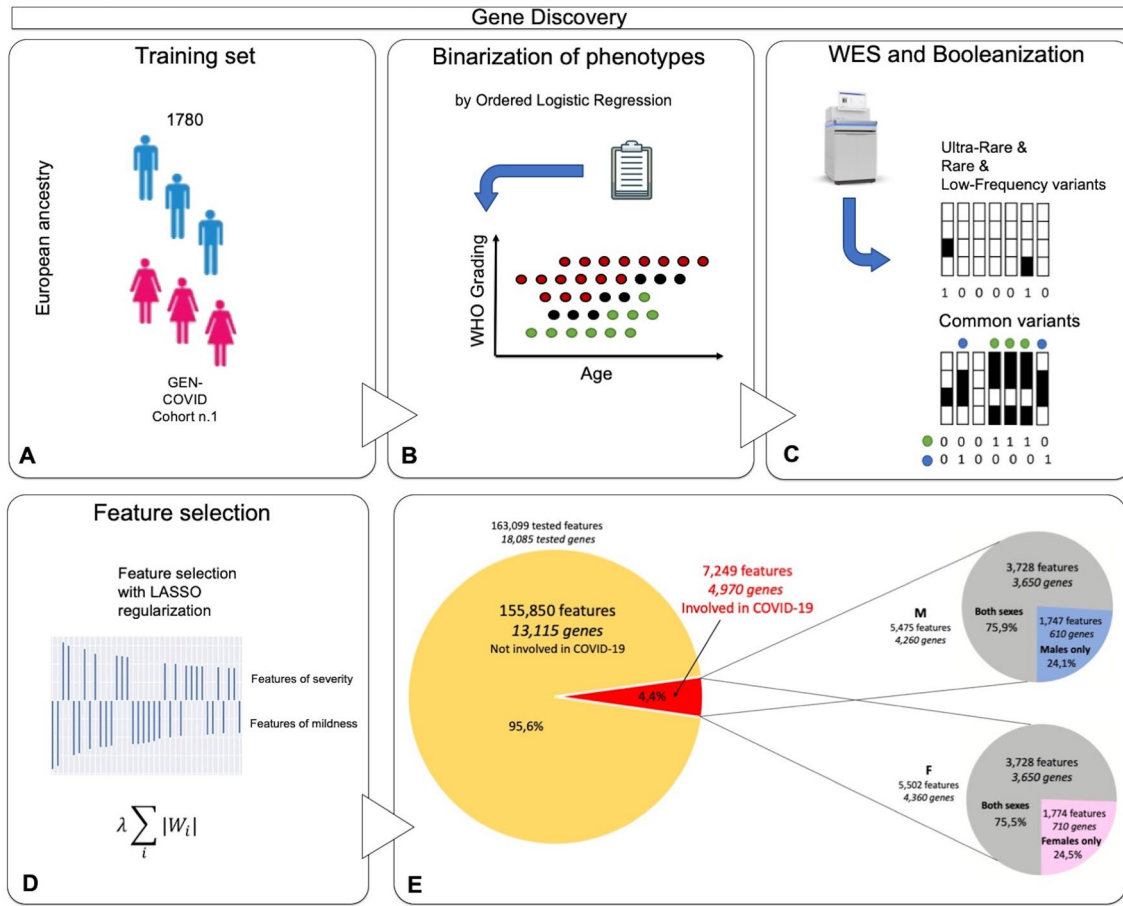
Fig. 1 Feature selection and gene discovery. **A** Whole-exome sequencing (WES) data stored in the Genetic Data Repository of the GEN-COVID Multicenter Study (GCGDR) and coming from biospecimens of 1780 SARS-CoV-2 PCR-positive subjects of European ancestry of different severity were used as the training set. **B** Clinical severity classification into severe and mild cases was performed by Ordered Logistic Regression (OLR) starting from the WHO grading and patient age classifications. **C** WES data were binarized into 0 or 1 depending on the absence (0) or the presence (1) of variants (or the combination of two or more variants only for common polymorphisms) in each gene. **D** LASSO logistic regression feature selec-tion methodology on multiple train-test splits of the cohort leads to the identification of the final set of features contributing to the clinical variability of COVID-19 (**E**). From the initial 163,099 cumulative features (divided into 36,540 ultra-rare, 23,470 rare, 13,056 low frequency and 90,033 common features) in 12 Boolean representations, the selected features contributing to COVID-19 clinical variability are 7249 and they are reported in the Supplementary Tables 3–6. The total number of genes contributing to COVID-19 clinical variability was 4260 in males and 4360 in females, 75% of which were in common

in opposite directions by androgens and estrogens when contributing with less penetrant common variants, such as p.L412F in *TLR3* and p.D603N in *SELP* gene (Fig. 2A–D).

Among the extracted ultra-rare variants there was a group of genes, such as *TLR3*, *TLR7* and *TICAM1*, already shown to be directly involved in the Mendelian-like forms of COVID-19 (Fig. 2A and Supplementary Table 3a, b). Furthermore, another group of genes are natural candidates because of their function: these include the *ACE2* shedding protein *ADAM17*, *CFTR*-related genes, genes involved in glycolipid metabolism, genes expressed by cells of the innate immune system, and genes involved in the coagulation pathway. Finally, a group of genes led by *ACE2* (if affected by ultra-rare variants) confers protection from the severe

disease. This group includes several genes whose mutations are responsible for auto-inflammatory disorders.

Among the rare variants extracted, we identified some genes as candidates for COVID-19 severity, including *TLR5* and *SLC26A9* as well as other genes involved in the inflammatory response (Fig. 2B and Supplementary Table 4a, b).

Among the low-frequency variants extracted, we identified some genes associated with either severity or protection from severe COVID-19 that are linked to the *CFTR* pathway (e.g., *PSMA6*) as well as specific genes involved in the immune response (e.g., *NOD2*) (Fig. 2C and Supplementary Table 5a, b).

The model was also able to identify a group of extracted common variants already shown to be linked to either severe

or mild COVID-19 (Fig. 2D and Supplementary Table 6a, b). Among them are the L412F *TLR3* and D603N *SELP* polymorphisms, already reported to be associated with the severe disease (Croci et al. 2021; Fallerini et al. 2021b) and several coding polymorphisms in Linkage Disequilibrium (LD) with already reported genomic SNP, such as the ABO blood group, *OAS1-3* genes, *PPP1R15A* gene and others (Elhabyan et al. 2020). In conclusion, considering their functions, genes involved in the immune and inflammatory responses, or those involved in the coagulation pathway and NK and T cell receptor, are to be considered natural candidates for severe or mild COVID-19.

Integrated PolyGenic Score definition

The Boolean features selected by the LASSO logistic models were used to calculate ten variables in Eq. (1) (Fig. 3A, B). The corresponding weights (*F* variables) were defined by optimizing the separation between severe and mild cases as offered by the *IPGS* formula. The optimization was measured using the Silhouette coefficient, and the optimal values were computed using a grid-search approach over a predefined grid ($F_{LF} \in [1, 4]$, $F_R \in [2, 8]$, and $F_{UR} \in [5, 100]$).

This optimization returned values of 2, 4, and 5 for the low-frequency, rare, and ultra-rare variants, respectively (Fig. 3C, D).

Pathway analysis

To understand the biological mechanisms underlying the variability of disease, we performed a pathways analysis of the genes carrying variants discovered in the feature selection described above. The features obtained with this approach do not have the same predicted impact and are not discovered with the same confidence. Therefore, we decided to perform a rank-based pathway analysis, with genes with the highest impact and confidence ranking highest in the list, rather than a simple over-representation approach. We ranked all the genes that were found to be significantly associated with severity/mildness in at least one bootstrap repetition, based on a score that takes into account three parameters: average coefficient in the LASSO models selecting the feature, number of significant bootstrap results, and the *F* correction factor for the frequency category used in the IPGS (detailed in the Methods section below). For genes with more than one significant Boolean feature, we summed up the scores of each feature. Gene Set Enrichment Analysis (GSEA) was then performed using two separate ranked gene lists (Supplementary Table 7) for females and males, followed by the generation of similarity networks using EnrichmentMap (Fig. 4A). The usage of rank-based search method allows to identify statistically significant pathways starting from extensive list genes, as each gene is associated with

its specific importance. Although no pathways satisfied the 0.25% FDR threshold normally required for standard GSEA analyses, the set of pathways considered significant using more relaxed thresholds on p values were shown to group in meaningful modules, providing useful information on pathogenetic mechanisms and on the genes that could explain how they can be affected. The network of all the pathways significantly enriched in both females and males ranked gene lists ($p < 0.01$, $n = 25$) is depicted in Fig. 4B, while the network of all the pathways enriched in either females or males with a more stringent *p* value ($p < 0.005$, $n = 100$) is shown in Fig. 4C. Detailed information on the names of the pathways and *p* values of enrichment is reported in Supplementary Figures 1 and 3. For the most representative pathways of each network, the heatmaps of the genes with their weights of association to disease variability are shown in Fig. 4D and Supplementary Figures 2 and 4, while gene lists and gene weights for all the significant pathways are reported in Supplementary Table 8.

COVID-19 post-Mendelian model predictivity

The functional interpretation of the variants identified by the feature selection approach, complemented by the strong link between the involved human biological pathways and COVID-19 pathogenicity, support the hypothesis that the IPGS score developed here may contribute significantly to predicting the severity of COVID-19. This hypothesis was tested using a logistic regression model that predicts COVID-19 severity based on age, sex and the IPGS (after percentile normalization). The training set is composed of 466 patients not included in the training set previously exploited for the IPGS feature engineering. The model's performance was then tested using three independent cohort sets of European ancestry (Fig. 5A). The model exhibited an overall accuracy of 0.73, precision equal to 0.78, with a sensitivity and specificity of 0.72 and 0.75, respectively. Noteworthy, all the aforementioned metrics are higher than the corresponding values obtained using a logistic model that adopted as input features only age and sex. The increase in performances of the model with IPGS suggests that this score indeed confers significant additional (genetic) information for predicting COVID-19 severity compared to only age and sex. The increase of the performances is statistically significant (*p* value < 0.05 for accuracy, precision, sensitivity, specificity) with respect to the distribution of performances for an ensemble of models where the IPGS feature has been randomized (Fig. 5C lower left). A third logistic regression model fitted with IPGS alone, shows performances well above the random guess. Furthermore, the empirical probability density function of IPGS scores (Fig. 5C right) has been estimated for the severe and non-severe patients of the cohort

Biological impact of features
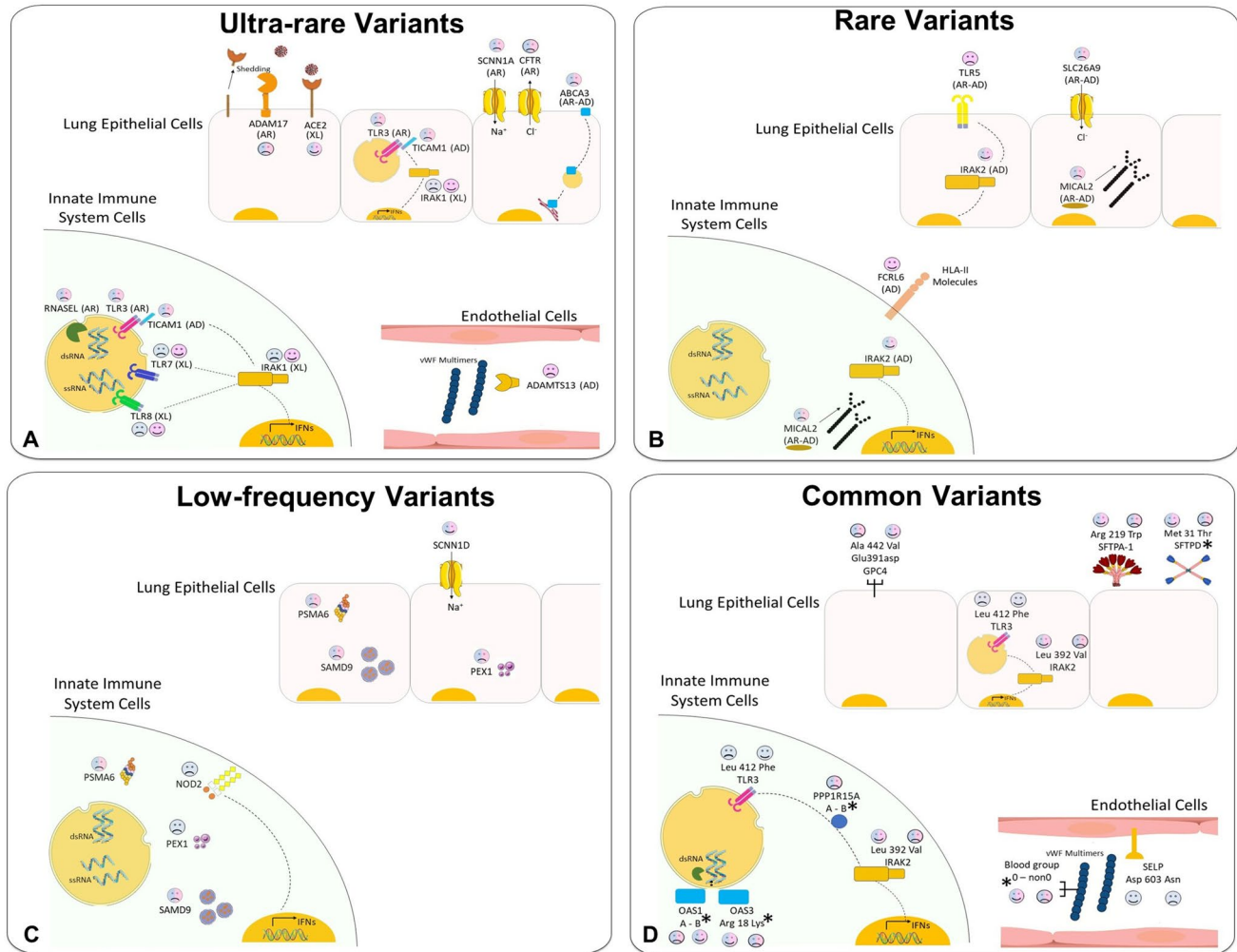
including both training and testing sets. It is worth noting the shift on the right of the *IPGS* distribution for the severe patients, with significant *p* value (< 0.001) for the *t* test of mean difference. This difference between severe and non-severe cases is preserved for the male and female cohorts when analyzed separately (*p* values < 0.001 and 0.024, respectively).

In line with the results obtained using the overall test set, the model including IPGS, age, and sex performed better than the model considering only sex and age as inputs, in each of the testing cohorts, separately (Fig. 5D). The increase in performance was systematically observed throughout all the cohorts: on average + 1.33% for accuracy, + 1% for precision, + 1.33% for sensitivity, + 1.67% for specificity. Considering the difference in phenotype classification inherent to a comparison among various international cohorts, and the genetic variability among different European sub-populations, the consistent increase in performances observed for the model with IPGS demonstrates that this score provides a robust

index for predicting COVID-19 severity. As a further test for the importance of the IPGS score for predicting COVID-19 severity, the univariate logistic models were used on the overall set including both train and test cohorts to estimate the OR of severe COVID-19 for IPGS, age, and sex, separately. The test confirmed that severity was associated with IPGS, showing an OR of 2.32 ($p < 0.001$, 95% confidence interval [1.79, 3.01]) with age, measured in decades, and sex, having OR of 1.89 ($p < 0.001$, 95% confidence interval [1.79, 2.00]) and 2.99 ($p < 0.001$, 95% confidence interval [2.58, 3.46]) respectively (Fig. 5E). The multivariate logistic regression using sex, age, and IPGS together, provided similar results reported in Supplementary Table 9 confirming the goodness of the regressors' OR. When adjusting for comorbidities, in the train cohort where the comorbidities were available, with a multivariable logistic model, OR of IPGS was 2.46 ($p = 0.05$, 95% confidence interval [1.15, 5.25]) as shown in Supplementary Table 10. This result further confirms that IPGS is a reliable predictor of COVID-19 clinical severity.

