medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Poking COVID-19: insights on genomic constraints among immune-related genes between Qatari and Italian populations

Hamdi Mbarek[1*†], Massimiliano Cocca[2*†], Yasser Al Sarraj[1†], Chadi Saad[1], Massimo Mezzavilla[2], Wadha AlMuftah[1], Dario Cocciadiferro[5], Antonio Novelli[5], Isabella Quinti[6], Azza AlTawashi[4], Salvino Salvaggio[4], Asma AlThani[1], Giuseppe Novelli[3] and Said Ismail[1]

Affiliations

[1]Qatar Genome Program, Qatar Foundation Research, Development and Innovation, Qatar Foundation, Doha, Qatar.

[2]Institute for Maternal and Child Health, IRCCS 'Burlo Garofolo', Trieste, Italy.

[3]Department of Biomedicine and Prevention, Tor Vergata University of Rome, Rome, Italy, Trieste, Italy.

[4]Qatar Foundation Research, Development and Innovation, Qatar Foundation, Doha, Qatar.

[5]Translational Cytogenomics Research Unit, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy.

[6]Department of Molecular Medicine, Sapienza University of Rome, Rome, Italy.

[*]Correspondence: Hamdi Mbarek (hmbarek@qf.org.qa); Massimiliano Cocca(massimiliano.cocca@burlo.trieste.it).

[†]These authors contributed equally to this work.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Abstract

Host genomic information, specifically genomic variations, may characterize susceptibility to disease and identify people with a higher risk of harm, leading to better targeting of care and vaccination. Italy was the epicentre for the spread of COVID-19 in Europe, the first country to go into a national lockdown and has one of the highest COVID-19 associated mortality rates. Qatar, on the other hand has a very low mortality rate. In this study, we compared whole-genome sequencing data of 14398 adults and Qatari-national to 925 Italian individuals. We also included in the comparison whole-exome sequence data from 189 Italian laboratory confirmed COVID-19 cases. We focused our study on a curated list of 3619 candidate genes involved in innate immunity and host-pathogen interaction. Two population-gene metric scores, the Delta Singleton-Cohort variant score (DSC) and Sum Singleton- Cohort variant score (SSC), were applied to estimate the presence of selective constraints in the Qatari population and in the Italian cohorts. Results based on DSC SSC metrics demonstrated a different selective pressure on three genes (MUC5AC, ABCA7, FLNA) between Qatari and Italian populations. This study highlighted the genetic differences between Qatari and Italian populations and identified a subset of genes involved in innate immunity and host-pathogen interaction.

Keywords: COVID-19; Population Genetics; Genetic Constraints; COVID-19 Severity

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Introduction

COVID-19 continues to spread worldwide, with over four million deaths to date and rising. However, this global spread is coupled with stark anomalies in morbidity and mortality. These differences can be seen not only between different populations but also within the same population[1–4].While most of these differences can be attributed to sociodemographic and clinical factors, this is also a unique opportunity to assess associations with host genomes. Host genomic information, specifically genomic variations, may characterize susceptibility to disease and identify people with a higher risk of harm, lead- ing to better targeting of care and vaccination[5–7]. In addition, characterizing these host factors may help identifying and development of adapted drugs and vaccines[8–10]. The scientific community came together with several efforts to investigate how the genomic variation in the host affects disease susceptibil- ity and progress[11, 12]. So far, these large consortia efforts have led to the identification of over 20 loci associated with susceptibility or severity of the disease[13].

Italy was the epicentre for the spread of COVID-19 in Europe and the first country to go into a national lockdown. It had one of the highest COVID-19 associated mortality rates in Europe[4]. At the time of writing, almost five million cases have been confirmed, with a death toll of more than 130 thousand people. Qatar, on the other hand, despite having one of the highest worldwide numbers of laboratory-confirmed cases (36,729 cases per million, by July 2020), has a very low mortality rate (infection fatality rate = 0.91 per 10,000 persons, by July 2020, per WHO COVID-19 mortality

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

classification)[14]. Studies have even suggested that some communities in Qatar have reached herd immunity for SARS-CoV-2 at a proportion of infection of 65-70% [15]. With the development of the first generation of RNA based vaccines[16, 17], along with the more standard adenovirus-based solutions[18, 19], the end of the pandemic seems to be in sight, though we are aware that this is just the beginningof a more bearable coexistence with the virus.

In this study, we focused on genes involved in the immune response, combining them with a dataset of 1500 proteins mostly involved in COVID-19 disease[20] and a subset of genes already identified as linked to COVID-19 susceptibility and progression[7]. We applied, on this set of genes, a prioritization method based on ultra-rare and population-specific variants. With this approach, we aim to identify a group of genes showing different signs of selective pressure in our study cohorts. Our hypothesis is that those genes can provide information to understand the pandemic progression and maybe help towards therapy.

Materials and Methods

Population description:

The Qatari Cohort: the Qatar Genome Program (QGP)[21] is a population-based project launched by the Qatar Foundation to generate a large-scale whole-genome sequence (WGS) dataset, in combination with comprehensive phenotypic information collected by the Qatar Biobank (QBB)[22]. All subjects included in the analysis were of Qatari Middle Eastern Arabian ancestry (Razali et al.in press). In this study, we use a cohort of 14,398 individuals with an average coverage of 30x. Data preprocessing and downstream quality control analyses for WGS data were conducted as recommended by the Covid19 Host Genetics Initiative study

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

protocol[23].

Italian Isolated cohorts: Three Italian cohorts belonging to the Italian network of Genetic Isolates (INGI) were involved in this study due to the availability of whole genome-sequence data. The selected populations localized in three different geographical areas of Italy: North-West (Val Borbera-VBI), North-East (Friuli Venezia Giulia-FVG) and South-East (Carlantino-CAR); In each cohort, a wide range of phenotypic data was collected for each participant (anthropometric traits, blood tests, sensory impairment, taste and food preferences, extensive personal and familial anamnesis). A total of 925 samples with low coverage (4x to 10x) WGS data were selected for the analyses[24].

Italian COVID-19 positive samples: a cohort of 189 samples which tested positive for the SARS-CoV-2 infection and collected at the Bambino Gesu' hospital in Rome was included in the study to provide information on the pattern of genetic variation in a group of selected genes in an outbred Italian cohort. Whole Exome Sequencing data was generated by the University of Tor Vergata. The samples are clustered in three groups, based on the disease severity: severe, extremely severe and asymptomatic[25].

All data analyzed was aligned to the reference genome's GRCh38 release, and functional annotations were obtained using the Ensembl VEP tool[26].

Principal component analysis

To highlight the study cohorts' population structure level, we performed a principal component analysis (PCA) using KING software[27]. Plink v1.9 software[28] was used to convert data from vcf to plink binary format. QGP and each INGI cohort results were projected into the 1000Genomes Project data[29]. To highlight the peculiar ancestry structure of the Qatari population, we also performed

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

an ancestry inference analysis using the software KING.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Genes selection and prioritization analyses

Literature curation process, GEL panel expert and IVA:

The candidate gene generation process, the initial candidate gene list ranking and curation are conveyed on the recent literature review to extract a list of genes involved in innate immunity and host-pathogen interaction. The primary gene list is curated according to the knowledge-literature base by the Ingenuity® Variant Analysis™ software from QIAGEN[30] and the viral gene panel expert from Genomics England (GEL)[31]. This list includes a total of 3617 genes (Table S1).

Candidate genes were annotated with the most common gene-ranking metrics using the loss of function intolerance score (pLI)[32] and the Residual Variation Intolerance Score (RVIS)[33]. In addition, we selected a list of 25 genes (Table S2) that were recently associated with COVID-19 susceptibility and severity[7] and overlapped with the primary list, extracting a subset of genes that underwent further analyses.

Population-based Gene constraints:

Two population-based gene metric scores, the Delta Singleton-Cohort variant score (DSC–accounting for the difference in singletons between coding and non-coding regions) and Sum Singleton-Cohort variant score (SSC– accounting for the sum of singletons variants in the coding and non-coding regions), were adopted to estimate the presence of specific pressures selection in the Qatari population as well as the Italian Isolated cohorts[34]. Only variants with a QUAL value above 30 were used in the calculation to limit the inclusion of genotyping error for variants with allele count (AC) equal to 1. Only scores calculated on canonical transcripts were selected. Only genes with scores values lower or equal to -2 and

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

greater or equal to 2 were retained in each population. These values represent the significant threshold that allows us to discriminate between a gene under constraint (DSC or SSC score ¡= -2) or under relaxation (DSC or SSC score ¿= 2). Since we aim to compare two populations with different structures and high levels of inbreeding, we also calculated the same set of scores for the closest ancestry populations of each study cohort. We used data from the gnomAD v3.1[35] call set, including 1000Genomes project samples and extracted information on the EUR, AFR and SAS superpopulations subset. The EUR subset was used as reference for the Italian samples[36], while the AFR and SAS subsets for the Qatari cohort [Razali et al, 2021, in press]. We defined two different levels of comparison: at the ancestry level, in which we selected all genes showing concordant selective signals between our study cohorts and their closest ancestry population, and at the population level, in which we selected only genes showing different behaviour between our target population and the reference. Since we were dealing with three Italian populations, comparing only with one reference, we selected genes that satisfied our criteria in at least one of the three target Italian populations.

On the other hand, we used two different references for the Qatari target population, so we selected all genes that meet our criteria in at least one reference population. Finally, we proceeded with the comparisons between our target populations, performing three sets of comparisons: population-specific, population-specific versus ancestry related and ancestry related comparisons (Table S3). Each comparison was performed separately for SSC and DSC scores. Finally, we generated a list of genes overlapping between SSC and DSC comparisons to select those genes that consistently showed opposite behaviour in terms of selection or relaxation in our target populations.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

We used the Fisher's test to compare DSC and SSC scores distributions between study cohorts, and reference populations. We performed a Shapiro-Wilk test to assess the normality of the score distribution in each cohort and an enrichment test to assess whether there was an enrichment in relaxed or constrained genes in our target populations versus the selected reference populations.

WES Covid-19 cohorts

Using Whole Exome Sequence data from a cohort of Covid-19 positive samples (n=189), we calculated singletons count and singletons density in the coding regions of the genes belonging to the shortlist generated, adjusting by sample size, and compared using Fisher's test against data from the other target and reference populations. In this cohort, each sample was characterized by a disease severity code. The disease severity classes are defined as follows: 1) Asymptomatic/Paucisymptomatic, 2) Severe, 3) Critical/life-threatening[37]. We used this information to investigate if we could identify any contribution of the singleton burden of the prioritized genes to the classification. A multinomial analysis with R was performed using age, gender and the singleton count as explanatory variables. We also analyzed the contribution of the prioritized genes to the outcome (Survived/Deceased) with a logistic regression model and the same covariates used in the multinomial analysis. We performed the analyses using both the whole-gene singleton count and the coding regions singleton count. A summary of the phenotype information is available in Table S4.

Results

Population stratification

As expected, the PCA analysis (Figure 1) showed a clear differentiation between QGP and INGI (European) ancestry. The

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Italian cohorts clustered with the European samples from the 1000Genomes Project reference data, while the QGP samples overlapped with clusters from different populations (AFR, SAS, AMR, EUR). Using the ancestry inference function provided by the KING software, we confirmed the presence of different sub-population clusters in the QGP cohort, highlighting that a considerable proportion of the analyzed samples (more than 4000 samples) belong to a 'missing' super population cluster (Figure S1). This is mainly due to the absence of population from the Near East in the 1000Genomes Project data. This outcome confirms results already obtained by different studies on the Italian populations[36] and on the first subset of nearly 6000 samples of the Qatari population[38].