◀**Fig. 2** Biological impact of ultra-rare, rare, low-frequency, and common features. Examples of ultra-rare (**A**), rare (**B**), low-frequency (**C**), and common (**D**) features are illustrated in panel **A–D**. The complete list of features is presented in Supplementary Tables 3–6. ☺=contributing to COVID-19 severity; ☺=contributing to COVID-19 mildness. Pink faces=contributing to females only; blue faces=contributing to males only; pink/blue faces=contribution in both sexes. In parentheses: AD=autosomal dominant inheritance; AR=autosomal recessive inheritance; XL=X-linked recessive inheritance. **A** Ultra-rare mutations in the RNA sensor *TLR7*, *TLR3*, and *TICAM1* (encoding TRIF protein), already reported associated with XL, AR and AD inheritance (Zhang et al. 2020a; Van der Made et al. 2020; Fallerini et al. 2021a; Solanich et al. 2021) impair interferon (IFNs) production in innate immune system cells. Mutations in *TLR8*, as well as of the signal transducer *IRAK1* also impair interferon production. The specific location of *TLR7/8* and *IRAK1* (on the X chromosome) as well as X-inactivation escaping are responsible for opposite effects in males and females. Mutation in *RNASEL* impair the antiviral effect of the gene. In lung epithelial cells, *ACE2* ultra-rare variants (on the X chromosome) exert protective effects (probably) due to lowering virus entrance, while ultra-rare variants in *ADAM17* (might) reduce the shedding of *ACE2* and induce a severe outcome. The same is true for *CFTR* and *SCNN1A* (encoding ENaCA protein and involved in a CFTR-related physiological pathway), and the lipid transporter *ABCA3* (Baldassarri et al. 2021b). Mutations of *ADAMTS13* in vessels reduce the cleavage of the multimeric von Willebrand Factor (VWF), leading to thrombosis; B) Rare variants of the estrogen regulated *TLR5* are associated with severity in females. Rare variants of the CFTR-related *SLC26A9* are associated with severity in both sexes. This ion transporter has three discrete physiological modes: nCl(–)-HCO(3)(–) exchanger, Cl(–) channel, and Na(+)-anion cotransporter. Other examples of rare mutations associated with severity are the NK and T cell receptor *FCRL6*, IFN signal transducer *IRAK2*, and the actin depolymerization *MICAL2*; **C** low-frequency variants in another CFTR-related gene, *SCNN1D* (encoding for ENaCD protein) are associated with mildness, while rare variants in the following genes are associated with severity: cargo protein SPMA6, vesicle formation PEX1, inflammatory protein *NOD2* (*CARD15*); **D** A number of coding polymorphisms, indicated with an asterisk, are in LD with genomic SNPs already associated with COVID-19 (The complete list is presented in Supplementary Table 11) (Severe Covid-19 GWAS Group (2020; Pairo-Castineira et al. 2020). In some cases, such as the case of *SFTDP*, the genomic SNP is the coding polymorphism itself. Of note are the genes of surfactant proteins associated with severe disease: *SFTDP* gene encoding for SP-D protein and *SFTPA1* gene encoding for SP-A protein; the signal transducer, *PPP1R15A* gene encoding for GADD34 protein. OAS1 and OAS3 related to RNA clearance of *RNASEL* (reported in panel A as having ultra-rare mutations; included here should also be the already reported *TLR3412* (Croci et al. 2021); the already reported SELP603 related to thrombosis (Fallerini et al. 2021a). Note: *OAS1* haplotype A=c.1039-1G>A (Wickenhagen et al. 2021), (p.(Gly162Ser)), (p.(Ala352Thr)), (p.(Arg361Thr)), (p.(Gly397Arg)), (p.(Thr358Profs*26)). *OAS1* haplotype B=haplotype without the variant combination in haplotype A

Advantages of IPGS and clinical interpretability of connected features

We then wanted to compare the clinical outcome with the probability of severity obtained from three different models: IPGS alone, sex-age alone or combined model (represented as heatmap in Fig. 6). It appears evident that in a subset of patients, the 2 models based on sex-age alone and IPGS alone have a discordant prediction (left and right end of dendrogram in Fig. 6A). In these cases, IPGS appears to be a relevant predictor of severity (Fig. 6A). This is in accordance with the above-presented logistic regression analysis (Fig. 5E) that shows IPGS having an OR of 2.32 for severity. Moreover, the list of features on which the IPGS score is built, represent a biological handle for pathophysiological mechanisms and possible personalized adjuvant treatments.

For example, three male patients, within two distinct age ranges (46–50, 51–55) (panels B, C and D) with severe outcome (intubation and CPAP) are imperfectly represented by the sex-age model (probability of severity from 0.52 to 0.66) and better represented by the IPGS model (probability of severity from 0.91 to 0.95). The detected genetic variants that would allow to clinically consider putative personalized treatments in similar cases are: (1) *TLR7* ultra-rare mutation indicating to consider possible adjuvant treatment with IFN gamma administration (Fallerini et al. 2021a); (2) homozygosity 603Asn in *SELP* gene suggesting putative adjuvant treatment with anti-selectin P autoantibodies (e.g. Crizanlizumab) (Fallerini et al. 2021b) and (3) polyQ longer than 23 in *AR* gene suggesting to consider possible adjuvant treatment with testosterone (Baldassarri et al. 2021a).

In a female patient, within age range 31–35, the sex-age model showed a probability of severity of 0.17 (panel D) while the IPGS score was 336 corresponding to a probability of severity of 0.95. The patient had no comorbidities except for hypothyroidism. She underwent steroid treatment and CPAP ventilation. She was found to be carrier of *ADAMTS13* ultra-rare mutation, being more susceptible to thrombosis (due to reduced capacity of cutting von Willebrand factor); she had indeed a high D-dimer value. Caplacizumab (an antibody anti-vWF) would be an option to consider as possible adjuvant treatment in the clinical management of similar cases.

Two male patients, within two distinct age ranges (81–85, 86–90) (panel F and G) with a relatively mild respiratory outcome (hospitalised with low-flow oxygen therapy) presented an IPGS score of − 258 and − 141, respectively. Their severity probabilities calculated on sex-age (0.9 and 0.94) do not mirror the relatively mild clinical outcome, which is instead better represented by the severity probability calculated in IPGS only (0.23 and 0.41). Those two patients presented ultra-rare variants in *ACE2* gene, likely responsible for reduced viral load (Benetti et al. 2020b), and in *AGTR2* gene, which reduced activity is known to prevent cystic fibrosis pulmonary manifestation (Darrah et al. 2019).
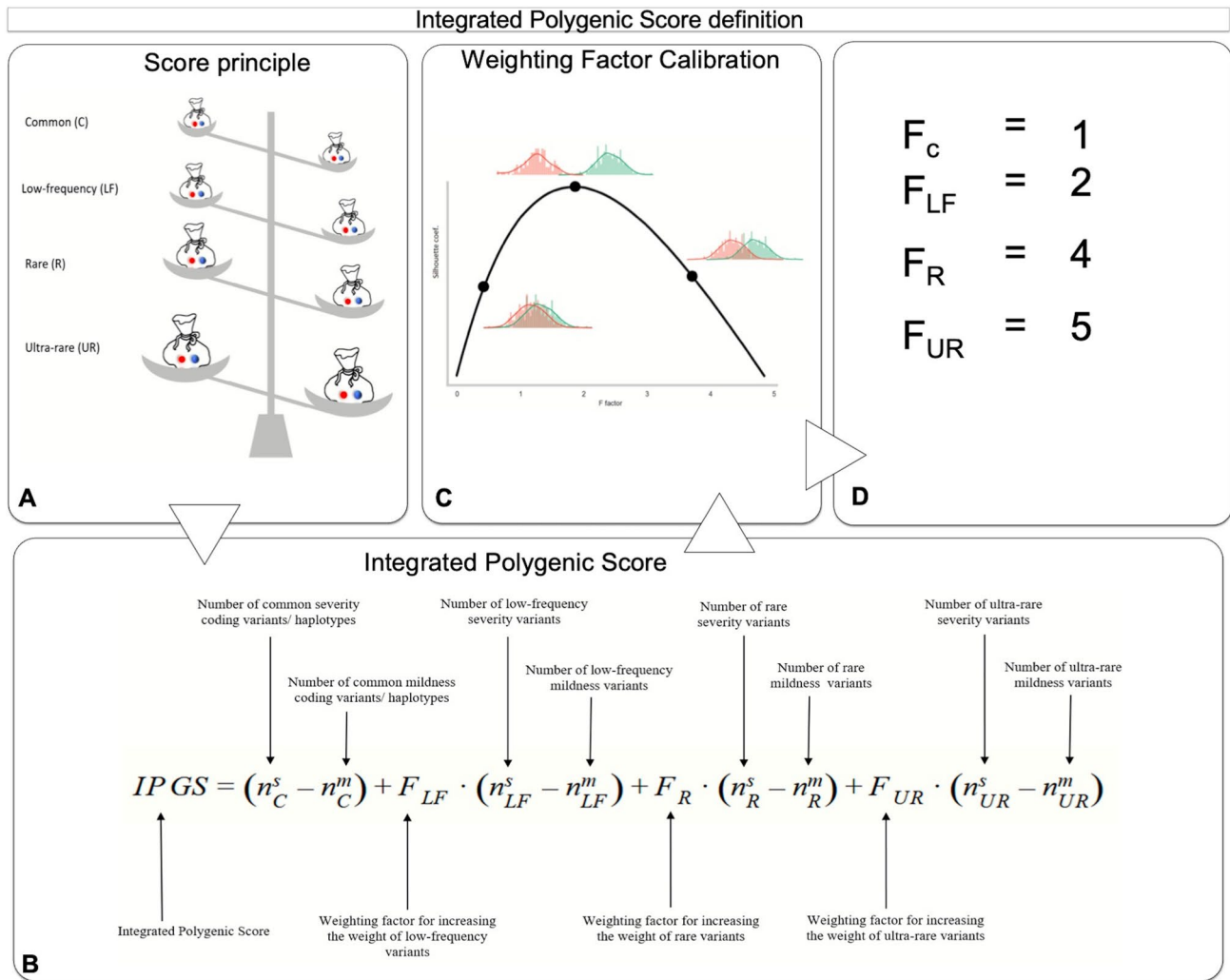
Fig. 3 Integrated PolyGenic Score Definition. **A** The model is based on the comparison of Boolean features of severity versus Boolean features of mildness. **B** Graphic representation of the IPGS formula used for this model. **C** Principle for the calibration of different weighting factors based on the separation of severe and mild cases. **D** The obtained value for low-frequency, rare, and ultra-rare, being $F = 1$ for common variants. Common variants are indicated as common haplotypes since they are intended as combinations of coding variants within a single gene (see Fig. 1C and the Material and methods section)

Discussion

The most robust and traditional method for feature selection in complex disorders is Genome-wide association studies (GWASs). GWAS focuses on common variants only whose effects are small. The method is based on the comparison of about 700,000 genomic SNPs, mostly non-coding, in cases/controls. The coverage of coding SNPs is usually only through imputed data by Linkage Disequilibrium. The method needs ten/hundred of thousands subjects. The missing heritability of this method is rare variants. Until now, 35 loci have been identified for COVID-19 severity (Severe Covid-19 GWAS Group 2020; Pairo-Castineira et al. 2020; COVID-19 Host Genetics Initiative et al. 2021). Another robust and traditional method is the Burden test. Burden

focuses on coding rare variants. The method is based on aggregation on a gene level of the variants and comparison between case and control subjects and likewise GWAS needs ten/hundred of thousands subjects. The missing heritability of this method is common variants. Few genes have been identified (Kosmicki et al. 2021) (WES/WGSHGI Working Group https://www.covid19hg.org/blog/2021-09-27-september-20-2021-meeting/). None of the above methods until now has been able to reach predictivity of COVID-19 severity useful in clinical practice.

The new proposed model takes into account both common and rare variants and has the ability to extract significant COVID-19 features (genes/variants) using a set of relatively low numbers of subjects. Furthermore, employment of oligo-asymptomatic SARS-CoV-2-infected subjects as
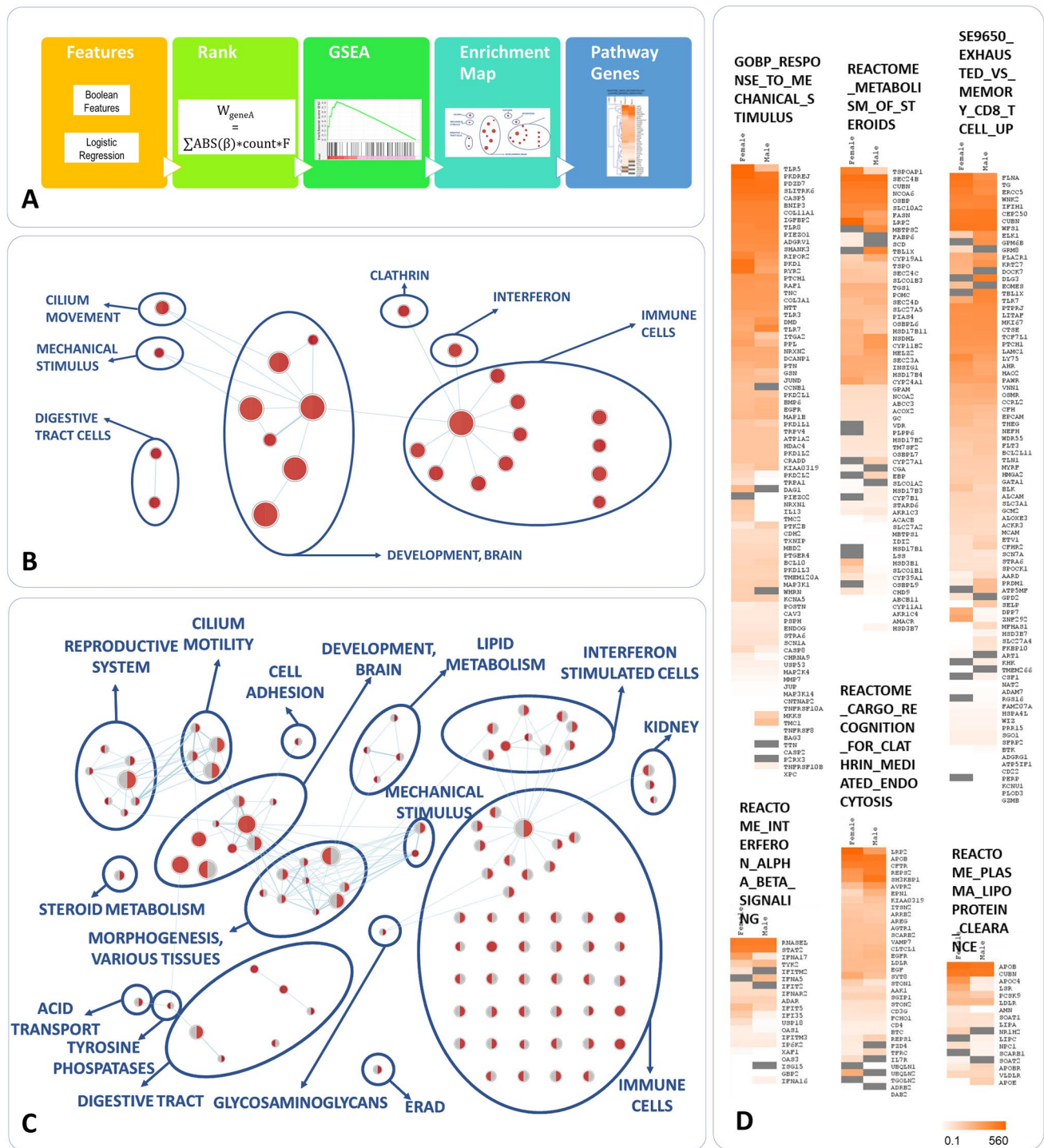
Fig. 4 Pathway enrichment analysis of the genes associated with disease severity/mildness. **A** Workflow of the analysis. Genes corresponding to Boolean features found to be associated at least once were ranked based on a composite score and subjected to Gene Set Enrichment Analysis. Two separate ranked gene lists for females (7317 genes, weight range $3 \times 10-5$-561) and males (7325 genes, weight range $7 \times 10-5$-452) were used. The list of significant pathways was analysed and presented as a similarity network: **B** Similarity network of the pathways with a significant enrichment both in females and males ($p < 0.01$). The size of the circles is proportional to the pathway size. Significance above threshold is indicated by the red color. **C** Similarity network of the pathways with a significant enrichment either in females (red left half of the circles) or males (red right half of the circles) ($p < 0.005$). **D** Heatmaps of the genes belonging to a selection of pathways of interest. The color gradient represents the weight of each gene, calculated and described in methods. Please note high ranking of TLR genes (*TLR5*, *TLR8*, *TLR3* and *TLR7*) in the pathway of Response to Mechanical Stimulus, *CFTR* gene in Recognition for Clathrin-mediated endocytosis, *RNASEL*, *TYK2*, *OAS1* and *OAS3* genes in Interferon alpha–beta signaling. Note also the presence of the relevant pathway of Exhaust vs Memory CD8 T cell Up that also includes *TLR7* gene
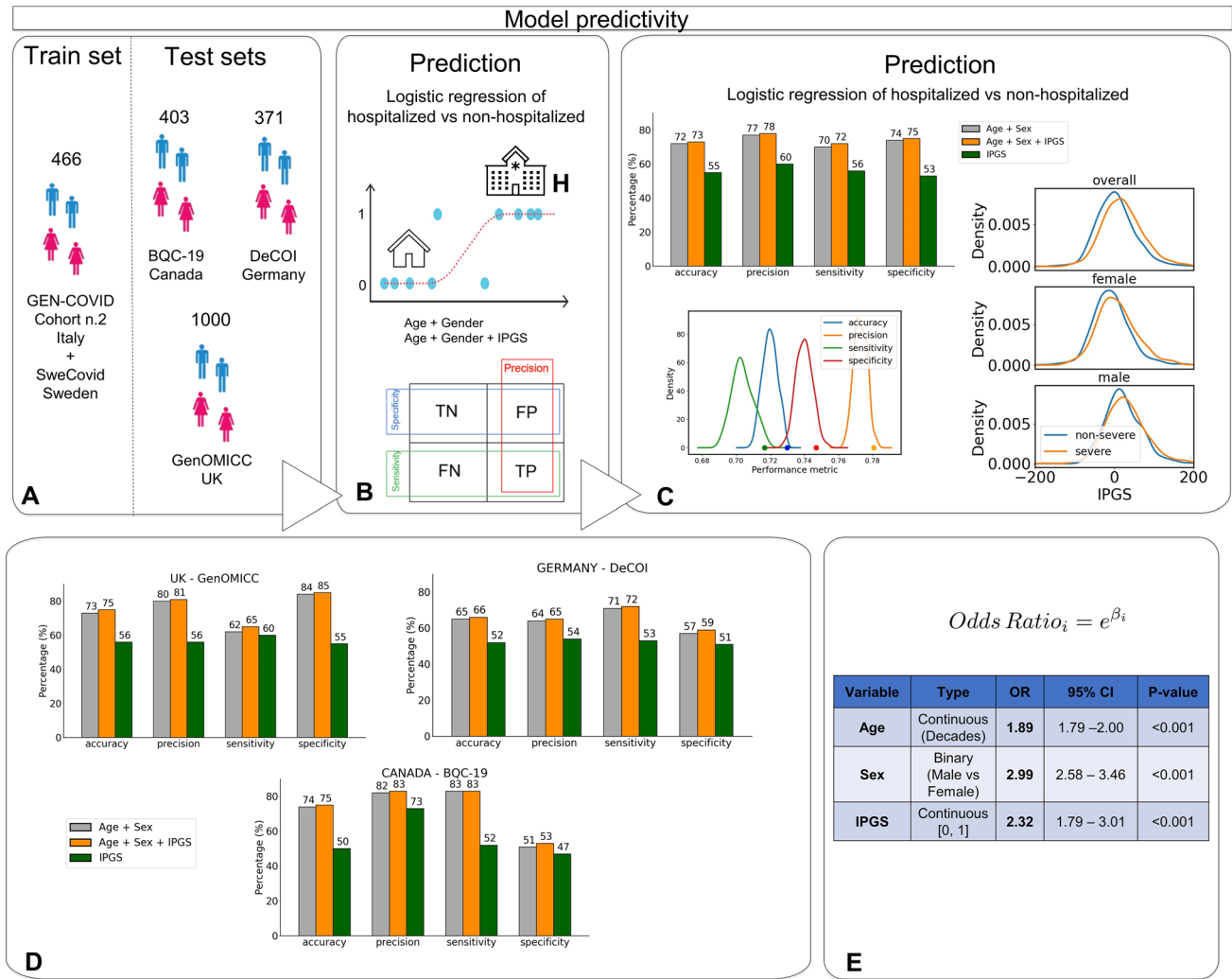
Fig. 5 Model predictivity. **A** The post-Mendelian model was trained using a sample of 466 patients from the GEN-COVID cohort n.2 and Swedish cohort (having cases only) and tested with three additional European cohorts from UK, Germany and Canada. **B** A logistic regression model was used for severity prediction. Severity was defined mainly on the basis of hospitalization versus not hospitalization. Hospitalized cases without respiratory support were included in controls. TN = true negative; TP = true positive; FN = false negative; FP = false positive. **C** When the IPGS is added to age and gender as a regressor, the performances of the model increase: accuracy + 1%, precision + 1%, sensitivity + 2%, specificity + 1%. These increases are statistically significant (p value < 0.05 for accuracy, precision, sensitivity and specificity) with respect to the null distribution obtained by randomizing the IPGS. The performances of the model built with IPGS alone are all above the random guess. In addition, on the right, we reported the distributions of the IPGS for severe and non-severe patients. **D** In the three tested cohorts, when the IPGS is added to age and sex as a regressor, all the performances increase: the accuracy up to + 2%, the precision up to + 1%, the sensitivity up to + 3%, and the specificity up to + 2%. We conclude that IPGS is able to improve prediction of clinical outcome in addition to the well-established powerful factors of age and sex. **E** The univariate logistic regression models fitted on the cohort including both train and test, confirmed that the IPGS is associated with severity with an odds-ratio (OR) of 2.32, while age (continuous in decades) and sex have an OR of 1.89 and 2.99, respectively