Population based gene prioritization

For each population and each gene in the selected subset, we calculated two scores related to the presence of cohort-singletons variants. In Figure 2, weshow the distribution, among the 3617 genes selected, of the DSC score (Figure 2 top panel) and the SSC score (Figure 2 bottom panel), in each study cohort and the selected reference groups (EUR, AFR and SAS). We compared each target population score distribution with the relevant reference population (Table S5). All the INGI populations are significantly different from the reference EUR population for both scores. In contrast, the Qatari population significantly differs from the reference populations (AFR and SAS) in DSC score distribution, but not for the SSC score distribution. This pattern is also confirmed in the enrichment analyses of relaxed and constrained genes, in target populations versus reference populations, of relaxed and constrained genes. Exact Fisher's tests show enrichment in constrained genes in all the target populations when comparing DSC scores (Table S6). Similarly, if we consider the

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

SSC scores, all target populations do not show significant enrichment in constrained genes. Regarding the relaxed genes, though, we detected a sig- nificant enrichment in both DSC and SSC scores for the Italian populations (CAR, VBI and FVG) but not for the QGP cohort (Table S6). We used a threshold of -2 to define significant constraint and a thresholdof +2 to define a significant relaxation signal[39]. Results for the comparisons between our target populations are summarized in Table 1 and Table 2. Regarding the DSC score, we identified six genes with a signature of constraint in the QGP population and relaxation in at least one of the Italian populations. Two of those genes (*TTN* and *LRP1B*) are results of population-specific comparisons, and one (RICTOR) is the outcome of an ancestry-related comparison. Eight genes showed an opposite pattern of relaxation in the Qatari cohort and constraint in at least one Italian cohort. Among them, *RYR3* is the result of an ancestry related comparison. When comparing our target populations based on the overall burden of singletons in each gene (SSC score), we identified a total of 35 genes that behave differently between the Qatari population and at least one Italian population (Table 2). Seventeen of those genes showed a pattern of constraint in the Qatar population and relaxation in at least one Italian cohort. The *HELZ* gene was the only one arising from a population-specific comparison. The remaining eighteen genes showed a pat- tern of relaxation in the QGP dataset but a significant constraint in at least one of the other targets. In this subset, the CELSR2 gene is the result of a population-specific comparison. Since our focus is to identify genes that consistently show different selection signals among our cohorts, we selected a subset of genes for which both DSC and SSC scores are concordant: *ABCA7*, *FLNA*, *MUC5AC* (Table 3). Those three genes showed a consistent relaxation pat- tern in the QGP cohort while being always characterized by strong

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

signals of constraints in at least one Italian cohort. Interestingly, *FLNA* shows a significant signal of constraint in CAR and VBI cohorts while remaining neutral in the FVG dataset. The trend for the constraint signal is also replicated in all the reference cohorts selected. Data from other outbred populations from the 1000Genome project (EAS and AMR) confirm the trend of constraints (Table S7). *ABCA7* repeats the pattern observed for *FLNA*, in terms of target populations, with a significant constraint signal in the FVG cohort, and a trend of constraint in the CAR cohort, while being neutral in the VBI cohort. This time though, we can see how the outbred reference populations, plus the remaining super populations of 1000Genomes, are all in agreement, showing relaxation signals. Lastly, the *MUC5AC* gene shows a consistent pattern of significant constraint signal in all the Italian cohorts, but conversely, always a significantly relaxed pattern in all other populations.

Covid-19 cohort analysis

Next, we included a cohort of 189 COVID-19 positive samples (TOV cohort) and calculated the number of singleton variants in this subset for the three genes of interest. Table 4 shows the results of the comparisons with the other study populations and the reference populations. If we consider the whole gene, we can see how, for the *FLNA* gene, the TOV cohort shows a small difference in the singleton density when compared to the FVG cohort and a more significant difference with the VBI and QGP cohorts, while *ABCA7* and *MUC5AC* genes have consistently a significantly different pattern when compared with all reference and target populations (Table S8). If we consider only the coding part of each gene, we confirm the minor differences in the *FLNA* gene between the TOV cohort and the VBI and QGP cohorts. We also confirm the results

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

for *ABCA7* and *MUC5AC* (Table S9).

To investigate if the burden of singletons in the prioritized genes could contribute to the disease severity classification, we performed multinomial logistic regression analyses. Disease severity class was the response variable, and age, gender and singleton count the explanatory variables. When we consider the contribution of the burden of singleton in the whole gene, the multinomial analyses showed that only age and gender are important predictors for the disease severity classification (Table S10, Figure S2). Using the burden of singletons in the coding regions as parameters in the regression model resulted in age and gender being significant predictors for being in class 2 versus class 1 (p-values: 6.09E-07 and 0.04463 respectively) and belonging to class 3 versus class 1 (p-values: 6.42E-10 and 0.02035). Age resulted in a significant predictor of being in class 2 versus class 3 (p-value: 0.008617). Regarding genes contribution, the model highlighted only the *FLNA* gene as a significant predictorfor being in class 2 versus class 1 and for being in class 3 versus class 2 (all p-values: ¡ 2.2e-16) (Table S11).

We also retrieved the disease outcome (Survived/Deceased) information and used the same parameters to perform a logistic regression analysis. As a result, we can still see that age is a major predictor for the outcome (p-value: 1.50e-10) (Figure S3). We can also see a contribution of the *ABCA7* gene, but only when we consider the number of singletons in the whole gene (p-value: 0.0228) (Table S12 and Table S13).

Discussion

Since the H1N1 influenza pandemic in 1918, the ongoing COVID-19 pandemicis the most severe emergency we have met globally.

For the scientific community, this emergency has been a wake-up

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

call to join forces to fight back, investigate the effect of the virus on patients' health, and understand the infection's molecular mechanisms. All this ongoing effort is producing knowledge that is driving therapy and vaccine development. In this context, we focused on population-based statistics to characterize a subset of genes involved in the inflammation/immune response biological process. These statistics were obtained by analyzing populations with different ancestries and levels of inbreeding and consanguinity.

We performed a comparison between our study cohorts and their matched reference populations, according to principal component analysis (PCA) results. The comparisons within our target populations allowed us to show the different patterns in genetic constraints for a large subset of genes involved in the immune response, leading to the prioritization of a group of genes that we could define as "the most differentiated" in terms of signatures of genetic constraints. These differences are most prominent for three genes (*MUC5AC*, *ABCA7*, *FLNA*), which harbor a pattern of relaxation in the QGP cohort with respect to other cohorts analyzed. This different pattern of relaxation could be a hint for a different impact of the role of these genes in different populations. Two of the genes, *MUC5AC* and *FLNA*, have already been linked to the COVID-19 host response to different degrees. The *MUC5AC* gene is a gelforming mucin expressed in the lungs in response to infectious agents. This protein plays a protective role against inhaled pathogens, like influenza[40]. A recent study[41] compared levels of MUC5AC, MUC1 and MUC1-CT between critical ill COVID-19 patients and healthy controls, finding a significantly higher level of those proteins in the patients' mucus. It is also worth noting that another recent work from Kousathanas et al. reported a significant genome-wide association between variants in the MUC1 gene and

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

critical illness caused by[41, 42].

The second gene, *FLNA*, codes for the Filamin A protein, which has been identified as a putative interaction candidate with coronaviruses S protein and is involved in the coronavirus replication cycle[43]. A recent study showed that the *FLNA* gene is part of the host protein-protein interaction (PPI) network for the SARS-CoV-2 virus and among the targets of different drugs under development[44]. A loss-of-function mutation of the *FLNA* gene was reported in family adults with emphysema[45].

There is no study showing a direct link between the I gene and COVID-19 yet, but it has been proven that it is highly expressed in the reticuloendothelial system and modulates the phagocytosis activity[46, 47], though its function, like many other ABC-transporters, has yet to be clarified. Interestingly the BioGRID interactome database[48] lists physical interactions of *ABCA7* with *ADBR2*, *C5AR2* and *SGTB*, among others. Each one of these genes has been linked to COVID-19 host response in previous studies, in terms of interaction[49], therapy[50] and severity in case of pre-existing health conditions[51].

Altogether with the information available on the prioritized genes and the knowledge of the different evolution of the pandemic between Qatar and Italy, we performed a proof-of-concept analysis. Using the information provided by a cohort of COVID-19 positive samples from Italy, to try to identify, if present, the contribution of the amount of ultra-rare variants in those genes to the outcome of the disease (Survived/Deceased) and the severity (Asymptomatic/Paucisymptomatic, Severe, Critical/life-threatening).

From a cohort-based perspective, we can see differences in the distribution of singletons in the COVID-19 positive samples

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

regarding our study populations and their reference populations. This outcome suggests that those three genes could play a role in the description of the cohort and that investigating rare genetic variations occurring in those genetic regions could be a starting point to complete the characterization of those samples. With a subsequent approach, we applied logistic regression analyses to investigate the impact of the singleton burden in the three prioritized genes on the disease outcome and disease severity. While the contribution of age and sex is explicit and expected, these analyses suggested that the burden of singletons carried by each patient in the *ABCA7* gene could predict a worse outcome together with age. From the point of view of the disease severity, the burden of singleton in the *FLNA* gene could help discriminate samples with distinct levels of disease severity. In this case, though, our findings seem to be inconsistent since we find that having a lower burden of singletons is a predictor of developing a severe reaction . However, a high singleton burden is a predictor of developing a critical reaction. This finding can be better explained by looking at the distribution of singletons in our cohort stratified by disease severity. For the *FLNA* gene, all samples belonging to class 2 of severity do not carry any singleton. This feature could be introduced by one of the limitations of this study: the sample size of the COVID-19 positive cohort. Increasing the number of cases will undoubtedly allow us to have a better estimate of the singleton distribution. Moreover, in our model, we didn't include any of the risk factors that are already linked to a diverse response to the infection.

One last limitation could be represented by the inclusion of only one cohort of COVID-19 positive samples, for which only Whole Exome sequence data was available. We chose to include this cohort due to the phenotypical characterization, which allowed us to investigate our

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

hypothesis of a genetic contribution to the disease severity prioritized genes. Nevertheless, for all the cohorts involved, information on the COVID-19 affected samples is already being collected. That will allow us to produce more precise results with further analyses. To our knowledge, this is the first study performing a whole-genome population-level comparison between Arabian and European populations, both differently affected by the pandemic. Recent similar studies focused only on the ACE2 receptor and populations from the 1000Genomes Project[52] or compared allele frequencies on covid-19 related genes in the Brazilian population with data from the 1000Genome and gnomAD datasets[53].

With the development of new vaccines against SARS-CoV-2 infection, we are bound to see a decrease in adverse disease outcomes and disease severity among the immunized populations. However, our work could be a starting point to better prioritize genes that could be therapeutical targets in different populations. Moreover, with the increased knowledge obtained thanks to the many studies that focused on understanding virus-host interaction, we could extend our method to any new similar threat that should arise in the future.

Conclusion

we were able to identify three candidate genes that could be further investigated for their role in the COVID-19 infection, and we want to stress the message that harnessing the information provided by rare genetic variants, in this still evolving context, is proving increasingly useful to explain the different outcomes of this disease.

Supplementary information.

Figure S1: Inferred ancestry. Using the king software inferred ancestry function, we were able to correctly infer the ancestry of

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

our target populations, matching the already available data.

Figure S2: Disease severity classification in the TOV cohort. The box plots show that age is still a major discriminant factor related to disease severity, in the TOV cohort. Disease severity classes are defined as 1) Asymptomatic/Paucisymptomatic, 2) Severe, 3) Critical/life-threatening

Figure S3: Disease outcome classification in the TOV cohort. The box plots show the clear contribution of age, in the definition of the disease outcome classes.

Table S1: Initial candidate gene list of 3617 genes curated according to the knowledge-literature base by the Ingenuity® Variant Analysis™ software (https://www.qiagenbioinformatics.com/products/ingenuity-variant-analysis) from QIAGEN, Inc. (IVA) and the viral gene panel expert from Genomics England (GEL).

Table S2: List of genes recently associated with COVID-19 susceptibility and severity

Table S3: Details of the comparisons defined between Target and Reference populations

Table S4: Summary statistics of the phenotype information available for the TOV cohort.

Table S5: Comparison between scores distribution of target populations versus reference populations

Table S6: Enrichment tests in target populations versus reference populations

Table S7: DSC and SSC scores for FLNA, MUC5AC and ABCA7 in all 1000Genomes super populations

Table S8: Comparison of Gene Singleton Density between the Covid-19 positive cohort (TOV) and other target and reference populations. All singleton counts have been adjusted considering the

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

sample size of each cohort.

Table S9: Comparison of Coding Singleton Density between the Covid-19 positive cohort (TOV) and other target and reference populations. All singleton counts have been adjusted considering the sample size of each cohort.

Table S10: Multinomial analysis results, comparing classes of disease severity, considering the burden of singletons in the whole gene. Class 1: Asymptomatic/Paucisymptomatic; Class 2: Severe; Class 3: Critical/life-threatening. Only significant predictors are reported.

Table S11: Multinomial analysis results, comparing classes of disease severity, considering the burden of singletons in the coding regions of each gene. Class 1: Asymptomatic/Paucisymptomatic; Class 2: Severe; Class 3: Critical/life-threatening. Only significant predictors are reported.

Table S12: Logistic analysis results for disease outcome classification.

Only significant predictors are reported.

Table S13: Logistic analysis results for disease outcome classification considering the burden of singletons in the coding regions of each gene. Only significant predictors are reported.

Author Contributions

Conceptualization, M.C. and H.M.; Methodology, M.C., M.M.; Formal Analysis, M.C, Y.AS and C.S., H.M; Resources, D.C., A.N., I.Q., G.N., W.A., A.AT., S.S., A.A., S.I.I.; Writing Original Draft Preparation, H.M., C.M.,Y.AS.,C.S., M.M.; Writing Review Editing, H.M.,C.M.,Y.AS.,C.S., M.M.

Funding

This study was also supported in part by a grant of Regione Lazio

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

(Italy, Progetti di Gruppi di Ricerca 2020 A0375-2020-36663 GecoBiomark) to G.N. and A.N. and Rome Foundation (Italy, Prot 317A/I) to GN.

Institutional Review

The recruitment and sequencing of participants from the Qatar Biobank (QBB) were approved by the Hamad Medical Corporation Ethics Committee in 2011 and continued with QBB IRB from 2017 onwards and renewed annually (protocol IRB-A-QBB-2019-0017). The genetic analyses were approved by the QBB IRB protocol E-2020-QBB-Res-ACC- 0226-0130.

Data Availability

Access to the QGP data used for this study is through a dedicated portal by QGP (Accession ID: QF-QGP-RES-PUB-0226). The informed consent given by the study participants does not cover posting of participant level phenotype and genotype data of Qatar Biobank (QBB)/Qatar Genome Project (QGP) in public databases. Access to QBB/QGP data can be obtained through an established ISO-certified process by submitting a project request at https://www.qatarbiobank.org.qa/research/ how-apply which is subject to approval by the QBB IRB committee.

The genetic data belonging to INGI cohorts and analyzed in this manuscript have been submitted to the European Variation Archive (EVA) are accessible in Variant Call Format (VCF) at the following link: https://www.ebi.ac.uk/ena/data/view/PRJEB33648

Acknowledgments

We would like to thank the Qatar Biobank for data collection and phenotyping. The Qatar Genome Program and Qatar Biobank are both Research, Development Innovation entities within Qatar

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Foundation for Education, Science and Community Development. We would like to thank the inhabitants and local administrations of Friuli Venezia Giulia Region, Carlantino and Val Borbera that made these studies possible.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] EG, P.-H., J, B., D, F., L, S.: Hospitalization and Mortality among Black Patients and White Patients with Covid-19. The New England journal of medicine **382**(26), 2534–2543 (2020). https://doi.org/10.1056/NEJMSA2011686

[2] HM, A.K., HF, A.R., LJ, A.-R., AB, A.-S., Z, A.K., A, A.K., E, A.K., S, A.M., M, A.M., HE, A.R., MH, A.T., PV, C., AN, L., R, O., R, B., AA, B.: Epidemiological investigation of the first 5685 cases of SARS-CoV-2 infection in Qatar, 28 February-18 April 2020. BMJ open **10**(10) (2020). https://doi.org/10.1136/BMJOPEN-2020-040428

[3] AS, O., MA, A., J, D., RA, A., M, E., SH, S., TBH, I., A, Z., D, B., AM, A., A, B., B, A., R, E., AAM, H., MN, B., FB, A., SH, A.J., S, A., A, K., A, A., HM, A.-K.: The first consecutive 5000 patients with Coronavirus Disease 2019 from Qatar; a nation-wide cohort study. BMC infectious diseases **20**(1) (2020). https://doi.org/10.1186/S12879-020-05511-8

[4] Weekly epidemiological update on COVID-19 - 31 August 2021. https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---31-august-2021 Accessed 2021-09-30

[5] Milne, R.: Societal considerations in host genome testing for COVID-19. Genetics in Medicine 2020 22:9 **22**(9), 1464–1466 (2020). https://doi.org/10.1038/S41436-020-0861-Y

[6] E, P.-C., S, C., L, K., AD, B., K, R., D, P., S, W., N, P., MH, F., CD, R.,J, F., A, R., E, G., N, W., D, H., B, W., Y, W., A, M., F, G., W, O., A, K.,L, M., Z, Y., R, Z., C, Z., G, G., R, B., J, M., B, S., S, K., M, Z., C, H., DJ,P., C, H., J, Y., J, K., C, S., M, S.-H., P, K., L, T., A, H., SC, M., C, H., P,H., A, N., D, M., L, L., D, M., H, M., T, W., AC, P., A, R., X, S., CP, P.,A, F., A, T., M, C., R, S., K, R., L, M., PJM, O., MG, S., A, L., V, V., JF,W., JK, B.: Genetic mechanisms of critical illness in COVID-19. Nature **591**(7848), 92–98 (2021). https://doi.org/10.1038/S41586-020-03065-Y

[7] COVID-19 Host Genetics Initiative: Mapping the human genetic architecture of COVID-19. Nature (2021). https://doi.org/10.1038/ s41586-021-03767-x

[8] T, A., B, B., F, O., D, M., M, M.-V., MRL, M.R., P, Z., L, M., A, B., M, M., S, K., A, G., E, T., B, B., Y, S., K, B., Y, Z., AL, N., M, O., SJ, P., T, L.V., J, R., Q, P., P, S.-P., R, C., A, M.-N., JG, R., Y, T.-L., K, C., M, S., IA, D., NA, O., D, M., N, H., F, P., T, O., G, N., A, N., G, C., A, A., P, C., S, B., F, B., P, R.-Q., C, T., JL, F., J, R., LF, R., IG, B., AA, A., G, M., K, C., J, T., L, P.-S., A, S., M, G., A, P., LM, A., C, R.-G., C, F., O, C.-M., DE, P., RP,

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

d.D., S, K., G, A., OM, A., YT, B., P, B., P, B., D, S., CIE, S., AC, N., TM, C., LE, C., L, H., Q, P.-H., H, A., S, M., N, M., M, A., F, A.A., T, K., AN, S., CL, D., P, B., A, B., S, T., C, B., R, N., A, K.- K., AL, S., J, B., A, P., S, B.-D., SY, Z., V, B., RP, L., P, B., LD, N., L, A.,HC, S., E, J., A, A., V, S., A, C., Q, Z., JL, C.: X-linked recessive TLR7 deficiency in 1old with life-threatening COVID-19. Science immunology **6**(62), 4348 (2021). https://doi.org/10.1126/SCIIMMUNOL.ABL4348

[9] Kousathanas, A., Pairo-Castineira, E., Rawlik, K., Stuckey, A., Odhams, C.A., Walker, S., Russell, C.D., Malinauskas, T., Millar, J., Elliott, K.S., Griffiths, F., Oosthuyzen, W., Morrice, K., Keating, S., Wang, B., Rhodes, D., Klaric, L., Zechner, M., Parkinson, N., Bretherick, A.D., Siddiq, A., Goddard, P., Donovan, S., Maslove, D., Nichol, A., Semple, M.G., Zainy, T., Maleady-Crowe, F., Todd, L., Salehi, S., Knight, J., Elgar, G., Chan, G., Arumugam, P., Fowler, T.A., Rendon, A., Shankar-Hari, M., Summers, C., Hinds, C., Horby, P., McAuley, D., Montgomery, H., Openshaw, P.J.M., Wu, Y., Yang, J., Elliott, P., Walsh, T., Investigators, G., 23andMe, Initiative, C.-.H.G., Fawkes, A., Murphy, L., Rowan, K., Ponting, C.P., Vitart, V., Wilson, J.F., Scott, R.H., Clohisey, S., Moutsianas, L., Law, A., Caulfield, M.J., Baillie, J.K.: Whole genome sequencing identifies multiple loci for critical illness caused by COVID-19. medRxiv **15**, 2021–090221262965 (2021). https://doi.org/10.1101/2021.09.02.21262965

[10] Degenhardt, F., Ellinghaus, D., Juzenas, S., Lerga-Jaso, J., Wendorff, M., Maya-Miles, D., Uellendahl-Werth, F., ElAbd, H., Arora, J., Özer, O., Lenning, O.B., Myhre, R., Vadla, M.S., Wienbrandt, L., Ortiz, A.B., de Salazar, A., Chercoles, A.G., Palom, A., Ruiz, A., Mantovani, A., Zanella, A., Holten, A.R., Mayer, A., Bandera, A., Cherubini, A., Protti, A., Aghemo, A., Gerussi, A., Popov, A., Ramirez, A., Braun, A., Nebel, A., Barreira, A., Lleo, A., Teles, A., Kildal, A.B., Biondi, A., Ganna, A., Gori, A., Glück, A., Lind, A., Hinney, A., Nolla, A.C., Fracanzani, A.L., Cavallero, A., Dyrhol-Riise, A.M., Ruello, A., Julià, A., Muscatello, A., Pesenti, A., Voza, A., Rando-Segura, A., Solier, A., Cortes, B., Mateos, B., Nafria-Jimenez, B., Schaefer, B., Jensen, B., Bellinghausen, C., Maj, C., Ferrando, C., de la Horra, C., Quereda, C., Skurk, C., Thibeault, C., Scollo, C., Herr, C., Spinner, C.D., Lange, C., Hu, C., Lehmann, C., Cappadona, C., Azuure, C., study Group, C., (COVAS), C.-.A.S., Bianco, C., Sancho, C., Lihaug Hoff, D.A., Galimberti, D., Prati, D., Haschka, D., Jiménez, D., Pestaña, D., Toapanta, D., Wacker, E.M., Azzolini, E., Scarpini, E., Helbig, E.T., Urrechaga, E., Paraboschi, E.M., Pontali, E., Reverter, E., Calderón, E.J., Navas, E., Solligård, E., Contro, E., Arana, E., Garcia, F., Sánchez, F.G., Ceriotti, F., Martinelli-Boneschi, F., Peyvandi, F., Kurth, F., Blasi, F., Malvestiti, F., Medrano, F.J., Mesonero, F., Rodriguez-Frias, F., Hanses, F., Müller, F., Bellani, G., Grasselli, G., Pezzoli, G., Costantino, G., Albano, G., Bellelli, G., Citerio, G., Foti, G., Lamorte, G., Neb, H., My, I., Kurth, I., Hernández, I., Pink, I., de Rojas, I., Galván-Femenia, I., Holter, J.C., Afset, J.E., Heyckendorf, J., Damås, J.K., Rybniker, J., Altmüller, J., Ampuero, J., Banales, J.M., Badia, J.R., Dopazo, J., Schneider, J., Hov, J.R., Bergan, J., Barretina, J., Walter, J., Quero, J.H., Goikoetxea, J., Delgado, J., Guerrero, J.M., Fazaal, J., Kraft, J., Schröder, J., Risnes, K., Banasik, K., Müller, K.E., Gaede, K.I., Garcia-Etxebarria, K., Tonby, K.,

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Heggelund, L., Izquierdo-Sanchez, L., Bettini, L.R., Sumoy, L., Sander, L.E., Lippert, L.J., Terranova, L., Nkambule, L., Knopp, L., Gustad, L.T., Garbarino, L., Santoro, L., Téllez, L., Roade, L., Ostadreza, M., Intxausti, M., Rühlemann, M.C., Kogevinas, M., Riveiro-Barciela, M., Berger, M.M., Niemi, M.E.K., Gutiérrez-Stampa, M.A., Valsecchi, M.G., Hernandez-Tejero, M., Vehreschild, M.J.G.T., Manunta, M., D'Angiò, M., Cazzaniga, M., Grimsrud, M.M., Grimsrud, M.M., Cornberg, M., Nöthen, M.M., Marquié, M., Castoldi, M., Cordioli, M., Cecconi, M., D'Amato, M., Augustin, M., Tomasi, M., Boada, M., Dreher, M., Seilmaier, M.J., Joannidis, M., Wittig, M., Mazzocco, M., Rodríguez-Gandía, M., Ayo, N.I., Blay, N., Chueca, N., Montano, N., Ludwig, N., Marx, N., Mart´ınez, N., Group, N.S.-C.-.S., Cornely, O.A., Witzke, O., Palmieri,O., Group, P.C.-.S., Faverio, P., Bonfanti, P., Tentorio, P., Castro, P., Rodrigues, P.M., España, P.P., Hoffmann, P., Rosenstiel, P., Schom- mers, P., Suwalski, P., de Pablo, R., Ferrer, R., Bals, R., Gualtierotti, R., Gallego-Durán, R., Nieto, R., Carpani, R., Morilla, R., Badalamenti, S., Haider, S., Ciesek, S., May, S., Bombace, S., Marsal, S., Pigazz- ini, S., Klein, S., Rolker, S., Pelusi, S., Wilfling, S., Bosari, S., Brunak,S., Raychaudhuri, S., Schreiber, S., Heilmann-Heimbach, S., Aliberti, S., Ripke, S., Dudman, S., Force, T.H.C.-.T., Force, T.H.G.C.-.T., Bah- mer, T., Eggermann, T., Illig, T., Brenner, T., Karlsen, T.H., Feldt, T., Folseraas, T., Cejudo, T.G., Landmesser, U., Protzer, U., Hehr, U., Rimoldi, V., Skogen, V., Keitel, V., Kopfnagel, V., Friaza, V., Andrade, V., Moreno, V., Poller, W., Farre, X., Wang, X., Khodamoradi, Y., Karad- eniz, Z., Latiano, A., Goerg, S., Bacher, P., Koehler, P., Tran, F., Zoller,H., Schulte, E.C., Heidecker, B., Ludwig, K.U., Fernández, J., Romero- Gómez, M., Albillos, A., Invernizzi, P., Buti, M., Duga, S., Bujanda, L., Lenz, T.L., Asselta, R., de Cid, R., Valenti, L., Cáceres, M., Franke, A.: New susceptibility loci for severe COVID-19 by detailed GWAS analy- sis in European populations. medRxiv **9**(11), 2021–072121260624 (2021). https://doi.org/10.1101/2021.07.21.21260624