controls instead of general population is providing a significant advantage in accuracy. Extracted genes were prioritized by matching with known viral susceptibility genes and transcriptomic data (Supplementary Table 12). Identity match was found for more than 25% of genes, even if the transcriptomic list is far to be representative of whole transcript alteration in the several tissues/organs involved in COVID-19 since most data comes from 2 tissues only (lung and swab). More than 55% of genes have a direct connection (degree 0 and 1 in HGC) with viral susceptibility genes, including those from GWAS loci and the genes in eQTL with them by GTEX analysis (Supplementary Table 13). Finally, up to 89% of genes are biologically significantly connected (p < 0.05) with the list above, supporting the validation of extracted genes.
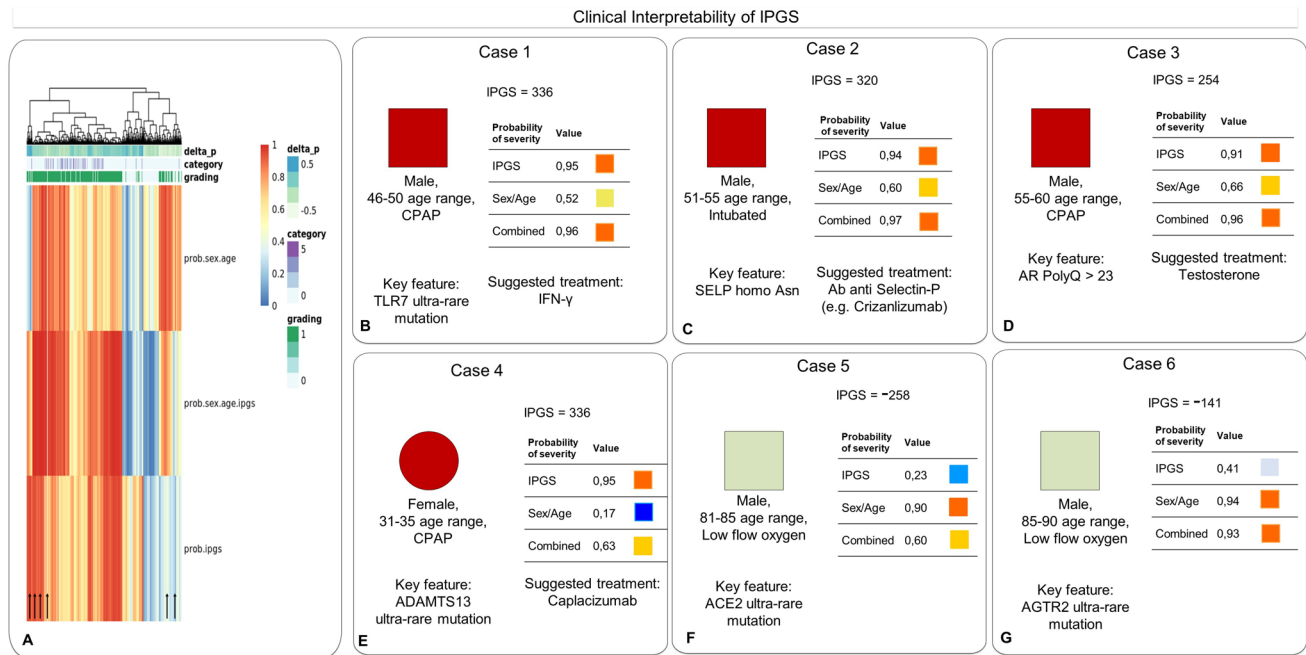
Fig. 6 Clinically interpretability of IPGS. Panel A shows the GEN-COVID cohort dendrogram and heatmaps of the probabilities of severity based on the 3 different models: sex-age alone, IPGS alone and combined model. In the extreme ends of dendrogram (left and right) the probability of severity based on sex-age alone and IPGS alone is highly discordant (different colors). Selected examples corresponding to the arrows are illustrated in panels B-G. In each panel IPGS score, probabilities of severity and key features useful for bedside clinical management are shown. B) Male patient, in the 46–50 age range, treated with CPAP ventilation, tocilizumab, enoxaparin, hydroxychloroquine and lopinavir/ritonavir; no comorbidities except for asthma have been reported. The patient presented a rare *TLR7* mutation that leads to an impaired production of interferon gamma (Made et al. 2020). C) Male patient, in the 51–55 age range, treated with invasive mechanical ventilation, steroids and enoxaparin. He had among comorbidities obesity, anxiety, hypertension and cerebral ischemia. He was found to be homozygous for the *SELP* rs6127 (p.Asp603Asn). Homozygosity of Asparagine in position 603 of Selectin P makes this endothelial protein more prone to clot formation and male patients more prone to COVID-19 thrombosis (Croci et al. 2021). Hence, the rationale for considering as putative adjuvant therapy in the management of similar cases the anti-Selectin P antibodies, a drug already approved for vascular events of sickle cell anemia. D) Male patient, in the 51–55 age range, treated with CPAP ventilation, tocilizumab, steroids, enoxaparin, hydroxychloroquine and lopinavir/ritonavir; no comorbidities except for diabe-

tes. He was found to have the androgen receptor polyQ repeats > 23. The regular function of the androgen receptor is correlated with a beneficial immunomodulatory effect in those male patients in whom the increase in testosterone levels may overcome the receptor resistance. The rationale is to consider giving testosterone to those male subjects who cannot, on their own, raise the levels enough to overcome the receptor resistance due to poly-glutamine stretch longer than 23 repeats (Daga et al. 2021). E) Female patient, in the 31–35 age range, treated with CPAP ventilation and steroids, enoxaparin and azithromycin; no comorbidities except for hypothyroidism. She was a carrier of an ultra-rare mutation in *ADAMTS13*. Impaired function of ADAMTS13 leads to reduced cleavage of von Willebrand factor (vWF) and enhanced clot formation. The effect is enhanced in females and responsible for SARS-CoV-2 related thrombosis. Anti-vWF immunoglobulins would be a putative therapeutic option to consider in similar cases. F-G) examples of low IPGS and related key features. F) Male patient, in the 81–85 age range, treated with low-flow oxygen. No information regarding pharmacological therapy during hospitalization is present. Among comorbidities: diabetes mellitus, congestive heart failure and bowel cancer and steroids. He presented an ultra-rare mutation in *ACE2*. G) Male patient, in the 86–90 age range, treated with low-flow oxygen, steroid, enoxaparin and ceftriaxone plus azithromycin. Among his comorbidities: colon diverticulosis with constipation?, benign prostatic hyperplasia?, anxious-depressive syndrome, sideropenic anemia. He was a carrier of an ultra-rare mutation in *AGTR2*

The importance of combination of rare and low-frequency variants has already been demonstrated to contribute to the prediction of susceptibility in other complex disorders (Marouli et al. 2017; International Multiple Sclerosis Genetics Consortium 2018). Here, we further expand this approach while demonstrating that ultra-rare, rare, low frequency, as well as common variants contribute to the likelihood of developing a severe form of COVID-19. Furthermore, we included in our analyses a calibration of the

relative weight of the variants vis-a-vis their impact on disease severity: a single ultra-rare variant might well by itself cause a severe phenotype of COVID-19, while this is less probable for a common polymorphism, one that is likely to have a markedly less direct effect on protein functionality. We performed a first modelization of COVID-19 genetics using both rare and common variants (Picchiotti et al. 2021). Since feature selection methodologies are generally sensitive to allele frequency, the extraction was performed separately

for rare (MAF < 1/100) and common (MAF > 1/100) variants. However, the methodology revealed the insight that low-frequency variants (MAF from 1 to 5%) were disadvantaged if selected together with common ones. Furthermore, for extracting Mendelian-like genes a threshold of MAF < 0.1% (ultra-rare variants) appeared more effective than MAF < 1% and all mutations in the *TLR7* gene that proved to have loss of function had indeed MAF < 1/1000 (Fallerini et al. 2021a). The model we arrived at, now considers separately ultra-rare, rare, low-frequency, and common variants.

Similar to the classical PRS (Polygenic Risk Score), the proposed IPGS (Integrated PolyGenic Score) may prove reliable for assessing the probability of severe COVID-19 following infection by SARS-CoV-2 (Mars et al. 2020). While PRS is based on common polymorphisms found at the genomic level with the majority of loci potentially conferring risk being not easily interpretable due to the uncertainty of linked genes, IPGS allows immediate biological interpretation because it only includes coding variants. Furthermore, as opposed to PRS, IPGS relies on both polymorphisms and rare variants is capable of differentially weighting features in an indirectly proportional way in respect to frequency, and therefore, to protein impact. Each patient indeed is assigned both a number and the list of her/his common and low-frequency polymorphisms relevant to COVID-19 supported by medically actionable information and of rare and ultra-rare variants conferring either risk of severity or protection from severe disease. Drawing on the entire picture presented through IPGS analysis, personalized adjuvant therapy could be envisaged. At the time of writing, a platform trial based on genetic markers is being discussed with the Italian Medicines Agency (EudraCT Number: 2021-002817-32).

Within 25 reported genomic SNPs demonstrably related to COVID-19 susceptibility/severity, 5 were reported to be in LD with coding variants (COVID-19 Host Genetics Initiative et al. 2021; Covid19hg.org 2021). The model presented here might provide useful information for uncovering the identity of the gene/coding variants responsible for COVID-19 susceptibility/severity linked to these genomic SNPs (Fig. 2D). For example, on chromosome 12, the genes mapping to the locus tagged by rs10774671 (COVID-19 Host Genetics Initiative et al. 2021) are both *OAS1* and *OAS3*. In *OAS3*, the coding variant is an Arginine to Lysine substitution (rs1859330) in high LD (0.8) with the tag SNP. This polymorphism was already associated with viral infection (Tan et al. 2017) based on the presence of Lysine having been shown to lead to a decreased INF-γ production. In *OAS1*, the haplotype (including 4 missense variants: G162S, A352T, R361T, and G397R), the splicing variant 1039-G>A (the reported genomic polymorphism itself), and the truncating mutation T359fs*26 are associated with severity and predicted to impair OAS1 function. Both OAS1 and OAS3 induce RNASEL, which in turn exerts antiviral activity.

Further support for the role of the OAS/RNASEL axis is indicated by the presence of ultra-rare recessive variants.

This innovative approach allowed us to better select genes located on the X chromosome related to COVID-19 that affect males and females in opposite ways (Fig. 2A and Supplementary Tables 3 and 6). Interestingly, many of these genes were previously confirmed or hypothesized to escape the X chromosome inactivation. With respect to these genes, females produce twice the levels of protein in comparison with males. Mutations in hemizygous state in males and heterozygous state in females appear silent until SARS-CoV-2 infection occurs. For example, *TLR7* and *TLR8* are selected for ultra-rare and associated with severity in males and with protection from severe disease in heterozygous females. We know that the activation of *TLR7/8* induces the production of type 1 and type 2 IFN as well as pro-inflammatory cytokines, where the production defect in hemizygous males leads to severe COVID-19. However, an excess of the sensor can also lead to damage from hyperinflammation. Therefore, the condition of carrier females is the more favorable state and has in fact been associated with mild COVID-19 (Subramanian et al. 2006).

Pathway analysis pointed to the relevance of obvious actors in COVID-19 pathology, such as immune cells and interferon signaling, but also to the important role of specific organs (brain, digestive tract, kidney, reproductive system) and functions (metabolism of lipids and steroids). The pathways identified through GSEA analyses reflected the multi-organ nature of the disease. In addition, our analyses reveal new candidate determinants of disease variability. The four pathways linked to cilium motility suggest a role for ciliated cells of the respiratory tract (and possibly others) in antiviral defense. The functionality of the clathrin-mediated endocytosis pathway may likely affect viral entry (Bayati et al. 2021). Likewise, endoplasmic reticulum associated protein degradation (ERAD), which is linked to autophagy and SARS-CoV-2 life-cycle (Reggiori et al. 2010), may also be relevant. Other pathways with a less obvious but potentially interesting role in the disease include cell adhesion and mechanical stimulus signaling.

The strong link between the involved human biological pathways and COVID-19 pathogenicity support the hypothesis that the proposed IPGS equation may contribute significantly to predicting the disease severity of COVID-19. Indeed, an overall significant increase of performance was obtained in comparison with the model based on solely on age and sex. Furthermore, the IPGS is significantly associated with severity, showing an OR of 2.46 after adjustment for age, sex, and comorbidities. This indicates that IPGS is a novel prognostic factor that should be considered in the management of COVID-19 patients.

Modelling precisely the role of the entire range of host genomics affecting disease susceptibility and severity in

COVID-19 is critical to obtaining a complete biological understanding of the aetiology and pathogenicity of COVID-19 as well as other severe complex diseases. The application of IPGS based on Machine Learning principles within a post-Mendelian model allows us to more precisely identify the gene variants at play in COVID-19 as well as their specific roles, individually and in combination. This deep dive into the genetic architecture that allows for, contributes to, or even helps prevent diseases while increasing or decreasing their impact is critical for, and directly translatable into, (personalized) medicines development as well as prevention and treatment protocols. An integrated modelling of genetic variants based on a limited patient cohort, even limited in its geographical spread, may be sufficient for the development of diagnosis, and therapeutics across a wider range of populations. The advantage of this IPGS post-Mendelian model is that it learns and continues to learn as well as being a model from which we can obtain insights on the fundamental architecture of human genomics when confronted with severe and complex diseases.