[11] The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. European journal of human genetics : EJHG **28**(6), 715–718 (2020). https://doi.org/10.1038/S41431-020-0636-6

[12] Bioscience Genomics involved in a genetic study on COVID-19 - Bioscience Institute. https://bioinst.com/en/ bioscience-genomics-involved-in-a-genetic-study-on-covid-19/

[13] VL, C., V, V., J, W., G, N., JKV, R.: Update on human genetic susceptibility to COVID-19: susceptibility to virus and response. Human genomics **15**(1) (2021). https://doi.org/10.1186/S40246-021-00356-X

[14] Abu-Raddad, L.J., Chemaitelly, H., Ayoub, H.H., Al Kanaani, Z., Al Khal, A., Al Kuwari, E., Butt, A.A., Coyle, P., Jeremijenko, A., Kaleeckal, A.H., Latif, A.N., Owen, R.C., Rahim, H.F.A., Al Abdulla, S.A., Al Kuwari, M.G., Kandy, M.C., Saeb, H., Ahmed, S.N.N., Al Romaihi, H.E., Bansal, D., Dalton, L., Al-Thani, M.H., Bertollini, R.: Characterizing the Qatar advanced-phase SARS-CoV-2 epidemic. Scientific Reports 2021 11:1 **11**(1), 1–15 (2021).

[15] Jeremijenko, A., Chemaitelly, H., Ayoub, H.H., Alishaq, M., Abou-Samra, A.-B., Ajmi, J.A.A.A.A., Ansari, N.A.A.A., Kanaani, Z.A., Khal, A.A., Kuwari, E.A., Al-Mohammed, A., Molawi, N.H.A.A., Naomi, H.M.A., Butt, A.A., Coyle, P., Kahlout, R.A.E., Gillani, I., Kaleeckal, A.H.,

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Masoodi, N.A., Thomas, A.G., Nafady-Hego, H., Latif, A.N., Shaik, R.M., Younes, N.B.M., Rahim, H.F.A., Yassine, H.M., Kuwari, M.G.A., Romaihi, H.E.A., Al-Thani, M.H., Bertollini, R., Abu-Raddad, L.J.: Herd Immunity against Severe Acute Respiratory Syndrome Coronavirus 2 Infection in 10 Communities, Qatar - Volume 27, Number 5—May 2021 - Emerging Infectious Diseases journal - CDC. Emerging Infectious Diseases **27**(5), 1343–1352 (2021). https://doi.org/10.3201/EID2705.204365

[16] FP, P., SJ, T., N, K., J, A., A, G., S, L., JL, P., G, P.M., ED, M., C, Z., R, B., KA, S., S, R., K, K., P, L., WV, K., D, C., RW, F., LL, H., Ö, T., H, N., A, S., S, Ü., DB, T., S, M., PR, D., U, , KU, J., WC, G.: Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. The New England journal of medicine **383**(27), 2603–2615 (2020). https://doi.org/10.1056/NEJMOA2034577

[17] LR, B., HM, E.S., B, E., K, K., S, F., R, N., D, D., SA, S., N, R., CB, C., J, M., S, K., N, S., J, S., A, B., C, F., H, S., K, N., L, C., P,G., H, J., D, F., M, M., J, M., L, P., J, L., BS, G., H, B., R, P., C, K., B,L., W, D., H, Z., S, H., M, I., J, M., T, Z.: Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. The New England journal of medicine **384**(5), 403–416 (2021). https://doi.org/10.1056/NEJMOA2035389

[18] Voysey, M., Clemens, S.A.C., Madhi, S.A., Weckx, L.Y., Folegatti, P.M., Aley, P.K., Angus, B., Baillie, V.L., Barnabas, S.L., Bhorat, Q.E., Bibi, S., Briner, C., Cicconi, P., Collins, A.M., Colin-Jones, R., Cutland, C.L., Darton, T.C., Dheda, K., Duncan, C.J.A., Emary, K.R.W., Ewer, K.J., Fairlie, L., Faust, S.N., Feng, S., Ferreira, D.M., Finn, A., Goodman, A.L., Green, C.M., Green, C.A., Heath, P.T., Hill, C., Hill, H., Hirsch, I., Hodgson, S.H.C., Izu, A., Jackson, S., Jenkin, D., Joe, C.C.D., Kerridge, S., Koen, A., Kwatra, G., Lazarus, R., Lawrie, A.M., Lelliott, A., Libri, V., Lillie, P.J., Mallory, R., Mendes, A.V.A., Milan, E.P., Minassian, A.M., McGregor, A., Morrison, H., Mujadidi, Y.F., Nana, A., O'Reilly, P.J., Padayachee, S.D., Pittella, A., Plested, E., Pollock, K.M., Ramasamy, M.N., Rhead, S., Schwarzbold, A.V., Singh, N., Smith, A., Song, R., Snape, M.D., Sprinz, E., Sutherland, R.K., Tarrant, R., Thomson, E.C., Török, M.E., Toshner, M., Turner, D.P.J., Vekemans, J., Villafana, T.L., Watson, M.E.E., Williams, C.J., Douglas, A.D., Hill, A.V.S., Lambe, T., Gilbert, S.C., Pollard, A.J., Aban, M., Abayomi, F., Abeyskera, K., Aboagye, J., Adam, M., Adams, K., Adamson, J., Adelaja, Y.A., Adewetan, G., Adlou, S., Ahmed, K., Akhalwaya, Y., Akhalwaya, S., Alcock, A., Ali, A., Allen, E.R., Allen, L., Almeida, T.C.D.S.C., Alves, M.P.S., Amorim, F., Andritsou, F., Anslow, R., Appleby, M., Arbe-Barnes, E.H.,Ariaans, M.P., Arns, B., Arruda, L., Azi, P., Azi, L., Babbage, G., Bailey, C., Baker, K.F., Baker, M., Baker, N., Baker, P., Baldwin, L., Baleanu, I., Bandeira, D., Bara, A., Barbosa, M.A.S., Barker, D., Barlow, G.D., Barnes, E., Barr, A.S., Barrett, J.R., Barrett, J., Bates, L., Batten, A., Beadon, K., Beales, E., Beckley, R., Belij-Rammerstorfer, S., Bell, J., Bel- lamy, D., Bellei, N., Belton, S., Berg, A., Bermejo, L., Berrie, E., Berry, L., Berzenyi, D., Beveridge, A., Bewley, K.R., Bexhell, H., Bhikha, S., Bhorat, A.E., Bhorat, Z.E., Bijker, E., Birch, G., Birch, S., Bird, A., Bird, O., Bisnauthsing, K., Bittaye, M., Blackstone, K., Blackwell, L., Bletchly, H., Blundell, C.L., Blundell, S.R., Bodalia, P., Boettger, B.C., Bolam, E., Boland, E., Bormans, D., Borthwick, N., Bowring, F., Boyd, A., Bradley, P., Brenner, T., Brown, P., Brown, C., Brown-O'Sullivan, C., Bruce, S., Brunt, E., Buchan, R., Budd, W.,

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Bulbulia, Y.A., Bull, M., Burbage, J., Burhan, H., Burn, A., Buttigieg, K.R., Byard, N., Puig, I.C., Calderon, G., Calvert, A., Camara, S., Cao, M., Cappuccini, F., Cardoso, J.R., Carr, M., Carroll, M.W., Carson-Stevens, A., Carvalho, Y.d.M., Carvalho, J.A.M., Casey, H.R., Cashen, P., Castro, T., Castro, L.C., Cathie, K.,Cavey, A., Cerbino-Neto, J., Chadwick, J., Chapman, D., Charlton, S., Chelysheva, I., Chester, O., Chita, S., Cho, J.-S., Cifuentes, L., Clark, E., Clark, M., Clarke, A., Clutterbuck, E.A., Collins, S.L.K., Conlon, C.P., Connarty, S., Coombes, N., Cooper, C., Cooper, R., Cornelissen, L., Cor- rah, T., Cosgrove, C., Cox, T., Crocker, W.E.M., Crosbie, S., Cullen, L., Cullen, D., Cunha, D.R.M.F., Cunningham, C., Cuthbertson, F.C., Guarda, S.N.F.D., da Silva, L.P., Damratoski, B.E., Danos, Z., Dantas, M.T.D.C., Darroch, P., Datoo, M.S., Datta, C., Davids, M., Davies, S.L., Davies, H., Davis, E., Davis, J., Davis, J., Nobrega, M.M.D.D., Kalid, L.M.D.O., Dearlove, D., Demissie, T., Desai, A., Marco, S.D., Maso, C.D., Dinelli, M.I.S., Dinesh, T., Docksey, C., Dold, C., Dong, T., Don- nellan, F.R., Santos, T.D., dos Santos, T.G., Santos, E.P.D., Douglas, N., Downing, C., Drake, J., Drake-Brockman, R., Driver, K., Drury, R., Dunachie, S.J., Durham, B.S., Dutra, L., Easom, N.J.W., van Eck, S., Edwards, M., Edwards, N.J., Muhanna, O.M.E., Elias, S.C., Elmore, M., English, M., Esmail, A., Essack, Y.M., Farmer, E., Farooq, M., Farrar, M., Farrugia, L., Faulkner, B., Fedosyuk, S., Felle, S., Feng, S., Silva, C.F.D., Field, S., Fisher, R., Flaxman, A., Fletcher, J., Fofie, H., Fok, H., Ford, K.J., Fowler, J., Fraiman, P.H.A., Francis, E., Franco, M.M., Frater, J., Freire, M.S.M., Fry, S.H., Fudge, S., Furze, J., Fuskova, M., Galian-Rubio, P., Galiza, E., Garlant, H., Gavrila, M., Geddes, A., Gibbons, K.A., Gilbride, C., Gill, H., Glynn, S., Godwin, K., Gokani, K., Goldoni, U.C., Goncalves, M., Gonzalez, I.G.S., Goodwin, J., Goondiwala, A., Gordon-Quayle, K., Gorini, G., Grab, J., Gracie, L., Greenland, M., Greenwood, N., Greffrath, J., Groenewald, M.M., Grossi, L., Gupta, G., Hackett, M., Hallis, B., Hamaluba, M., Hamilton, E., Hamlyn, J., Hammersley, D., Hanrath, A.T., Hanumunthadu, B., Harris, S.A., Harris, C., Harris, T., Harrison, T.D., Harrison, D., Hart, T.C., Hartnell, B., Hassan, S., Haughney, J., Hawkins, S., Hay, J., Head, I., Henry, J., Herrera, M.H., Hettle, D.B., Hill, J., Hodges, G., Horne, E., Hou, M.M., Houlihan, C., Howe, E., Howell, N., Humphreys, J., Humphries, H.E., Hurley, K., Huson, C., Hyder-Wright, A., Hyams, C., Ikram, S., Ishwarbhai, A., Ivan, M., Iveson, P., Iyer, V., Jackson, F., Jager, J.D., Jaumdally, S., Jeffers, H., Jesudason, N., Jones, B., Jones, K., Jones, E., Jones, C., Jorge, M.R., Jose, A., Joshi, A., Júnior, E.A.M.S., Kadziola, J., Kailath, R., Kana, F., Karampatsas, K., Kasanyinga, M., Keen, J., Kelly, E.J., Kelly, D.M., Kelly, D., Kelly, S., Kerr, D., Kfouri, R.d.Á., Khan, L., Khozoee, B., Kidd, S., Killen, A., Kinch, J., Kinch, P., King, L.D.W., King, T.B., Kingham, L., Klenerman, P., Knapper, F., Knight, J.C., Knott, D., Kol- eva, S., Lang, M., Lang, G., Larkworthy, C.W., Larwood, J.P.J., Law, R., Lazarus, E.M., Leach, A., Lees, E.A., Lemm, N.-M., Lessa, A., Leung, S., Li, Y., Lias, A.M., Liatsikos, K., Linder, A., Lipworth, S., Liu, S., Liu, X., Lloyd, A., Lloyd, S., Loew, L., Ramon, R.L., Lora, L., Lowthorpe, V., Luz, K., MacDonald, J.C., MacGregor, G., Madhavan, M., Mainwaring, D.O., Makambwa, E., Makinson, R., Malahleha, M., Malamatsho, R., Mallett, G., Mansatta, K., Maoko, T., Mapetla, K., Marchevsky, N.G., Marinou, S., Marlow, E., Marques,