*GEN-COVID Multicenter Study (*https://sites.google.com/dbm.unisi.it/gen-covid*)* Francesca Montagnani[1,27], Mario Tumbarello[1,27], Ilaria Rancan[1,27], Massimiliano Fabbiani[27], Barbara Rossetti[27], Laura Bergantini[28], Miriana D'Alessandro[28], Paolo Cameli[28], David Bennett[28], Federico Anedda[29], Simona Marcantonio[29], Sabino Scolletta[29], Federico Franchi[29], Maria Antonietta Mazzei[30], Susanna Guerrini[30], Edoardo Conticini[31], Luca Cantarini[31], Bruno Frediani[31], Danilo Tacconi[32], Chiara Spertilli Raffaelli[32], Marco Feri[33], Alice Donati[33], Raffaele Scala[34], Luca Guidelli[34], Genni Spargi[35], Marta Corridi[35], Cesira Nencioni[36], Leonardo Croci[36], Gian Piero Caldarelli[37], Maurizio Spagnesi[38], Davide Romani[38], Paolo Piacentini[38], Maria Bandini[38], Elena Desanctis[38], Silvia Cappelli[38], Anna Canaccini[39], Agnese Verzuri[39], Valentina Anemoli[39], Manola Pisani[39], Agostino Ognibene[40], Alessandro Pancrazzi[40], Maria Lorubbio[40], Massimo Vaghi[41], Antonella D'Arminio Monforte[42], Federica Gaia Miraglia[42], Mario U. Mondelli[43,44],, Massimo Girardis[45], Sophie Venturelli[45], Stefano Busani[45], Andrea Cossarizza[46], Andrea Antinori[47], Alessandra Vergori[47], Arianna Emiliozzi[47], Stefano Rusconi[48,49], Matteo Siano[49], Arianna Gabrieli[49], Agostino Riva[48,49], Daniela Francisci[50], Elisabetta Schiaroli[50], Francesco Paciosi[50], Andrea Tommasi[50], Pier Giorgio Scotton[51], Francesca Andretta[51], Sandro Panese[52], Stefano Baratti[52], Renzo Scaggiante[53], Francesca Gatti[53], Saverio Giuseppe Parisi[54], Francesco Castelli[55], Eugenia Quiros-Roldan[55], Melania Degli Antoni[55], Isabella Zanella[56,57], Matteo Della Monica[58], Carmelo Piscopo[58], Mario Capasso[59,60,61], Roberta Russo[59,60], Immacolata Andolfo[59,60], Achille Iolascon[59,60], Giuseppe Fiorentino[62], Massimo Carella[63], Marco Castori[63], Filippo Aucella[64], Pamela Raggi[65], Rita Perna[65], Matteo Bassetti[66,67], Antonio Di Biagio[66,67], Maurizio Sanguinetti[68,69], Luca Masucci[68,69], Alessandra Guarnaccia[68], Serafina Valente[70], Oreste De Vivo[70], Gabriella Doddato[1,2], Rossella Tita[5], Annarita Giliberti[1,2], Maria Antonietta Mencarelli[5], Caterina Lo Rizzo[5], Anna Maria Pinto[5], Valentina Perticaroli[1,2,5], Francesca Ariani[1,2,5], Miriam Lucia Carriero[1,2], Laura Di Sarno[1,2], Diana Alaverdian[1,2], Elena Bargagli[28], Marco Mandalà[71], Alessia Giorli[71], Lorenzo Salerni[71], Patrizia Zucchi[72], Pierpaolo Parravicini[72], Elisabetta Menatti[73], Tullio Trotta[74], Ferdinando Giannattasio[74], Gabriella Coiro[74], Fabio Lena[75], Leonardo Gianluca Lacerenza[75], Domenico A. Coviello[76], Cristina Mussini[77], Enrico Martinelli[78], Sandro Mancarella[79], Luisa Tavecchia[79], Mary Ann Belli[79], Lia Crotti[80,81,82,83,84], Gianfranco Parati[80,81], Maurizio Sanarico[85], Francesco Raimondi[86], Filippo Biscarini[87], Alessandra Stella[87], Marco Rizzi[88], Franco Maggiolo[88], Diego Ripamonti[88], Claudia Suardi[89], Tiziana Bachetti[90], Maria Teresa La Rovere[91], Simona Sarzi-Braga[92], Maurizio Bussotti[93], Katia Capitani[94], Simona Dei[95], Sabrina Ravaglia[96], Rosangela Artuso[97], Elena Andreucci[97], Giulia Gori[97], Angelica Pagliazzi[97], Erika Fiorentini[97], Antonio Perrella[98], Francesco Bianchi[98], Paola Bergomi[99], Emanuele Catena[99], Riccardo Colombo[99], Sauro Luchi[100], Giovanna Morelli[100], Paola Petrocelli[100], Sarah Iacopini[100], Sara Modica[100], Silvia Baroni[101], Francesco Vladimiro Segala[102], Francesco Menichetti[103], Marco Falcone[103], Giusy Tiseo[103], Chiara Barbieri[103], Tommaso Matucci[103], Davide Grassi[104], Claudio Ferri[105], Franco Marinangeli[106], Francesco Brancati[107], Antonella Vincenti[108], Valentina Borgo[108], Lombardi Stefania[108], Mirco Lenzi[108], Massimo Antonio Di Pietro[109], Francesca Vichi[109], Benedetta Romanin[109], Letizia Attala[109], Cecilia Costa[109], Andrea Gabbuti[109], Menè Roberto[80,81], Umberto Zuccon[110], Lucia Vietri[110], Stefano Ceri[111], Pietro Pinoli[111], Patrizia Casprini[112], Giuseppe Merla[113,114], Gabriella Maria Squeo[113], Marcello Maffezzoni[115], Raffaele Bruno[116,117], Marco Vecchia[116], Marta Colaneri[116], Serena Ludovisi[118]. *WES/WGS working group within the HGI (*https://www.covid19hg.org/projects/*)* The genetic predisposition to severe COVID-19 (SweCovid):

Yanara Marincevic-Zuniga[119], Jessica Nordlund[119], Tomas Luther[16], Anders Larsson[120], Katja Hanslin[16], Anna Gradin[16], Sarah Galien[16], Sara Bulow Anderberg[16], Jacob Rosén[16], Sten Rubertsson[16], Hugo Zeberg[15],, Robert Frithiof[16], Miklós Lipcsey[16,17], Michael Hultström[16,18]. *GenOMICC Consortium* GenOMICC co-investigators: Sara Clohisey[25], Peter Horby[121], Johnny Millar[25], Julian Knight[122], Hugh Montgomery[123], David Maslove[124], Lowell Ling[125], Alistair Nichol[126], Charlotte Summers[127], Tim Walsh[26], Charles Hinds[128], Malcolm G. Semple[129,130], Peter J.M. Openshaw[131,132], Manu Shankar-Hari[133], Antonia Ho[134], Danny McAuley[135,136], Chris Ponting[24], Kathy Rowan[137], J. Kenneth Baillie[24,25,26]. Central management and laboratory team: Fiona Griffiths[25], Wilna Oosthuyzen[25] Jen Meikle[25], Paul Finernan[25], James Furniss[25], Ellie Mcmaster[25], Andy Law[25], Sara Clohisey[25], J. Kenneth Baillie[24,25,26], Trevor Paterson[25], Tony Wackett[25], Ruth Armstrong[25], Lee Murphy[138], Angie Fawkes[138], Richard Clark[138], Audrey Coutts[138], Lorna Donnelly[138], Tammy Gilchrist[138], Katarzyna Hafezi[138], Louise Macgillivray[138], Alan Maclean[138], Sarah McCafferty[138], Kirstie Morrice[138], Jane Weaver[25], Ceilia Boz[25], Ailsa Golightly[25], Mari Ward[25], Hanning Mal[25], Helen Szoor-McElhinney[25], Adam Brown[25], Ross Hendry[25], Andrew Stenhouse[25], Louise Cullum[25], Dawn Law[25], Sarah Law[25], Rachel Law[25], Max Head Fourman[25], Maaike Swets[25], Nicky Day[25], Filip Taneski[25], Esther Duncan[25], Marie Zechner[25], Nicholas Parkinson[25]. Data analysis team: Erola Pairo-Castineira[24,25], Sara Clohisey[25], Lucija Klaric[24], Andrew D. Bretherick[24], Konrad Rawlik[25], Dorota Pasko[139], Susan Walker[139], Nick Parkinson[25], Max Head Fourman[25], Clark D. Russell[25,140], James Furniss[25], Anne Richmond[24], Elvina Gountouna[141], David Harrison[137], Bo Wang[24], Yang Wu[142], Alison Meynert[24], Athanasios Kousathanas[139], Loukas Moutsianas[139], Zhijian Yang[143], Ranran Zhai[143], Chenqing Zheng[143], Graeme Grimes[24], Jonathan Millar[25], Barbara Shih[25], Marie Zechner[25], Jian Yang[144,145], Xia Shen[143,146,147], Chris P. Ponting[2], Albert Tenesa[2,25,146], Kathy Rowan[137], Andrew Law[25], Veronique Vitart[24], James F. Wilson[24,146], J. Kenneth Baillie[24,25,26]. Barts Health NHS Trust, London, UK: D. Collier[149], S. Wood[148], A. Zak[148], C. Borra[148], M. Matharu[148], P. May[148], Z. Alldis[148], O. Mitchelmore[148], R. Bowles[148], A. Easthorpe[148], F. Bibi[148], I. Lancoma-Malcolm[148], J. Gurasashvili[148], J. Pheby[148], J. Shiel[148], M. Bolton[148], M. Patel[148], M. Taylor[148], O. Zongo[148], P. Ebano[148], P. Harding[148], R. Astin-Chamberlain[148], Y. Choudhury[148], A. Cox[148], D. Kallon[148], M. Burton[148], R. Hall[148], S. Blowes[148], Z. Prime[148], J. Biddle[148], O. Prysyazhna[148], T. Newman[148], C. Tierney[148], J. Kassam[148]. Guys and St Thomas' Hospital, London, UK: M. Shankar-Hari[149], M. Ostermann[149], S. Campos[149], A. Bociek[149], R. Lim[149], N. Grau[149], T. O. Jones[149], C. Whitton[149], M. Marotti[40], G. Arbane[149]. James Cook University Hospital, Middlesburgh, UK: S. Bonner[150], K. Hugill[150], J. Reid[150]. The Royal Liverpool University Hospital, Liverpool, UK: I. Welters[151], V. Waugh[151], K. Williams[151], D. Shaw[151], J. Fernandez Roman[151], M. Lopez Martinez[151], E. Johnson[151], A. Waite[151], B. Johnson[151], O. Hamilton[151], S. Mulla[151]. King's College Hospital, London, UK: M. McPhail[152], J. Smith[152]. Royal Infirmary of Edinburgh, Edinburgh, UK: J. K. Baillie[23,24,25], L. Barclay[153], D. Hope[153], C. McCulloch[153], L. McQuillan[153], S. Clark[153], J. Singleton[153], K. Priestley[153], N. Rea[153], M. Callaghan[153], R. Campbell[153], G. Andrew[153], L. Marshall[153]. John Radcliffe Hospital, Oxford, UK: S. McKechnie[154], P. Hutton[154], A. Bashyal[154], N. Davidson[154]. Addenbrooke's Hospital, Cambridge, UK: C. Summers[155], P. Polgarova[155], K. Stroud[155], N. Pathan[155], K. Elston[155], S. Agrawal[155]. Morriston Hospital, Swansea, UK: C. Battle[156], L. Newey[156], T. Rees[156], R. Harford[156], E. Brinkworth[156], M. Williams[156], C. Murphy[156]. Ashford and St Peter's Hospital, Surrey, UK: I. White[157], M. Croft[157]. Royal Stoke University Hospital, Staffordshire, UK: N. Bandla[158], M. Gellamucho[158], J. Tomlinson[158], H. Turner[158], M. Davies[158], A. Quinn[158], I. Hussain[158], C. Thompson[158], H. Parker[158], R. Bradley[158], R. Griffiths[158]. Queen Elizabeth Hospital, Birmingham, UK: J. Scriven[159], J. Gill[159]. Glasgow RoyalInfirmary, Glasgow, UK: A. Puxty[160], S. Cathcart[160], D. Salutous[160], L. Turner[160], K. Duffy[160], K. Puxty[160]. Kingston Hospital, Surrey, UK: A. Joseph[161],

R. Herdman-Grant[161], R. Simms[161], A. Swain[161], A. Naranjo[161], R. Crowe[161], K. Sollesta[161], A. Loveridge[161], D. Baptista[161], E. Morino[161]. The Tunbridge Wells Hospital and Maidstone Hospital, Kent, UK: M. Davey[162], D. Golden[162], J. Jones[162]. North Middlesex University Hospital NHS trust, London, UK: J. Moreno Cuesta[163], A. Haldeos[163], D. Bakthavatsalam[163], R. Vincent[163], M. Elhassan[163], K. Xavier[163], A. Ganesan[163], D. Purohit M. Abdelrazik[163]. Bradford Royal Infirmary, Bradford, UK: J. Morgan[164], L. Akeroyd[164], S. Bano[164], D. Warren[164], M. Bromley[164], K. Sellick[164], L. Gurr[164], B. Wilkinson[164], V. Nagarajan[164], P. Szedlak[164]. Blackpool Victoria Hospital, Blackpool, UK: J. Cupitt[165], E. Stoddard[165], L. Benham[165], S. Preston[165], N. Slawson[165], Z. Bradshaw[165], J. brown[165], M. Caswell[165], SMelling[165]. Countess of Chester Hospital, Chester, UK: P. Bamford[165], M. Faulkner[166], K. Cawley[166], H. Jeffrey[166], E. London[166], H. Sainsbury[166], I. Nagra[166], F. Nasir[166], Ce Dunmore[166], R. Jones[166], A. Abraheem[166], M. Al-Moasseb[166], R. Girach[166]. Wythenshawe Hospital, Manchester, UK: C. Brantwood[167], P. Alexander[167], J. Bradley-Potts[167], S. Allen[167], T. Felton[167]. St George's Hospital, London, UK: S. Manna[168], S. Farnell-Ward[168], S. Leaver[168], J. Queiroz[168], E. Maccacari[168], D. Dawson[168], C. Castro Delgado[168], R. Pepermans Saluzzio[168], O. Ezeobu[168], L. Ding[168], C. Sicat[168], R. Kanu[168], G. Durrant[168], J. Texeira[168], A. Harrison[168], T. Samakomva[168]. Good Hope Hospital, Birmingham, UK: J. Scriven[169], H. Willis[169], B. Hopkins[169], L. Thrasyvoulou[169]. Stepping Hill Hospital, Stockport, UK: M. Jackson[170], A. Zaki[170], C. Tibke[170], S. Bennett[170], W. Woodyatt[170], A. Kent[170], E. Goodwin[170]. Manchester Royal Infirmary, Manchester, UK: C. Brandwood[171], R. Clark[171], L. Smith[171]. Royal Alexandra Hospital, Paisley, UK: K. Rooney[172], N. Thomson[172], N. Rodden[172], E. Hughes[172], D. McGlynn[172], C. Clark[172], P. Clark[172], L. Abel[172], R. Sundaram[172], L. Gemmell[172], M. Brett[172], J. Hornsby[172], P. MacGoey[172], R. Price[172], B. Digby[172], P. O'Neil[172], P. McConnell[172], P. Henderson[172]. Queen Elizabeth University Hospital, Glasgow, UK: S. Henderson[173], M. Sim[173], S. Kennedy-Hay[173], C. McParland[173], L. Rooney[173], N. Baxter[173]. Queen Alexandra Hospital, Portsmouth, UK: D. Pogson[174], S. Rose[174], Z. Daly[174], L. Brimfield[174]. BHRUT (Barking Havering)—Queens Hospital and King George Hospital, Essex, UK: M. K. Phull[175], M. Hussain[175], T. Pogreban[175], L. Rosaroso[175], E. Salciute L. Grauslyte[175]. University College Hospital, London, UK: D. Brealey[176], E. Wraith[176], N. MacCallum[176], G. Bercades[176], I. Hass[176], D. Smyth[176], A. Reyes[176], G. Martir[176]. Royal Victoria Infirmary, Newcastle Upon Tyne, UK. I. D. Clement[177], K. Webster[177], C. Hays[177], A. Gulati[177]. Western Sussex Hospitals, West Sussex, UK. L. Hodgson[178], M. Margarson[178], R. Gomez[178], Y. Baird[178], Y. Thirlwall[178], L. Folkes[178], A. Butler[178], E. Meadows[178], S. Moore[178], D. Raynard[178], H. Fox[178], L. Riddles[178], K. King[178], S. Kimber[178], G. Hobden[178], A. McCarthy[178], V. Cannons[178], I. Balagosa[178], I. Chadbourn[178], A. Gardner[178]. Salford Royal Hospital, Manchester, UK. D. Horner[179], D. McLaughlanv[179], B. Charles[179], N. Proudfoot[179], T. Marsden[179], L. Mc Morrow[179], B. Blackledge[179], J. Pendlebury[179], A. Harvey[179], E. Apetri[179], C. Basikolo[179], L. Catlow[179], R. Doonan[179], K. Knowles[179], S. Lee[179], D. Lomas[179], C. Lyons[179], J. Perez[179], M. Poulaka[179], M. Slaughter[179], K. Slevin[179], M. Taylor[179], V. Thomas[179], D. Walker[179], J. Harris[179]. The Royal Oldham Hospital, Manchester, UK. A. Drummond[180], R. Tully[180], J. Dearden[180], J. Philbin[180], S. Munt[180], C. Rishton[180], G. O'Connor[180], M. Mulcahy[180], E. Dobson[180], J. Cuttler[180], M. Edward[180]. Pinderfields General Hospital, Wakefield, UK: A. Rose[181], B. Sloan[181], S. Buckley[181], H. Brooke[181], E. Smithson[181], R. Charlesworth[181], R. Sandu[181], M. Thirumaran[181], V. Wagstaff[181], J. Cebrian Suarez[181]. Basildon Hospital, Basildon, UK: A. Kaliappan[182], M. Vertue[182], A. Nicholson[182], J. Riches[182], A. Solesbury[182], L. Kittridge[182], M. Forsey[182], G. Maloney[182]. University Hospital of Wales, Cardiff, UK: J. Cole[183], M. Davies[183], R. Davies[183], H. Hill[183], E. Thomas[183], A.Williams[183], D. Duffin[183], B. Player[183]. Broomfield Hospital, Chelmsford, UK: J. Radhakrishnan[184], S. Gibson[184], A. Lyle[184], F. McNeela[184]. Royal Brompton Hospital, London, UK: B. Patel[185], M. Gummadi[185], G. Sloane[185], N. Dormand[185], S. Salmi[185], Z.

Farzad[185], D. Cristiano[185], K. Liyanage[185], V. Thwaites[185], M. Varghese[185]. Nottingham University Hospital, Nottingham, UK: M. Meredith[186]. Royal Hallamshire Hospital and Northern General Hospital, Sheffield, UK. G. Mills[187], J. Willson[187], K. Harrington[187], B. Lenagh[187], K. Cawthron[187], S. Masuko[187], A. Raithatha[187], K. Bauchmuller[187], N. Ahmad[187], J. Barker[187], Y. Jackson[187], F. Kibutu[187], S. Bird[187]. Royal Hampshire County Hospital, Hampshire, UK. G. Watson[188], J. Martin[188], E. Bevan[188], C. Wrey Brown[188], D. Trodd[188]. Queens Hospital Burton, Burton-On-Trent, UK: K. English[189], G. Bell[189], L. Wilcox[189], A. Katary[189], New Cross Hospital, Wolverhampton, UK: S. Gopal[190], V. Lake[190], N. Harris[190], S. Metherell[190], E. Radford[190]. Heartlands Hospital, Birmingham, UK: J. Scriven[191], F. Moore[191], H. Bancroft[191], J. Daglish[191], M. Sangombe[191], M. Carmody[191], J. Rhodes[191], M. Bellamy[191]. Walsall Manor Hospital, Walsall, UK: A. Garg[192], A. Kuravi[192], E. Virgilio[192], P. Ranga[192], J. Butler[192], L. Botfield[192], C. Dexter[192], J. Fletcher[192]. Stoke Mandeville Hospital, Buckinghamshire, UK: P. Shanmugasundaram[193], G. Hambrook[193], I. Burn[193], K. Manso[193], D. Thornton[193], J. Tebbutt[193], R. Penn[193]. Sandwell General Hospital, Birmingham, UK: J. Hulme[194], S. Hussain[194], Z. Maqsood[194], S. Joseph[194], J. Colley[194], A. Hayes[194], C. Ahmed[194], R. Haque[194], S. Clamp[194], R. Kumar[194], M. Purewal[194], B. Baines[194]. Royal Berkshire NHS Foundation Trust, Berkshire, UK: M. Frise[195], N. Jacques[195], H. Coles[195], J. Caterson[195], S. Gurung Rai[195], M. Brunton[195], E. Tilney[195], L. Keating[195], A. Walden[195]. Charing Cross Hospital, St Mary's Hospital and Hammersmith Hospital, London, UK: D. Antcliffe[196], A. Gordon[196], M. Templeton[196], R. Rojo[196], D. Banach[196], S. Sousa Arias[196], Z. Fernandez[196], P. Coghlan[196]. Dumfries and Galloway Royal Infirmary, Dumfries, UK: D. Williams[197], C. Jardine[197]. Bristol Royal Infirmary, Bristol, UK: J. Bewley[198], K. Sweet[198], L. Grimmer[198], R. Johnson[198], Z. Garland[198], B. Gumbrill[198]. Royal Sussex County Hospital, Brighton, UK: C. Phillips[199], L. Ortiz-Ruiz de Gordoa[199], E. Peasgood[199]. Whiston Hospital, Prescot, UK: A. Tridente[200], K. Shuker S. Greer[200]. Royal Glamorgan Hospital, Cardiff, UK: C. Lynch[201], C. Pothecary[201], L. Roche[201], B. Deacon[201], K. Turner[201], J. Singh[201], G. Sera Howe[201]. King's Mill Hospital, Nottingham, UK: P. Paul[202], M. Gill[202], I. Wynter[202], V. Ratnam[202], S. Shelton[202]. Fairfield General Hospital, Bury, UK: J. Naisbitt[203], J. Melville[203]. Western General Hospital, Edinburgh, UK: R. Baruah[204], S. Morrison[204]. Northwick Park Hospital, London, UK: A. McGregor[205], V. Parris[205], M. Mpelembue[205], S. Srikaran[205], C. Dennis[205], A. Sukha[205]. Royal Preston Hospital, Preston, UK: A. Williams[206], M. Verlande[206]. Royal Derby Hospital, Derby, UK: K. Holding[207], K. Riches[207], C. Downes[207], C. Swan[207]. Sunderland Royal Hospital, Sunderland, UK: A. Rostron[208], A. Roy[208], L. Woods[208], S. Cornell[208], F. Wakinshaw[208]. Royal Surrey County Hospital, Guildford, UK. B. Creagh-Brown[209], H. Blackman[209], A. Salberg[209], E. Smith[209], S. Donlon[209], S. Mtuwa[209], N. Michalak-Glinska[209], S. Stone[209], C. Beazley[209], V. Pristopan[209]. Derriford Hospital, Plymouth, UK: N. Nikitas[210], L. Lankester[210], C. Wells[210]. Croydon University Hospital, Croydon, UK: A. S. Raj[211], K. Fletcher[211], R. Khade[211], G. Tsinaslanidis[211]. Victoria Hospital, Kirkcaldy, UK: M. McMahon[212], S. Fowler[212], A. McGregor[212], T. Coventry[212]. Milton Keynes University Hospital, Milton Keynes, UK: R. Stewart[213], L. Wren[213], E. Mwaura[213], L. Mew[213], A. Rose[213], D. Scaletta[213], F. Williams[213]. Barnsley Hospital, Barnsley, UK: K. Inweregbu[214], A. Nicholson[214], N. Lancaster[214], M. Cunningham[214], A. Daniels[214], L. Harrison[214], S. Hope[214], S. Jones[214], A. Crew[214], G. Wray[214], J. Matthews[214], R. Crawley[214]. York Hospital, York, UK: J. Carter[215], I. Birkinshaw[215], J. Ingham[215], Z. Scott[215], K. Howard[215], R. Joy[215], S. Roche[215]. University Hospital of North Tees, Stockton on Tees, UK: M. Clark[216], S. Purvis[216]. University Hospital Wishaw, Wishaw, UK: A. Morrison[217], D. Strachan[217], M. Taylor[217], S. Clements[217], K. Black[217]. Whittington Hospital, London, UK: C. Parmar[218], A. Altabaibeh[218], K. Simpson[218], L. Mostoles[218], K. Gilbert[218], L. Ma[218], A. Alvaro[218]. Southmead Hospital, Bristol, UK: M. Thomas[219], B. Faulkner[219], R. Worner[219], K. Hayes[219], E. Gendall[219], H. Blakemore[219], B. Borislavova[219], E. Goff[219].

The Royal Papworth Hospital, Cambridge, UK: A. Vuylsteke[220], L. Mwaura[220], J. Zamikula[220], L. Garner[220], A. Mitchell[220], S. Mepham[220], L. Cagova[220], A. Fofano[220], H. Holcombe[220], K. Praman[220]. Royal Gwent Hospital, Newport, UK: T. Szakmany[221], A. E. Heron[221], S. Cherian[221], S. Cutler[221], A. Roynon-Reed[221]. Norfolk and Norwich University hospital (NNUH), Norwich, UK: G. Randell[222], K. Convery[222], K. Stammers D. Fottrell-Gould[222], L. Hudig[222], J. Keshetprice[222]. Great Ormond St Hospital and UCL Great Ormond St Institute of Child Health NIHR Biomedical Research Centre, London, UK: M. Peters[223], L. O'Neill[223], S. Ray[223], H. Belfield[223], T. McHugh[223], G. Jones[223], O. Akinkugbe[223], A. Tomas[223], E. Abaleke[223], E. Beech[223], H. Meghari[223], S. Yussuf[223], A. Bamford[223]. Airedale General Hospital, Keighley, UK: B. Hairsine[224], E. Dooks[224], F. Farquhar[224], S. Packham[224], H. Bates[224], C. McParland[224], L. Armstrong[224]. Aberdeen Royal Infirmary, Aberdeen, UK: C. Kaye[225], A. Allan[225], J. Medhora[225], J. Liew[225], A. Botello[225], F. Anderson[225]. Southampton General Hospital, Southampton, UK: R. Cusack[226], H. Golding[226], K. Prager[226], T. Williams[226], S. Leggett[226], K. Golder[226], M. Male[226], O. Jones[226], K. Criste[226], M. Marani[226]. Russell's Hall Hospital, Dudley, UK: V. Anumakonda[227], V. Amin[227], K. Karthik[227], R. Kausar[227], E. Anastasescu[227], K. Reid[227], M. Jacqui[227]. Rotherham General Hospital, Rotherham, UK: A. Hormis[228], R. Walker[228], D. Collier[228]. North Manchester General Hospital, Manchester, UK: T. Duncan[229], A. Uriel[229], A. Ustianowski[229], H. T-Michael[229], M. Bruce[229], K. Connolly[229], K. Smith[229]. Basingstoke and North Hampshire Hospital, Basingstoke, UK: R. Partridge[230], D. Griffin[230], M. McDonald[230], N. Muchenje[230]. Royal Free Hospital, London, UK: D. Martin[231], H. Filipe[231], C. Eastgate[231], C. Jackson[231]. Hull Royal Infirmary, Hull, UK: A. Gratrix[232], L. Foster[232], V. Martinson[232], E. Stones[232], Caroline Abernathy[232], P. Parkinson[232]. Harefield Hospital, London, UK: A. Reed[233], C. Prendergast[233], P. Rogers[233], M. Woodruff[233], R. Shokkar[233], S. Kaul[233], A. Barron[233], C. Collins[233]. Chesterfield Royal Hospital Foundation Trust, Chesterfield, UK: S. Beavis[234], A. Whileman[234], K. Dale[234], J. Hawes[234], K. Pritchard[234], R. Gascoyne[234], L. Stevenson[234]. Barnet Hospital, London, UK: R. Jha[235], L. Lim[235], V. Krishnamurthy[235]. Aintree University Hospital, Liverpool, UK: R. Parker[236], I. Turner-Bone[236], L. Wilding[236], A. Reddy[236]. St James's University Hospital and Leeds General Infirmary, Leeds, UK: S. Whiteley[237], E. Wilby[237], C. Howcroft[237], A. Aspinwall[237], S. Charlton[237], B. Ogg[237]. Glan Clwyd Hospital, Bodelwyddan, UK: D. Menzies[238], R. Pugh[238], E. Allan[238], R. Lean[238], F. Davies[238], J. Easton[238], X. Qiu[238], S. Kumar[238], K. Darlington[238]. University Hospital Crosshouse, Kilmarnock, UK: G. Houston[239], P. O'Brien[239], T. Geary[239], J. Allan[239], A. Meikle[239]. Royal Bolton Hospital, Bolton, UK: G. Hughes[240], M. Balasubramaniam[240], S. Latham[240], E. McKenna[240], R. Flanagan[240]. Princess of Wales Hospital, Llantrisant, UK: S. Sathe[24], E. Davies[24], L. Roche[241]. Pilgrim Hospital, Lincoln, UK: M. Chablani[242], A. Kirkby[242], K. Netherton[242], S. Archer[242]. Northumbria Healthcare NHS Foundation Trust, North Shields, UK: B. Yates[243], C. Ashbrook-Raby[243]. Ninewells Hospital, Dundee, UK: S. Cole[244], M. Casey[244], L. Cabrelli[244], S. Chapman[244], M. Casey[244], P. Austin[244], A. Hutcheon[244], C. Whyte[244], C. Almaden-Boyle[244]. Lister Hospital, Stevenage, UK: N. Pattison[245], C. Cruz[245]. Bedford Hospital, Bedford, UK: A. Vochin[246], H. Kent[246], A. Thomas[246], S. Murdoch[246], B. David[246], M. Penacerrada[246], G. Lubimbi[246], V. Bastion[246], R. Wulandari[246], J. Valentine[246], D. Clarke[246]. Royal United Hospital, Bath, UK: A. Serrano-Ruiz[247], S. Hierons[247], L. Ramos[247], C. Demetriou[247], S. Mitchard[247], K. White[247]. Royal Bournemouth Hospital, Bournemouth, UK: N. White[248], S. Pitts[248], D. Branney[248], J. Frankham[248]. The Great Western Hospital, Swindon, UK: M. Watters[249], H. Langton[249], R. Prout[249]. Watford General Hospital, Watford, UK: V. Page[250], T. Varghes[250]. University Hospital North Durham, Darlington, UK: A. Cowton[251], A.Kay[251], K. Potts[251], M. Birt[251], M. Kent[251], A. Wilkinson[251]. Tameside General Hospital, Ashton Under Lyne, UK: E. Jude[252], V. Turner[252], H. Savill[252], J. McCormick[252], M. Clark[252], M. Coulding[252], S. Siddiqui[252], O. Mercer[252], H. Rehman[252],

D. Potla[252]. Princess Royal Hospital, Telford and Royal Shrewsbury Hospital, Shrewsbury, UK: N. Capps[253], D. Donaldson[253], J. Jones[253], H. Button[253], T. Martin[253], K. Hard[253], A. Agasou[253], L. Tonks[253], T. Arden[253], P. Boyle[253], M. Carnahan[253], J. Strickley[253], C. Adams[253], D. Childs[253], R. Rikunenko[253], M. Leigh[253], M. Breekes[253], R. Wilcox[253], A. Bowes[253], H. Tiveran[253], F. Hurford[253], J. Summers[253], A. Carter[253], Y. Hussain[253], L. Ting[253], A. Javaid[253], N. Motherwell[253], H. Moore[253], H. Millward[253], S. Jose[253], N. Schunki[253], A. Noakes[253], C. Clulow[253]. Arrowe Park Hospital, Wirral, UK: G. Sadera[254], R. Jacob[254], C. Jones[254]. The Queen Elizabeth Hospital, King's Lynn, UK: M. Blunt[255], Z. Coton[255], H. Curgenven[255], S. Mohamed Ally[255], K. Beaumont[255], M. Elsaadany[255], K. Fernandes[255], I. Ali Mohamed Ali[255], H. Rangarajan[255], V. Sarathy[255], S. Selvanayagam[255], D. Vedage[255], M. White[255]. Royal Blackburn Teaching Hospital, Blackburn, UK: M. Smith[256], N. Truman[256], S. Chukkambotla[256], S. Keith[256], J. Cockerill-Taylor[256], J. Ryan-Smith[256], R. Bolton[256], P. Springle[256], J. Dykes[256], J. Thomas[256], M. Khan[256], M. T. Hijazi[256], E. Massey[256], G. Croston[256]. Poole Hospital, Poole, UK: H. Reschreite r[258], J. Camsooksai[257], S. Patch[257], S. Jenkins[257], C. Humphrey[257], B. Wadams[257], J. Camsooksai[257]. Medway Maritime Hospital, Gillingham, UK: N. Bhatia[258], M. Msiska[258], O. Adanini[258]. Warwick Hospital, Warwick, UK: B. Attwood[259], P. Parsons[259]. The Royal Marsden Hospital, London, UK: K. Tatham[260], S. Jhanji[260], E. Black[260], A. Dela Rosa[260], R. Howle[260], B. Thomas[260], T. Bemand[260], R. Raobaikady[260]. The Princess Alexandra Hospital, Harlow, UK: R. Saha[261], N. Staines[261], A. Daniel[261], J. Finn[261]. Musgrove Park Hospital, Taunton, UK: J. Hutter[262], P. Doble[262], C. Shovelton[262], C. Pawley[262]. George Eliot Hospital NHS Trust, Nuneaton, UK: T. Kannan[263], M. Hill[263]. East Surrey Hospital, Redhill, UK: E. Combes[264], S. Monnery[264], T. Joefield[264]. West Middlesex Hospital, Isleworth, UK: M. Popescu[265], M. Thankachen[265], M. Oblak[265]. Warrington General Hospital, Warrington, UK: J. Little[266], S. McIvor[266], A. Brady[266], H. Whittle[266], H. Prady[266], R. Chan[266]. Southport and Formby District General Hospital, Ormskirk, UK: A. Ahmed[267], A. Morris[267]. Royal Devon and Exeter Hospital, Exeter, UK: C. Gibson[268], E. Gordon[268], S. Keenan[268], H. Quinn[268], S. Benyon[268], S. Marriott[268], L. Zitter[268], L. Park[268], K. Baines[268]. Macclesfield District General Hospital, Macclesfield, UK: M. Lyons[269], M. Holland[269], N. Keenan[269], M. Young[269]. Borders General Hospital, Melrose, UK: S. Garrioch[270], J. Dawson[270], M. Tolson[270]. Birmingham Children's Hospital, Birmingham, UK: B. Scholefield[271], R. Bi[271]. William Harvey Hospital, Ashford, UK: N. Richardson[272], N. Schumacher[272], T. Cosier[272], G. Millen[272]. Royal Lancaster Infirmary, Lancaster, UK: A. Higham[273], K. Simpson[273]. Queen Elizabeth the Queen Mother Hospital, Margate, UK: S. Turki[274], L. Allen[274], N. Crisp[274], T. Hazleton[274], A. Knight[274], J. Deery[274], C. Price[274], S. Turney[274], S. Tilbey[274], E. Beranova[274]. Liverpool Heart and Chest Hospital, Liverpool, UK: D. Wright[275], L. Georg[275], S. Twiss[275]. Darlington Memorial Hospital, Darlington, UK: A. Cowton[276], S. Wadd[276], K. Postlethwaite[276]. Southend University Hospital, Westcliff-on-Sea, UK: P. Gondo[277], B. Masunda[277], A. Kayani[277], B. Hadebe[277]. Raigmore Hospital, Inverness, UK: J. Whiteside[278], R. Campbell[278], N. Clarke[278]. Salisbury District Hospital, Salisbury, UK: P. Donnison[278], F. Trim[279], I. Leadbitter[279]. Peterborough City Hospital, Peterborough, UK: D. Butcher[280], S. O'Sullivan[280]. Ipswich Hospital, Ipswich, UK: B. Purewal[281], S. Bell[281], V. Rivers'[281]. Hereford County Hospital, Hereford, UK: R. O'Leary[282], J. Birch[282], E. Collins[282], S. Anderson[282], K. Hammerton[282], E. Andrews[282]. Furness General Hospital, Barrow-in-Furness, UK: A. Higham[283], K. Burns[283]. Forth Valley Royal Hospital, Falkirk, UK: I. Edmond[284], D. Salutous[284], A. Todd[284], J. Donnachie[284], P. Turner[284], L. Prentice[284], L. Symon[284], N. Runciman[284], F. Auld[284]. Torbay Hospital, Torquay, UK: M. Halkes[285], P. Mercer[285], L. Thornton[285]. St Mary's Hospital, Newport, UK: G. Debreceni[286], J. Wilkins[286], A. Brown[286], V. Crickmore[286]. Royal Manchester Children's Hospital, Manchester, UK: G. Subramanian[287], R. Marshall[287], C. Jennings[287], M. Latif[278], L. Bunni[287]. Royal Cornwall Hospital, Truro, UK: M. Spivey[288], S. Bean[288], K. Burt[288]. Queen Elizabeth Hospital Gateshead, Gateshead, UK: V. Linnett[289], J. Ritzema[289], A. Sanderson[289], W. McCormick[289], M. Bokhari[289]. Kent & Canterbury Hospital, Canterbury, UK: R. Kapoor[290], D. Loader[290]. James Paget University Hospital NHS Trust, Great Yarmouth, UK: A. Ayers[291], W. Harrison[291], J. North[291]. Darent Valley Hospital, Dartford, UK: Z. Belagodu[292], R. Parasomthy[292], O. Olufuwa[292], A. Gherman[292], B. Fuller[292], C. Stuart[292]. The Alexandra Hospital, Redditch and Worcester Royal Hospital, Worcester, UK: O. Kelsall[293], C. Davis[293], L. Wild[293], H. Wood[293], J. Thrush[293], A. Durie[293], K. Austin'[293], K. Archer[293], P. Anderson[2934], C. Vigurs[293]. Ysbyty Gwynedd, Bangor, UK: C. Thorpe[294], A. Thomas[294], E. Knights[294], N. Boyle[294], A. Price[294]. Yeovil Hospital, Yeovil, UK: A. Kubisz-Pudelko[295], D. Wood[295], A. Lewis[295], S. Board[295], L. Pippard[295], J. Perry[295], K. Beesley[295]. University Hospital Hairmyres, East Kilbride, UK: A. Rattray[296], M. Taylor[296], E. Lee[296], L. Lennon[296], K. Douglas[296], D. Bell[296], R. Boyle[296], L. Glass[296]. Scunthorpe General Hospital, Scunthorpe, UK: M. Nauman Akhtar[297], K. Dent[297], D. Potoczna[297], S. Pearson[297], E. Horsley[297], S. Spencer[297]. Princess Royal Hospital Brighton, West Sussex, UK: C. Phillips[298], D. Mullan[298], D. Skinner[298], J. Gaylard[298], L. Ortiz-Ruizdegordoa[298]. Lincoln County Hospital, Lincoln, UK: R. Barber[299], C. Hewitt[299], A. Hilldrith[299], S. Shepardson[299], M. Wills[299], K. Jackson-Lawrence[299]. Homerton University Hospital, London, UK: A. Gupta[300], A. Easthope[300], E. Timlick[300], C. Gorman[300]. Glangwili General Hospital, Camarthen, UK: I. Otaha[301], A. Gales[301], S. Coetzee[301], M. Raj[301], M. Peiu[301]. Ealing Hospital, Southall, UK: V. Parris[302], S. Quaid[302], E. Watson[302]. Scarborough General Hospital, Scarborough, UK: K. Elliott[303], J. Mallinson[303], B. Chandler[303], A. Turnbull[303]. Royal Albert Edward Infirmary, Wigan, UK: A. Quinn[304], C. Finch[304], C. Holl[304], J. Cooper[304], A. Evans[304]. Queen Elizabeth Hospital, Woolwich, London, UK: W. Khaliq[305], A. Collins[305], E. Treus Gude[305]. North Devon District Hospital, Barnstaple, UK: N. Love[306], L. van Koutrik[306], J. Hunt[306], D. Kaye[306], E. Fisher[306], A. Brayne[306], V. Tuckey[306], P. Jackson[306], J. Parkin[306]. National Hospital for Neurology and Neurosurgery, London, UK: D. Brealey[307], E. Raith[307], A. Tariq[307], H. Houlden[307], A. Tucci[307], J. Hardy[307], E. Moncur[307]. Eastbourne District General Hospital, East Sussex, UK and Conquest Hospital, East Sussex, UK: J. Highgate[308], A. Cowley[308]. Diana Princess of Wales Hospital, Grimsby, UK: A. Mitra[309], R. Stead[309], T. Behan[309], C. Burnett[309], M. Newton[309], E. Heeney[309], R. Pollard[309], J. Hatton[309]. The Christie NHS Foundation Trust, Manchester, UK: A. Patel[310], V. Kasipandian[310], S. Allibone[310], R. M. Genetu[310]. Prince Philip Hospital, Lianelli, UK: I. Otahal[311], L. O'Brien[311], Z. Omar[311], E. Perkins[311], K. Davies[311]. Prince Charles Hospital, Merthyr Tydfil, UK: D. Tetla[312], C. Pothecary[312], B. Deacon[312]. Golden Jubilee National Hospital, Clydebank, UK: B. Shelley[313], V. Irvine[313]. Dorset County Hospital, Dorchester, UK: S. Williams[314], P. Williams[314], J. Birch[314], J. Goodsell[314], R. Tutton[314], L. Bough[314], B. Winter-Goodwin[314]. Calderdale Royal Hospital, Halifax, UK: R. Kitson[315], J. Pinnell[315], A. Wilson[315], T. Nortcliffe[315], T. Wood[315], M. Home[315], K. Holdroyd[315], M. Robinson[315], R. Shaw[315], J. Greig[315], M. Brady[315], A. Haigh[315], L. Matupe[315], M. Usher[315], S. Mellor[315], S. Dale[315], L. Gledhill[315], L. Shaw[315], G. Turner[315], D. Kelly[315], B. Anwar[315], H. Riley[315], H. Sturgeon[315], A. Ali[315], L. Thomis[315], D. Melia[315], A. Dance[315], K. Hanson[315]. West Suffolk Hospital, Suffolk, UK: S. Humphreys[316], I. Frost[316], V. Gopal[316], J. Godden[316], A. Holden[316], S. Swann[316]. West Cumberland Hospital, Whitehaven, UK: T. Smith[317], M. Clapham[317], U. Poultney[317], R. Harper[317], P. Rice[317]. University Hospital Lewisham, London, UK: W. Khaliq[318], R. Reece-Anthony[318], B. Gurung[318]. St John's Hospital Livingston, Livingston, UK: S. Moultrie[319], M. Odam[319]. Sheffield Children's Hospital, Sheffield, UK: A. Mayer[320], A. Bellini[320], A. Pickard[320], J. Bryant[320], N. Roe[320], J. Sowter[320]. Hinchingbrooke Hospital, Huntingdon, UK: D. Butcher[321], K. Lang[321], J. Taylor[321]. Glenfield Hospital, Leicester, UK: P. Barry[322]. Bronglais General Hospital, Aberystwyth, UK: M. Hobrok[323], H. Tench[323], R. Wolf-Roberts[323], H. McGuinness[323], R. Loosley[323]. Alder Hey Children's Hospital, Liverpool, UK: D. Hawcutt[324], L. Rad[324], L.