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

G.N., Marriott, P., Marshall, R.P., Marshall, J.L., Martins, F.J., Masenya, M., Masilela, M., Masters, S.K., Mathew, M., Matlebjane, H., Matshidiso, K., Mazur, O., Mazzella, A., McCaughan, H., McEwan, J., McGlashan, J., McInroy, L., McIntyre, Z., McLenaghan, D., McRobert, N., McSwiggan, S., Megson, C., Mehdipour, S., Meijs, W., Mendonça, R.N.Á., Mentzer, A.J., Mirtorabi, N., Mitton, C., Mnyakeni, S., Moghaddas, F., Molapo, K., Moloi, M., Moore, M., Moraes-Pinto, M.I., Moran, M., Morey, E., Morgans, R., Morris, S., Morris, S., Morris, H.C., Morselli, F., Morshead, G., Morter, R., Mottal, L., Moultrie, A., Moya, N., Mpelembue, M., Msomi, S., Mugodi, Y., Mukhopadhyay, E., Muller, J., Munro, A., Munro, C., Murphy, S., Mweu, P., Myasaki, C.H., Naik, G., Naker, K., Nastouli, E., Nazir, A., Ndlovu, B., Neffa, F., Njenga, C., Noal, H., Noé, A., Novaes, G., Nugent, F.L., Nunes, G., O'Brien, K., O'Connor, D., Odam, M., Oelofse, S., Oguti, B., Olchawski, V., Oldfield, N.J., Oliveira, M.G., Oliveira, C., Oosthuizen, A., O'Reilly, P., Osborne, P., Owen, D.R.J., Owen, L., Owens, D., Owino, N., Pacurar, M., Paiva, B.V.B., Palhares, E.M.F., Palmer, S., Parkinson, S., Parracho, H.M.R.T., Parsons, K., Patel, D., Patel, B., Patel, F., Patel, K., Patrick-Smith, M., Payne, R.O., Peng, Y., Penn, E.J., Pennington, A., Alvarez, M.P.P., Perring, J., Perry, N., Perumal, R., Petkar, S., Philip, T., Phillips, D.J., Phillips, J., Phohu, M.K., Pickup, L., Pieterse, S., Piper, J., Pipini, D., Plank, M., Plessis, J.D., Pollard, S., Pooley, J., Pooran, A., Poulton, I., Powers, C., Presa, F.B., Price, D.A., Price, V., Primeira, M., Proud, P.C., Provstgaard-Morys, S., Pueschel, S., Pulido, D., Quaid, S., Rabara, R., Radford, A., Radia, K., Rajapaska, D., Rajeswaran, T., Ramos, A.S.F., Lopez, F.R., Rampling, T., Rand, J., Ratcliffe, H., Rawlinson, T., Rea, D., Rees, B., Reiné, J., Resuello-Dauti, M., Pabon, E.R., Ribiero, C.M., Ricamara, M., Richter, A., Ritchie, N., Ritchie, A.J., Robbins, A.J., Roberts, H., Robinson, R.E., Robinson, H., Rocchetti, T.T., Rocha, B.P., Roche, S., Rollier, C., Rose, L., Russell, A.L.R., Rossouw, L., Royal, S., Rudiansyah, I., Ruiz, S., Saich, S., Sala, C., Sale, J., Salman, A.M., Salvador, N., Salvador, S., Sampaio, M., Samson, A.D., Sanchez-Gonzalez, A., Sanders, H., Sanders, K., Santos, E., Guerra, M.F.S.S., Satti, I., Saunders, J.E., Saunders, C., Sayed, A., van der Loeff, I.S., Schmid, A.B., Schofield, E., Screaton, G., Seddiqi, S., Segireddy, R.R., Senger, R., Serrano, S., Shah, R., Shaik, I., Sharpe, H.E., Sharrocks, K., Shaw, R., Shea, A., Shepherd, A., Shepherd, J.G., Shiham, F., Sidhom, E., Silk, S.E., Moraes, A.C.d.S., Silva-Junior, G., Silva-Reyes, L., Silveira, A.D., Silveira, M.B.V., Sinha, J., Skelly, D.T., Smith, D.C., Smith, N., Smith, H.E., Smith, D.J., Smith, C.C., Soares, A., Soares, T., Solórzano, C., Sorio, G.L., Sorley, K., Sosa-Rodriguez, T., Souza, C.M.C.D.L., Souza, B.S.D.F., Souza, A.R., Spencer, A.J., Spina, F., Spoors, L., Stafford, L., Stamford, I., Starinskij, I., Stein, R., Steven, J., Stockdale, L., Stockwell, L.V., Strickland, L.H., Stuart, A.C., Sturdy, A., Sutton, N., Szigeti, A., Tahiri-Alaoui, A., Tanner, R., Taoushanis, C., Tarr, A.W., Taylor, K., Taylor, U., Taylor, I.J., Taylor, J., Naude, R.t.W., Themistocleous, Y., Themistocleous, A., Thomas, M., Thomas, K., Thomas, T.M., Thombrayil, A., Thompson, F., Thompson, A., Thompson, K., Thompson, A., Thomson, J., Thornton-Jones, V., Tighe, P.J., Tinoco, L.A., Tiongson, G., Tladinyane, B., Tomasicchio, M., Tomic, A., Tonks, S., Towner, J., Tran, N., Tree, J., Trillana, G., Trinham, C., Trivett, R., Truby, A., Tsheko, B.L., Turabi, A., Turner, R., Turner, C., Ulaszewska, M., Underwood, B.R., Varughese, R., Ver-

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

bart, D., Verheul, M., Vichos, I., Vieira, T., Waddington, C.S., Walker, L., Wallis, E., Wand, M., Warbick, D., Wardell, T., Warimwe, G., Warren, S.C., Watkins, B., Watson, E., Webb, S., Webb-Bridges, A., Webster, A., Welch, J., Wells, J., West, A., White, C., White, R., Williams, P., Williams, R.L., Winslow, R., Woodyer, M., Worth, A.T., Wright, D., Wroblewska, M., Yao, A., Zimmer, R., Zizi, D., Zuidewind, P.: Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. The Lancet **397**(10269), 99–111 (2021). https://doi.org/10.1016/S0140-6736(20)32661-1

[19] Logunov, D.Y., Dolzhikova, I.V., Shcheblyakov, D.V., Tukhvatulin, A.I., Zubkova, O.V., Dzharullaeva, A.S., Kovyrshina, A.V., Lubenets, N.L., Grousova, D.M., Erokhova, A.S., Botikov, A.G., Izhaeva, F.M., Popova, O., Ozharovskaya, T.A., Esmagambetov, I.B., Favorskaya, I.A., Zrelkin, D.I., Voronina, D.V., Shcherbinin, D.N., Semikhin, A.S., Simakova, Y.V., Tokarskaya, E.A., Egorova, D.A., Shmarov, M.M., Nikitenko, N.A., Gushchin, V.A., Smolyarchuk, E.A., Zyryanov, S.K., Borisevich, S.V., Naroditsky, B.S., Gintsburg, A.L.: Safety and efficacy of an rAd26 and rAd5 vector-based heterologous prime boost COVID-19 vaccine an interim analysis of a randomised controlled phase 3 trial in Rus- sia. The Lancet **397**(10275), 671–681 (2021). https://doi.org/10.1016/ S0140-6736(21)00234-8

[20] Stolfi, P., Manni, L., Soligo, M., Vergni, D., Tieri, P.: Designing a Network Proximity-Based Drug Repurposing Strategy for COVID-19. Frontiers in Cell and Developmental Biology **0**, 1021 (2020). https://doi.org/10.3389/FCELL.2020.545089

[21] Mbarek, H., Gandhi, G.D., Selvaraj, S., Al-Muftah, W., Badji, R., Al-Sarraj, Y., Saad, C., Darwish, D., Alvi, M., Fadl, T., Yasin, H., Al-Kuwari, F., Razali, R.M., Aamer, W., Abbaszadech, F., Ahmed, I., Mokrab, Y., Suhre, K., Albagha, O., Fakhro, K., Badii, R., Ismail, S.I., Althani, A., Consortium, T.Q.G.P.R.: Qatar Genome: Insights on Genomics from the Middle East. medRxiv, 2021–091921263548 (2021). https://doi.org/10.1101/2021.09.19.21263548

[22] A, A.T., E, F., S, P., A, A.M., Z, S., F, Q., N, A.: Qatar Biobank Cohort Study: Study Design and First Results. American journal of epidemiology **188**(8), 1420–1433 (2019). https://doi.org/10.1093/AJE/KWZ084

[23] Mapping the human genetic architecture of COVID-19. Nature 2021, 1–8 (2021). https://doi.org/10.1038/S41586-021-03767-X

[24] Cocca, M., Barbieri, C., Concas, M.P., Robino, A., Brumat, M., Gandin, I., Trudu, M., Sala, C.F., Vuckovic, D., Girotto, G., Matullo, G., Polasek, O., Kolčić, I., Gasparini, P., Soranzo, N., Toniolo, D., Mezzavilla, M.: A bird's-eye view of Italian genomic variation through whole-genome sequencing. European Journal of Human Genetics **28**(4), 435–444 (2020). https://doi.org/10.1038/s41431-019-0551-x

[25] Latini, A., Agolini, E., Novelli, A., Borgiani, P., Giannini, R., Gravina, P., Smarrazzo, A., Dauri, M., Andreoni, M., Rogliani, P., Bernardini, S., Helmer-citterich, M., Biancolella, M., Novelli, G.: COVID-19 and genetic variants of protein involved in the SARS-CoV-2 entry into the host cells. Genes **11**(9), 1–8 (2020). https://doi.org/10.3390/genes11091010

[26] McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., Cunningham, F.: The Ensembl Variant Effect Predictor. Genome Biology 2016 17:1 **17**(1), 1–14 (2016). https://doi.org/10.1186/

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

S13059-016-0974-4

[27] Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., Chen, W.-M.: Robust relationship inference in genome-wide association studies. Bioinformatics **26**(22), 2867–2873 (2010). https://doi.org/10.1093/BIOINFORMATICS/BTQ559

[28] Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J.: Second-generation PLINK: Rising to the challenge of larger and richer datasets. GigaScience **4**(1), 7 (2015) arXiv:1410.4803. https://doi.org/10.1186/s13742-015-0047-8