O'Malley[324], P. Saunderson[324], G. Seddon[324], T. Anderson[324], N. Rogers[324]. University Hospital Monklands, Airdrie, UK: J. Ruddy[325], M. Harkins[325], M. Taylor[325], C. Beith[325], A. McAlpine[325], L. Ferguson[325], P. Grant[325], S. MacFadyen[325], M. McLaughlin[325], T. Baird[325], S. Rundell[32], L. Glass[325], B. Welsh[325], R. Hamill[325], F. Fisher[325]. Cumberland Infirmary, Carlisle, UK: T. Smith[326], J. Gregory[326], A. Brown[326]. *German COVID-19 OMICS Initiative (DeCOI)—Host genetics subgroup* Axel Schmidt[20], Kerstin U. Ludwig[20], Selina Rolker[20], Markus M. Nöthen[20], Julia Fazaal[20], Verena Keitel[328], Björn Jensen[327], Torsten Feldt[327], Lisa Knopp[327], Julia Schröder[21], Carlo Maj[32], Fabian Brand[328], Marc M. Berger[328], Thorsten Brenner[329], Anke Hinney[330], Oliver Witzke[331], Robert Bals[332], Christian Herr[332], Nicole Ludwig[333], Jörn Walter[334], Jochen Schneider[335], Johanna Erber[335], Christoph D. Spinner[335,336], Clemens M. Wendtner[337], Christof Winter[338], Ulrike Protzer[339], Nicolas Casadei[340], Stephan Ossowski[341], Susanne Motameny[342], Olaf H. Riess[342], Eva C. Schulte[21,22,23]. Quebec COVID-19 Biobank (BQC-19): J. Brent Richards[12,14,15], Guillaume Butler-Laporte[12,13]. POLCOVID GEnomika: Mirosław Kwasniewski[343], Urszula Korotko[343], Karolina Chwialkowska[343], Magdalena Niemira[344], Jerzy Jaroszewicz[345], Barbara Sobala-Szczygiel[345], Beata Puzanowska[346], Anna Parfieniuk-Kowerda[347], Diana Martonik[347], Anna Moniuszko-Malinowska[348], Sławomir Pancewicz[348], Dorota Zarębska-Michaluk[349], Krzysztof Simon[350],Monika Pazgan-Simon[350], Iwona Mozer-Lisewska[351], Maciej Bura[351], Agnieszka Adamek[351], Krzysztof Tomasiewicz[352] Małgorzata Pawłowska[353], Anna Piekarska[354], Aleksandra Berkan-Kawinska[354], Andrzej Horban[355], Justyna Kowalska[355], Regina Podlasin[356], Piotr Wasilewski[356], Arsalin Azzadin[357], Miroslaw Czuczwar[358], Slawomir Czaban[359], Paweł Olszewski[360], Jacek Bogocz[360], Magdalena Ochab[360], Anna Kruk[360], Sandra Uszok[360], Agnieszka Bielska[360], Anna Szałkowska[361], Justyna Raczkowska[361], Gabriela Sokołowska[361], Joanna Chorostowska-Wynimko[362], Aleksandra Jezela-Stanek[362], Adriana Roży[362], Urszula Lechowicz[362], Urszula Polowianiuk[363], Kamil Grubczak[364], Aleksandra Starosz[364], Andrzej Eljaszewicz[364], Wiktoria Izdebska[364], Adam Krętowski[365], Robert Flisiak[366], Marcin Moniuszko[367]. Saudi GENOME: Malak Abedalthagafi[368], Manal Alaamery[369,370,371], Salam Massadeh[369,370,371], Mohamed Fawzy[368], Hadeel AlBardis[368], Nora Aljawini[369,370], Moneera Alsuwailm[369], Faisal Almalki[369], Serghei Mangul[372], Junghyun Jung[372]. QATAR GENOME: Hamdi Mbarek[373], Chadi Saad[373], Yaser Al-Sarraj[373], Wadha Al-Muftah[373], Radja Badji[373], Asma Al Thani[373], Said I. Ismail[373].

(1) Med Biotech Hub and Competence Center, Department of Medical Biotechnologies, University of Siena, Italy. (2) Medical Genetics, University of Siena, Italy. (3) University of Siena, DIISM-SAILAB, Siena, Italy. (4) Department of Mathematics, University of Pavia, Pavia, Italy. (5) Genetica Medica, Azienda Ospedaliero-Universitaria Senese, Italy. (6) Istituto di Tecnologie Biomediche – Consiglio Nazionale delle Ricerche, Segrate (MI), Italy. (7) Pharmacogenomics Unit, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy. (8) Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden. (9) Department of Medicine, Clinical Immunology and Infectious Diseases, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy. (10) Models and Algorithms for Artificial Intelligence (MAASAI) Research Group, Université Côte d'Azur, Inria, CNRS, I3S, France. (11) Good Clinical Practice Alliance-Europe (GCPA) and Strategic Initiative for Developing Capacity in Ethical Review (SIDCER), Leuven, Belgium. (12) Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, Québec, Canada. (13) Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada. (14) Department of Human Genetics, McGill University, Montréal, Québec, Canada. (15) Department of Twin Research, King's College London, London, United Kingdom. (16) Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden. (17) Anaesthesiology and Intensive Care Medicine, Department of Surgical Sciences, Uppsala University, Uppsala, Sweden. (18) Hedenstierna Laboratory, CIRRUS, Anaesthesiology and Intensive Care Medicine, Department of Surgical Sciences, Uppsala University, Uppsala, Sweden. (19) Integrative Physiology, Department of Medical Cell Biology, Uppsala University, Uppsala, Sweden. (20) Institute of Human Genetics, University of Bonn, School of Medicine & University Hospital Bonn, Bonn, Germany. (21) Institute of Psychiatric Phenomics and Genomics (IPPG), University Hospital, LMU Munich, Munich, 80336, Germany. (22) Department of Psychiatry and Psychotherapy, University Hospital, LMU Munich, Munich, 80336, Germany. (23) Institute of Virology, Technical University Munich/Helmholtz Zentrum München, Munich, Germany. (24) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK. (25) Roslin Institute, University of Edinburgh, Easter Bush, Edinburgh, EH25 9RG, UK. (26) Intensive Care Unit, Royal Infirmary of Edinburgh, 54 Little France Drive, Edinburgh, EH16 5SA, UK. (27) Department of Medical Sciences, Infectious and Tropical Diseases Unit, Azienda Ospedaliera Universitaria Senese, Siena, Italy. (28) Unit of Respiratory Diseases and Lung Transplantation, Department of Internal and Specialist Medicine, University of Siena, Italy. (29) Dept of Emergency and Urgency, Medicine, Surgery and Neurosciences, Unit of Intensive Care Medicine, Siena University Hospital, Italy. (30) Department of Medical, Surgical and Neuro Sciences and Radiological Sciences, Unit of Diagnostic Imaging, University of Siena, Italy. (31) Rheumatology Unit, Department of Medicine, Surgery and Neurosciences, University of Siena, Policlinico Le Scotte, Italy. (32) Department of Specialized and Internal Medicine, Infectious Diseases Unit, San Donato Hospital Arezzo, Italy. (33) Dept of Emergency, Anesthesia Unit, San Donato Hospital, Arezzo, Italy. (34) Cardio-thoraco-neuro-vascular Department, Pneumology Unit and UTIP, San Donato Hospital, Arezzo, Italy. (35) Department of Emergency, Anesthesia Unit, Misericordia Hospital, Grosseto, Italy. (36) Department of Specialized and Internal Medicine, Infectious Diseases Unit, Misericordia Hospital, Grosseto, Italy. (37) Clinical Chemical Analysis Laboratory, Misericordia Hospital, Grosseto, Italy. (38) Dipartimento di Prevenzione, Azienda USL Toscana Sud Est, Italy. (39) Dipartimento Tecnico-Scientifico Territoriale, Azienda USL Toscana Sud Est, Italy. (40) Clinical Chemical Analysis Laboratory, San Donato Hospital, Arezzo, Italy. (41) Chirurgia Vascolare, Ospedale Maggiore di Crema, Italy. (42) Department of Health Sciences, Clinic of Infectious Diseases, ASST Santi Paolo e Carlo, University of Milan, Italy. (43) Division of Infectious Diseases and Immunology, Department of Medical Sciences and Infectious Diseases, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy. (44) Department of Internal Medicine and Therapeutics, University of Pavia, Italy. (45) Department of Anesthesia and Intensive Care, University of Modena and Reggio Emilia, Modena, Italy. (46) Department of Medical and Surgical Sciences for Children and Adults, University of Modena and Reggio Emilia, Modena, Italy. (47) HIV/AIDS Department, National Institute for Infectious Diseases, IRCCS, Lazzaro Spallanzani, Rome, Italy. (48) III Infectious Diseases Unit, ASST-FBF-Sacco, Milan, Italy. (49) Department of Biomedical and Clinical Sciences Luigi Sacco, University of Milan, Milan, Italy. (50) Clinica di Malattie Infettive, Dipartimento di Medicina e Chirurgia, Università di Perugia, Perugia, Italy. (51) Department of Infectious Diseases, Treviso Hospital, Local Health Unit 2 Marca Trevigiana, Treviso, Italy. (52) Clinical Infectious Diseases, Mestre Hospital, Venezia, Italy. (53) Infectious Diseases Clinic, ULSS1, Belluno, Italy. (54) Department of Molecular Medicine, University of Padova, Italy. (55) Department of Infectious and Tropical Diseases, University of Brescia and ASST Spedali Civili Hospital, Brescia, Italy. (56) Department of Molecular and Translational Medicine, University of Brescia, Italy. (57) Clinical Chemistry Laboratory, Cytogenetics and Molecular Genetics Section, Diagnostic Department, ASST Spedali Civili di Brescia, Italy. (58) Medical Genetics and Laboratory of Medical Genetics Unit, A.O.R.N. "Antonio Cardarelli", Naples, Italy. (59) Department of Molecular Medicine and Medical Biotechnology, University of

Naples Federico II, Naples, Italy. (60) CEINGE Biotecnologie Avanzate, Naples, Italy. (61) IRCCS SDN, Naples, Italy. (62) Unit of Respiratory Physiopathology, AORN dei Colli, Monaldi Hospital, Naples, Italy. (63) Division of Medical Genetics, Fondazione IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Italy. (64) Department of Medical Sciences, Fondazione IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Italy. (65) Clinical Trial Office, Fondazione IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Italy. (66) Department of Health Sciences, University of Genova, Genova, Italy. (67) Infectious Diseases Clinic, Policlinico San Martino Hospital, IRCCS for Cancer Research Genova, Italy. (68) Microbiology, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Catholic University of Medicine, Rome, Italy. (69) Department of Laboratory Sciences and Infectious Diseases, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy. (70) Department of Cardiovascular Diseases, University of Siena, Siena, Italy. (71) Otolaryngology Unit, University of Siena, Italy. (72) Department of Internal Medicine, ASST Valtellina e Alto Lario, Sondrio, Italy. (73) Study Coordinator Oncologia Medica e Ufficio Flussi Sondrio, Italy. (74) First Aid Department, Luigi Curto Hospital, Polla, Salerno, Italy. (75) Department of Pharmaceutical Medicine, Misericordia Hospital, Grosseto, Italy. (76) U.O.C. Laboratorio di Genetica Umana, IRCCS Istituto G. Gaslini, Genova, Italy. (77) Infectious Diseases Clinics, University of Modena and Reggio Emilia, Modena, Italy. (78) Department of Respiratory Diseases, Azienda Ospedaliera di Cremona, Cremona, Italy. (79) U.O.C. Medicina, ASST Nord Milano, Ospedale Bassini, Cinisello Balsamo (MI), Italy. (80) Istituto Auxologico Italiano, IRCCS, Department of Cardiovascular, Neural and Metabolic Sciences, San Luca Hospital, Milan, Italy. (81) Department of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy. (82) Istituto Auxologico Italiano, IRCCS, Center for Cardiac Arrhythmias of Genetic Origin, Milan, Italy. (83) Istituto Auxologico Italiano, IRCCS, Laboratory of Cardiovascular Genetics, Milan, Italy. (84) Member of the European Reference Network for Rare, Low Prevalence and Complex Diseases of the Heart-ERN GUARD-Heart. (85) Independent Data Scientist, Milan, Italy. (86) Scuola Normale Superiore, Pisa, Italy. (87) CNR-Consiglio Nazionale delle Ricerche, Istituto di Biologia e Biotecnologia Agraria (IBBA), Milano, Italy. (88) Unit of Infectious Diseases, ASST Papa Giovanni XXIII Hospital, Bergamo, Italy. (89) Fondazione per la ricerca Ospedale di Bergamo, Bergamo, Italy. (90) Direzione Scientifica, Istituti Clinici Scientifici Maugeri IRCCS, Pavia, Italy. (91) Istituti Clinici Scientifici Maugeri IRCCS, Department of Cardiology, Institute of Montescano, Pavia, Italy. (92) Istituti Clinici Scientifici Maugeri, IRCCS, Department of Cardiac Rehabilitation, Institute of Tradate (VA), Italy. (93) Istituti Clinici Scientifici Maugeri IRCCS, Department of Cardiology, Institute of Milan, Milan, Italy. (94) Core Research Laboratory, ISPRO, Florence, Italy. (95) Health Management, Azienda USL Toscana Sudest, Tuscany, Italy. (96) IRCCS C. Mondino Foundation, Pavia, Italy. (97) Medical Genetics Unit, Meyer Children's University Hospital, Florence, Italy. (98) Department of Medicine, Pneumology Unit, Misericordia Hospital, Grosseto, Italy. (99) Department of Anesthesia and Intensive Care Unit, ASST Fatebenefratelli Sacco, Luigi Sacco Hospital, Polo Universitario, University of Milan, Milan. (100) Infectious Disease Unit, Hospital of Lucca, Italy. (101) Department of Diagnostic and Laboratory Medicine, Unity of Chemistry, Biochemistry and Clinical Molecular Biology, Fondazione Policlinico Universitario A. Gemelli IRCCS, Catholic University of the Sacred Heart, Rome, Italy. (102) Clinic of Infectious Diseases, Catholic University of the Sacred Heart, Rome, Italy. (103) Department of Clinical and Experimental Medicine, Infectious Diseases Unit, University of Pisa, Pisa, Italy. (104) Dept. of Clinical Medicine, PublicHealth, Life and Environment Sciences, University of L'Aquila, L'Aquila, Italy. (105) Department of Clinical Medicine, Public Health, Life and Environment Sciences, University of L'Aquila, Italy. (106) Anesthesiology and Intensive Care, University of L'Aquila, L'Aquila, Italy. (107) Medical Genetics Unit, Department of Life, Health and Environmental Sciences, University of L'Aquila, L'Aquila, Italy. (108) Infectious Disease Unit, Hospital of Massa, Italy. (109) Infectious Diseases Unit, Santa Maria Annunziata Hospital, USL Centro, Florence, Italy. (110) Respiratory Diseases Unit, "Santa Maria degli Angeli" Hospital, Pordenone, Italy. (111) Department of Electronics, Information and Bioengineering (DEIB), Politecnico di Milano, Milano, Italy. (112) Laboratory of Clinical Pathology and Immunoallergy, Florence-Prato, Italy. (113) Laboratory of Regulatory and Functional Genomics, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (Foggia), Italy. (114) Department of Molecular Medicine and Medical Biotechnology, University of Naples Federico II, Naples, Italy. (115) University of Pavia, Pavia, Italy. (116) Division of Infectious Diseases I, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy. (117) Department of Clinical, Surgical, Diagnostic, and Pediatric Sciences, University of Pavia, Pavia, Italy. (118) Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. (119) Department of Medical Sciences, Science for Life Laboratory, Uppsala University, Uppsala, Sweden. (120) Department of Medical Sciences, Clinical Chemistry, Uppsala University, Uppsala, Sweden. (121) Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Old Road Campus, Roosevelt Drive, Oxford, OX3 7FZ, UK. (122) Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. (123) UCL Centre for Human Health and Performance, London, W1T 7HA, UK. (124) Department of Critical Care Medicine, Queen's University and Kingston Health Sciences Centre, Kingston, ON, Canada. (125) Department of Anaesthesia and Intensive Care, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, China. (126) Clinical Research Centre at St Vincent's University Hospital, University College Dublin, Dublin, Ireland. (127) Department of Medicine, University of Cambridge, Cambridge, UK. (128) William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK. (129) NIHR Health Protection Research Unit for Emerging and Zoonotic Infections, Institute of Infection, Veterinary and Ecological Sciences University of Liverpool, Liverpool, L69 7BE, UK. (130) Respiratory Medicine, Alder Hey Children's Hospital, Institute in The Park, University of Liverpool, Alder Hey Children's Hospital, Liverpool, UK. (131) National Heart and Lung Institute, Imperial College London, London, UK. (132) Imperial College Healthcare NHS Trust: London, London, UK. (133) Department of Intensive Care Medicine, Guy's and St. Thomas NHS Foundation Trust, London, UK. (134) MRC-University of Glasgow Centre for Virus Research, Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK. (135) Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast, Northern Ireland, UK. (136) Department of Intensive Care Medicine, Royal Victoria Hospital, Belfast, Northern Ireland, UK. (137) Intensive Care National Audit & Research Centre, London, UK. (138) Edinburgh Clinical Research Facility, Western General Hospital, University of Edinburgh, EH4 2XU, UK. (139) Genomics England, London, UK. (140) Centre for Inflammation Research, The Queen's Medical Research Institute, University of Edinburgh, 47 Little France Crescent, Edinburgh, UK. (141) Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK. (142) Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia. (143) Biostatistics Group, School of Life Sciences, Sun Yat-sen University, Guangzhou, China. (144) School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310024, China. (145) Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310024, China. (146) Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, Teviot Place, Edinburgh EH8 9AG, UK. (147) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. (148) Barts Health NHS Trust, London, UK. (149)

Guys and St Thomas' Hospital, London, UK. (150) James Cook University Hospital, Middlesburgh, UK. (151) The Royal Liverpool University Hospital, Liverpool, UK. (152) King's College Hospital, London, UK. (153) Royal Infirmary of Edinburgh, Edinburgh, UK. (154) John Radcliffe Hospital, Oxford, UK. (155) Addenbrooke's Hospital, Cambridge, UK. (156) Morriston Hospital, Swansea, UK. (157) Ashford and St Peter's Hospital, Surrey, UK. (158) Royal Stoke University Hospital, Staffordshire, UK. (159) Queen Elizabeth Hospital, Birmingham, UK. (160) Glasgow Royal Infirmary, Glasgow, UK. (161) Kingston Hospital, Surrey, UK. (162) The Tunbridge Wells Hospital and Maidstone Hospital, Kent, UK. (163) North Middlesex University Hospital NHS trust, London, UK. (164) Bradford Royal Infirmary, Bradford, UK. (165) Blackpool Victoria Hospital, Blackpool, UK. (166) Countess of Chester Hospital, Chester, UK. (167) Wythenshawe Hospital, Manchester, UK. (168) St George's Hospital, London, UK. (169) Good Hope Hospital, Birmingham, UK. (170) Stepping Hill Hospital, Stockport, UK. (171) Manchester Royal Infirmary, Manchester, UK. (172) Royal Alexandra Hospital, Paisley, UK. (173) Queen Elizabeth University Hospital, Glasgow, UK. (174) Queen Alexandra Hospital, Portsmouth, UK. (175) BHRUT (Barking Havering)—Queens Hospital and King George Hospital, Essex, UK. (176) University College Hospital, London, UK. (177) Royal Victoria Infirmary, Newcastle Upon Tyne, UK. (178) Western Sussex Hospitals, West Sussex, UK. (179) Salford Royal Hospital, Manchester, UK. (180) The Royal Oldham Hospital, Manchester, UK. (181) Pinderfields General Hospital, Wakefield, UK. (182) Basildon Hospital, Basildon, UK. (183) University Hospital of Wales, Cardiff, UK. (184) Broomfield Hospital, Chelmsford, UK. (185) Royal Brompton Hospital, London, UK. (186) Nottingham University Hospital, Nottingham, UK. (187) Royal Hallamshire Hospital and Northern General Hospital, Sheffield, UK. (188) Royal Hampshire County Hospital, Hampshire, UK. (189) Queens Hospital Burton, Burton-On-Trent, UK. (190) New Cross Hospital, Wolverhampton, UK. (191) Heartlands Hospital, Birmingham, UK. (192) Walsall Manor Hospital, Walsall, UK. (193) Stoke Mandeville Hospital, Buckinghamshire, UK. (194) Sandwell General Hospital, Birmingham, UK. (195) Royal Berkshire NHS Foundation Trust, Berkshire, UK. (196) Charing Cross Hospital, St Mary's Hospital and Hammersmith Hospital, London, UK. (197) Dumfries and Galloway Royal Infirmary, Dumfries, UK. (198) Bristol Royal Infirmary, Bristol, UK. (199) Royal Sussex County Hospital, Brighton, UK. (200) Whiston Hospital, Prescot, UK. (201) Royal Glamorgan Hospital, Cardiff, UK. (223) King's Mill Hospital, Nottingham, UK. (203) Fairfield General Hospital, Bury, UK. (204) Western General Hospital, Edinburgh, UK. (205) Northwick Park Hospital, London, UK. (206) Royal Preston Hospital, Preston, UK. (207) Royal Derby Hospital, Derby, UK. (208) Sunderland Royal Hospital, Sunderland, UK. (209) Royal Surrey County Hospital, Guildford, UK. (210) Derriford Hospital, Plymouth, UK. (211) Croydon University Hospital, Croydon, UK. (212) Victoria Hospital, Kirkcaldy, UK. (213) Milton Keynes University Hospital, Milton Keynes, UK. (214) Barnsley Hospital, Barnsley, UK. (215) York Hospital, York, UK. (216) University Hospital of North Tees, Stockton on Tees, UK. (217) University Hospital Wishaw, Wishaw, UK. (218) Whittington Hospital, London, UK. (219) Southmead Hospital, Bristol, UK. (220) The Royal Papworth Hospital, Cambridge, UK. (221) Royal Gwent Hospital, Newport, UK. (222) Norfolk and Norwich University hospital (NNUH), Norwich, UK. (223) Great Ormond St Hospital and UCL Great Ormond St Institute of Child Health NIHR Biomedical Research Centre, London, UK. (224) Airedale General Hospital, Keighley, UK. (225) Aberdeen Royal Infirmary, Aberdeen, UK. (226) Southampton General Hospital, Southampton, UK. (227) Russell's Hall Hospital, Dudley, UK. (228) Rotherham General Hospital, Rotherham, UK. (229) North Manchester General Hospital, Manchester, UK. (230) Basingstoke and North Hampshire Hospital, Basingstoke, UK. (231) Royal Free Hospital, London, UK. (232) Hull Royal Infirmary, Hull, UK. (233) Harefield Hospital, London, UK. (234) Chesterfield Royal Hospital Foundation Trust, Chesterfield, UK.

(235) Barnet Hospital, London, UK. (236) Aintree University Hospital, Liverpool, UK. (237) St James's University Hospital and Leeds General Infirmary, Leeds, UK. (238) Glan Clwyd Hospital, Bodelwyddan, UK. (239) University Hospital Crosshouse, Kilmarnock, UK. (240) Royal Bolton Hospital, Bolton, UK. (241) Princess of Wales Hospital, Llantrisant, UK. (242) Pilgrim Hospital, Lincoln, UK. (243) Northumbria Healthcare NHS Foundation Trust, North Shields, UK. (244) Ninewells Hospital, Dundee, UK. (245) Lister Hospital, Stevenage, UK. (246) Bedford Hospital, Bedford, UK. (247) Royal United Hospital, Bath, UK. (248) Royal Bournemouth Hospital, Bournemouth, UK. (249) The Great Western Hospital, Swindon, UK. (250) WatfordGeneral Hospital, Watford, UK. (251) University Hospital North Durham, Darlington, UK. (252) Tameside General Hospital, Ashton Under Lyne, UK. (253) Princess Royal Hospital Shrewsbury and Royal Shrewsbury Hospital, Shrewsbury, UK. (254) Arrowe Park Hospital, Wirral, UK. (255) The Queen Elizabeth Hospital, King's Lynn, UK. (256) Royal Blackburn Teaching Hospital, Blackburn, UK. (257) Poole Hospital, Poole, UK. (258) Medway Maritime Hospital, Gillingham, UK. (259) Warwick Hospital, Warwick, UK. (260) The Royal Marsden Hospital, London, UK. (261) The Princess Alexandra Hospital, Harlow, UK. (262) Musgrove Park Hospital, Taunton, UK. (263) George Eliot Hospital NHS Trust, Nuneaton, UK. (264) East Surrey Hospital, Redhill, UK. (265) West Middlesex Hospital, Isleworth, UK. (266) Warrington General Hospital, Warrington, UK. (267) Southport and Formby District General Hospital, Ormskirk, UK. (268) Royal Devon and Exeter Hospital, Exeter, UK. (269) Macclesfield District General Hospital, Macclesfield, UK. (270) Borders General Hospital, Melrose, UK. (271) Birmingham Children's Hospital, Birmingham, UK. (272) William Harvey Hospital, Ashford, UK. (273) Royal Lancaster Infirmary, Lancaster, UK. (274) Queen Elizabeth the Queen Mother Hospital, Margate, UK. (275) Liverpool Heart and Chest Hospital, Liverpool, UK. (276) Darlington Memorial Hospital, Darlington, UK. (277) Southend University Hospital, Westcliff-on-Sea, UK. (278) Raigmore Hospital, Inverness, UK. (279) Salisbury District Hospital, Salisbury, UK. (280) Peterborough City Hospital, Peterborough, UK. (281) Ipswich Hospital, Ipswich, UK. (282) Hereford County Hospital, Worcester, UK. (283) Furness General Hospital, Barrow-in-Furness, UK. (284) Forth Valley Royal Hospital, Falkirk, UK. (285) Torbay Hospital, Torquay, UK. (286) St Mary's Hospital, Newport, UK. (287) Royal Manchester Children's Hospital, Manchester, UK. (288) Royal Cornwall Hospital, Truro, UK. (289) Queen Elizabeth Hospital Gateshead, Gateshead, UK. (290) Kent & Canterbury Hospital, Canterbury, UK. (291) James Paget University Hospital NHS Trust, Great Yarmouth, UK. (292) Darent Valley Hospital, Dartford, UK. (293) The Alexandra Hospital, Redditch and Worcester Royal Hospital, Worcester, UK. (294) Ysbyty Gwynedd, Bangor, UK. (295) Yeovil Hospital, Yeovil, UK. (296) University Hospital Hairmyres, East Kilbride, UK. (297) Scunthorpe General Hospital, Scunthorpe, UK. (298) Princess Royal Hospital Brighton, West Sussex, UK. (299) Lincoln County Hospital, Lincoln, UK. (300) Homerton University Hospital, London, UK. (301) Glangwili General Hospital, Camarthen, UK. (303) Ealing Hospital, Southall, UK. (303) Scarborough General Hospital, Scarborough, UK. (304) Royal Albert Edward Infirmary, Wigan, UK. (305) Queen Elizabeth Hospital, Woolwich, London, UK. (306) North Devon District Hospital, Barnstaple, UK. (307) National Hospital for Neurology and Neurosurgery, London, UK. (308) Eastbourne District General Hospital, East Sussex, UK and Conquest Hospital, East Sussex, UK. (309) Diana Princess of Wales Hospital, Grimsby, UK. (310) The Christie NHS Foundation Trust, Manchester, UK. (311) Prince Philip Hospital, Lianelli, UK. (312) Prince Charles Hospital, Merthyr Tydfil, UK. (313) Golden Jubilee National Hospital, Clydebank, UK. (314) Dorset County Hospital, Dorchester, UK. (315) Calderdale Royal Hospital, Halifax, UK. (316) West Suffolk Hospital, Suffolk, UK. (317) West Cumberland Hospital, Whitehaven, UK. (318) University Hospital Lewisham, London, UK. (319) St John's Hospital Livingston, Livingston, UK. (320) Sheffield Children's Hospital, Sheffield, UK. (321) Hinchingbrooke Hospital,