[29] Auton, A., Abecasis, G.R., Altshuler, D.M., Durbin, R.M., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., Flicek, P., Gabriel, S.B., Gibbs, R.A., Green, E.D., Hurles, M.E., Knoppers, B.M., Korbel, J.O., Lander, E.S., Lee, C., Lehrach, H., Mardis, E.R., Marth, G.T., McVean, G.A., Nickerson, D.A., Schmidt, J.P., Sherry, S.T., Wang, J., Wilson, R.K., Boerwinkle, E., Doddapaneni, H., Han, Y., Korchina, V., Kovar, C., Lee, S., Muzny, D., Reid, J.G., Zhu, Y., Chang, Y., Feng, Q., Fang, X., Guo, X., Jian, M., Jiang, H., Jin, X., Lan, T., Li, G., Li, J., Li, Y., Liu, S., Liu, X., Lu, Y., Ma, X., Tang, M., Wang, B., Wang, G., Wu, H., Wu, R., Xu, X., Yin, Y., Zhang, D., Zhang, W., Zhao, J., Zhao, M., Zheng, X., Gupta, N., Gharani, N., Toji, L.H., Gerry, N.P., Resch, A.M., Barker, J., Clarke, L., Gil, L., Hunt, S.E., Kelman, G., Kulesha, E., Leinonen, R., McLaren, W.M., Radhakrishnan, R., Roa, A., Smirnov, D., Smith, R.E., Streeter, I., Thormann, A., Toneva, I., Vaughan, B., Zheng-Bradley, X., Grocock, R., Humphray, S., James, T., Kingsbury, Z., Sudbrak, R., Albrecht, M.W., Amstislavskiy, V.S., Borodina, T.A., Lienhard, M., Mertes, F., Sultan, M., Timmermann, B., Yaspo, M.L., Fulton, L., Ananiev, V., Belaia, Z., Beloslyudtsev, D., Bouk, N., Chen, C., Church, D., Cohen, R., Cook, C., Garner, J., Hefferon, T., Kimelman, M., Liu, C., Lopez, J., Meric, P., O'Sullivan, C., Ostapchuk, Y., Phan, L., Ponomarov, S., Schneider, V., Shekhtman, E., Sirotkin, K., Slotta, D., Zhang, H., Balasubramaniam, S., Burton, J., Danecek, P., Keane, T.M., Kolb-Kokocinski, A., McCarthy, S., Stalker, J., Quail, M., Davies, C.J., Gollub, J., Webster, T., Wong, B., Zhan, Y., Campbell, C.L., Kong, Y., Marcketta, A., Yu, F., Antunes, L., Bainbridge, M., Sabo, A., Huang, Z., Coin, L.J.M., Fang, L., Li, Q., Li, Z., Lin, H., Liu, B., Luo, R., Shao, H., Xie, Y., Ye, C., Yu, C., Zhang, F., Zheng, H., Zhu, H., Alkan, C., Dal, E., Kahveci, F., Garrison, E.P., Kural, D., Lee, W.P., Leong, W.F., Stromberg, M., Ward, A.N., Wu, J., Zhang, M., Daly, M.J., DePristo, M.A., Handsaker, R.E., Banks, E., Bhatia, G., Del Angel, G., Genovese, G., Li, H., Kashin, S., McCarroll, S.A., Nemesh, J.C., Poplin, R.E., Yoon, S.C., Lihm, J., Makarov, V., Gottipati, S., Keinan, A., Rodriguez-Flores, J.L., Rausch, T., Fritz, M.H., Stütz, A.M., Beal, K., Datta, A., Herrero, J., Ritchie, G.R.S., Zerbino, D., Sabeti, P.C., Shlyakhter, I., Schaffner, S.F., Vitti, J., Cooper, D.N., Ball, E.V., Stenson, P.D., Barnes, B., Bauer, M., Cheetham, R.K., Cox, A., Eberle, M., Kahn, S., Murray, L., Peden, J., Shaw, R., Kenny, E.E., Batzer, M.A., Konkel, M.K., Walker, J.A., MacArthur, D.G., Lek, M., Herwig, R., Ding, L., Koboldt, D.C., Larson, D., Ye, K., Gravel, S., Swaroop, A., Chew, E., Lappalainen, T., Erlich, Y., Gymrek, M., Willems, T.F., Simpson, J.T., Shriver, M.D., Rosenfeld, J.A., Bustamante, C.D., Montgomery, S.B., De La Vega, F.M., Byrnes,J.K., Carroll, A.W., DeGorter, M.K., Lacroute, P., Maples, B.K., Mar- tin, A.R., Moreno-Estrada, A., Shringarpure, S.S., Zakharia,

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

F., Halperin, E., Baran, Y., Cerveira, E., Hwang, J., Malhotra, A., Plewczynski, D., Radew, K., Romanovitch, M., Zhang, C., Hyland, F.C.L., Craig, D.W., Christoforides, A., Homer, N., Izatt, T., Kurdoglu, A.A., Sinari, S.A., Squire, K., Xiao, C., Sebat, J., Antaki, D., Gujral, M., Noor, A., Ye, K., Burchard, E.G., Hernandez, R.D., Gignoux, C.R., Haussler, D., Katzman, S.J., Kent, W.J., Howie, B., Ruiz-Linares, A., Dermitzakis, E.T., Devine, S.E., Kang, H.M., Kidd, J.M., Blackwell, T., Caron, S., Chen, W., Emery,S., Fritsche, L., Fuchsberger, C., Jun, G., Li, B., Lyons, R., Scheller, C.,Sidore, C., Song, S., Sliwerska, E., Taliun, D., Tan, A., Welch, R., Wing,M.K., Zhan, X., Awadalla, P., Hodgkinson, A., Li, Y., Shi, X., Quitadamo, A., Lunter, G., Marchini, J.L., Myers, S., Churchhouse, C., Delaneau, O., Gupta-Hinch, A., Kretzschmar, W., Iqbal, Z., Mathieson, I., Menelaou, A., Rimmer, A., Xifara, D.K., Oleksyk, T.K., Fu, Y., Liu, X., Xiong, M., Jorde, L., Witherspoon, D., Xing, J., Browning, B.L., Browning, S.R., Hormozdiari, F., Sudmant, P.H., Khurana, E., Tyler-Smith, C., Albers, C.A., Ayub, Q., Chen, Y., Colonna, V., Jostins, L., Walter, K., Xue, Y., Gerstein, M.B., Abyzov, A., Balasubramanian, S., Chen, J., Clarke, D., Fu, Y., Harmanci, A.O., Jin, M., Lee, D., Liu, J., Mu, X.J., Zhang, J., Zhang, Y., Hartl, C., Shakir, K., Degenhardt, J., Meiers, S., Raeder, B., Casale, F.P., Stegle, O., Lameijer, E.W., Hall, I., Bafna, V., Michaelson, J., Gardner, E.J., Mills, R.E., Dayama, G., Chen, K., Fan, X., Chong, Z., Chen, T., Chaisson, M.J., Huddleston, J., Malig, M., Nelson, B.J., Parrish, N.F., Blackburne, B., Lindsay, S.J., Ning, Z., Zhang, Y., Lam, H., Sisu,C., Challis, D., Evani, U.S., Lu, J., Nagaswamy, U., Yu, J., Li, W., Habegger, L., Yu, H., Cunningham, F., Dunham, I., Lage, K., Jespersen, J.B., Horn, H., Kim, D., Desalle, R., Narechania, A., Sayres, M.A.W., Mendez, F.L., Poznik, G.D., Underhill, P.A., Mittelman, D., Banerjee, R., Cerezo, M., Fitzgerald, T.W., Louzada, S., Massaia, A., Yang, F., Kalra, D., Hale,W., Dan, X., Barnes, K.C., Beiswanger, C., Cai, H., Cao, H., Henn, B.,Jones, D., Kaye, J.S., Kent, A., Kerasidou, A., Mathias, R., Ossorio, P.N.,Parker, M., Rotimi, C.N., Royal, C.D., Sandoval, K., Su, Y., Tian, Z.,Tishkoff, S., Via, M., Wang, Y., Yang, H., Yang, L., Zhu, J., Bodmer, W.,Bedoya, G., Cai, Z., Gao, Y., Chu, J., Peltonen, L., Garcia-Montero, A., Orfao, A., Dutil, J., Martinez-Cruzado, J.C., Mathias, R.A., Hennis, A., Watson, H., McKenzie, C., Qadri, F., LaRocque, R., Deng, X., Asogun, D., Folarin, O., Happi, C., Omoniwa, O., Stremlau, M., Tariyal, R., Jal- low, M., Joof, F.S., Corrah, T., Rockett, K., Kwiatkowski, D., Kooner, J., Hien, T.T., Dunstan, S.J., ThuyHang, N., Fonnie, R., Garry, R., Kanneh,L., Moses, L., Schieffelin, J., Grant, D.S., Gallo, C., Poletti, G., Saleheen, D., Rasheed, A., Brooks, L.D., Felsenfeld, A.L., McEwen, J.E., Vaydyle- vich, Y., Duncanson, A., Dunn, M., Schloss, J.A.: A global reference forhuman genetic variation. Nature 2015 526:7571 **526**(7571), 68–74 (2015). https://doi.org/10.1038/NATURE15393QIAGEN-Inc.: Ingenuity Variant Analysis (IVA). https://www.qiagenbioinformatics.com/products/ingenuity-variant-analysis

[30] Martin, A.R., Williams, E., Foulger, R.E., Leigh, S., Daugherty, L.C., Niblock, O., Leong, I.U.S., Smith, K.R., Gerasimenko, O., Haraldsdottir, E., Thomas, E., Scott, R.H., Baple, E., Tucci, A., Brittain, H., de Burca, A., Ibañez, K., Kasperaviciute, D., Smedley, D., Caulfield, M., Rendon, A., McDonagh, E.M.: PanelApp crowdsources expert knowledge to establish

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

consensus diagnostic gene panels. Nature Genetics 2019 51:11 **51**(11), 1560–1565 (2019). https://doi.org/10.1038/s41588-019-0528-2

[31] Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., Tukiainen, T., Birnbaum, D.P., Kosmicki, J.A., Duncan, L.E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D.N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M.I., Moonshine, A.L., Natarajan, P., Orozco, L., Peloso, G.M., Poplin, R., Rivas, M.A., Ruano-Rubio, V., Rose, S.A., Ruderfer, D.M., Shakir, K., Stenson, P.D., Stevens, C., Thomas, B.P., Tiao, G., Tusie-Luna, M.T., Weisburd, B., Won, H.-H., Yu, D., Altshuler, D.M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J.C., Gabriel, S.B., Getz, G., Glatt, S.J., Hultman, C.M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Neale, B.M., Palotie, A., Purcell, S.M., Saleheen, D., Scharf, J.M., Sklar, P., Sullivan, P.F., Tuomilehto, J., Tsuang, M.T., Watkins, H.C., Wilson, J.G., Daly, M.J., MacArthur, D.G.: Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016 536:7616 **536**(7616), 285–291 (2016). https://doi.org/10.1038/NATURE19057

[32] Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S., Goldstein, D.B.: Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes. PLOS Genetics **9**(8), 1003709 (2013). https://doi.org/10.1371/JOURNAL.PGEN.1003709

[33] Mezzavilla, M., Cocca, M., Guidolin, F., Gasparini, P.: A population-based approach for gene prioritization in understanding complex traits. Human Genetics 2020 139:5 **139**(5), 647–655 (2020). https://doi.org/10.1007/S00439-020-02152-4

[34] li, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., Gauthier, L.D., Brand, H., Solomonson, M., Watts, N.A., Rhodes, D., Singer-Berk, M., England, E.M., Seaby, E.G., Kosmicki, J.A., Wal- ters, R.K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J.X., Samocha, K.E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A.H., Minikel, E.V., Weisburd, B., Lek, M., Ware, J.S., Vittal, C., Armean, I.M., Bergelson, L., Cibul- skis, K., Connolly, K.M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltz- man, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M.E., Neale, B.M., Daly, M.J., MacArthur, D.G.: The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020 581:7809 581(7809), 434–443 (2020). https://doi.org/10. 1038/S41586-020-2308-7

[35] Esko, T., Mezzavilla, M., Nelis, M., Borel, C., Debniak, T., Jakkula, E., Julia, A., Karachanak, S., Khrunin, A., Kisfali, P., Krulisova, V., Kučinskiené, Z.A., Rehnström, K., Traglia, M., Nikitina-Zake, L., Zimprich, F., Antonarakis, S.E., Estivill, X., Glavač, D., Gut, I., Klovins, J., Krawczak, M., Kučinskas, V., Lathrop, M., Macek, M., Marsal, S., Meitinger, T., Melegh, B., Limborska, S., Lubinski, J., Paolotie, A., Schreiber, S., Toncheva, D., Toniolo, D., Wichmann, H.E., Zimprich, A., Metspalu, M., Gasparini, P., Metspalu, A., D'Adamo, P.: Genetic characterization of northeastern Italian population isolates in the context of

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

broader European genetic diversity. European Journal of Human Genetics **21**(6), 659–665 (2013). https://doi.org/10.1038/ejhg.2012.229

[36] Latini, A., Agolini, E., Novelli, A., Borgiani, P., Giannini, R., Gravina, P., Smarrazzo, A., Dauri, M., Andreoni, M., Rogliani, P., Bernardini, S., Helmer-citterich, M., Biancolella, M., Novelli, G.: COVID-19 and genetic variants of protein involved in the SARS-CoV-2 entry into the host cells. Genes **11**(9), 1–8 (2020). https://doi.org/10.3390/genes11091010