Huntingdon, UK. (322) Glenfield Hospital, Leicester, UK. (323) Bronglais General Hospital, Aberystwyth, UK. (324) Alder Hey Children's Hospital, Liverpool, UK. (325) University Hospital Monklands, Airdrie, UK. (326) Cumberland Infirmary, Carlisle, UK. (327) Department of Gastroenterology, Hepatology and Infectious Diseases, University Hospital Duesseldorf, Medical Faculty Heinrich Heine University, Duesseldorf, Germany. (328) Institute of Genomic Statistics and Bioinformatics, University Hospital Bonn, Medical Faculty University of Bonn, Venusberg-Campus 1, Bonn Germany. (329) Department of Anesthesiology and Intensive Care Medicine, University Hospital Essen, University Duisburg-Essen, Essen, Germany. (330) Department of Child and Adolescent Psychiatry, University Hospital Essen, University of Duisburg-Essen, Essen, Germany. (331) Department of Infectious Diseases, University Hospital Essen, University Duisburg-Essen, Essen, Germany. (332) Department of Internal Medicine V—Pneumology, Allergology and Intensive Care Medicine, University Hospital Saarland, Homburg/Saar, Germany. (333) Center of Human and Molecular Biology, Department of Human Genetics, University Hospital Saarland, Homburg/Saar, Germany. (334) Department of Genetics & Epigenetics, Saarland University, Saarbrücken, Germany. (335) Department of Internal Medicine II, School of Medicine, University Hospital Rechts der Isar, Technical University of Munich, Munich, Germany. (338) German Center for Infection Research (DZIF), Partner Site Munich, Munich, Germany. (339) Department of Hematology, Oncology, Immunology, Palliative Medicine, Infectious Diseases and Tropical Medicine, Munich Clinic Schwabing, Academic Teaching Hospital (LMU), Kölner Platz 1, 80804, Munich, Germany. (338) Institute of Clinical Chemistry and Pathobiochemistry, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany. (339) Institute of Virology, School of Medicine, Technical University of Munich/Helmholtz Zentrum München, 81675 Munich, Germany; DZIF, partner sites Munich and Cologne/Bonn, Germany. (340) Institute of Medical Genetics and Applied Genomics, The University of Tübingen, Germany; NGS Competence Center Tübingen, The University of Tübingen, Germany. (341) Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany. (342) Cologne Center for Genomics (CCG), University of Cologne and University Hospital Cologne, Cologne, Germany. (343) Centre for Bioinformatics and Data Analysis, Medical University of Bialystok, Bialystok, Poland. (344) Clinical Research Centre, Medical University of Bialystok, Bialystok, Poland. (345) Department of Infectious Diseases in Bytom, Medical University of Silesia, Silesia, Poland. (346) Department of Infectious Diseases, Megrez Hospital in Tychy, Tychy, Poland. (347) Department of Infectious Diseases and Hepatology, Medical University of Bialystok, Bialystok, Poland. (348) Department of Infectious Diseases and Neuroinfection, Medical University of Bialystok, Biastolyk, Poland. (349) Department of Infectious Diseases, Jan Kochanowski University, Kielce, Poland. (350) Department of Infectious Diseases and Hepatology, Wroclaw Medical University, Wrocław, Poland. (351) Department of Infectious Diseases and Hepatology, Poznan University of Medical Sciences, Poznan, Poland. (352) Department of Infectious Diseases, Medical University of Lublin, Lublin, Poland. (353) Department of Infectious Diseases and Hepatology, Nicolaus Copernicus University, Toruń, Poland. (354) Department of Infectious Diseases and Hepatology, Medical University of Lodz, Lodz, Poland. (355) Department of Adult Infectious Diseases, Warsaw Medical University, Warsaw, Poland. (356) IV-th Department, Hospital for Infectious Diseases, Poland. (357) District Hospital in Bielsk Podlaski, Poland. (358) Department of Anaesthesiology and Intensive Therapy, Medical University of Lublin, Lublin, Poland. (359) Department of Anaesthesiology and Intensive Therapy, Medical University of Bialystok, Bialystok, Poland. (360) IMAGENE.ME SA, Bialystok, Poland. (361) Clinical Research Centre, Medical University of Bialystok, Biastolyk, Poland. (362) Department of Genetics and Clinical Immunology, National Institute of Tuberculosis and Lung Diseases in Warsaw, Warsaw, Poland. (363) District Sanitary Inspectorate in Bialystok, Biastolyk, Poland. (364) Department of Regenerative Medicine and Immune Regulation, Medical University of Bialystok, Biastolyk, Poland. (365) Clinical Research Centre, Medical University of Bialystok, Bialystok, Poland. (366) Department of Infectious Diseases and Hepatology, Medical University of Bialystok, Bialystok, Poland. (367) Department of Allergology and Internal Medicine; Department of Regenerative Medicine and Immune Regulation, Medical University of Bialystok, Bialystok, Poland. (368) Genomics Research Department, Saudi Human Genome Project, King Fahad Medical City and King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia. (369) Developmental Medicine Department, King Abdullah International Medical Research Center, King Saud Bin Abdulaziz University for Health Sciences, King Abdulaziz Medical City, Ministry of National Guard- Health Affairs. (370) KACST-BWH Centre of Excellence for Biomedicine, Joint Centers of Excellence Program, King Abdulaziz City for Science and Technology (KACST), Riyadh, Saudi Arabia. (371) King Abdulaziz City for Science and Technology (KACST)-Saudi Human Genome Satellite Lab at Abdulaziz Medical City, Ministry of National Guard Health Affairs (MNGHA), Riyadh, Saudi Arabia. (372) Department of Clinical Pharmacy, School of Pharmacy, University of Southern California, Los Angeles, California, USA. (373) Qatar Genome Program, Qatar Foundation Research, Development and Innovation, Qatar Foundation, Doha, Qatar.

and Compute Canada is acknowledged. TwinsUK is funded by the Welcome Trust, Medical Research Council, European Union, the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. These funding agencies had no role in the design, implementation or interpretation of this study. JBR has served as an advisor to GlaxoSmithKline and Deerfield Capital. His institution has received investigator-initiated grant funding from Eli Lilly, GlaxoSmithKline and Biogen for projects unrelated to this research. He is the founder of 5 Prime Sciences. SweCovid received funding from the European Union's Horizon 2020 research and innovation program under Grant agreement No 824110, the SciLifeLab/KAW national Covid-19 research program project grants to MH (KAW 2020.0182 and KAW 2020.0241) and the Swedish Research Council Grants to RF (2014-02569 and 2014-07606). Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

Declarations

Conflict of interest The authors declare no conflict interests.

Informed consent The patients were informed of this research and agreed to it through the informed consent process.

Institutional review board statement The study was conducted according to the guidelines of the Declaration of Helsinki. The GEN-COVID is a multicentre academic observational study that was approved by the Internal Review Boards (IRB) of each participating centre (protocol code 16917, dated March 16, 2020 for GEN-COVID at the University Hospital of Siena). BQC19 received ethical approval from the Jewish General Hospital (JGH) research ethics board (2020-2137). The Swedish part of the study was performed at the general intensive care unit (ICU) of Uppsala University Hospital, Sweden (a tertiary hospital) and approved by the National Ethical Review Agency (No. 2020-01623). The Declaration of Helsinki and its subsequent revisions were followed. DeCOI received ethical approval by the Ethical Review Board (ERB) of the participating hospitals/centres for CoMRI (Technical University Munich, Munich, Germany) and BoSCO (Medical Faculty Bonn, Bonn, Germany; Medical Board of the Saarland, Germany; University Duisburg-Essen, Germany; Medical Faculty Duesseldorf, Düsseldorf, Germany). GenOMICC and ISARIC4C were approved by the appropriate research ethics committees (Scotland, 15/SS/0110; England, Wales and Northern Ireland, 19/WM/0247).

Data availability and data sharing statement The data and samples referenced here are housed in the GEN-COVID Patient Registry and the GEN-COVID Biobank and are available for consultation. You may contact the corresponding author, Prof. Alessandra Renieri (e-mail: alessandra.renieri@unisi.it).

References

Baldassarri M, Picchiotti N, Fava F et al (2021a) Shorter androgen receptor polyQ alleles protect against life-threatening COVID-19 disease in European males. EBioMedicine 65:103246

Baldassarri M, Fava F, Fallerini C et al (2021b) Severe COVID-19 in hospitalized carriers of single CFTR pathogenic variants. J Person Med 11(6):558

Bayati A, Kumar R, Francis V, McPherson PS (2021) SARS-CoV-2 infects cells after viral entry via clathrin-mediated endocytosis. J Biol Chem 296:100306

Benetti E, Giliberti A, Emiliozzi A et al (2020a) Clinical and molecular characterization of COVID-19 hospitalized patients. PLoS ONE 15(11):e0242534

Benetti E, Tita R, Spiga O et al (2020b) ACE2 gene variants may underlie interindividual variability and susceptibility to COVID-19 in the Italian population. Eur J Hum Genet 28(11):1602–1614

Chen N, Zhou M, Dong X et al (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 395:507–551

COVID-19 Host Genetics Initiative. https://www.covid19hg.org/. Accessed 13 Apr 2020

COVID-19 Host Genetics Initiative et al (2021) Mapping the human genetic architecture of COVID-19. Published online ahead of print. Nature. https://doi.org/10.1038/s41586-021-03767-x

COVID-19 Therapeutic Trial Synopsis (2020) WHO R&D blueprint novel coronavirus. Covid 19 therapeutic trial synopsis. R&D Blueprint

Covid19hg.org (2021) June 14, 2021 meeting. https://www.covid19hg.org/blog/2021-06-17-june-14-2021-meeting/. Accessed 13 Apr 2020

Croci S, Venneri MA, Mantovani S et al (2021) The polymorphism L412F in TLR3 inhibits autophagy and is a marker of severe COVID-19 in males. Autophagy (in press)

Daga S, Fallerini C, Baldassarri M et al (2021) Employing a systematic approach to biobanking and analyzing clinical and genetic data for advancing COVID-19 research. Eur J Hum Genet 29(5):745–759

Danecek P, Bonfield JK, Liddle J et al (2021) Twelve years of SAM tools and BCF tools. GigaScience 10(2):giab00

Darrah RJ, Jacono FJ, Joshi N et al (2019) AGTR2 absence or antagonism prevents cystic fibrosis pulmonary manifestations. J Cyst Fibros 18(1):127–134

Elhabyan A, Elyaacoub S, Sanad E et al (2020) The role of host genetics in susceptibility to severe viral infections in humans and insights into host genetics of severe COVID-19: a systematic review. Virus Res 289:198163

Fallerini C et al (2021a) Association of Toll-like receptor 7 variants with life-threatening COVID-19 disease in males: findings from a nested case-control study. eLife 10:e67569

Fallerini C, Daga S, Benetti E et al (2021b) *SELP* Asp603Asn and severe thrombosis in COVID-19 males: implication for anti P-selectin monoclonal antibodies treatment. J Hematol Oncol 14(1):123

Gordon DE, Jang GM, Bouhaddou M et al (2020) A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 583(7816):459–468. https://doi.org/10.1038/s41586-020-2286-9

International Multiple Sclerosis Genetics Consortium (2018) Low-frequency and rare-coding variation contributes to multiple sclerosis risk. Cell 175(6):1679-1687.e7

Islam MR, Hoque MN, Rahman MS et al (2020) Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. Sci Rep 10(1):14004

Kosmicki JA, Horowitz JE, Banerjee N et al (2021) Pan-ancestry exome-wide association analyses of COVID-19 outcomes in 586,157 individuals. Am J Hum Genet 108(7):1350–1355. https://doi.org/10.1016/j.ajhg.2021.05.017

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26(5):589–595

Li X, Zhong X, Wang Y et al (2021) Clinical determinants of the severity of COVID-19: a systematic review and meta-analysis. PLoS ONE 16(5):e0250602

Liaw A, Wiener M (2002) Classification and regression by randomForest. R News 2(3):18–22

Livingston E, Bucher K (2020) Coronavirus disease 2019 (COVID-19) in Italy. JAMA 323(14):1335

Marouli E, Graff M, Medina-Gomez C et al (2017) Rare and low-frequency coding variants alter human adult height. Nature 542(7640):186–190

Mars N, Widén E, Kerminen S et al (2020) (2020) The role of polygenic risk and susceptibility genes in breast cancer over the course of life. Nat Commun 11(1):6383

Merico D, Isserlin R, Bader GD. Visualizing gene-set enrichment results using the cytoscape plug-in enrichment map. In: Cagney G, Emili A (eds) (2011) Network biology. Methods in molecular biology (methods and protocols), vol 781. Humana Press

Mootha VK, Lindgren CM, Eriksson K et al (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 34(3):267–273

O'Connor BD, Van Auwera G (2020) Genomics in the cloud: using Docker, Gatk, and WDL in Terra. O'Reilly Media

Pairo-Castineira E, Clohisey S, Klaric L et al (2020) Genetic mechanisms of critical illness in COVID-19. Nature 591(7848):92–98

Pathak GA, Singh K, Miller-Fleming TW et al (2021) Integrative genomic analyses identify susceptibility genes underlying COVID-19 hospitalization. Nat Commun 12(1):4569. https://doi.org/10.1038/s41467-021-24824-z

Pedregosa F, Varoquaux G, Gramfort A et al (2011) Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830

Picchiotti N, Benetti E, Fallerini C et al (2021) Post-Mendelian genetic model in COVID-19. Cardiol Cardiovas Med (in press)

Purcell S, Neale B, Todd-Brown K et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81(3):559–575

Reggiori F, Monastyrska I, Verheije MH et al (2010) Coronaviruses Hijack the LC3-I-positive EDEMosomes, ER-derived vesicles exporting short-lived ERAD regulators, for replication. Cell Host Microbe 7:500–508

Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP (2006) GenePattern 2.0. Nat Genet 38:500–501

Severe Covid-19 GWAS Group, Ellinghaus D, Degenhardt F et al (2020) Genomewide association study of severe covid-19 with respiratory failure. N Engl J Med 383(16):1522–1534

Shannon P, Markiel A, Ozier O et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11):2498–2504

Solanich X, Vargas-Parra G, van der Made CI et al (2021) Genetic screening for *TLR7* variants in young and previously healthy men with severe COVID-19: a case series. Front Immunol 12:719115

Subramanian A, Tamayo P, Mootha VK et al (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102(43):15545–15550

Subramanian S, Tus K, Li QZ et al (2006) A Tlr7 translocation accelerates systemic autoimmunity in murine lupus. PNAS 103(26):9970–9975

Tan Y, Yang T, Liu P et al (2017) Association of the OAS3 rs1859330 G/A genetic polymorphism with severity of enterovirus-71 infection in Chinese Han children. Arch Virol 162:2305–2313

Van der Made CI, Simons A, Schuurs-Hoeijmakers J et al (2020) Presence of genetic variants among young men with severe COVID-19. JAMA 324(7):663–673

Virtanen P, Gommers R, Oliphant TE et al (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17(3):261–272

Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38(16):e164

Wickenhagen A, Sugrue E, Lytras S et al (2021) A prenylated dsRNA sensor protects against severe COVID-19. Science. https://doi.org/10.1126/science.abj3624

Williams F, Freidin MB, Mangino M et al (2020) Self-reported symptoms of COVID-19, including symptoms most predictive of SARS-CoV-2 infection, are heritable. ISTS 23(6):316–321

Zhang X, Tan Y, Ling Y et al (2020a) Viral and host factors related to the clinical outcome of COVID-19. Nature 583:7816:437–440

Zhang Q, Bastard P, Liu Z et al (2020b) Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. Science (New York, N.Y.) 370(6515)

Authors and Affiliations

Chiara Fallerini[1,2] · Nicola Picchiotti[3,4] · Margherita Baldassarri[1,2] · Kristina Zguro[1] · Sergio Daga[1,2] · Francesca Fava[1,2,5] · Elisa Benetti[1] · Sara Amitrano[5] · Mirella Bruttini[1,2,5] · Maria Palmieri[1,2] · Susanna Croci[1,2] · Mirjam Lista[1,2] · Giada Beligni[1,2] · Floriana Valentino[1,2] · Ilaria Meloni[1,2] · Marco Tanfoni[3] · Francesca Minnai[6] · Francesca Colombo[6] · Enrico Cabri[7] · Maddalena Fratelli[7] · Chiara Gabbi[8] · Stefania Mantovani[9] · Elisa Frullanti[1,2] · Marco Gori[3,10] · Francis P. Crawley[11] · Guillaume Butler-Laporte[12,13] · Brent Richards[12,14,15] · Hugo Zeberg[16] · Miklos Lipcsey[17,18] · Michael Hultström[17,19] · Kerstin U. Ludwig[20] · Eva C. Schulte[21,22,23] · Erola Pairo-Castineira[24,25] · John Kenneth Baillie[24,25,26] · Axel Schmidt[20] · Robert Frithiof[17] · WES/WGS Working Group Within the HGI · GenOMICC Consortium · GEN-COVID Multicenter Study · Francesca Mari[1,2,5] · Alessandra Renieri[1,2,5,27] (iD) · Simone Furini[1]

1. Department of Medical Biotechnologies, Med Biotech Hub and Competence Center, University of Siena, Siena, Italy

2. Medical Genetics, University of Siena, Siena, Italy

3. University of Siena, DIISM-SAILAB, Siena, Italy

4. Department of Mathematics, University of Pavia, Pavia, Italy

5. Genetica Medica, Azienda Ospedaliero-Universitaria Senese, Siena, Italy

6. Istituto di Tecnologie Biomediche-Consiglio Nazionale delle Ricerche, Segrate, MI, Italy

7. Pharmacogenomics Unit, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy

8. Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden

9. Department of Medicine, Clinical Immunology and Infectious Diseases, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy

10. Models and Algorithms for Artificial Intelligence (MAASAI) Research Group, Université Côte d'Azur, Inria, CNRS, I3S, Biot, France

11. Good Clinical Practice Alliance-Europe (GCPA) and Strategic Initiative for Developing Capacity in Ethical Review (SIDCER), Leuven, Belgium

12. Lady Davis Institute, Jewish General Hospital, McGill University, Montreal, QC, Canada

13. Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada

14. Department of Human Genetics, McGill University, Montreal, QC, Canada

15. Department of Twin Research, King's College London, London, UK

16. Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden

17. Anaesthesiology and Intensive Care Medicine, Department of Surgical Sciences, Uppsala University, Uppsala, Sweden

18. Hedenstierna Laboratory, CIRRUS, Anaesthesiology and Intensive Care Medicine, Department of Surgical Sciences, Uppsala University, Uppsala, Sweden

19. Integrative Physiology, Department of Medical Cell Biology, Uppsala University, Uppsala, Sweden

20. Institute of Human Genetics, School of Medicine and University Hospital Bonn, University of Bonn, Bonn, Germany

21. Institute of Psychiatric Phenomics and Genomics (IPPG), University Hospital, LMU Munich, 80336 Munich, Germany

22. Department of Psychiatry and Psychotherapy, University Hospital, LMU Munich, 80336 Munich, Germany

23. Institute of Virology, Technical University Munich/Helmholtz Zentrum München, Munich, Germany

24. MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Western General Hospital, University of Edinburgh, Crewe Road, Edinburgh EH4 2XU, UK

25. Roslin Institute, University of Edinburgh, Easter Bush, Edinburgh EH25 9RG, UK

26. Intensive Care Unit, Royal Infirmary of Edinburgh, 54 Little France Drive, Edinburgh H16 5SA, UK

27. Medical Genetics Unit, University of Siena, Policlinico Le Scotte, Viale Bracci, 2, 53100 Siena, Italy