[37] Thareja, G., Al-Sarraj, Y., Belkadi, A., Almotawa, M., Ismail, S., Al-Muftah, W., Badji, R., Mbarek, H., Darwish, D., Fadl, T., Yasin, H., Ennaifar, M., Abdellatif, R., Alkuwari, F., Alvi, M., Al-Sarraj, Y., Saad, C., Althani, A., Fethnou, E., Qafoud, F., Alkhayat, E., Afifi, N., Tomei, S., Liu, W., Lorenz, S., Syed, N., Almabrazi, H., Vempalli, F.R., Temanni, R., Saqri, T.A., Khatib, M., Hamza, M., Zaid, T.A., El Khouly, A., Pathare, T., Poolat, S., Al-Ali, R., Albagha, O., Al-Khodor, S., Alshafai, M., Badii, R., Chouchane, L., Estivill, X., Fakhro, K., Mokrab, Y., Puthen, J., Tatari, Z., Suhre, K., Albagha, O.M.E.: Whole genome sequencing in the Middle Eastern Qatari population identifies genetic associations with 45 clinically relevant traits. Nature Communications **12**(1) (2021). https://doi.org/10.1038/s41467-021-21381-3

[38] Mezzavilla, M., Cocca, M., Guidolin, F., Gasparini, P.: A population-based approach for gene prioritization in understanding complex traits. Human Genetics **139**(5), 647–655 (2020). https://doi.org/10.1007/s00439-020-02152-4

[39] Ehre, C., Worthington, E.N., Liesman, R.M., Grubb, B.R., Barbier, D., O'Neal, W.K., Sallenave, J.M., Pickles, R.J., Boucher, R.C.: Overexpressing mouse model demonstrates the protective role of Muc5ac in the lungs. Proceedings of the National Academy of Sciences of the United States of America **109**(41), 16528–16533 (2012). https://doi.org/10.1073/pnas.1206552109

[40] Lu, W., Liu, X., Wang, T., Liu, F., Zhu, A., Lin, Y., Luo, J., Ye, F., He, J., Zhao, J., Li, Y., Zhong, N.: Elevated MUC1 and MUC5AC mucin protein levels in airway mucus of critical ill COVID-19 Patients. Journal of Medical Virology **93**(2), 582–584 (2021). https://doi.org/10.1002/JMV.26406

[41] Kousathanas, A., Pairo-Castineira, E., Rawlik, K., Stuckey, A., Odhams, C.A., Walker, S., Russell, C.D., Malinauskas, T., Millar, J., Elliott, K.S., Griffiths, F., Oosthuyzen, W., Morrice, K., Keating, S., Wang, B., Rhodes, D., Klaric, L., Zechner, M., Parkinson, N., Bretherick, A.D., Siddiq, A., Goddard, P., Donovan, S., Maslove, D., Nichol, A., Semple, M.G., Zainy, T., Maleady-Crowe, F., Todd, L., Salehi, S., Knight, J., Elgar, G., Chan, G., Arumugam, P., Fowler, T.A., Rendon, A., Shankar-Hari, M., Summers, C., Hinds, C., Horby, P., McAuley, D., Montgomery, H., Openshaw, P.J.M., Wu, Y., Yang, J., Elliott, P., Walsh, T., Investigators, G., Investigators, a., Initiative, C.-.H.G., Fawkes, A., Murphy, L., Rowan, K., Ponting, C.P., Vitart, V., Wilson, J.F., Scott, R.H., Clohisey, S., Moutsianas, L., Law, A., Caulfield, M.J., Baillie, J.K.: Whole genome sequencing identifies multiple loci for critical illness caused by COVID-19. medRxiv **15**, 2021–090221262965 (2021). https://doi.org/10.1101/2021.09.02.21262965

[42] Trincone, A., Schwegmann-Weßels, C.: Looking for a needle in a haystack: Cellular proteins that may interact with the tyrosine-based sorting signal of the TGEV S protein. Virus Research **202**, 3–11 (2015). https://doi.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

org/10.1016/j.virusres.2014.11.029

[43] Chakrabarty, B., Das, D., Bulusu, G., Roy, A.: Network-Based Analysis of Fatal Comorbidities of COVID-19 and Potential Therapeutics. IEEE/ACM Transactions on Computational Biology and Bioinformatics **18**(4), 1271–1280 (2021). https://doi.org/10.1109/TCBB.2021.3075299

[44] Valentin, V., Bervar, J.F., Vincent-Delorme, C., Smol, T., Wemeau, L., Remy, M., Le Rouzic, O., Chenivesse, C.: Filamin A Mutations: A New Cause of Unexplained Emphysema in Adults? Chest **159**(3), 131–135 (2021). https://doi.org/10.1016/j.chest.2020.10.003

[45] Jehle, A.W., Gardai, S.J., Li, S., Linsel-Nitschke, P., Morimoto, K., Janssen, W.J., Vandivier, R.W., Wang, N., Greenberg, S., Dale, B.M.Qin, C., Henson, P.M., Tall, A.R.: ATP-binding cassette transporter A7 enhances phagocytosis of apoptotic cells and associated ERK sig- naling in macrophages. Journal of Cell Biology **174**(4), 547–556 (2006). https://doi.org/10.1083/jcb.200601030

[46] Iwamoto, N., Abe-Dohmae, S., Sato, R., Yokoyama, S.: ABCA7 expression is regulated by cellular cholesterol through the SREBP2 pathway and associated with phagocytosis. Journal of Lipid Research **47**(9), 1915–1927 (2006). https://doi.org/10.1194/jlr.M600127-JLR200

[47] Oughtred, R., Rust, J., Chang, C., Breitkreutz, B.J., Stark, C., Willems, A., Boucher, L., Leung, G., Kolas, N., Zhang, F., Dolma, S., Coulombe-Huntington, J., Chatr-aryamontri, A., Dolinski, K., Tyers, M.: The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein Science **30**(1), 187–200 (2021). https://doi.org/10.1002/pro.3978

[48] Zhou, Y., Hou, Y., Shen, J., Mehra, R., Kallianpur, A., Culver, D.A., Gack, M.U., Farha, S., Zein, J., Comhair, S., Fiocchi, C., Stappenbeck, T., Chan, T., Eng, C., Jung, J.U., Jehi, L., Erzurum, S., Cheng, F.: A network medicine approach to investigation and population-based validation of disease manifestations and drug repurposing for COVID-19. PLoS Biology **18**(11) (2020). https://doi.org/10.1371/journal.pbio.3000970

[49] Woodruff, T.M., Shukla, A.K.: The Complement C5a-C5aR1 GPCR Axis in COVID-19 Therapeutics. Trends Immunol (2020). https://doi.org/10.1016/j.it.2020.09.008. https://pubmed.ncbi.nlm.nih.gov/33023856/

[50] Lee, M.E., Chang, Y., Ahmadinejad, N., Johnson-Agbakwu, C.E., Bailey, C., Liu, L.: COVID-19 Mortality is Associated with Impaired Innate Immunity in Pre-existing Health Conditions. bioRxiv, 2021–0531446476 (2021). https://doi.org/10.1101/2021.05.31.446476

[51] Cao, Y., Li, L., Feng, Z., Wan, S., Huang, P., Sun, X., Wen, F., Huang, X., Ning, G., Wang, W.: Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. Cell Discov (2020). https://doi.org/10.1038/s41421-020-0147-1. https://pubmed.ncbi.nlm.nih.gov/32133153/

[52] Secolin, R., de Araujo, T.K., Gonsales, M.C., Rocha, C.S., Naslavsky, M., Marco, L.D., Bicalho, M.A.C., Vazquez, V.L., Zatz, M., Silva, W.A., Lopes-Cendes, I.: Genetic variability in COVID-19-related genes in the Brazilian population. Human Genome Variation **8**(1) (2021). https://doi.org/10.1038/s41439-021-00146-w

Table 1: Results from comparison of DSC scores between target cohorts (CAR, FVG, VBI, QGP) and the relevant reference superpopulations from the 1000 Genomes Project (EUR, AFR, SAS). The last column refers to the nature of the comparison carried out, as detailed in Supplementary table 3.

		DSC score							
Transcript id	**Gene name**	**QGP**	**CAR**	**FVG**	**VBI**	**EUR**	**AFR**	**SAS**	**Comparison**
ENST00000369850	*FLNA*	3.854	-2.435	0.272	-2.510	-2.399	-2.166	-1.879	C5
ENST00000350763	*TNC*	3.370	-3.792	1.388	-0.631	2.666	1.838	2.187	C4
ENST00000389048	*ALK*	2.575	3.651	0.290	-4.098	3.388	3.212	2.852	C4
ENST00000263094	*ABCA7*	2.566	-0.433	-2.168	0.071	3.020	2.681	2.150	C4
ENST00000647814	*ABCC2*	2.528	-3.466	0.467	3.004	2.562	2.877	2.508	C4
ENST00000621226	*MUC5AC*	2.435	-2.404	-2.017	-2.892	3.477	3.500	3.032	C4
ENST00000634891	*RYR3*	2.229	-3.377	-1.554	-3.586	-2.431	2.639	-3.449	C8
ENST00000542267	*FBXL17*	2.026	-1.232	0.180	-2.658	3.086	0.266	2.477	C4
ENST00000589042	*TTN*	-2.242	-2.595	3.411	4.584	-2.388	-1.965	3.498	C1
ENST00000357387	*RICTOR*	-2.369	-2.206	-0.033	2.407	2.181	1.284	-4.070	C7
ENST00000561890	*MUC22*	-2.472	-1.682	-1.632	2.156	-2.562	-2.191	-2.425	C3
ENST00000336596	*EPHA3*	-3.001	2.139	-3.421	-1.535	-3.704	1.921	-2.814	C3
ENST00000648947	*INO80*	-3.444	-1.309	2.424	-2.928	-3.439	-2.692	-0.727	C3
ENST00000389484	*LRP1B*	-4.888	-4.916	3.192	2.602	-4.245	-2.136	3.231	C1

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 2: Results from comparison of SSC scores between target cohorts (CAR, FVG, VBI, QGP) and the relevant reference superpopulations from the 1000 Genomes Project (EUR, AFR, SAS). The last column refers to the nature of the comparison carried out, as detailed in Supplementary table 3.

		SSC score							
Transcript id	**Gene name**	**QGP**	**CAR**	**FVG**	**VBI**	**EUR**	**AFR**	**SAS**	**Comparison**
ENST00000378473	*PLCB4*	-4.524	2.917	-3.907	-0.413	-4.272	-2.873	-3.079	C3
ENST00000366574	*RYR2*	-4.347	3.792	-4.694	-5.053	-3.704	-2.318	-2.087	C3
ENST00000315872	*ROCK2*	-3.680	3.822	-2.372	-1.149	-4.160	-3.248	-3.665	C3
ENST00000361445	*MTOR*	-3.371	-0.975	0.100	2.659	-3.378	-4.276	-3.888	C3
ENST00000358691	*HELZ*	-3.131	4.249	1.658	-3.729	-3.437	-2.929	3.413	C1
ENST00000355286	*EYA4*	-3.000	-1.900	2.671	-1.486	-2.087	-2.589	-0.756	C3
ENST00000381501	*TEC*	-2.996	-2.615	-1.656	3.085	-2.497	-2.427	-0.767	C3
ENST00000265382	*PIP5K1B*	-2.952	2.574	-0.576	-2.746	-3.246	-3.197	-1.583	C3
ENST00000359015	*MAP3K5*	-2.758	2.108	0.850	1.431	-3.555	-2.472	-3.293	C3
ENST00000335670	*RORA*	-2.586	-2.953	1.176	2.228	-2.499	-2.743	-0.526	C3
ENST00000370056	*VAV3*	-2.523	3.467	1.312	1.038	-3.282	-2.646	-1.406	C3
ENST00000381298	*IL6ST*	-2.522	-1.224	3.859	2.542	-2.120	-1.466	-2.644	C3
ENST00000432237	*CD163*	-2.506	-1.404	2.793	-2.306	-2.419	-0.629	-2.156	C3
ENST00000392552	*GPR155*	-2.338	-1.261	-1.336	2.338	-2.417	-1.608	-2.585	C3
ENST00000382292	*SACS*	-2.324	-4.408	3.917	2.284	-3.530	-2.726	-2.082	C3
ENST00000392132	*XRCC5*	-2.176	-2.147	2.722	-1.257	-2.673	-2.107	-1.787	C3
ENST00000313708	*EBF1*	-2.068	2.253	-1.422	-0.914	-2.980	-1.665	-3.222	C3
ENST00000400841	*CRLF2*	2.036	-1.347	-1.496	-2.083	2.581	2.185	1.052	C4
ENST00000369850	*FLNA*	2.058	-3.158	-0.860	-4.025	-3.073	-3.351	-3.097	C5
ENST00000344327	*TRPC6*	2.062	-3.776	-2.671	-2.711	-3.382	0.278	-2.242	C5
ENST00000263317	*NOX4*	2.225	-2.716	-2.554	-2.717	2.134	3.532	3.770	C4
ENST00000403662	*CSF2RB*	2.237	-2.363	1.702	0.620	2.613	0.319	2.782	C4
ENST00000297494	*NOS3*	2.243	1.436	-2.178	-0.851	2.109	2.460	2.455	C4
ENST00000295598	*ATP1A1*	2.258	-2.679	0.547	0.930	-2.204	-1.886	-2.448	C5

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

ENST00000085219	*CD22*	2.288	0.576	0.028	-2.311	2.368	-0.600	2.142	C4
ENST00000305877	*BCR*	2.397	-1.338	2.028	-3.021	3.994	2.923	3.631	C4
ENST00000333149	*TRIM50*	2.501	2.275	1.138	-2.022	2.271	1.372	3.197	C4
ENST00000271332	*CELSR2*	2.522	3.651	-2.581	2.455	2.443	-2.129	2.936	C2
ENST00000447648	*TECPR1*	2.666	-2.351	1.822	-1.669	2.777	3.213	-0.027	C4
ENST00000324856	*ARID1A*	3.434	-3.120	-2.683	-1.845	-2.085	-2.053	0.835	C5
ENST00000263094	*ABCA7*	3.796	-1.601	-2.581	1.004	2.591	3.325	3.998	C4
ENST00000372923	*DNM1*	3.941	-2.514	-1.222	-0.710	-2.066	-1.575	-2.077	C5
ENST00000621226	*MUC5AC*	3.965	-3.705	-3.751	-4.601	3.679	3.267	4.244	C4
ENST00000533211	*SPTBN2*	4.531	-2.266	1.209	-1.893	2.589	2.191	2.778	C4
ENST00000529681	*MUC5B*	4.744	3.085	1.483	-2.070	4.884	4.396	5.095	C4

Table 3: List of genes with a concordant signature of selection between DSC and SSC scores, after the comparison between target cohorts (CAR, FVG, VBI, QGP) and the relevant reference superpopulations from the 1000 Genomes Project (EUR, AFR, SAS).

		DSC score							**SSC score**						
Transcript id	**Gene name**	**QGP**	**CAR**	**FVG**	**VBI**	**EUR**	**AFR**	**SAS**	**QGP**	**CAR**	**FVG**	**VBI**	**EUR**	**AFR**	**SAS**
ENST00000369850	*FLNA*	3.854	-2.435	0.272	-2.510	-2.399	-2.166	-1.879	2.058	-3.158	-0.860	-4.025	-3.073	-3.351	-3.095
ENST00000263094	*ABCA7*	2.566	-0.433	-2.168	0.071	3.020	2.681	2.150	3.796	-1.601	-2.581	1.004	2.591	3.325	3.998
ENST00000621226	*MUC5AC*	2.435	-2.404	-2.017	-2.892	3.477	3.500	3.032	3.965	-3.705	-3.751	-4.601	3.679	3.267	4.244

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

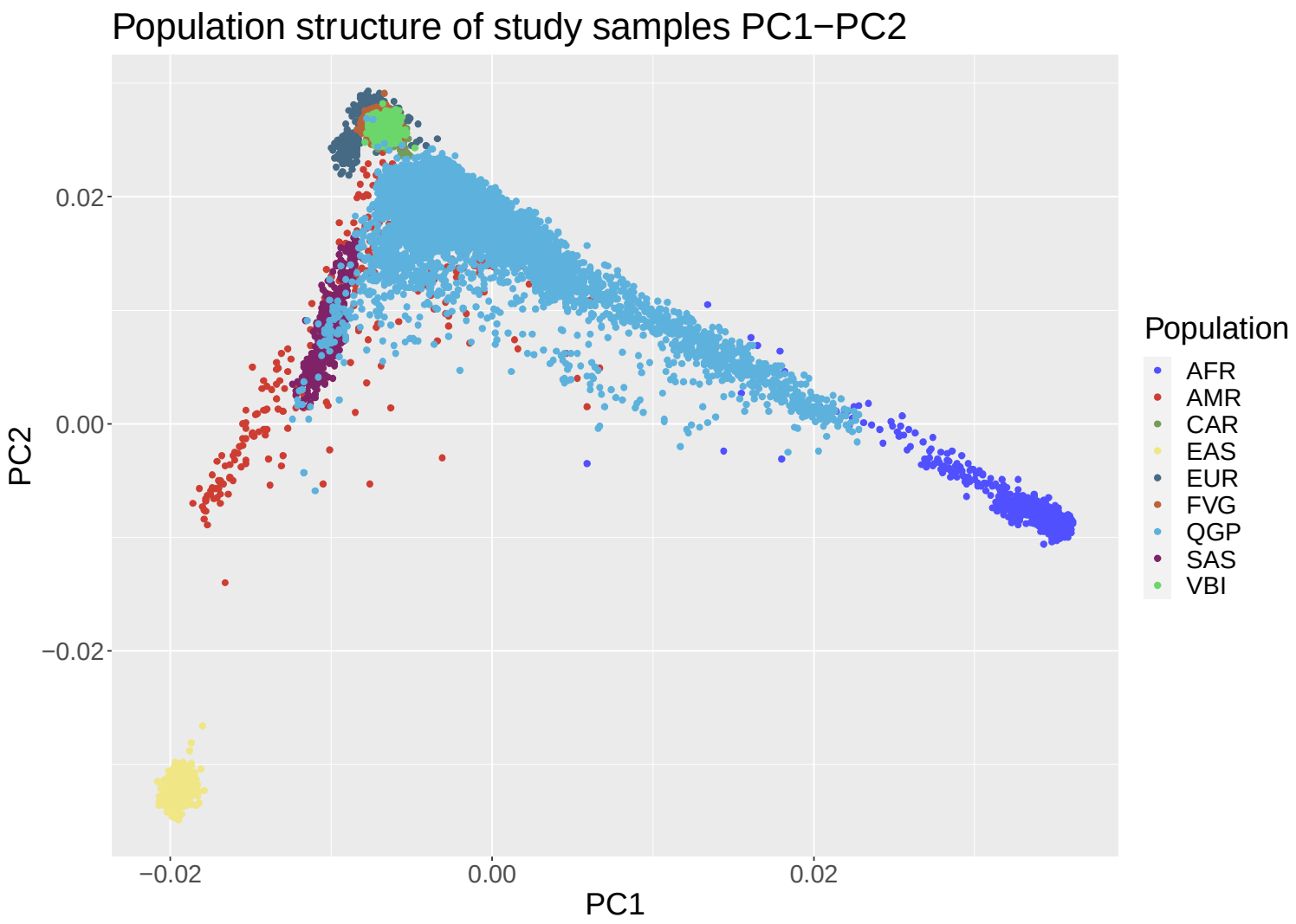
Table 4: Comparison of Singleton burden between the Covid-19 positive cohort (TOV) and other target and reference populations. The reported p-values refer to the comparison between whole gene singleton burden ("P-value whole gene" column) and coding regions singletons burden ("P-value CDS region" column). All singleton counts have been adjusted considering the sample size of each cohort.

Transcript ID	Gene name	Cohort	P-value whole gene	P-value CDS regions
		CAR	0.630140	0.409653
		FVG	0.046901	0.316565
		VBI	0.000458	0.013015
ENST00000369850	***FLNA***	**QGP**	0.000028	0.039803
		EUR	0.312323	0.786342
		AFR	0.878006	0.787767
		SAS	0.200408	0.813561
		CAR	3.2746E-11	2.5959E-07
		FVG	3.1278E-23	1.0535E-17
		VBI	7.1607E-21	2.0060E-16
ENST00000263094	***ABCA7***	**QGP**	6.2413E-63	1.5606E-40
		EUR	4.4467E-10	1.0966E-08
		AFR	1.7360E-08	5.3713E-09
		SAS	2.3435E-04	3.7924E-06
		CAR	7.4274E-12	8.4692E-11
		FVG	1.8836E-31	2.7623E-24
		VBI	2.3148E-36	3.2118E-30
ENST00000621226	***MUC5AC***	**QGP**	1.5512E-76	1.2101E-64
		EUR	3.3142E-06	4.0071E-08
		AFR	2.9701E-08	1.0241E-10
		SAS	7.2316E-03	9.9471E-05

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

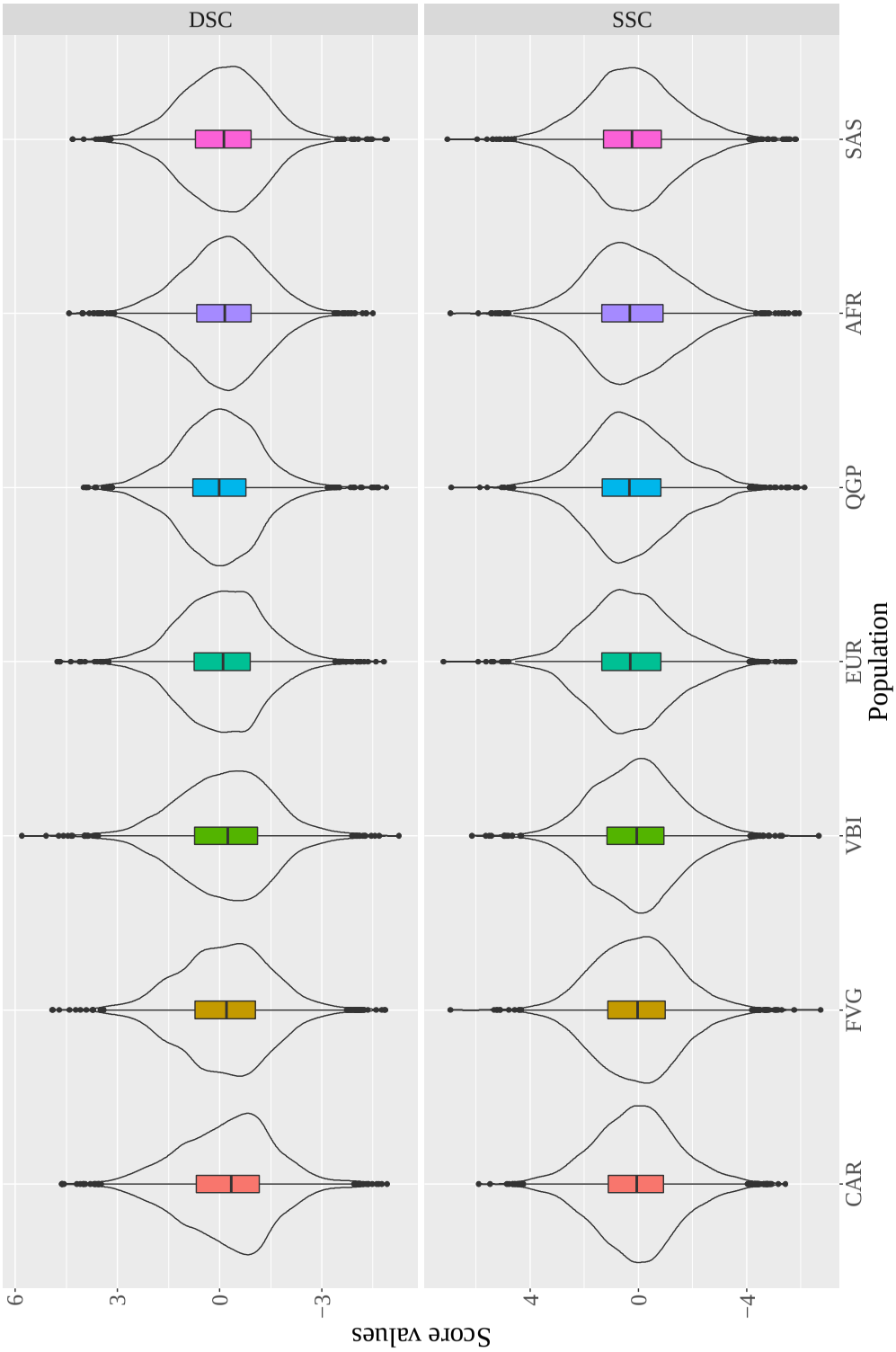
Figure 1: PCA plot of the QGP and INGI cohorts projected onto 1000Genomes Project data. As expected, the first two principal components already show the separation between the QGP and the INGI cohorts and the overlap with the selected populations for the ancestry-related comparisons.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure 2: Distributions of the prioritization scores. Violin plots of the distributions of DSC (top panel) and SSC (bottom panel) scores in the subset of selected genes for all target populations (CAR, FVG, VBI, QGP) and all reference outbred populations (AFR, EUR, SAS) from 1000Genomes project.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264507; this version posted October 7, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .